

University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

---

Electronic Theses and Dissertations

---

5-2009

# Speech perception and the McGurk effect : a cross cultural study using event-related potentials.

Jia Wu

*University of Louisville*

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

---

## Recommended Citation

Wu, Jia, "Speech perception and the McGurk effect : a cross cultural study using event-related potentials." (2009). *Electronic Theses and Dissertations*. Paper 1597.

<https://doi.org/10.18297/etd/1597>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact [thinkir@louisville.edu](mailto:thinkir@louisville.edu).

SPEECH PERCEPTION AND THE MCGURK EFFECT--- A CROSS CULTURAL  
STUDY USING EVENT-RELATED POTENTIALS

By

Jia Wu

B.S. Fudan University, 2001  
M.A. University of Louisville, 2005

A Dissertation  
Submitted to the Faculty of the  
Graduate School of University of Louisville  
in Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Psychological and Brain Sciences  
University of Louisville  
Louisville, Kentucky

May 2009

Copyright 2009 by Jia Wu

All rights reserved

SPEECH PERCEPTION AND THE MCGURK EFFECT--- A CROSS CULTURAL  
STUDY USING EVENTS RELATED POTENTIALS

By

Jia Wu

B.S. Fudan University 2001

M.A. University of Louisville 2005

A Dissertation Approved on

March 10, 2009

By the following Dissertation Committee:

---

Dissertation Director

---

---

---

---

---

## **DEDICATION**

This dissertation is dedicated to my parents

Mr. Mingyuan Wu

and

Mrs. Weiqin Liu

Who have given me invaluable educational opportunities

## ACKNOWLEDGMENTS

I would like to thank my major advisor, Dr. Dennis Molfese, for his great guidance. I would also like to thank the other committee members, Dr. Roscoe Dykman, Dr. Enda Ross, Dr. Richard Lewine and Dr. Pavel Zahorik for their comments, assistances and patience throughout the procedures of completing the paper. Thanks to my lab members Alysia Rue, Adrienne Roman, Chris Warren, Jill Jacobi-Vessels, Christy Osborne, Jennifer Beswick, Rachel Waford, Matt Stone, Erin Shannon, Jeanne Johnson and others for helping me throughout the experiments. Thanks my friend Ron McGavic for presenting the stimuli. Thanks my colleagues Michael Crowley, James McPartland, Helena Rutherford and Christopher Bailey in Yale Child Study Center for their valuable advices. Thanks my friend Jingping Xu and Andrew Haun for printing and delivering the paper. Thanks to all my friends for participating the study. Finally I want to reference Gregory Brown for the kind and loving help he provided in discussing the proposal with me; I would also like to thank all my family members in China for their constant support.

## ABSTRACT

### SPEECH PERCEPTION AND THE MCGURK EFFECT--- A CROSS CULTURAL STUDY USING EVENT-RELATED POTENTIALS

Jia Wu

March 10, 2009

Previous research has indicated the important role of visual information in the speech perception process. These studies have elucidated the areas of the brain involved in the processing of audiovisual stimuli. The McGurk effect, an audiovisual illusion, has been demonstrated to be a useful tool in the study of audiovisual integration. Brain imaging research suggests that the McGurk effect is modulated by brain structures in the Superior Temporal Gyrus, Supplemental Temporal Cortex and Broadman area 41. Electrophysiological studies suggest that the McGurk effect generates a different brainwave than the standard audiovisual congruent condition in frontal, central and parietal regions among N1, P2, N300 and later ERP components. The magnitude of the McGurk effect is influenced by lexical information, cultural difference and language experience, although these effects have not been investigated by electrophysiological methods. The current study compares three experimental groups, native Chinese speakers

who had less than three years of residency in the United States; native Chinese speakers with more than three years of residency in the United States; and native English speakers. A high density (256 channels) hydrocel system was used to collect the ERP data from the participants in three experiments. The main effect of McGurk was studied in Experiment 1, by dubbing an audio /ba/ onto a video /ga/. The behavioral responses and the characteristic brainwave of the McGurk effect were examined. The semantic and culturally related McGurk effects were investigated in Experiment 2 and 3. The three groups were contrasted in each of the experiments. The results showed brainwave differences for the McGurk and the standard audiovisual congruent condition in the left hemisphere. These results also revealed a word/nonword separation in the McGurk effect, which provides electrophysiological evidence for an interaction of the McGurk condition and semantic/cognitive processing. Contrary to hypothesis, significant group differences were not found between the two Chinese groups. Limitations and implications of this study are discussed.

## TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS.....	iv
ABSTRACT.....	v
LIST OF TABLES.....	xi
LIST OF FIGURES.....	xii
CHAPTERS	
I. INTRODUCTION.....	1
1.1 Speech Perception and the McGurk effect.....	3
1.2 The integration theory of AV speech.....	4
1.3 Neural correlates.....	7
1.3.1 Functional MRI.....	7
1.3.2 Event-Related Potentials (ERPs).....	9
1.4 Cross-cultural language studies.....	18
1.5 Phonological categorization and second language.....	21
1.6 Summary.....	24
II. HYPOTHESES.....	26
2.1 Experiment 1.....	26
2.2 Experiment 2 and 3.....	27

III. METHODS.....	30
3.1 Participants.....	30
3.2 Stimulus.....	31
3.3 Procedure.....	32
3.4 Equipment.....	34
3.5 Experiment Design.....	35
3.5.1 Experiment 1: Consonant identification of syllables .....	35
3.5.2 Experiment 2: Consonant identification of Chinese words/non- words .....	35
3.5.3 Experiment 3: Consonant identification of English words/nonwords.....	37
IV. RESULTS.....	39
4.1 Experiment 1.....	39
4.1.1 Results for Hypothesis 1.....	39
4.1.2 Results for Hypothesis 2 .....	40
4.1.3 Results for Hypothesis 3.....	42
4.1.3.1 ERP data pre-processing.....	43
4.1.3.2 ERP data analysis: factor analysis and repeated measures ANOVA.....	44
4.1.4 Hypothesis 4.....	49
4.1.5 Hypothesis 5.....	52

4.1.6 Hypothesis 6.....	52
4.2 Experiment 2 and 3 Results.....	54
4.2.1 Results for Hypothesis 1 and 2.....	54
4.2.2 Results for Hypothesis 3.....	58
4.2.3 Results for Hypothesis 4.....	63
V. DISCUSSION.....	70
5.1 Experiment 1.....	72
5.1.1 Hypothesis 1.....	73
5.1.2 Hypothesis 2.....	73
5.1.3 Hypothesis 3.....	74
5.1.4 Hypothesis 4.....	75
5.1.5 Hypothesis 5.....	78
5.1.6 Hypothesis 6.....	78
5.2 Experiment 2 and 3.....	81
5.2.1 Hypothesis 1.....	81
5.2.2 Hypothesis 2.....	83
5.2.3 Hypothesis 3.....	85
5.2.4 Hypothesis 4.....	86
5.3 Summary and Conclusions.....	89
REFERENCES.....	92
APPENDIX.....	97

CURRICULUM VITAE.....100

## LIST OF TABLES

TABLE	PAGE
1. Pairwise Comparisons of Reaction Time (ms).....	41
2. Factor analysis for Experiment 1 (n = 36).....	46
3. Significant results for the behavioral part of Experiment 2 and 3.....	55
4. Factor analysis For Experiment 2 (N = 21).....	60
5. Significant ERP results for Experiment 2.....	62
6. Factor analysis for Experiment 3.....	65
7. Significant ERP results for Experiment 3.....	66
8. The distribution of the different levels of the McGurk effects in three experimental groups.....	79

## LIST OF FIGURES

FIGURES	PAGE
1. Impact of the McGurk effect on Reaction Time .....	42
2. Clustering of adjacent scalp electrode channels for subsequent data analysis.....	45
3. Plot of the Factor loadings for Experiment 1.....	46
4. ERPs of Experiment 1 over the left central area. ....	48
5. ERPs of Experiment 1 over the right central area.....	49
6. The mean and the standard error of the McGurk effect in Experimental Groups.....	51
7. The interaction of LANGUAGE and AUDITORY LEXICAL INFLUENCE on the McGurk rate. ....	56
8. The interaction of LANGUAGE and VISUAL LEXICAL INFLUENCE on the McGurk rate.....	57
9. The main effects of LANGUAGE and GROUP on the McGurk effect. ....	59
10. Plot of the factor loadings for Experiment 2. ....	61
11. ERPs elicited in Experiment 2 over the left centro-parietal region. ....	63
12. Plot of the factor loadings for Experiment 3.....	65
13. The ERPs of Experiment 3 over the left centro-parietal region .....	68
14. The ERPs of Experiment 3 over the right parietal region .....	69

## **CHAPTER I**

### **INTRODUCTION**

Most linguistic interactions occur within a face-to-face context, involving both auditory and visual inputs. Although people are able to understand words without visual input, congruent visual and auditory information is important in understanding a conversation (Sumbly and Pollack 1954). When visual and auditory information are incongruent, people sometimes fuse or combine the two types of information erroneously, perceiving a new sound. This phenomenon is known as the McGurk effect (McGurk and MacDonald 1976).

Findings from both behavioral (the McGurk effect) and neuroimaging studies support the view that Audiovisual speech perception is an integrated procedure. Studies using functional Magnetic Resonance Imaging (fMRI) show that the primary auditory cortex is active during pure lip reading, while the primary visual cortex is active with just auditory inputs (Calvert, Bullmore et al. 1997). These studies as well as those reported below strongly support the notion that visual information is very important in speech perception.

While fMRI can be used to locate the source of brain sites involved in the integration of speech perception, it fails to provide accurate information about the precise timing of neural events. Event-related potentials (ERPs), by contrast, provide very

accurate temporal information, but are generally regarded as less accurate for localizing the participating brain structures. However, high density EEGs with its broad head coverage, in conjunction with a new electrical source analysis program just released by Electrical Geodesics, Inc. may prove also to be a very valuable tool in locating the sources within the brain that generate the potentials recorded from the surface of the head. Ultimately, these new developments could provide information about the timing of neural events, the brain sources involved in mediating neural-cognitive processes, and the way in which these source structures communicate with each other.

A review of the literature indicates that several brain waves (components) of the ERP are associated with the McGurk effect, including N1, P2, N400, P600 and Mismatch Negativity (MMN) (Friederici 1995; Friederici 2002; Lebib, Papo et al. 2004; van Wassenhove, Grant et al. 2005). The amplitude and latency of these components reflect differences in multi-sensory perception versus single modality perception as early as 150 ms following auditory stimulus onset (van Wassenhove, Grant et al. 2005).

Studies of cultural differences indicate that people with different language backgrounds exhibit different levels of the McGurk effect. Some languages are not sufficiently ambiguous to clearly show the McGurk effect (Sekiyama and Tohkura 1993; Calvert, Bullmore et al. 1997; Sekiyama 1997; Gentilucci and Cattaneo 2005; Sekiyama and Burnham 2008). In these cases, visual input has little effect on the perception of speech sounds. It is hypothesized that different levels of language experience require the activation of different parts of the brain (Sekiyama and Tohkura 1993; Sekiyama and Burnham 2008). However, most of the cross-cultural studies conducted to date report only behavioral measures (Sekiyama and Tohkura 1993; Sekiyama 1997; Gentilucci and

Cattaneo 2005; Sekiyama and Burnham 2008).

Overall, the McGurk effect is a valuable phenomenon for research on speech perception because it can be used to study the binding of visual and auditory percepts in behavioral and neuroscience contexts. Previous studies suggest that positive associations exist between language background, brain processing and the McGurk effect. The present study tests participants with different backgrounds in language acquisition in a McGurk eliciting paradigm: native born English-speaking US citizens, native born Chinese who acquired English as a second language and have been in the US for longer than three years, and native born Chinese who have been in the US for less than three years. ERPs were recorded from 256 different electrode sites on the head in paradigms designed to evaluate group differences in the McGurk effect across language backgrounds. ERPs generated from three different speech paradigms were analyzed in relation to concurrent behavioral measures.

## **1.1 Speech Perception and the McGurk effect**

It has been understood for many years that speech perception occurs not only on what is heard but also upon what is seen (Sumby and Pollack 1954). Speech is more intelligible when visual information is congruent with auditory information (MacDonald, Andersen et al. 2000). Seeing the speaker facilitates the understanding of a conversation in a noisy environment (Sumby and Pollack 1954; MacDonald, Andersen et al. 2000). However, prior to 1970, it was believed that visual information aids speech perception only when auditory information is distorted. This explanation remained unchanged until McGurk and MacDonald discovered what is now simply referred to as “the McGurk effect” (McGurk and MacDonald 1976). Audio and visual information are no longer

treated as separate phenomena. Instead, speech perception is now discussed in terms of the binding of auditory and visual percepts.

McGurk & MacDonald (1976) found that when a visually presented syllable such as /ga/ and the auditory stimulus /ba/ were dubbed together, 98% of the participants perceived the syllable as /da/. Surprisingly, participants perceived a syllable that differed from both the visual and auditory input when both were given at the same time. The same effect was also found in other syllable pairs; e.g., the visual /ka/ dubbed into an audio /pa/ created a /ta/ perception. McGurk fusion effect occurs between speech sounds differing in places-of-articulation (POA); i.e., places in the mouth where articulators such as the tongue and mandible move relative to the lips or roof of the mouth. For example, the misperception of an alveolar constant /d/ occurs when an audio signal of a bilabial stop-consonant speech sound /b/ is dubbed onto the visual velar /g/. However, when the audio and visual information are combined in an opposite way with the visual bilabial dubbed onto an audio velar, the fused perception does not occur. Instead, a McGurk combination is perceived in which participants report a combined syllable such as /daba/ (McGurk and MacDonald 1976; MacDonald and McGurk 1978). The McGurk effect is observed in both English speaking children and adults (McGurk and MacDonald 1976; Rosenblum, Schmuckler et al. 1997; Cienkowski and Carney 2002).

## **1.2 The integration theory of AV speech**

Facial movements during speech provide a complex and dynamic structure of visual information that influences auditory speech processing. Multisensory integration in speech occurs at both the perceptual and neural processing levels. Although Audiovisual (AV) speech is a natural component of speech signal processing, classic models of speech

processing are either based on silent reading of words/sentences or focusing on the acoustic raw inputs (Hagoort 2008). However, the accumulating evidence for multisensory processing raises questions about the validity of traditional speech processing models (Calvert, Bullmore et al. 1997; Cienkowski and Carney 2002; Friederici 2002; van Wassenhove, Grant et al. 2005; Stekelenburg and Vroomen 2007; van Wassenhove, Grant et al. 2007). Given this dilemma between traditional and emerging views, multisensory processing theories that address the interaction and integration of different modalities are needed to gain a more comprehensive understanding of speech perception.

The ‘Analysis-by-Synthesis’ (ABS) framework is a well-known model purporting to explain the integration of AV speech (van Wassenhove, Grant et al. 2007). This approach was based on the empirical findings that visual speech plays a crucial role in predicting subsequent auditory speech inputs. The dynamics of the facial articulators perceived by the listener enable the brain to narrow down the possible meanings of the auditory signals. Consequently, if the auditory signal is congruent with visual information (such as facial and lip movements), perception is quicker and more accurate. In contrast, if the two inputs are contradictory, perception is impaired.

The Motor Theory of Speech Perception is another theory purporting to explain how the integration of AV information influences the perception of speech (Liberman and Mattingly 1985). This theory hypothesizes that AV speech creates a ‘Common Currency’ in the brain activating a Motor System controlling congruent lip movements. The perception of AV speech is achieved by the activation of the Common Currency; i.e., the activation of one modality by input from another. The Abs Theory also involves the

motor system, but the Motor theory additionally emphasizes a large-scale brain network in recognizing the visual information as well as an abstract speech representation (Common Currency).

In addition to conceptual models for the integration of AV speech, there is a broader body of knowledge afforded by cognitive neuroscience (Puce, Allison et al. 1998; Beauchamp, Lee et al. 2004; Gentilucci and Cattaneo 2005). For instance, the posterior superior temporal sulcus (pSTS) region appears to be strongly involved in multisensory processing (Ghazanfar, Maier et al. 2005). Information about visual speech is delivered to the pSTS by an adjacent movement sensitive areas MT/V5. In addition, face motion processing occurring in the superior temporal cortex is sensitive to gaze and mouth movements (Puce, Allison et al. 1998). This region also contains networks of multisensory neurons with intricate connections to different sensory areas (Beauchamp, Lee et al. 2004).

From a social-cognitive neuroscience perspective, the idea of a ‘mirror neuron system’ also bears on our thinking about the integration of AV speech. Specifically, the idea is that a cortical neuronal system XYZ activates a mirror neuron system in the frontal region of the brain. Neurons in the mirror region are activated in a similar manner whether a person performs an activity or simply sees its performance. Rizzolatti and Craighero describe the mirror neuron system role in the integration of AV speech (Rizzolatti and Craighero 2004). They propose that Broca’s area is the neural structure responsible for the mirror integration of visual speech (Rizzolatti and Craighero 2004). This area is well connected to pSTS enabling it to feed information about the visual speech properties of speech to the auditory cortex. (Skipper, van Wassenhove et al.

2007).

### **1.3 Neural correlates**

With the development of brain imaging techniques, researchers can investigate the neural correlates of language processing in the brain. The most widely used techniques record fMRI or ERP. Neuroimaging and electrophysiology methods provide neurological evidence supporting integration theories of AV speech.

#### **1.3.1 Functional MRI**

Functional MRI is based on the changes in the levels of hemoglobin in the different brain areas activated by speech and other cognitive events. The spatial resolution of fMRI can be as small as a square millimeter of brain tissue. Researchers have used fMRI research to study the function of the motor system in the integration of natural AV speech. Reports have demonstrated an appreciable overlap of AV speech perception and speech production, especially in the frontal motor area, an integration that does not occur to isolated audio stimuli (van Wassenhove, Grant et al. 2007). Specifically, speech production and perception overlap appreciably in anterior and posterior superior temporal lobes, and the ventral premotor cortex. The involvement of the motor area in AV speech processing, but not in auditory alone speech, provides strong evidence for both the Abs and Motor theories.

Numerous fMRI studies have reported an interaction between visual and auditory modalities. Calvert and colleagues (Calvert, Bullmore et al. 1997) observed that auditory cortex could be activated without actual auditory input. Brain activity was observed while

subjects listened to either speech or observed silent lip-reading. Surprisingly, silent lip-reading activated both primary and associate regions of auditory cortex. However, the auditory cortex was not activated if the mouth was closed and the speech was not observable from the face.

Hall and colleagues (Hall, Fussell et al. 2005) found that the associate auditory cortex was activated during silent lip reading. They also described other cortical areas potentially associated with lip reading. For example, the left inferior frontal gyrus was associated with poor lip reading ability. They conjectured that a central executive ability was required to a greater degree in poor speech readers than in controls, suggesting speech perception was modulated in part by executive ability.

The interaction of visual and auditory modalities is also supported from other studies, which show one type of perceptual input engages cortical regions commonly associated with another type of input. For instance, it appears that visual lip reading activates auditory cortex and auditory input activates the visual cortex. Giraud and colleagues (Giraud and Truy 2002) report that the fusiform gyrus and the primary visual cortex are activated by listening to speech in both normal and cochlear-implant subjects. They assume that because the visual cortex is “expecting” some input associated with the audio inputs, it is activated through a top-down pathway. More interesting, the visual area activation is much greater in cochlear-implant patients than in subjects with normal hearing, suggesting a greater reliance on vision in the patients. These results demonstrate the plasticity of the Audiovisual integration system. The interaction between auditory and visual information was altered when one modality is degraded.

As mentioned earlier, the McGurk effect has been shown to be a useful tool for

probing the interaction of visual and auditory modalities. An fMRI study by van Wassenhove and colleagues (van Wassenhove, Grant et al. 2007), provides additional interesting insights into the McGurk effect. In this experiment, a visual /k/ (Vk) was dubbed onto auditory /p/ (Ap). The brain activation of the resulting fused perception /t/ correlated more closely with the activation of the actual perception of an AV/t/ than with either of the sensory inputs. It was found that when the fused effect was not perceived and the visual /k/ was reported as the outcome, the brain activation was focused in the middle and inferior frontal gyri; but when the fused effect /t/ occurred, the activation was moved near to the left somatosensory areas, premotor cortex and the primary motor cortices (van Wassenhove, Grant et al. 2007). Different areas of the brain were activated in sequence in a fused McGurk perception and followed a different time course. At first, sensory cortices were observed to correlate with their congruent counterparts during a McGurk perception. Next, motor, auditory and somatosensory cortices were activated, and found to correlate with the actual perception or the phonological categorization. These findings suggest an active role of the motor area during AV speech perception.

### **1.3.2 Event-Related Potentials (ERPs)**

More recent research indicates that event-related potentials (ERPs) can provide important information about the time course of AV speech integration. ERPs are electroencephalograph (EEG) recordings that are time locked to the onset of a stimulus. ERPs vary with the nature of perceptual/cognitive tasks as evidenced in a large number of experiments (Picton and Taylor 2007; Hagoort 2008). Unlike behavioral measures, ERPs represent on-line cognitive and physical processing with both spatial and temporal

resolution that can be measured in fractions of a millisecond. ERP signals can be modeled to detect regions within the brain that produce the changes in voltages seen from electrodes placed on the surface of the head. As demonstrated in past studies, the ERP is very sensitive to phonetic variations (Molfese 1978; Molfese and Molfese 1988; Molfese, Alexandra et al. 2005).

In the domain of ERP research, the Mismatch Negativity (MMN) paradigm has come to be widely used to study speech perception. MMN has been identified as a pre-attentive ERP component (Näätänen 2003) that is usually easily observable in an oddball paradigm, where a high ratio of similar sounds (standards) are mixed with a low ratio of deviant sounds (deviants). It is a very good paradigm in the sense that it is completely passive (requires no response) and produces a reliable average by allowing the delivery of a large number of trials in a short time (5 – 7 minutes). While MMN will not be studied in this proposal, it has been used in studies of the McGurk effect.

Colin and colleagues ((Colin, Radeau et al. 2002) studied the MMN in an audio, visual and audiovisual paradigm (N = 8), using an Ag-AgCl 6 electrode recording system (Fz, Oz, C3, C4, M1 and M2). In the audiovisual conditions, the standard and the deviant stimuli shared the same auditory information, but differed in the visual information. They reported an MMN in Fz at the P1-N1-P2 component around 50 to 170 ms in the audiovisual and audio only condition. No MMN was found in the visual only condition. The results indicated that visual information interfered with audiovisual accuracy in an early pre-attentive stage.

In another study by Kislyuk and colleagues (Kislyuk, Mottonen et al. 2008) observed MMN in an auditory only condition, presenting /va/ as the frequent or standard

stimulus and /ba/ as the infrequent or target stimulus. In their McGurk condition, the target stimulus was the video /va/ dubbed with an auditory /ba/ (perceived as /va/); and the standard stimulus was the audiovisual /va/. Perceptually the stimuli were the same even though they had different acoustic components. No MMN occurred during the Audio-Only condition. MMN started around 130 to 170 ms and was at various brain regions including frontal, central, occipital and central-parietal regions. In this study they used 30 electrodes and tested 11 subjects.

Saint-Amour and colleagues used a high-density array (Saint-Amour, De Sanctis et al. 2007). They recorded ERPs from audiovisual incongruent conditions in which the standard Audiovisual /ba/ and deviant stimulus was dubbed with audio /ba/ and visual /va/. Visual-only /ba/ and visual-only /va/ were also presented in addition to standard and deviant multisensory stimuli. The MMN was obtained by subtracting the correspondent visual-only ERP from the Audiovisual ERP. An oddball paradigm was used with deviant stimuli occurring on about 20% of the 1420 trials per modality. They found a larger left hemisphere MMN response over the frontal region that occurred at 175 to 225 ms following auditory onset, which originated from the left transverse gyrus (Brodmann area 41). This was followed at 290 ms by a bilateral frontal-central activation, which was generated by a dipole in the right superior temporal gyrus (STG). After 350 ms the left hemisphere dominance (i.e., larger response over the left hemisphere) of MMN appeared again, but the dipole for wave was located in the left superior temporal gyrus. The results indicated that STG was activated during the McGurk condition, but not during the audiovisual congruent condition.

As indicated above, the mismatch response (MMN) is recognized as a pre-

attentive process completely automatic, even occurring in sleep (Sallinen, Kaartinen et al. 1994). While it may correlate with or be predictive of higher order cognitive processing, in much the same way as arousal, it does not convey direct information about higher order cognitive processing. While the paradigm is much the same as the one used to study P300, it does not require as does P300 any preparation for making a discriminative motor response or any updating of memory (working memory). It is doubtful, therefore, that it could reveal any reliable information about either phonological or nonphonological variables or the way they interact in the comprehension of speech or reading. Complicated learning is involved in the perception of speech sounds. For example, the McGurk effect would appear to depend upon complex learning of the meaning of facial movements occurring over a number of years, perhaps all of life. Thus it is doubtful that this process could ever be completely automatic and it would seem to require a constant search of working memory. Because of the complexity of the McGurk effect, the MMN paradigm does not appear to be a good option. Instead, each condition needs to be presented the same number of trials. In those types of studies, various components of ERPs were investigated as below.

Van Wassenhove and colleagues (van Wassenhove, Grant et al. 2005) recorded EEG to the syllable /ka/, /pa/ and /ta/ from 16 subjects in an audio, visual and audiovisual paradigms using a 32-electrode system. The audiovisual paradigm contained both congruent and incongruent stimuli (audio /pa/ dubbed onto visual /ka/). They found that for the congruent stimuli, N1 and P2 in the audiovisual condition had decreased amplitude and shortened latency compared to the Audio Only condition. For the incongruent stimuli, N1 and P2 in the audiovisual condition had the same amplitude

reduction, but with the absence of the temporal facilitation. The writers suggested that visual information systematically influenced the key timing properties of the auditory responses, also indicating that auditory processing was facilitated when auditory information was reliably predicted from visual information. This research supported the 'Analysis-by-Synthesis' theory of auditory-visual speech perception. This theory would predict that the detection of a syllable is facilitated when the visual cue is a strong predictor of a sound, such as in the lip movements involved in the production of the syllable, /pa/; however, little or no facilitation would occur when the visual cue is a weak predictor of the sound such as in the minimal lip movements involved in the production of the syllable, /ka/. In their study, the brainwaves of different modalities were compared. However they did not compare the brainwaves from congruent and incongruent audiovisual stimuli directly. The brainwave from congruent and incongruent audiovisual conditions should have had different latencies at N1 and P2. Moreover, they only analyzed 6 locations: FC3, FC4, FCz, CPz, P7 and P8. It is possible that the high-density array system to be used in this study will show other important information. In this study, brainwaves of the two stimuli are compared directly in all 256 channels.

Stekelenburg and Vroomen (Stekelenburg and Vroomen 2007) indicated that audiovisual processing interfered with the auditory processing by suppressing the N1 component when the visual information is relevant; but this suppressing effect disappears when the visual information is not relevant. They also found the phenomenon replicated in a non-speech environment.

Other studies used a similar model to study components occurring later than N1 and P2 (Kutas and Hillyard 1980; Friederici 1995; Friederici 2002; Lebib, Papo et al.

2004). Lebib and colleagues (Lebib, Papo et al. 2004) tested 13 French speakers on four types of stimuli consisting of French vowels. The stimuli were either congruent or incongruent, either easy or hard to distinguish. The brain response was recorded with a 62-electrode system. The results indicated that stimuli categorized as incongruent audiovisual perceptions elicited a larger N300 than the congruent condition at frontal, central and parietal regions. In addition, the amplitude of N300 component was modulated by the degree of audiovisual discrepancy, in that when the discrepancy was small (ones that were harder to distinguish), the negativity disappeared. They suggested that the N300 was associated with the classic N400 component which reflected semantic deviation (Kutas and Hillyard 1983).

The P600 is also related with speech processing. P600 is the positive occurring at about 600 ms after stimulus onset. It was confirmed in many studies to relate to the syntactic complexity of words (van den Brink and Hagoort 2004; Kotz, Holcomb et al. 2008). Previous research suggests that the P600 is related to cerebral repair and reanalysis processes (Friederici 1995; Friederici 2002). These studies reported that brainwave differences occurred when subjects were unaware of stimuli, suggesting automatic processing.

Proofs from behavioral studies indicated that the later components in ERP appeared in the McGurk effect might be related to cognitive functioning. Brancazio and Brancazio (Brancazio and Brancazio 2004) conducted a study testing the McGurk effect on pairs of words and nonwords. The video of a /d/ sound (such as “dish”) was dubbed to the audio of a /b/ sound in the pair (such as “bish”), and the participants were required to judge whether it was a /b/ or /d sound/. It was found that when the video represented a

word, McGurk effect was more likely to occur; however when the audio was a word, the McGurk effect was less likely to occur. It was consistent with the fact that McGurk effect was a special case of visual lexical influence.

While no study has shown any electrophysiological proof of the lexical influence on the McGurk effect, the N400 research is relevant. It was first described by Kutas and Hillyard (1984) as a negative going wave that occurs approximately 350-500 ms after stimulus onset over the centro-parietal region (Kutas and Hillyard 1980; Kutas and Hillyard 1984). It reflects verbal incongruities of many kinds including semantic and syntax errors, or the insertion of an improper word into a sentence (Kutas and Hillyard 1980; Kutas and Hillyard 1984). N400 was not only detected in a semantic environment in which a full sentence was presented (Kutas and Hillyard 1980; Kutas and Hillyard 1984), but also in the single word mode (Bentin, McCarthy et al. 1985; Holcomb 1988; Kounios and Holcomb 1994). As many researches on N400 used visually presented sequences of letters, it was also identified in studies using spoken words (Friederici 2002; Liu, Shu et al. 2006; Schiller, Horemans et al. 2008). N400 was also detected in language other than English (Liu, Shu et al. 2006). McPherson and colleagues (McPherson, Ackerman et al. 1996; McPherson, Newton et al. 1997; McPherson, Ackerman et al. 1998) asked children to discriminate words and non-words (pseudo-words), and found that the pseudo-words produced a larger N400 in groups in which IQ was controlled. Therefore N400 should show up in the McGurk effect if the fused replications were perceived as a non-word, even though the audio stimulus was a word.

Across a number of ERP studies of the McGurk effect, different ERP components have been analyzed, making it difficult to compare their results. None of the studies have

conducted a complete analysis of the brain processing using the entire stimulus interval (i.e., all the components/waves that occur after stimulus onset until the EEG recording return to baseline) (Friederici 1995; Friederici 2002; Lebib, Papo et al. 2004; van Wassenhove, Grant et al. 2005; Saint-Amour, De Sanctis et al. 2007). Moreover, all of the studies cited above analyzed specific peaks defined by some arbitrary peak-selection method. This approach ignores other peaks in the wave and fails to take into account the correlations of the data points within and across peaks.

To counteract these problems, the present proposal uses a Principal Component Analysis (PCA) to identify peaks and investigate their relationships. PCA has several advantages over traditional peak analyses (Chapman and McCrary 1995). First, PCA is able to disentangle temporally overlapping electrical brain processes. Second, the factor loading scores, when rotated, identify optimal peaks to investigate, so important information is not missed. And third, PCA criteria are mathematically defined, thereby enabling results to be more consistent across researchers. The subjective aspects of baseline-to-peak analysis are replaced by a less biased method, thus enhancing replicability.

However, PCA analysis of ERP data has its limitations. Wood and McCarthy (Wood and McCarthy 1984) suspected that PCA was vulnerable to variance misallocation. But Beauducel and Debener (Beauducel and Debener 2003) found that the misallocation of variance was acceptable when an appropriate test power was simulated. They also found that misallocation of variance was also a problem in the baseline-to-peak analysis. The parameters about how to conduct PCA and rotations were also argued. The general sources of disagreement among PCA-users are similar to those among

statisticians and are concerned with the type of data to be analyzed (correlations or covariance), the number of factors to be extracted, and the type of rotation to be employed (Varimax or one of the oblique methods). Nonetheless, whichever PCA method is chosen, it is more objective and easier to replicate than any of the current baseline-to-peak approaches.

Besides investigating changes in ERP components, investigators also studied the additivity of brain waves. Stein and Meredith (Stein, Meredith et al. 1993) developed the principle of 'spatio-temporal coincidence' (STC) and introduced the 'supra-additive' property. They suggested that in some areas of the brain the neurons respond to multisensory events with a greater firing rate than the expected summation of neuron's response at each modality. This phenomenon could contribute to a Gestalt-like effect where all incoming information is utilized by the nervous system to produce an integrated perceptual experience. Electrophysiological recordings suggest early supra-additive effects originating from both multisensory areas and single modality areas in the cortex in response to non-meaningful Audiovisual events (Giard and Peronnet 1999).

Since audio and visual signals may occur at the same time, supra-additivity also has been applied to the model of Audiovisual integration in speech perception. In support of this approach, an fMRI study of Audiovisual speech indicated that auditory and multisensory cortices especially superior temporal sulcus and superior temporal gyrus were increasingly activated relative to unimodal speech (Calvert, Brammer et al. 1999; Calvert, Campbell et al. 2000). It was suggested that the involvement of multisensory areas validated the computational model of Audiovisual speech integration: multimodal signals were integrated in multisensory cortices sending feedback to the primary sensory

cortices (Calvert, Campbell et al. 2000). Consequently, the detection of supra-additivity in electrophysiological data may be an indicator of Audiovisual integration. However, the study by van Wassenhove and colleagues (van Wassenhove, Grant et al. 2005) failed to support the hypothesis of supra-additivity. In their study, the auditory-only condition elicited the largest N1 and P2, followed, by the audiovisual condition and last the visual condition that did not induce any auditory signal at central-parietal electrode sites. They proposed a deactivation mechanism, arguing that, in the audiovisual conditions, the auditory cortex was actually inhibited by preceding visual inputs. Consequently, there would be less activation in audiovisual condition than in the auditory condition. At this time, the notion that activation follows a supra-additive pattern or a deactivation pattern remains speculative as a supposition to be tested.

#### **1.4 Cross-cultural language studies**

Most studies of the McGurk effect recruited English-speaking subjects. However, it appears logical to propose that visual input is important in comprehending speech in all languages. In an important early study, Sekiyama and Tohkura (Sekiyama and Tohkura 1993) reported that Japanese participants used visual information much less than United States participants. Japanese and English syllables were presented to individuals speaking both languages. The order of the McGurk effect from largest to smallest was as follows: 1) American English speakers deciphering Japanese syllables, 2) American English speakers deciphering English syllables, 3) Japanese decoding English syllables, and 4) Japanese subjects listening to Japanese syllables. These results suggest that the McGurk effect depends upon both the language of the listener as well as the type of language. Sekiyama & Tohkura explained their results in two ways. First, they argued that the

effects resulted in part from the large cultural difference between Japanese and western cultures. For example it is impolite to look directly at the speaker in Japan, whereas the opposite is true for western cultures. So the Japanese speakers would have less experience with visual and audio inputs in speech context because of it. Second, they advanced the notion that the Japanese pitch accent is more readily perceptible in the auditory modality than in the visual modality.

However, if Japanese pitches and tones depend more on auditory than visual information and produce a smaller McGurk effect than English, American English speakers should show their strongest effects during the perception of English syllables. Language experience could be responsible for such an effect. As noted above, when auditory information is ambiguous, people tend to rely on visual information. Unfortunately, these analyses did not test whether American English speakers in the study were poorer in listening to Japanese than they were to English. If in fact they were, then they might be expected to rely much more on visual cues in understanding Japanese than English. Thus language experience could play an important role in producing the McGurk effect. Based on this notion, it was hypothesized that one's language experience can significantly influence the magnitude of the McGurk effect.

A subsequent study from this same laboratory also supported the association of language competency and the McGurk effect. Sekiyama (Sekiyama 1997) investigated the McGurk effect in Chinese subjects who had been in Japan for a period of time. The stimulus set included 10 syllables (/ba/, /pa/, /ma/, /wa/, /da/, /ta/, /na/, /ga/, /ka/, /ra/) pronounced by two speakers, one Japanese and one American English speaker. Subjects reported what they heard after listening to an audio stimulus while at the same time

looking at a visual stimulus. The Chinese subjects exhibited a weaker McGurk effect than Japanese subjects. Importantly, the magnitude of the effect was positively correlated with the length of time the Chinese subjects had lived in Japan.

A review of McGurk effect studies in different languages suggests that the McGurk effect occurs with different strengths in different languages. For Japanese, visual information is not accurate enough to identify pitches and tones in the Japanese language (Sekiyama and Tohkura 1993). Consequently, Japanese speakers do not attend to visual information. However, for Italian speakers, visual information is not necessary and is consequently neglected (Gentilucci and Cattaneo 2005). They studied the McGurk effect in Italians and found that 70% of their participants relied exclusively upon auditory information instead of visual. They suggested that the Italian phonemic repertoire and phonetic realization of syllables are simpler than those in English. Consequently, auditory information should be sufficient for sound identification in the Italian language.

Taken together, it appears that English may be the best language to test the McGurk effect because auditory discrimination is neither too easy nor too difficult. As noted above, 98% of English speakers show the McGurk effect to syllable pairs (visual /ga/, auditory /ba/).

However, care must be taken in viewing the Japanese study (Sekiyama and Tohkura 1993) and the Italian study (Gentilucci and Cattaneo 2005) together because these studies were not completely parallel. In the studies of Japanese, Japanese and English were directly compared in the two cultures. Consequently the contrast of both language and subjects were apparent. However, in the Italian study, only Italian was used with Italian-speaking subjects. There was no control of another language or another

group of subjects. Consequently, it was not clear whether the smaller McGurk effect ‘70%’ reported by Gentilucci and Cattaneo resulted from the influence of language or from the biological/cultural differences between subjects with different language backgrounds. In light of such findings, we argue that it is important in cross-cultural studies to include control groups speaking different languages.

Sekiyama & Burnham (Sekiyama and Burnham 2008) completed a study showing that English-speaking adults and older children seem to require visual information from conversations, but not Japanese subjects of the same age. In addition, visual reliance increases with age in English-speaking children. This again suggests that the prevalence of the McGurk effect depends upon characteristics of the language. The results of these studies, however, were confounded by differences in countries of origin: English speakers were tested in Australia while the Japanese speakers were tested in Japan. It is possible that the country of origin, instead of language background, leads to different visual dependencies. This possibility will be investigated in the present study.

All of the cross-cultural studies described above used only behavioral measures. Hence, there was no way to determine whether differences in brain processing were associated with different languages and differences in subjects who produced the McGurk effect. It is important to include neuropsychological and neurophysiologic measurements in studies of different languages in order to obtain more information concerning group characteristics and then use such information in order to isolate the bases for differences in the McGurk effect.

## **1.5 Phonological categorization and second language**

People distrust their perception when they hear incongruent visual and audio

signals, as indicated by slowed response (Sekiyama and Burnham 2008). If a person relies more on vision, s/he is more likely to rely on visual information and the opposite is true for a person who relies more on audition. However, a person is more likely to show the McGurk effect or combined effect when s/he relies on both visual and auditory modality equally. As mentioned above, the difference can come from either cultural or linguistic differences. This proposal will attempt to clarify this issue.

Based upon the review presented above, it appears that when a person receives incongruent Audiovisual information s/he would either hear those sounds or something different. This effect depends on size of the boundaries between the two sounds. In some languages, there is an intermediate sound between the two sounds presented, and subjects may perceive the intermediate sound even when it is not presented. However in other languages, the intermediate sound may not be part of their phonetic system, so those subjects should not show the McGurk effect.

Phonological categorization is very different from one language to another. Since Chinese and English are very different languages, we argue that native Chinese and native English speakers have learned very different linguistic categories and consequently they will behave differently on the McGurk test.

Another important consideration when discussing phonological categorization is that language acquisition changes during development. According to the 'Critical period hypothesis' proposed by Eric Lenneberg (Lenneberg 1967), children before the age of puberty are able to master a second language to a level of competency equal to that of their native language. However, once past this critical period, the acquisition of a new language is much more difficult. Hence, we argue that Chinese subjects beyond the age

of 12 who learn English as a second language will demonstrate ERPs that differ in many ways from native born American English speakers in generating the McGurk effect. In addition, one would expect large differences in ERPs to also occur in English when it the Chinese learn it as a second language.

When people learn a second language, new categories must form to facilitate the comprehension of the new language. For example, a native Chinese speaker who does not learn English is never able to pronounce /th/ as in 'they' and 'theta'. Consequently, s/he would not distinguish the two sounds visually or acoustically. But native Chinese speakers learning English as a second language do show some ability to make this distinction compared to those who have not studied English. Consequently, exposing people to a new language environment changes their acoustic perception overall. As stated above, of the languages studied to date, English is the one that appears to show the largest McGurk effect (Sekiyama and Tohkura 1993; Gentilucci and Cattaneo 2005; Sekiyama and Burnham 2008). Thus English is an appropriate language that suits to examine the above questions.

Amount of language exposure has not been directly examined in an AV speech perception study. For instance, a potential problem of Sekiyama & Burnham's study (Sekiyama and Burnham 2008) can be investigated by recruiting two groups of Chinese speakers, one with more English-speaking experience than the other. Following Sekiyama and Burnham's study, the number of years of residence in the United States was used in the present study to reflect different levels of experience in an English-speaking experience. Those with longer periods of residence were expected to have more experience with English than those who were in the United States for a shorter time

period. As reported above, Sekiyama & Burnham found that an English environment makes English speakers more visually dependent upon AV speech perception than Japanese speakers in a Japanese environment. They attributed this to differences in language structures and environment. However it could also be due to the differences in the physiological structures of the two different ethnicities.

In the present study we address the issue of degrees of language exposure. The two Chinese groups to be studied share a similar ethnicity background (less heterogeneous than American English speakers), but will differ in the duration of their English experiences. Experience would seem to be a more important causal factor than ethnicity if higher visual dependency is found in the group with more experience. We hypothesized that the magnitude of the McGurk effect with English syllables is positively associated with the amount of experience one has in an English-speaking environment. Since I was investigating second language learners after the Critical Period, the age of my participants was restricted to the period after puberty, between 17 and 37 years old.

## **1.6 Summary**

Speech perception involves both audio and visual information. Research indicates that the audio and visual information are integrated in speech perception, particularly in the English language (McGurk and MacDonald 1976; Sekiyama and Tohkura 1993; Gentilucci and Cattaneo 2005; Sekiyama and Burnham 2008). The McGurk effect provides strong supporting evidence for the phenomenon of multisensory integration. Two theoretical approaches, ‘Analysis-by-Synthesis’ and ‘Motor Theory’ have been examined in previous studies in order to explain the function of visual speech in AV speech processing (Lieberman and Mattingly 1985; van Wassenhove, Grant et al. 2005).

Functional imaging studies show that brain regions such as the frontal motor area and the posterior superior temporal sulcus are involved in multisensory integration (van Wassenhove, Grant et al. 2007). These studies provide neural evidence for Audiovisual integration in speech perception and for the McGurk effect. For instance, it was found that the brain activation matches that of the perceived phonological categorization instead of that of the actual inputs (van Wassenhove, Grant et al. 2005). ERP studies indicate the visual influence on auditory process occurs as early as 100 ms after auditory onset (van Wassenhove, Grant et al. 2005). It is also found that the difference of the McGurk condition and the standard condition can be detected from both the early (P1, N1, P2 (Colin, Radeau et al. 2002)) and late (N300 and even later (Lebib, Papo et al. 2004; Saint-Amour, De Sanctis et al. 2007)) ERP components over various brain regions such as Fz, Cz, Pz. In a semantic environment, the McGurk effect is positively correlated with the visual Lexical influence (Brancazio and Brancazio 2004). The effect might be related to a late ERP component N400 (Friederici 2002; Lebib, Papo et al. 2004). Finally, cross-language studies, very important for the present research, indicate that the characteristics of languages as well the experience with a language impacts the degree to which a person relies upon visual cues in interpreting dependency (Sekiyama and Tohkura 1993; Sekiyama 1997; Gentilucci and Cattaneo 2005; Sekiyama and Burnham 2008). Therefore the McGurk and semantic influences on auditory inputs might be modulated by different experience with the language in which the McGurk effect is presented.

## CHAPTER II

### HYPOTHESES

#### 2.1 Experiment 1

1). Combined incongruent Audiovisual (AV) input (McGurk effect) will impair the accuracy of speech perception relative to Audio-Only input when subjects are instructed to respond to the auditory stimuli (McGurk and MacDonald 1976; MacDonald and McGurk 1978; Sekiyama and Burnham 2008).

2). The McGurk effect will have a negative impact on the reaction time (RT) relative to the Audio-Only condition and the audiovisual condition (van Wassenhove, Grant et al. 2005; Sekiyama and Burnham 2008).

3). The ERP indicator of the McGurk effect: the McGurk effect would generate a larger amplitude from the standard condition for the N1 component (peak amplitude occurring between 50 ms to 150 ms), P2 (peak amplitude occurring between 150 ms to 250 ms), N300 (peak amplitude occurring between 250 ms to 350 ms) and later component (a slow wave occurring between 350 to 600 ms) recorded at frontal, central and parietal regions (Colin, Radeau et al. 2002; Lebib, Papo et al. 2004; van Wassenhove, Grant et al. 2005; Saint-Amour, De Sanctis et al. 2007; Stekelenburg and Vroomen 2007). This expectation was based on previous ERP studies that compared the McGurk effect to the standard condition.

4). According to the behavior indicator of the McGurk effect, the degree of the McGurk effect will be largest in the group of native English speakers, followed by the Chinese participants who have been in United States of America more than three years. The effect will be least pronounced in Chinese participants who have been in the United States of America less than three years (Sekiyama 1997; Sekiyama and Burnham 2008).

5). The impact on the reaction time for the McGurk condition compared with the Audio-Only condition and the congruent audiovisual condition will be in the same order as described for the McGurk effect in hypothesis 4 (Sekiyama 1997; Sekiyama and Burnham 2008).

6). The McGurk markers found in Hypothesis 4 should be greatest in American English group, followed by Chinese group who has been in the United States for longer than three years. The Chinese group who has been in the states for less than three years would have the lowest peak amplitudes in these markers. (Sekiyama 1997; van Wassenhove, Grant et al. 2005; Sekiyama and Burnham 2008).

## **2.2 Experiment 2 and 3**

1). For the Chinese LANGUAGE, the Chinese Groups (Group 1 and Group 2) would generate a higher McGurk rate than the American Group (Group 3); for the English LANGUAGE, the American Group (Group 3) would produce the highest McGurk rate; a lower would be generated by the Chinese Group that resided in the United States of America for longer than three years (Group 2); and the lowest rate by the Chinese Group that was in the United States of America for less than three years (Group 1) have the least degree of the McGurk effects (Sekiyama 1997; Sekiyama and Burnham 2008).

2). The mixed stimuli that have a fused perception of word would induce a higher McGurk rate than the mixed stimuli that have a fused perception of non-words. The word/non-word effect would affect the Chinese groups in the Chinese LANGUAGE and would affect all three groups in the English LANGUAGE with the same order of degrees as in hypothesis 1: the American Group (Group 3), the Chinese Group who had been in the United States of America for longer than three years (Group 2), and the Chinese Group who had been in the United States of America for less than three years (Group 1) (Brancazio and Brancazio 2004).

3). This hypothesis tests the McGurk lexicon interaction in Chinese stimuli in Chinese groups and American group. If the auditory source of the McGurk condition is a word (such as /bench/) and the visual source generates a nonword (such as /dench/), the McGurk nonword condition would have a more negative peak around 400 ms (N400 with the peak latency during 350 to 450 ms) than the audiovisual word condition (such as audiovisual /bench/). This should only have happened in the Chinese groups (Group 1 and 2) when the language is Chinese (Kutas and Hillyard 1980; Kutas and Hillyard 1984; Friederici 2002).

4). This hypothesis tests the McGurk lexicon interaction in English stimuli in three Experimental groups. If the auditory source of the McGurk condition is a word (such as /bench/) and the visual source generates a nonword (such as /dench/), the McGurk nonword condition would elicit a more negative peak around 400 ms (N400 with the peak latency during 350 to 450 ms) than the audiovisual word condition (such as audiovisual /bench/). This effect would occur in all three groups when the language is English, but the degree of the negativity would be the largest in the American group

(Group 3), followed by the Chinese group who have been in America for more than three years (Group 2); and the smallest in the Chinese group who have been in the United States of America for less than three years (Group 1) (Kutas and Hillyard 1980; Kutas and Hillyard 1984; Sekiyama and Tohkura 1993; Sekiyama 1997; Friederici 2002; Sekiyama and Burnham 2008).

## CHAPTER III

### METHODS

#### 3.1 Participants

Thirty-six subjects were tested for this study (17 females, mean age 27.06 yrs, s.d. = 3.80 yrs., 19 males, mean age = 29.42 yrs., s.d. = 5.00 yrs.). All Chinese participants were recruited through the CSSA (Chinese Student and Scholar Association) email list of University of Louisville. Three groups of subjects participated. Group one included 12 native Chinese (6 females) speakers who had been in the US for no more than three years; group two included 12 native Chinese speakers (6 females) who had been in the US for more than three years; and group three included 12 native English speakers (5 females) recruited from University of Louisville. All Chinese participants had at least six years of standard education of English. Of the American participants, one had one year of experience in learning Chinese while the Chinese LANGUAGE was novel to the remaining participants. Gender was matched across the three groups. All participants were screened to ensure that they had no visual and auditory ability defects using a Snellen vision chart and a standard auditory air conduction-screening test. In addition, all were right handed as determined by the Edinburgh Handedness Battery (Oldfield, 1978). All participants signed an informed consent approved by the University of Louisville IRB committee and received a \$20.00 honorarium for their participation.

## 3.2 Stimulus

Audiovisual stimuli were recorded simultaneously from the natural articulations of a Caucasian male English speaker using a SONY DCR-PC100 digital camcorder (frame rate: 30 images/s; audio sample rate: 44.1 KHz at 16 bits). Prior to the recording session, the speaker practiced the articulation of syllables, words and non-words for Experiment 1 and 3. In addition, the speaker practiced for two hours to pronounce accurately the Chinese words and non-words for Experiment 2. The speaker's accuracy and intelligibility of pronunciation of the Chinese stimuli were verified by a native Chinese speaker. The recording camera only focused on the portion of the face below the eyes. The visual stimuli were edited using iMovie 5.0.2 ©. The auditory stimuli were edited using the Amadeus II 3.8.7 © software package. In order to construct the final movies for the experiments, the movies of various articulations and vocal movements were cut into small one-second segments based upon the mouth opening. Multiple segments were studied in order to determine an auditory lag time (time between visual input and the later auditory input) that would fit every syllable and sound naturally when dubbed with different visual stimulus. A delay of 200 ms was chosen, which was the observed average delay between the audio and visual onset of the recorded clips. The exact time of auditory onset was determined by a rapid increase in the amplitude of the sound wave. In later editing, the initial, sub-threshold slow amplitude increase in the early portion of the auditory waveform was deleted to ensure a clear auditory onset while avoiding any click noise that might occur if the transition was too abrupt. The auditory onset of each sound was marked in the movie; a visual onset was set at 200 ms before the

auditory onset and lasted for 1 second. The duration of all auditory stimuli was 500 ms. By this way, the auditory onset time relative to the visual onset time was the same even though the mouth articulations of different syllables were not. Finally, from multiple versions of each movie clip, the ones that had similar mouth opening times were chosen so that when the auditory information was changed from congruent to incongruent, there was no abnormal time lag or confound between lip movements and pronunciations. Software TagMovieX 1.2 [developed by Mr. Ron Crans, Department of Psychology, Colgate University] was used to combine the audio and visual stimuli. To create an Audiovisual dubbed stimulus (AbVg), the video clip that contained the original matching Audiovisual information (AgVg) was opened in TagMovieX. Both the audio (Ag) and visual (Vg) tracks were viewed simultaneously. The old audio track (Ag) was then deleted; and the designed audio track (Ab) added at the desired offset (200ms).

### **3.3 Procedure**

Each participant was tested individually in a sound dampened room at the University of Louisville, Developmental Neuroscience Laboratory. At the start of a test session and while adapting to the room, each participant read and signed an informed consent form. Next, their hearing and vision were measured.

Throughout testing, each participant sat in a room facing a CRT display. A speaker was positioned 1 meter above the center of the head. Before applying the electrophysiological net, the participant's head was measured. Using our standard application procedure, the 256 HydroCel net was immersed for 5 minutes in a saline solution containing 1.5 teaspoons of potassium-chloride warmed to body temperature. The electrode net was placed on the head following measurement and marking of Cz

(vertex) and nasion electrode positions to aid in the proper alignment of the net on the participant's head. Scalp impedances were measured using Net Station 4.2 and adjusted to under 40 kiloOhms prior to the start of the test. During subsequent testing, the ERP signals were digitized at 4 ms intervals over a one-second period for each of the 256 electrode sites. Filter settings were maintained at 0.1 Hz for the high pass and 100 Hz for the low pass with a gain setting of 10K. The level of the sound was adjusted to 75 dB (A). The visual stimuli were presented at the center of a 17-inch computer CRT monitor that was positioned in front of and centered in the middle of participants' visual field at a distance of 36" (Dell, 800 x 600 pixels, 75Hz). The visual angle for the visual stimuli was 8.7° x 8.7°.

An experimenter instructed the participant to concentrate on the stimulus presentations of both the auditory and visual stimuli. They also were told to press different buttons (1, 2, or 4) in response to the different consonant sounds, /b/, /d/, /g/. The order of the response buttons was counterbalanced across participants. A total of 330 trials were presented in Experiment 1 and 360 trials each in Experiments 2 and 3. Each experiment lasted approximately 20 minutes. The details of the experimental design are described in the section on Experiment Design (session 3.5).

Once the electrophysiological recording session ended, Geodesic Photogrammetry System (GPS) recordings were obtained in order to record the exact location of each electrode on the participant's head. These scalp electrode locations were then saved for subsequent use to align electrodes across subjects, thereby reducing some of the variability in electrode placement across participants.

In addition to the electrophysiological measure, a basic information questionnaire

(BIQ) and the PPVT (Peabody Picture Vocabulary Test Third edition, PPVT-III) (Dunn and Dunn 1997) were administered to all participants either before or after the test session. The BIQ requested information concerning the length of time that the subject had been in US and the scores of the subjects on the standard tests (TOEFL and GRE). The PPVT was used to evaluate English receptive word knowledge.

### **3.4 Equipment**

The EEG data were acquired using a 256 HydroCel net and sampled at a 250 Hz digitizing rate with the low pass filter at 100Hz. The Geodesic Sensor Net consisted of electrodes positioned in a geodesic configuration interconnected with elastic connectors between nylon pedestals spaced evenly across the head surface. These pedestals contained a silver chloride coated carbon pellet within an electrolytic sponge in each pedestal that, in turn, was connected via carbon wiring to an electrode holder. Each electrode was connected via a shielded cable to its own high-impedance low-noise physiological amplifier that provided fast Fourier transformation and analogue-to-digital conversion of the EEG signals obtained during a testing session. The low-power recording circuits were insulated from any higher power electrical circuits by optical coupled devices that ensured the safety of each participant. Quick Time movie clips were presented during data collection. Software StimTaskX 1.2 [developed by Ron Crans, Department of Psychology, Golgate University] enabled stimulus presentation and data collection to occur at the same time on the same computer, which was connected with two monitors. The video signal was split so that one monitor could serve for stimulus presentation and the other one for data monitoring/recording. Cz served as the reference during data acquisition. All data were collected using Net Station 4.2 on a Macintosh

PowerPC G5 running system O.S.X (10.4).

### **3.5 Experiment Design**

#### **3.5.1 Experiment 1: Consonant identification of syllables**

The experiment contained 11 different stimuli. For each syllable of /ba/, /da/ and /ga/, three types of stimuli were created: (1) visual only, (2) audio with still picture and (3) Audiovisual congruent. The symbol /AbVb/ is used to represent a stimulus with an auditory /ba/ dubbed onto a visual movie depicting the actor producing the syllable /ba/; /Ab/ represents the stimulus of an auditory /ba/ dubbed onto a movie of a closed-mouth, inactive face. /Vb/ represents a movie clip of lip movement of /ba/ but with no sound. In addition to the nine stimuli, two additional stimuli were included in the incongruent condition representing the McGurk effect. An auditory /ba/ was dubbed onto a visual movie of an actor producing the syllables, /ga/ and /da/, respectively. In summary, the 11 stimuli were as follow: /AbVb/, /Ab/, /Vb/, /AdVd/, /Ad/, /Vd/, /AgVg/, /Ag/, /Vg/, /AbVd/, /AbVg/. Stimuli were presented in a random order with each one presented 30 times. The inter-stimulus interval ranged from 1.9 to 2.4 seconds.

During testing, participants were asked to make a three-alternative forced choice decision when a stimulus appeared on the screen. Participants needed to decide whether the syllable started with the /b/, /d/ or /g/ consonant based on their perception.

#### **3.5.2 Experiment 2: Consonant identification of Chinese words/non-words**

A series of Audiovisual stimuli was presented in a random order to each participant. The incidence rates were all controlled for three factors: word versus non-

words, consonant /b/ versus consonant /d/, and the interaction of word/non-word with consonants.

1. One set of Chinese non-word/word stimuli included: congruent Audiovisual /be2/, congruent Audiovisual /de2/ and simultaneous audio /be2/ and visual /de2/ (would produce fused perception of /de2/). Each occurred 30 times. In this case, the auditory influence (/b/ sound) formed a non-word and the visual influence (/d/) formed a word.

Every stimulus in Experiment 2 and 3 was coded using 4 letters. The first letter recognized whether it was a Chinese or an English stimulus, by using letter 'c' or letter 'e'. The second letter informed about whether the auditory influence (/b/ sound stimulus) formed a word ('w') or a non-word ('n'). The third letter had the information about whether the visual influence (/d/ sound stimulus) formed a word ('w') or a non-word ('n'). The fourth letter represented whether it was a /b/ sound ('b') or /d/ sound ('d') or mixed one ('m') with auditory /b/ stimulus and visual /d/ stimulus. Thus, the first set of stimuli was represented by the acronyms: cnwb, cnwd and cnwm.

2. A second set of stimuli included: congruent Audiovisual /bie4/, congruent Audiovisual /die4/ and simultaneous audio /bie4/ and visual /die4/ (would produce fused perception of /die4/), and each occurred 30 times. In this case, the auditory influence /b/ formed a word and the visual influence /d/ formed a non-word. The second set of stimuli was represented by the acronyms: cwnb, cwnd and cwnm.

3. A third set of stimuli consisted of the following elements: congruent Audiovisual /bie1/, congruent Audiovisual /die1/ and simultaneous audio /bie1/ and visual /die1/ (would produce fused perception of /die1/), and each occurred 30 times. In this case, both the auditory influence /b/ and the visual influence /d/ sounds formed

words. The third set of stimuli was represented by the acronyms: cwwb, cwwd and cwwm.

4. The last set included: congruent Audiovisual /biong3/, congruent Audiovisual /diong3/ and simultaneous audio /biong3/ and visual /diong3/ (would produce fused perception of /diong3/), and each occurred 30 times. In this case, both the auditory influence /b/ and the visual influence /d/ formed non-words. The fourth set of stimuli was represented by the acronyms: cnnb, cnnd and cnnm.

### **3.5.3 Experiment 3: Consonant identification of English words/nonwords**

Experiment 3 was structured similar to Experiment 2, except it used English words/non-words. The conditions were as followed:

1. The first set of stimuli used in this experiment included the following: congruent Audiovisual /bish/, congruent Audiovisual /dish/ and simultaneous audio /bish/ and visual /dish/ (would produce fused perception of /dish/) and each occurred 30 times. In this case, the auditory influence /b/ sound formed a non-word while the visual influence /d/ sound formed a word. The first set of stimuli could be represented by the acronyms: enwb, enwd and enwm.

2. A second set of stimuli used in this experiment included: congruent Audiovisual /bench/, congruent Audiovisual /dench/ and simultaneous audio /bench/ and visual /dench/ (would produce fused perception of /dench/), and each occurred 30 times. In this case, the auditory influence /b/ sound formed a word while the visual influence /d/ sound formed a non-word. The second stimuli were represented by the acronyms: ewnb, ewnd and ewnm.

3. A third set of stimuli included: congruent Audiovisual /belt/, congruent

Audiovisual /delt/, simultaneous audio /belt/ and visual /delt/ (would produce fused perception of /delt/), and each occurred 30 times. In this case, both the auditory and visual influences (/b/ sound and /d/ sound) formed words. The third set of stimuli was represented by the acronyms: ewwb, ewwd and ewwm.

4. A fourth set of stimuli included the following: congruent Audiovisual /barve/, congruent Audiovisual /darve/, simultaneous audio /barve/ and visual /darve/ (would produce fused perception of /darve/), and each occurred 30 times. In this case, both the auditory and visual influences (/b/ sound and /d/ sound) formed non-words. The last set of stimuli was represented by the acronyms: ennb, ennd and ennm.

## CHAPTER IV

### RESULTS

#### 4.1 Experiment 1

##### 4.1.1 Results for Hypothesis 1

Hypothesis 1: It was hypothesized that the combined incongruent Audiovisual (AV) input (McGurk effect) would impair the accuracy of speech perception relative to Audio-Only input when subjects were instructed to respond to the correct answer (McGurk and MacDonald 1976; MacDonald and McGurk 1978; Sekiyama and Burnham 2008).

Based on the hypothesis, the following analysis was conducted. For all repeated measures ANOVAs presented in the current study, Greenhouse-Geisser corrections were employed.

In testing the McGurk effect, the auditory stimulus was defined as the correct answer. Consequently, if the response elicited during the McGurk condition was different from its auditory source, it was judged as incorrect. The button response of the McGurk conditions (AbVg) and the Audio-Only conditions (Ab) were collected for each subject. Each response was marked as inaccurate if it was different from that of the audio stimulus.

To address this hypothesis, paired samples t-tests (McGurk /AbVg/ vs. Audio-

Only /Ab/) of the McGurk effect and Accuracy interaction identified a significant difference of the two conditions:  $t = -8.380\%$ ,  $df = 35$ ,  $p < .001$ , Std.Error Mean = 6.166%, Mean Difference = - 51.67% (the mean for McGurk is 37.13%, the mean for the Audio-Only condition is 88.80%). The negative t-value indicated that the accuracy for the McGurk condition was significantly less than the Audio-Only condition. Therefore, the hypothesis was supported.

#### **4.1.2 Results for Hypothesis 2**

Hypothesis 2: it was hypothesized that the McGurk effect would have a negative impact on the reaction time (RT) relative to the Audio-Only condition and the audiovisual condition (van Wassenhove, Grant et al. 2005; Sekiyama and Burnham 2008).

For all subjects, the reaction time (RT) data were collected and averaged for each of the three conditions: the McGurk condition (AbVg), the Audio-Only conditions (Ab) and the Audiovisual condition (AbVb) for the /ba/ sound. A repeated measures ANOVA was conducted on the RT using the following model: CONDITIONS (3: McGurk /AbVg/ vs. Audio-Only /Ab/ vs. Audiovisual /AbVb/).

A significant effect in the main condition was found:  $F(2, 70) = 18.366$ ,  $p < .001$ , Observed Power  $> .999$ . Pairwise comparisons indicated that the McGurk condition produced a significantly longer reaction time than the Audio-Only condition, while the Audio-Only condition had produced longer reaction times than the Audiovisual condition (as listed in Table 1 and illustrated in Figure 1). Thus the hypothesis was supported.

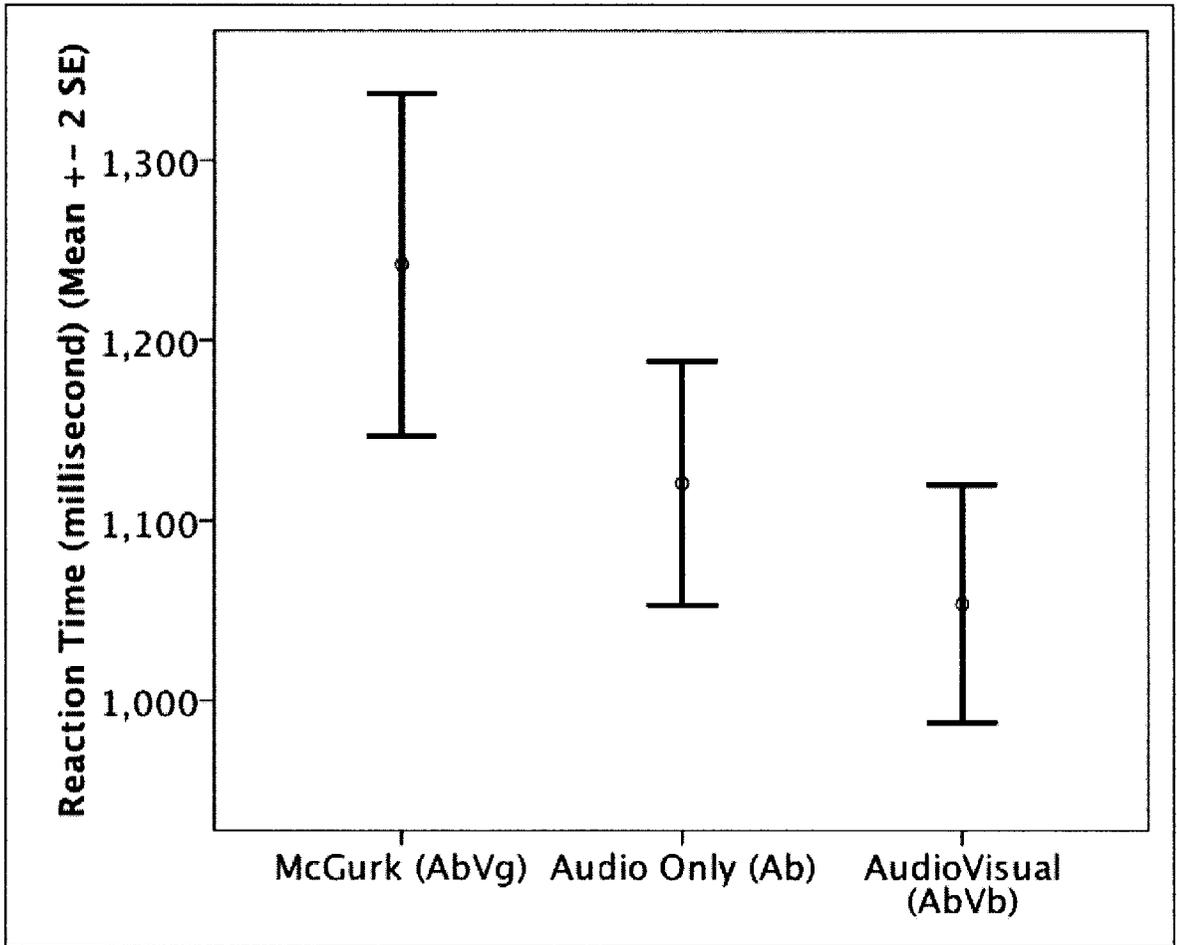
**Table 1**Pairwise Comparisons of Reaction Time (ms)

		Mean	Std. Error	Sig.	95% Confidence Interval for Difference	
		Difference			Lower Bound	Upper Bound
(I) COND	(J) COND	(I-J)				
McGurk	Audio-Only	121.466	38.319	.009	25.384	217.548
	Audiovisual	188.352	35.092	.000	100.361	276.343
Audio-Only	McGurk	-121.466	38.319	.009	-217.548	-25.384
	Audiovisual	66.886	16.698	.001	25.017	108.755
Audiovisual	McGurk	-188.352	35.092	.000	-276.343	-100.361
	Audio-Only	-66.886	16.698	.001	-108.755	-25.017

Based on estimated marginal means

\* The mean difference is significant at the .05 level.

a Adjustment for multiple comparisons: Sidak.



**Figure 1.** Impact of the McGurk effect on Reaction Time. The McGurk condition produced significantly longer reaction times than the Audio-Only condition; the Audio-Only condition produced significantly longer reaction times than the standard Audiovisual condition. The error bar represents two standard errors.

**4.1.3 Results for Hypothesis 3**

Hypothesis 3: It was hypothesized that the McGurk effect would generate a larger amplitude than the standard condition for the N1 (peak amplitude occurring between 50 ms to 150 ms), P2 (peak amplitude occurring between 150 ms to 250 ms), N300 (peak amplitude occurring between 250 ms to 350 ms) and a later component (a slow wave

occurring between 350 to 600 ms) recorded at frontal, central and parietal regions (Colin, Radeau et al. 2002; Lebib, Papo et al. 2004; van Wassenhove, Grant et al. 2005; Saint-Amour, De Sanctis et al. 2007; Stekelenburg and Vroomen 2007).

#### **4.1.3.1 ERP data pre-processing**

The recorded ERP data were first low pass filtered at 30 Hz. Next, the data were segmented from 100 ms pre to 700 ms post the onset of the auditory stimulus. The next step involved artifact detection and rejection. Segments with eye blinks or eye movements (eye channels containing recordings over 100 uV) or that contained more than 15 bad channels (i.e., channels that have over 90% of recordings over 150 uV) were excluded from the final average. Following this step, the bad channels were replaced through a standard interpolation process and an average reference applied to the data. The data were then baseline corrected based on the 100 ms interval immediately preceding stimulus onset. Finally, the data were averaged together by averaging together segments from the same category.

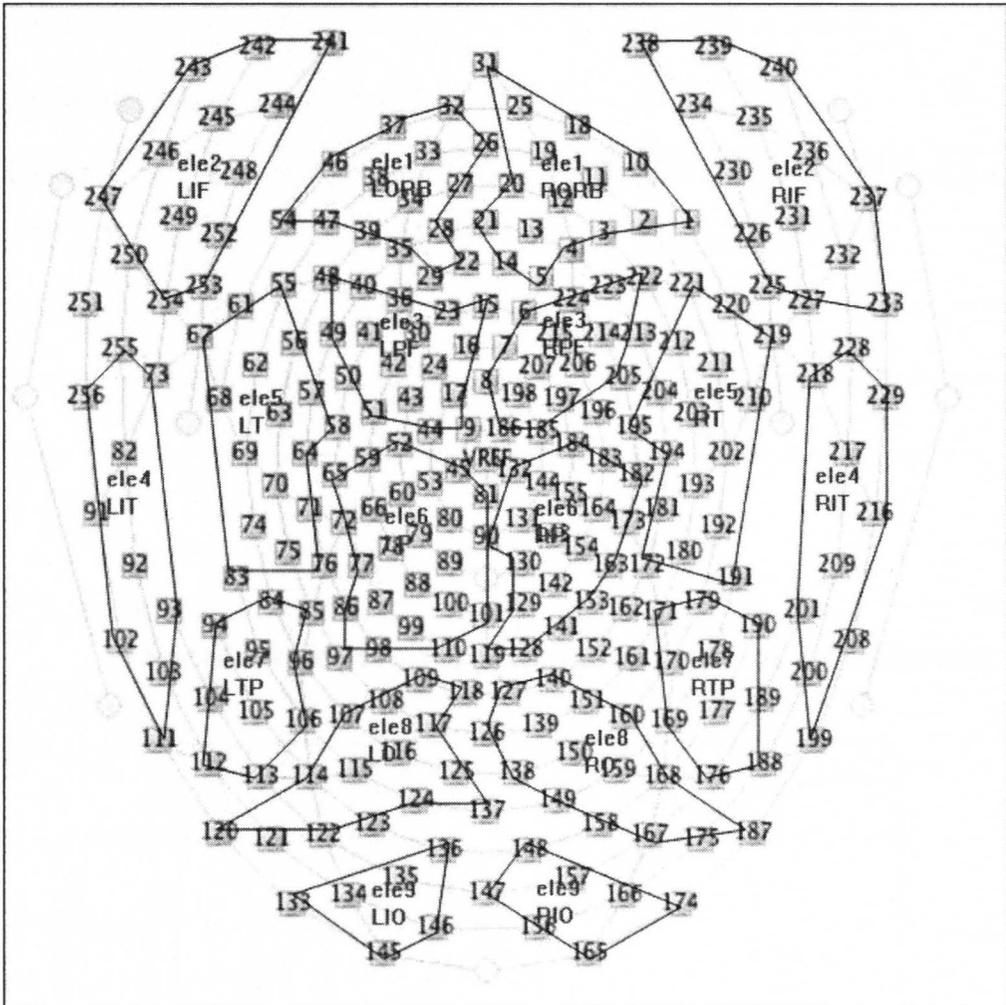
Two of 36 subjects (one from Group 1 and one from Group 2) were excluded from the ERP analysis due to artifacts. As a consequence, 34 participants were included in the final ERP analysis for Experiment 1.

Averaged ERPs for each condition and each electrode channel were obtained for each participant. The ERPs of the McGurk condition (AbVg) and the ERPs of the audiovisual congruent condition (AbVb) were submitted to further statistic analysis. Data from different electrode regions on the scalp were grouped into 9 electrode regions for each hemisphere based on distinct scalp anatomical regions (Orbitofrontal, Inferior Frontal, Posterior Frontal, Inferior Temporal, Temporal, Parietal, Temporal-Parietal,

Occipital and Inferior Occipital, abbreviated as ORB, IF, PF, IT, T, P, TP, O, IO) as illustrated in Figure 2.

#### **4.1.3.2 ERP data analysis: factor analysis and repeated measures ANOVA.**

The original ERP data included 175 data points (digitized at 250 Hz over a 700 ms interval) and the adjacent data points are likely to be highly correlated. To account for the correlation between adjacent data points, a Factor analysis procedure that employed a Varimax rotation was applied to the time dimension of the entire dataset. In this way, 175 data points are reduced to a few independent temporal regions that accounted for the majority of total variance in the data set. An effective time and peak latency was generated for each resulted factor. The effective time was defined as the time period during which the particular factor has a higher than 0.4 correlation (the y-axis in Figure 3). Peak latency was defined as the time point when the particular factor has the highest correlation. For each factor, the averaged amplitude during that time was calculated for each electrode cluster, condition and subject. The effective region of each factor was determined by choosing the none-overlapping portion of the factors.



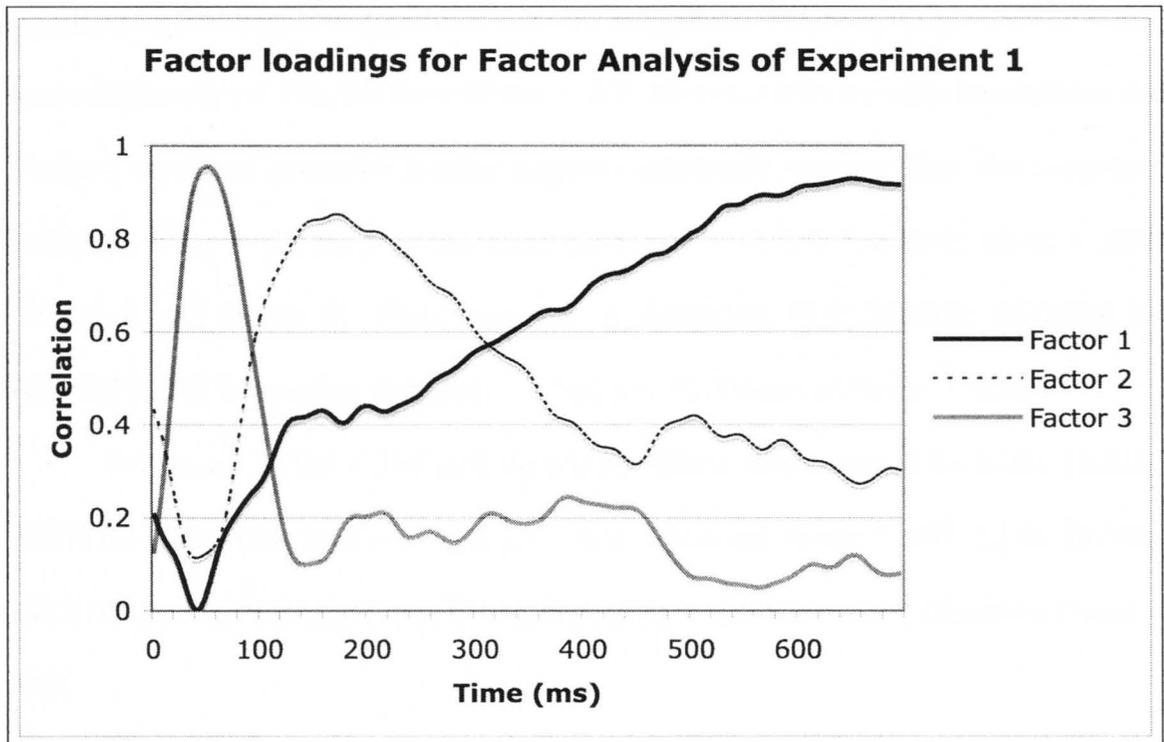
**Figure 2.** Clustering of adjacent scalp electrode channels for subsequent data analysis. 256 channels were averaged into 9 electrode regions (Orbitofrontal, Inferior Frontal, Posterior Frontal, Inferior Temporal, Temporal, Parietal, Temporal-Parietal, Occipital and Inferior Occipital) for each of the two hemispheres (Left and Right).

The results of Factor analysis results are illustrated in Table 2 and Figure 3.

**Table 2**

Factor analysis for Experiment 1 (n = 36)

Factor	Variance accounted (%)	Effective Time (ms)	Peak Latency (ms)
1	74.33	320 - 700	656
2	9.49	100 - 316	176
3	5.00	20 - 96	56



**Figure 3.** Plots of the Factor loadings for Experiment 1. Factor 1 was a slow wave with the effective time range from 320 to the end of the segment and the highest amplitude at 656 ms; Factor 2 was the second peak with the effective time range from 100 to 316 ms, with the highest amplitude at 176 ms; Factor 3 characterized the first peak effective between 20 and 96 ms, with the highest amplitude at 56 ms.

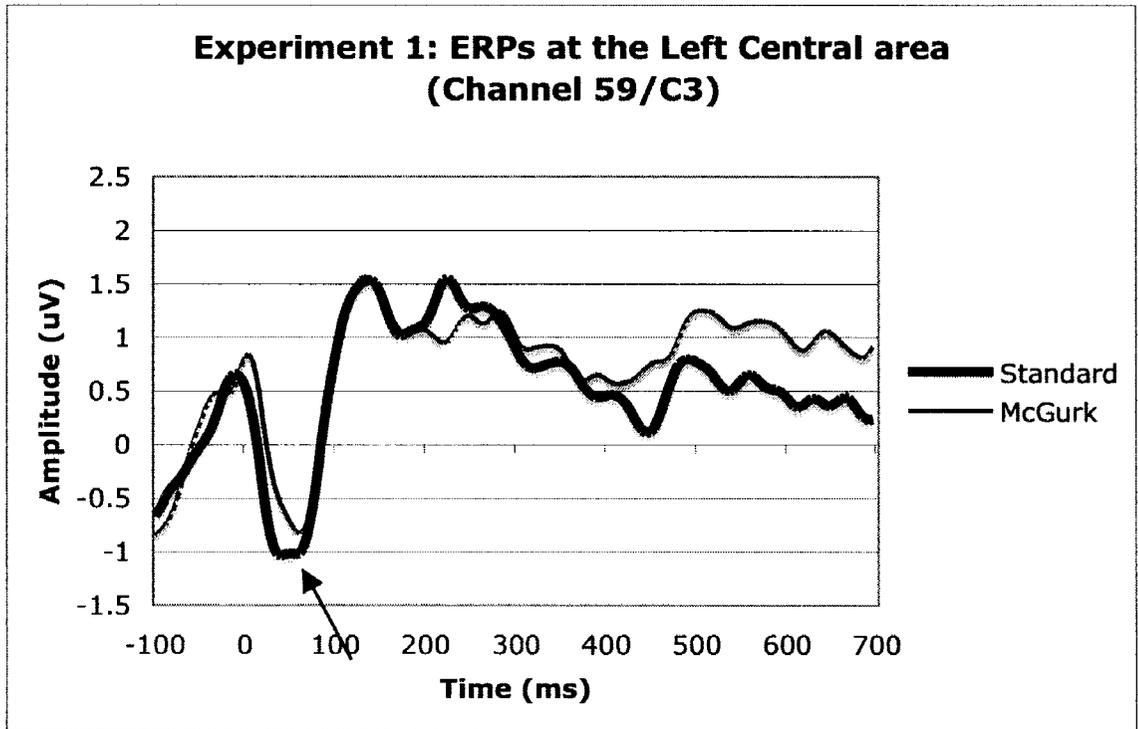
ERPs for each condition, cluster and subject were averaged on the effective time range for each factor. The averaged amplitudes served as the dependent measure in separate repeated measures ANOVAs for each factor. The repeated measures ANOVA followed the design: ELECTRODE REGIONS (9) x HEMISPHERES (2: left vs. right) x CONDITIONS (2: McGurk /AbVg/ vs. congruent /AbVb/).

For Factor 3 (20 - 96 ms), a significant HEMISPHERE by CONDITION interaction occurred,  $F(1, 33) = 5.777, p < .05$ , Observed Power = .645. Paired samples T-Test indicated that over the left hemisphere, the McGurk condition generated significantly less negative amplitudes than the congruent condition,  $t(33) = .397, p < .05$ , mean difference = 1.778, Std Error Mean = .742. However over the right hemisphere, the McGurk condition generated a more negative amplitude response than the congruent condition,  $t(33) = -2.136, p < .05$ , mean difference = -1.420, Std Error Mean = .664. (Figure 4 and Figure 5). There was also a significant ELECTRODE REGION by HEMISPHERE interaction,  $F(8, 264) = 4.284, p < .01$ , Observed Power = .854.

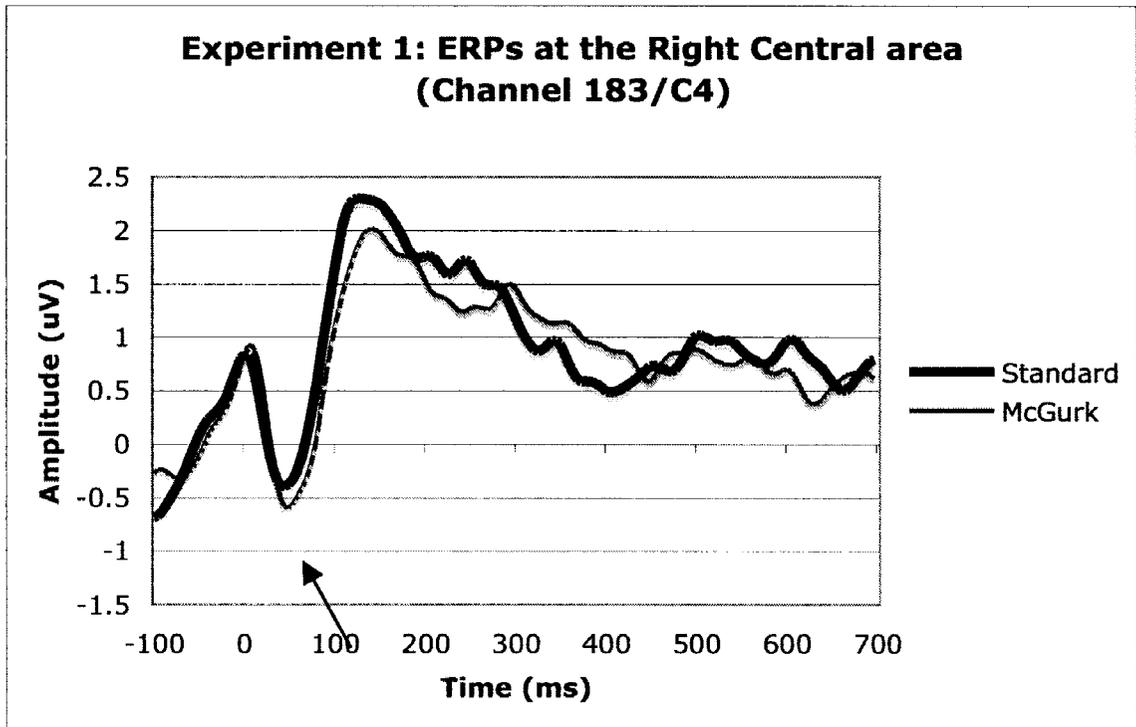
For Factor 2 (100 – 316 ms), significant effects only occurred for ELECTRODE REGION effect,  $F(8, 264) = 17.313, p < .001$ , Observed Power > .999; ELECTRODE REGION by HEMISPHERE interaction,  $F(8, 264) = 5.845, p < .001$ , Observed Power = .949.

In Factor 1 (320 – 700 ms), only a significant ELECTRODE REGION effect was found,  $F(8, 264) = 21.764, p < .001$ , Observed Power > .999.

The hypothesis was partially supported for the early component N1 (Factor 3), but was not supported for other components.



**Figure 4.** ERPs of Experiment 1 over the left central area. The figure shows the brainwaves over the left central area, channel 59, comparable to C3 in a 10-10 system. The black line represents the standard /AbVb/ condition and the gray line represents the McGurk condition. The arrow points to the first negative peak (N1) occurred between 20 and 96 ms. The McGurk condition produced a smaller negative amplitude response than the Standard condition during this time period.



**Figure 5.** ERPs of Experiment 1 over the right central area. The figure shows the brainwaves over the right central area, channel 183, comparable to C4 in a 10-10 system. The black line represents the standard /AbVb/ condition and the gray line represents the McGurk condition. The arrow points to the first negative peak (N1) occurred between 20 and 96 ms. The McGurk condition generated a larger negative peak amplitude than the Standard condition.

#### 4.1.4 Hypothesis 4

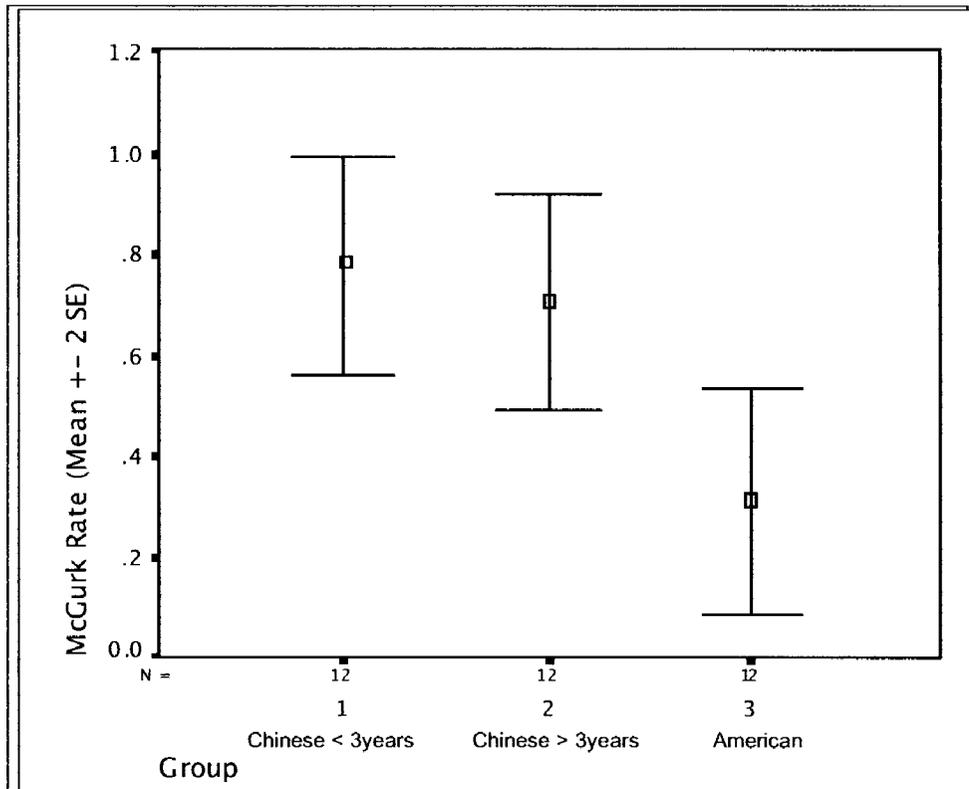
Hypothesis 4: It was hypothesized that the degree of the McGurk effect would be largest in the group of native English speakers, followed by the Chinese participants who have been in United States of America more than three years. The effect was expected to be least pronounced in Chinese participants who have been in the United States of America less than three years (Sekiyama 1997; Sekiyama and Burnham 2008).

Before the analysis of McGurk effect, the PPVT scores were calculated for each group. Because of scheduling conflicts, a total of 7 participants (3 of them belonged to Group 1; 3 of them belonged to Group 2 and 1 belonged to Group 3) did not complete the PPVT exam and their scores were not included in the PPVT comparison. The results indicated that Group 1 Chinese had a mean PPVT score of 65.56, Std.Error = 5.660; Group 2 Chinese had a mean score of PPVT of 82.78, Std.Error = 3.692 and Group 3 American English speaker had a mean PPVT score of 116.18, Std.Error = 3.413. A One-way ANOVA indicated that there was a significant group effect,  $F(2, 26) = 38.087, p < .001$ . Scheffe post hoc tests indicated that there were significant group differences between each pair-wise comparison, Group 1 differed from Group 2: mean difference = -17.22,  $p < .05$ ; Group 2 differed from Group 3: mean difference = -33.44,  $p < .001$ ; Group 1 differed from Group 3: mean difference = -50.66,  $p < .001$ . So the PPVT score was consistent with the length of time spent in the United States of America in that high scores were correlated with longer residency in the United States of America.

The rate of the McGurk effect was an indicator of the degree that the McGurk effect occurred. Similar to that in Experiment 1, in Experiment 2 and 3, the McGurk rate was calculated by subtracting the accuracy of the McGurk condition (AbVg) from that of the congruent condition (AbVb) (Sekiyama & Burnham, 2008).

A One-way ANOVA was conducted on the McGurk rate in the three groups (including all the 36 subjects). There was a significant group difference,  $F(2,33) = 5.296, p < .05$ . Scheffe post hoc tests indicated that Group 1 had a significantly higher McGurk rate than Group 3 (Mean Difference = .467,  $p < .05$ ). However, there was no significant difference between Group 1 and Group 2 (Mean Difference = .075, n.s.) or between

Group 2 and Group 3 (Mean Difference = .392, n.s.). Overall, these results did not support the hypothesis. The means and the standard error of the means of each group are plotted in Figure 6.



**Figure 6.** The mean and the standard error of the McGurk effect in Experimental Groups. The x-axis lists three groups with 12 subjects in each group; the y-axis plots the mean and the standard error of the McGurk rate for each group. As indicated in the text, Group 3 produced a lower McGurk rate than Group 1. Group 1 represents the Chinese group who has less than three years of residence in the states; Group 2 represents the Chinese group who has more than three years of residency in the states; Group 3 included the native American English speakers.

#### **4.1.5 Hypothesis 5**

Hypothesis 5: it was hypothesized that the impact on the reaction time for the McGurk condition compared with the Audio-Only condition and the congruent audiovisual condition would be in the same order as described for the McGurk effect in hypothesis 4 (Sekiyama 1997; Sekiyama and Burnham 2008).

A repeated measures ANOVA was applied to the reaction time data using the model: CONDITIONS (3: McGurk /AbVg/ vs. Audio-Only /Ab/ vs. Congruent /AbVb/) x GROUPS (3: Group 1 vs. Group 2 vs. Group 3). Post-hoc tests (Scheffe) were conducted when there were significant GROUP effects. T-tests were used to assess significant GROUP by CONDITION interactions.

A significant CONDITION effect was found,  $F(2, 60) = 17.453, p < .001$ , Observed Power  $> .999$ . It was the same effect as tested in Hypothesis 2, session 4.1.2. The GROUP by CONDITION interaction was not significant, with an observed power .240. Thus the hypothesis was not supported.

#### **4.1.6 Hypothesis 6**

Hypothesis 6: it was hypothesized that the McGurk markers found in Hypothesis 4 were expected to be the most pronounced in the American English speakers, followed by Chinese group who has been in the United States of America for longer than three years. The Chinese group who has been in the states for less than three years would generate the smallest peak amplitudes in these markers. (Sekiyama 1997; van Wassenhove, Grant et al. 2005; Sekiyama and Burnham 2008). To address this hypothesis, a different repeated measures ANOVA design was applied to the data that

involved ELECTRODE REGIONS (9) x HEMISPHERES (2: left vs. right) x CONDITIONS (2: McGurk /AbVg/ vs. Congruent /AbVb/) x GROUPS (3: Group 1 vs. Group 2 vs. Group 3). Post hoc tests were conducted using T-Tests.

During the time period of Factor 3 (20 – 96 ms), a significant ELECTRODE REGION by HEMISPHERE interaction was noted,  $F(8, 224) = 3.631, p < .05$ , Observed Power = .762; a significant HEMISPHERE by CONDITION interaction was also found,  $F(1, 28) = 5.615, p < .05$ , Observed Power = .629. No GROUP related effects were found.

During the time period of Factor 2 (100 – 316 ms), a significant ELECTRODE REGION by CONDITION by GENDER by GROUP interaction occurred,  $F(16, 224) = 2,878, p < .05$ , Observed Power = .789. To conduct the Post-hoc test, the ERP data were averaged together across hemispheres. In each ELECTRODE REGION, CONDITION and GENDER, the GROUP effect was tested using One-way ANOVAs. However, no significant result was found.

During the same time period, a significant ELECTRODE REGION effect occurred,  $F(8, 224) = 16.769, p < .001$ , Observed Power = .999; as well as a significant ELECTRODE REGION by GENDER interaction,  $F(8, 224) = 3.664, p < .05$ , Observed Power = .653; a significant ELECTRODE REGION by HEMISPHERE interaction,  $F(8, 224) = 5.265, p < .01$ , Observed Power = .914; and a significant GENDER effect,  $F(1, 28) = 7.948, p < .01$ , Observed Power = .777.

During the time period of Factor 1 (320 – 700 ms), a significant ELECTRODE REGION effect occurred,  $F(8, 224) = 20.910, p < .001$ , Observed Power > .999. No GROUP related effects were found.

The summary of these results indicated that the hypothesis was not supported.

## **4.2 Experiment 2 and 3 Results**

### **4.2.1 Results for Hypothesis 1 and 2**

Hypothesis 1: it was hypothesized that for the Chinese LANGUAGE, the Chinese Groups (Group 1 and Group 2) would have a higher McGurk rate than the American Group (Group 3); In the English LANGUAGE, the American Group (Group 3) was expected to produce the highest McGurk rate; followed by the Chinese Group who had been in the United States of America for longer than three years (Group 2). The Chinese Group who had been in the United States of America for less than three years (Group 1) were expected to produce the least degree of the McGurk effect (Sekiyama 1997; Sekiyama and Burnham 2008).

Hypothesis 2: it was hypothesized that the mixed stimuli that had a fused perception of words would induce a higher McGurk rate than the mixed stimuli that had a fused perception of non-words. The word/non-word effect was expected to effect both Chinese groups in the Chinese LANGUAGE and all three groups in the English LANGUAGE with the same order of degrees as in hypothesis 1: the American Group (Group 3), the Chinese Group who had been in the United States of America for longer than three years (Group 2), and the Chinese Group who had been in the United States of America for less than three years (Group 1) (Brancazio and Brancazio 2004).

Hypothesis 1 and 2 were analyzed together. The McGurk rate in Experiments 2 and 3 followed the same calculation rule of that for Experiment 1, the accuracy of the mixed condition (/b/ as the correct responses) was subtracted from the accuracy of the /b/

word condition of the same set of word. Analysis involved a repeated measures ANOVA for LANGUAGE (2: Chinese vs. English) x AUDITORY LEXICAL INFLUENCE (2: word vs. nonword) x VISUAL LEXICAL INFLUENCE (2: word vs. nonword) x GROUPS (3: Group 1, Group 2 vs. Group 3). Significant results are listed in the following table.

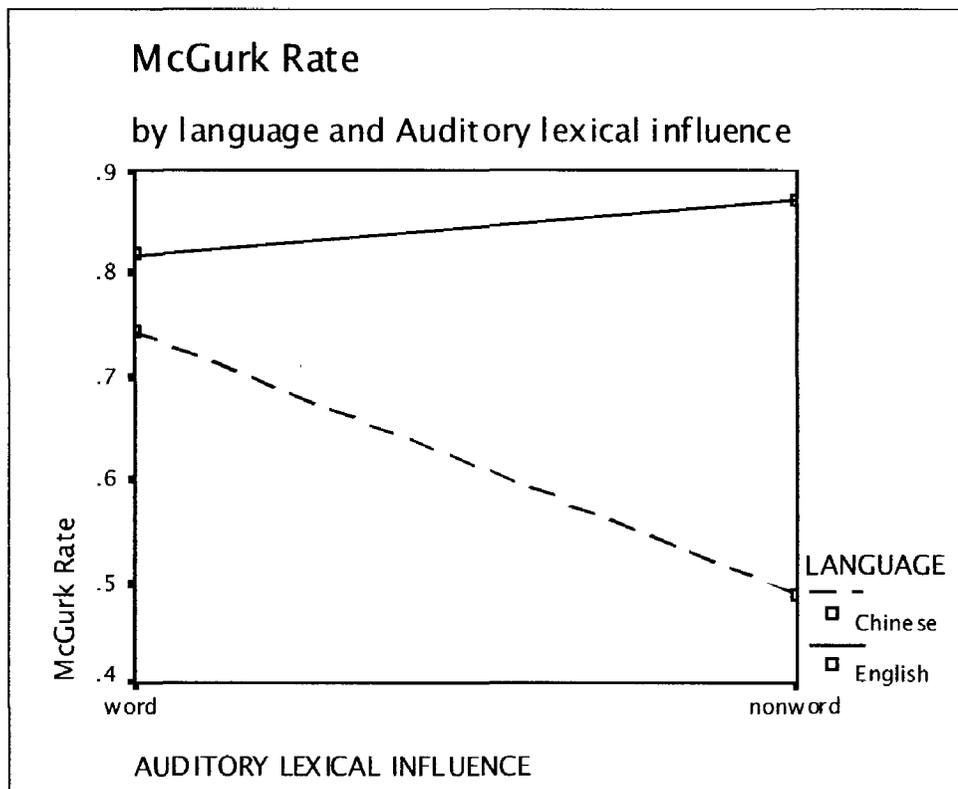
**Table 3**

Significant results for the behavioral part of Experiment 2 and 3

Effect	Significance
LANGUAGE	$F(1, 33) = 30.823, p < .001, \text{Observed Power} > .999.$
AUDITORY LEXICAL INFLUENCE	$F(1, 33) = 14.083, p < .01, \text{Observed Power} = .954.$
LANGUAGE by AUDITORY LEXICAL INFLUENCE	$F(1, 33) = 47.562, p < .001, \text{Observed Power} > .999.$
LANGUAGE by VISUAL LEXICAL INFLUENCE	$F(1, 33) = 11.392, p < .01, \text{Observed Power} = .906.$
AUDITORY LEXICAL INFLUENCE by VISUAL LEXICAL INFLUENCE	$F(1, 33) = 5.313, p < .05, \text{Observed Power} = .609.$
GROUP	$F(1, 33) = 5.923, p < .01, \text{Observed Power} = .846.$

Following the significant interaction of LANGUAGE by AUDITORY LEXICAL INFLUENCE, the McGurk rate was averaged across visual lexical conditions. For each

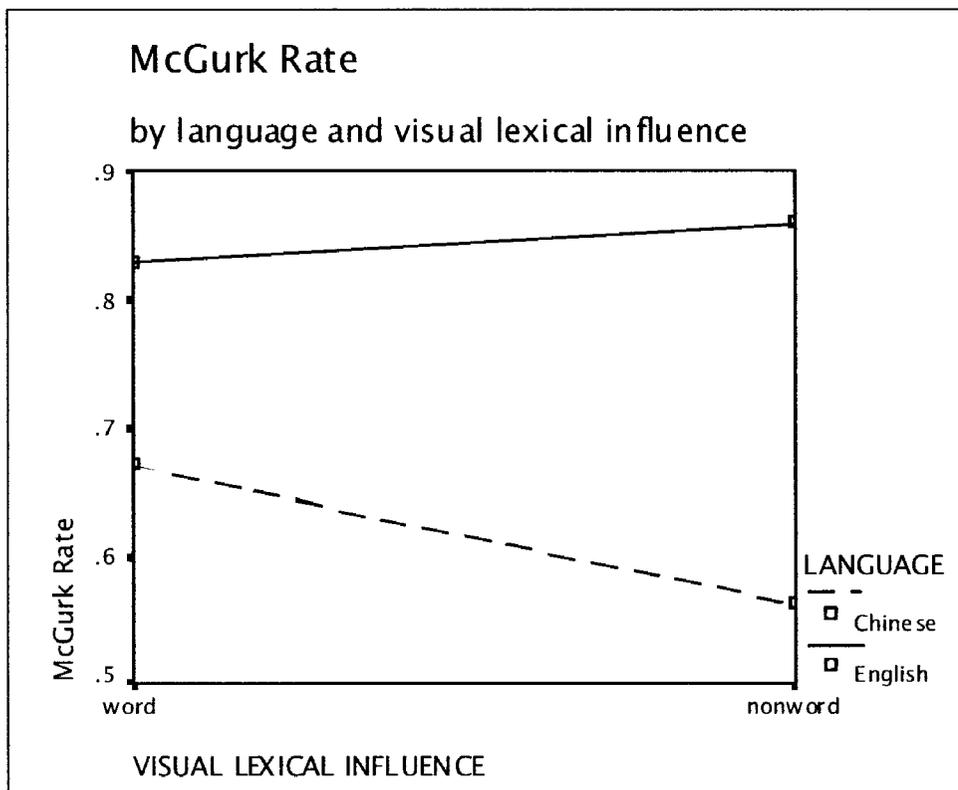
language, the effect of AUDITORY LEXICAL INFLUENCE was compared using paired-sample T-tests. For Chinese, the AUDITORY LEXICAL INFLUENCE had a positive effect on the McGurk Rate,  $t(35) = 5.832, p < .001$ , indicating that the word condition generated a higher McGurk rate than the nonword condition. However in English,  $t(35) = -2.363, p < .05$ , indicating that the word condition generated lower McGurk rate than the nonword condition. The effect was illustrated in Figure 7.



**Figure 7.** The interaction of LANGUAGE and AUDITORY LEXICAL INFLUENCE on the McGurk rate. The word condition generated a higher McGurk rate than the nonword condition in Chinese, but an opposite pattern in English.

Following the significant interaction of LANGUAGE by VISUAL LEXICAL INFLUENCE, the McGurk rate was averaged across auditory lexical conditions. Then for each language, the effect of VISUAL LEXICAL INFLUENCE was compared using

paired-sample T-tests. The results indicated that in Chinese, the VISUAL LEXICAL INFLUENCE had a positive effect on the McGurk rate,  $t(35) = 3.979, p < .001$ , indicating that the word condition generated a higher McGurk rate than the nonword condition; however, no significant effect occurred for English,  $t(35) = -1.076, ns$ . This result is illustrated in Figure 8.



**Figure 8.** The interaction of LANGUAGE and VISUAL LEXICAL INFLUENCE on the McGurk rate. The word condition generated a higher McGurk rate than the non-word condition in Chinese; but in English, the effect was not significant.

Pairwise Comparisons of the main effect for GROUP with Sidak correction indicated that Group 3 produced a significantly lower McGurk Rate than Groups 1 and 2, Mean Difference between Group 3 and Group 1 =  $-.191, p < .05$ ; Mean Difference between Group 3 and Group 2 =  $-.185, p < .05$ ; however Group 1 and Group 2 did not

differ significantly from each other.

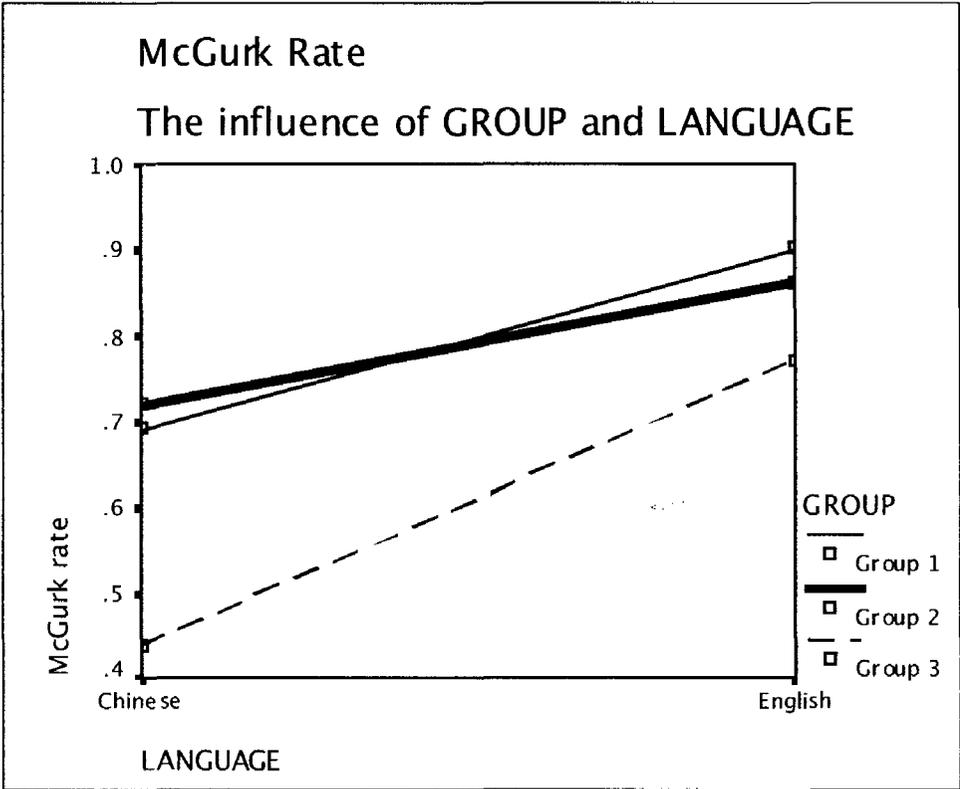
Following the main effect of LANGUAGE, a Pairwise Comparison with Sidak correction indicated that the McGurk rate was lower when the language was Chinese than when the language was English, Mean Difference Chinese versus English =  $-.228$ ,  $p < .001$ . No significant GROUP by LANGUAGE interaction was observed, suggesting that the different languages influenced the different groups in the same way. Thus the hypothesis was not supported. The main effects of LANGUAGE and GROUP are illustrated in Figure 9.

#### **4.2.2 Results for Hypothesis 3**

Hypothesis 3: This hypothesis addressed the McGurk lexicon interaction in Chinese stimuli. It was hypothesized that if the auditory source of the McGurk condition was a word (such as /bench/) and the visual source generated a nonword (such as /dench/), the McGurk nonword condition would produce a negative peak centered near 400 ms (N400 with the peak latency during 350 to 450 ms) than the audiovisual word condition (such as audiovisual /bench/). This should only have occurred in the Chinese groups (Group 1 and 2) when the language was Chinese (Kutas and Hillyard 1980; Kutas and Hillyard 1984; Friederici 2002).

According to the hypothesis, the cwnb and cwnm conditions were analyzed, in which the /b/ sound was a Chinese word /bie3/ and the McGurk condition was a nonword /die3/. Only the trials that reflected a McGurk response were included in the final average. Fourteen subjects were excluded due to not having a strong McGurk effect (McGurk Rate  $< .65$ ). One subject was excluded because of artifacts. After exclusion, 15

subjects from Group 1 and Group 2 (9 female), and 6 subjects from Group 3 (2 female) qualified for further analysis.



**Figure 9. The main effects of LANGUAGE and GROUP on the McGurk effect. The thin solid line represents Group 1; the thick solid line represents Group 2 and the dashed line represents Group 3. On the left side of the x-axis, the values represent the McGurk rates that occurred in the conditions when the stimulus language was Chinese for the three groups; on the right side of the x-axis, the values represent the McGurk rates of the conditions when the stimulus language was English for the three groups. The McGurk rate was lower in Chinese language than English language; and the McGurk rate was lower for Group 3 than Group 1 and Group 2.**

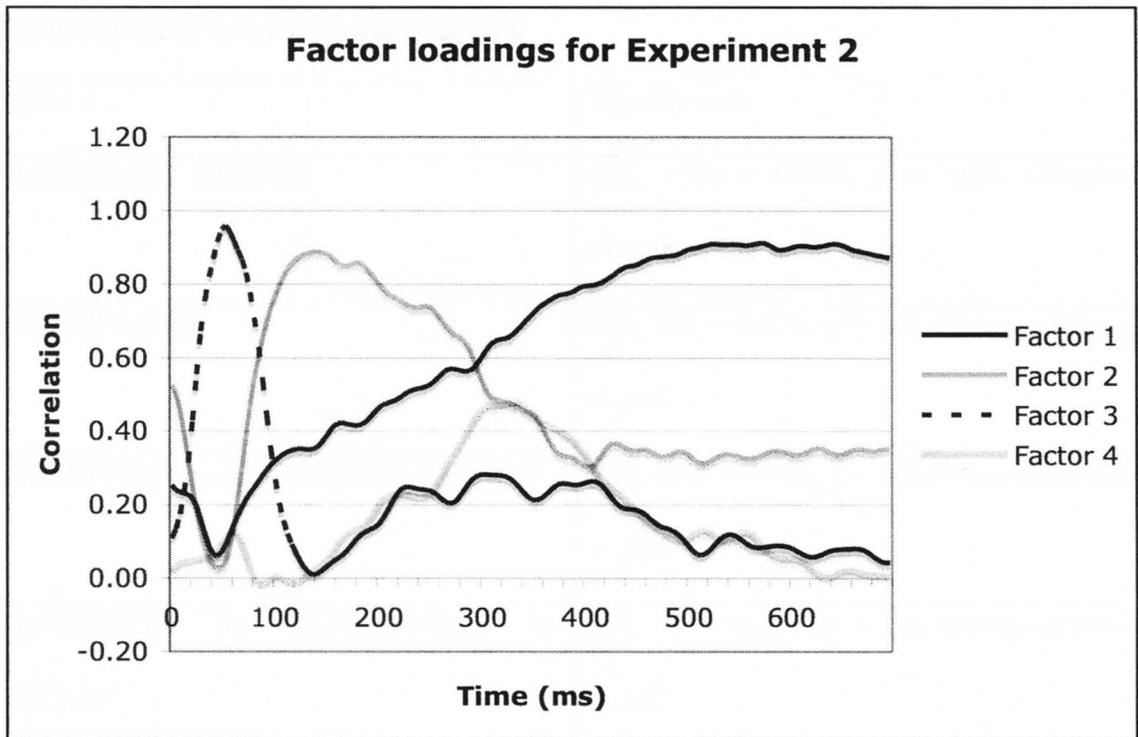
The preprocessing followed the same procedure as in Experiment 1. The averaged

brainwaves for the two conditions, cwnb and cwnm, were obtained for each subject; next, a Factor analysis followed by Varimax rotation was conducted on the time dimension of these ERP data. Factor selection for subsequent data analyses was based on a factor loading greater than 0.4. The results of Factor analysis are illustrated in Table 4 and Figure 10. Factor 4 was picked as the Factor of interest, because it was the one that had a peak latency close to 400 ms. Brainwave data were averaged during the 292 to 380 ms period and then submitted to a repeated measures ANOVA: CONDITION (2: standard vs. McGurk) x ELECTRODE REGIONS (9) x HEMISPHERE (2) x GROUP (2: the Chinese groups vs. the American group). The significant effects are listed in Table 5.

**Table 4**

Factor analysis for Experiment 2 (N = 21)

Factor	Variance accounted (%)	Effective Time (ms)	Peak Time (ms)
1	74.18	384 – 700	560
2	8.51	188 – 288	140
3	5.88	24 – 84	56
4	3.22	292 – 380	336

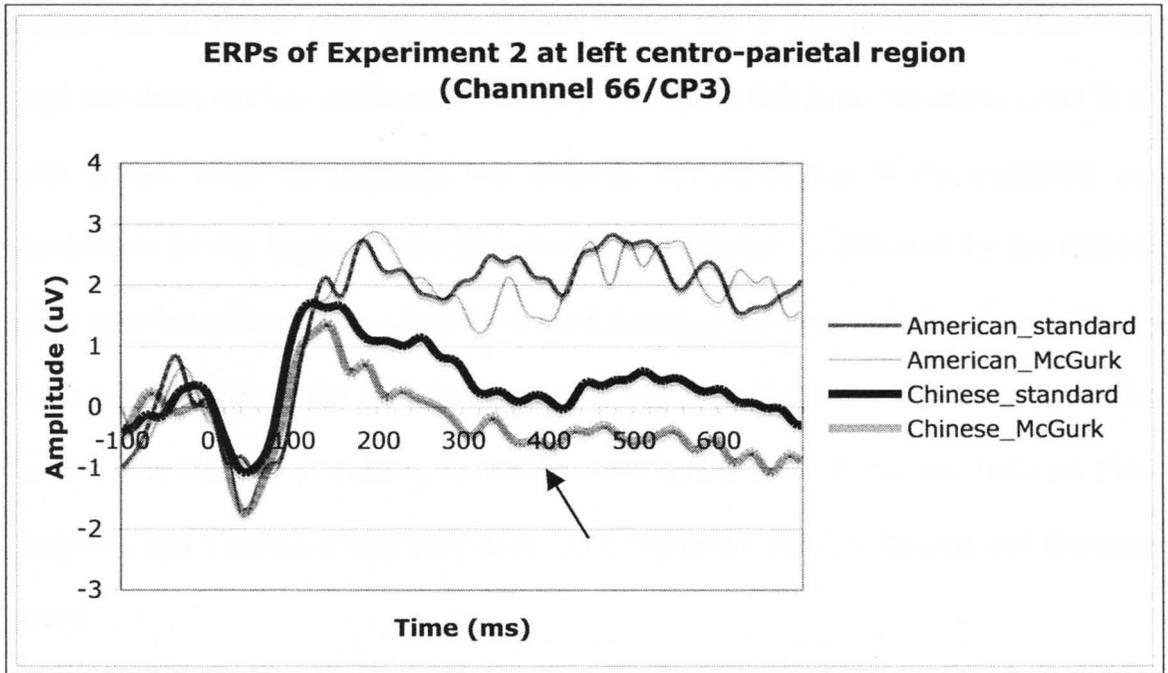


**Figure 10.** Plot of the factor loadings for Experiment 2. Factor 1 characterized a slow wave starting at 384 ms and that continued throughout the remainder of the segment, with the maximal amplitude reached at 560 ms; Factor 2 was the second peak, effective from 88 to 288 ms with the highest amplitude reached at 140 ms; Factor 3 characterized the first ERP peak with the effective time range from 24 to 84 ms, and with the highest amplitude occurring at 56 ms; and Factor 4 characterized the third peak between 292 and 380 ms, with the highest amplitude reached at 336 ms.

**Table 5**Significant ERP results for Experiment 2

Effect	Significance
ELECTRODE REGION	$F(8, 152) = 9.065, p < .001$ , Observed Power = .991
HEMISPHERE	$F(1, 19) = 5.971, p < .05$ , Observed Power = .640
CONDITION by ELECTRODE REGION	$F(8, 152) = 4.951, p < .05$ , Observed Power = .736
CONDITION by HEMISPHERE by GROUP	$F(1, 19) = 6.359, p < .05$ , Observed Power = .667
ELECTRODE REGION by HEMISPHERE	$F(8, 152) = 3.054, p < .05$ , Observed Power = .724

Following the significant CONDITION by HEMISPHERE by GROUP interaction, brainwaves in different electrode regions were averaged; then the CONDITION effect was tested in each hemisphere and each group. Paired samples T-Test indicated that in the left hemisphere the Chinese group generated a more negative amplitude in the McGurk condition than in the standard condition:  $t(14) = 2.274, p < .05$ , Mean difference (Standard – McGurk) = 3.293, Std. Error Mean = 1.45. However the condition effect was not significant for the right hemisphere electrodes; neither was it significant for the American group. Thus the hypothesis was supported. The effect is illustrated in Figure 11.



**Figure 11.** ERPs elicited in Experiment 2 over the left centro-parietal region. (channel 66, equivalent to CP3 from the 10-20 system). The thin dark line represents the American group in the standard (cwnb) condition; the thin gray line represents the American group in the McGurk (cwnm) condition; the thick dark line represents the Chinese group in the standard (cwnb) condition and the thick gray line represents the Chinese group in the McGurk (cwnm) condition. The Chinese group generated a negative slow wave that overlaps with 400 ms post stimulus onset (indicated by the arrow) during the McGurk condition versus the standard condition but it was not observed in the American group.

#### 4.2.3 Results for Hypothesis 4

Hypothesis 4: it was hypothesized that when the auditory source of the McGurk condition was a word (such as /bench/) and the visual source generated a nonword (such as /dench/), then the McGurk nonword condition would generate a more negative peak

around 400 ms (N400 with the peak latency during 350 to 450 ms) than the audiovisual word condition (such as audiovisual /bench/). This effect was hypothesized to occur in all three groups when the language was English, but the degree of the negativity was expected to be the largest in the American group (Group 3), followed by the Chinese group who have been in the United States of America for more than three years (Group 2); and the smallest in the Chinese group who had been in the United States of America for less than three years (Group 1) (Kutas and Hillyard 1980; Kutas and Hillyard 1984; Sekiyama and Tohkura 1993; Sekiyama 1997; Friederici 2002; Sekiyama and Burnham 2008).

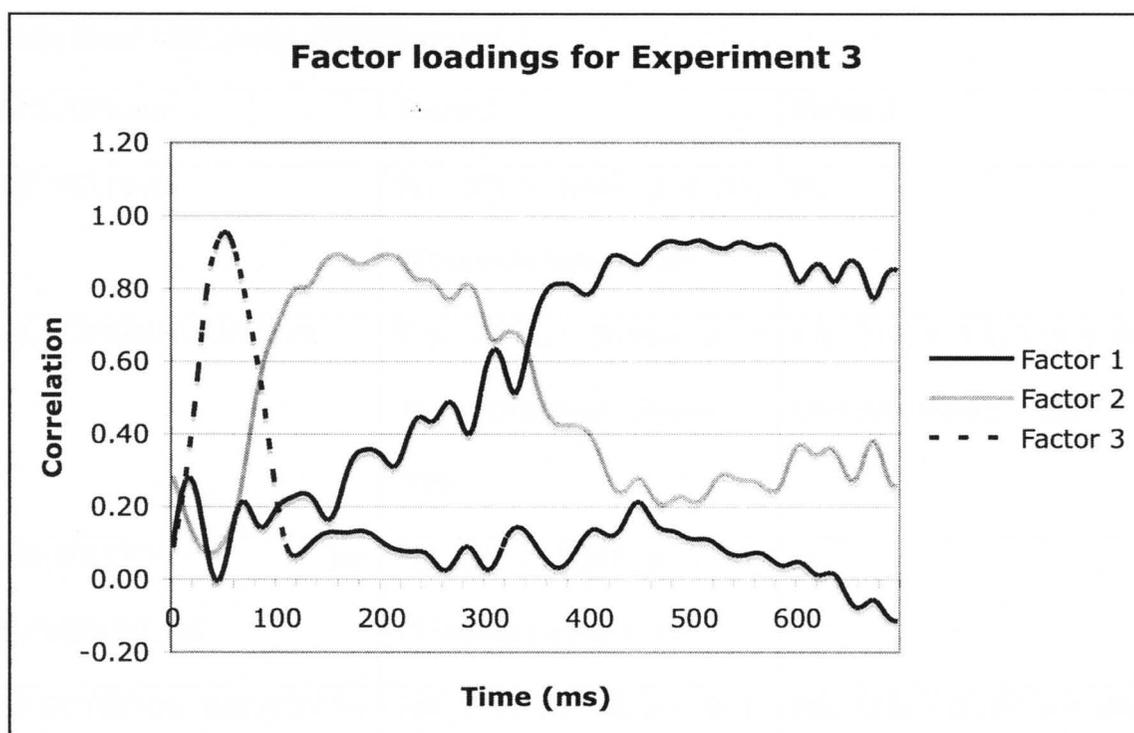
To test this hypothesis, the ewnb and ewnm conditions were analyzed in which the /b/ sound was an English word /bench/ and the McGurk condition was a nonword /dench/. Only the trials that reflected a McGurk response were included in the final average. Six subjects were excluded due to not having a strong McGurk effect (McGurk Rate < .65). After exclusion, 11 subjects from Group 1 (6 female), 9 subjects from Group 2 (4 female), and 10 subjects from Group 3 (3 female) qualified for further analysis.

The preprocessing followed the same procedure as in Experiment 1. The averaged brainwaves for the two conditions, ewnb and ewnm, were obtained for each subject. A Factor analysis followed by Varimax rotation then was conducted on the time dimension of the ERP data. The effective time was determined by two criteria in order: first, it was not overlapped with other factors; second, it had a correlation over 0.4 during that time. The result of Factor analysis is illustrated in Table 6 and Figure 11.

**Table 6**

Factor analysis for Experiment 3

Factor	Variance accounted (%)	Effective Time (ms)	Peak Time (ms)
1	65.13	344 – 700	508
2	12.02	88 – 340	164
3	6.83	24 – 84	56



**Figure 12.** Plot of the factor loadings for Experiment 3. Factor 1 was a slow wave starting at 333 ms and ran throughout the whole segment, with the highest amplitude at 508 ms; Factor 2 was the second peak, effective from 88 to 340 ms with the highest amplitude at 164 ms; Factor 3 was the first peak with the effective time range from 24 to 84 ms, and with the highest amplitude at 56 ms.

There was no factor that had a peak value around 400 ms, so the negativity around 400 ms was not prominent. In case that that negativity might shift a little bit earlier or later, both Factor 2 and Factor 3 were studied in the experiment. Brainwave data were averaged during the effective time range of both factors and then submitted to two repeated measures ANOVAs for each factor: CONDITION (2: standard vs. McGurk) x ELECTRODE REGIONS (9) x HEMISPHERE (2) x GROUP (3). The significant effects were listed in Table 7.

**Table 7**

Significant ERP results for Experiment 3

Effect/Factor	Factor 2	Factor 3
CONDITION	F(1, 27) = 4.642, $p < .05$ , Observed Power = .547	ns.
ELECTRODE REGION	F(8, 216) = 16.066, $p < .001$ , Observed Power > .999	F(8, 216) = 5.372, $p < .01$ , Observed Power = .827
CONDITION by HEMISPHERE	F(1, 27) = 8.358, $p < .01$ , Observed Power = .796	ns.
ELECTRODE REGION by HEMISPHERE	F(8, 216) = 9.005, $p < .001$ , Observed Power = .996	F(8, 216) = 6.104, $p < .001$ , Observed Power = .966

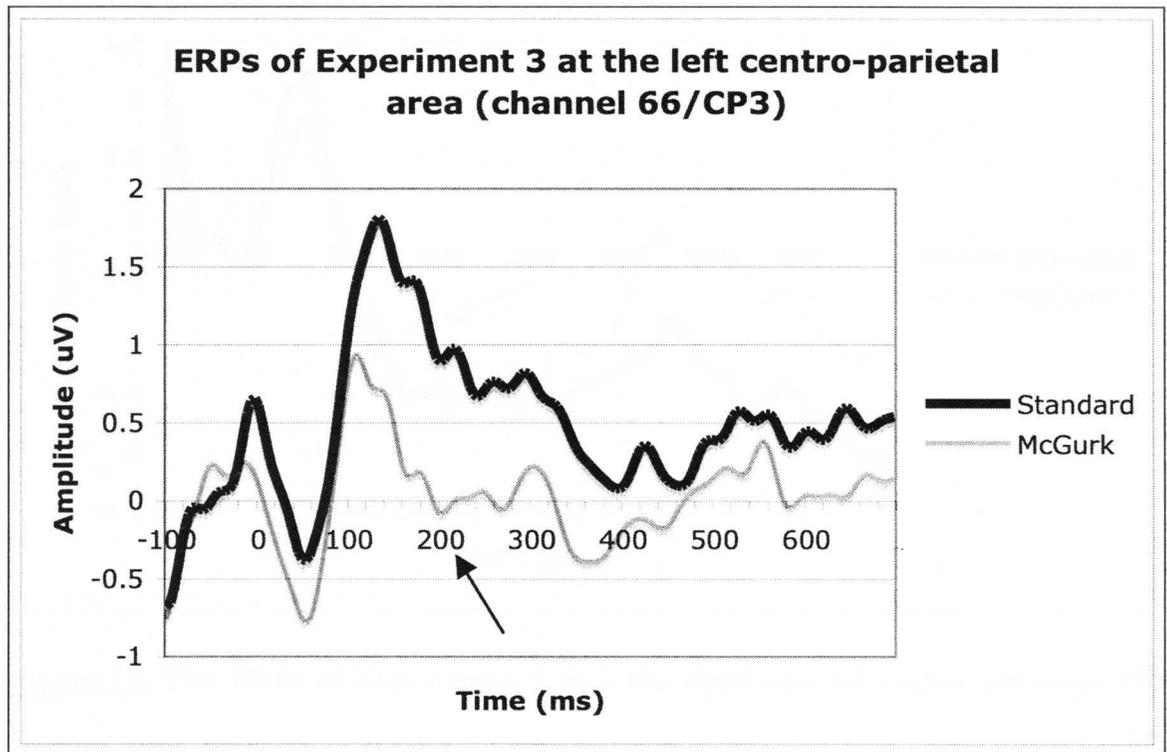
Following the significant CONDITION by HEMISPHERE interaction in Factor 2 (88-340 ms), brainwaves at different electrode regions were averaged. Next, the CONDITION effect on the averaged brainwave was tested at each hemisphere using

Paired Samples T-Tests. The results indicated that over the left hemisphere, the McGurk condition produced a less positive amplitude than the standard condition,  $t(29) = 2.418$ ,  $p < .05$ , Mean difference (standard – McGurk) = 2.157, Std.Error Mean = .892. The results also indicated that the McGurk condition generated a smaller negative amplitude ERP than the standard condition over the right hemisphere,  $t(29) = -3.305$ ,  $p < .01$ , Mean difference (standard – McGurk) = -3.451, Std.Error Mean = 1.044. The CONDITION by HEMISPHERE did not involve the GROUP effect, suggesting that the effect was true for all the groups. It supported the first part of the hypothesis. However, since GROUP did not modulate the effect, the second part of the hypothesis was not supported.

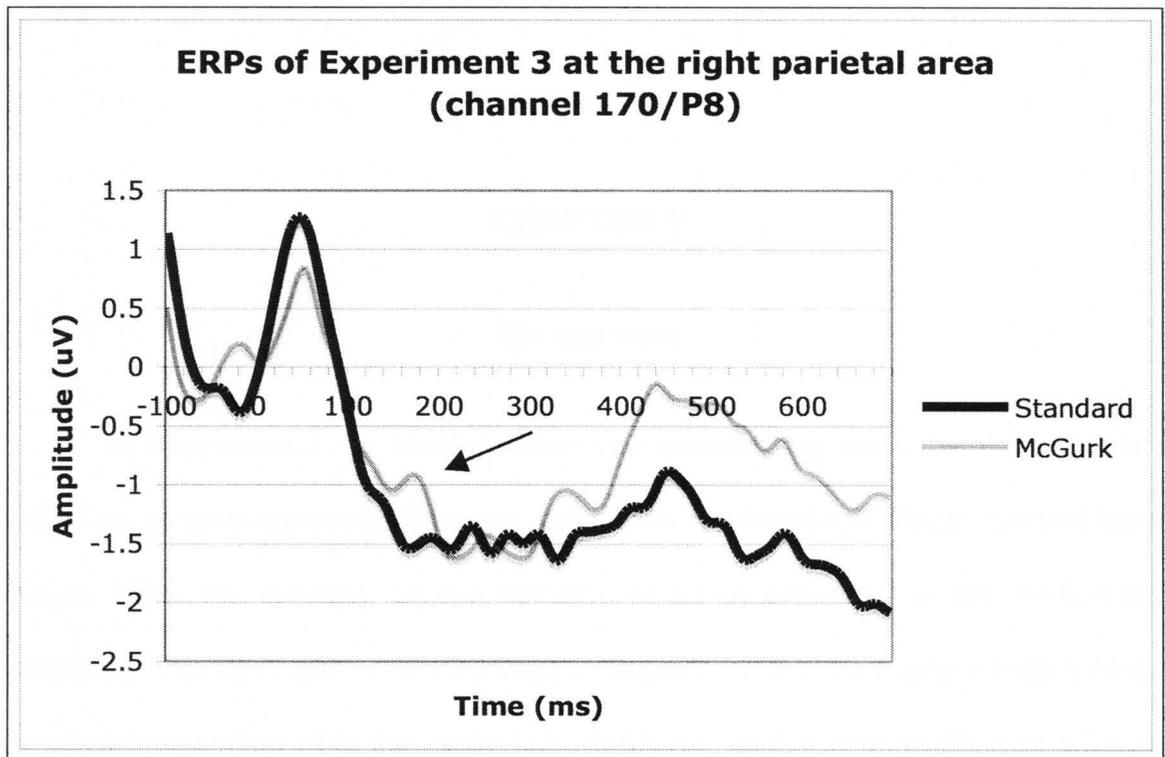
These results are illustrated in Figure 13 and Figure 14. From the grand average of ERPS across the entire scalp of the head, the effect over the left hemisphere was most pronounced at the central and parietal areas. Channel 66 in the left centro-parietal region (CP3 in a 10-10 system) was selected to illustrate this effect in Figure 13. Because the effect was most pronounced over the more posterior area of the right hemisphere, channel 170 in the right parietal region (P8 in a 10-10 system) was selected to illustrate this effect in Figure 14.

Experiments 2 and 3 studied the McGurk effect as a function of different language backgrounds. The American English speakers produced a greater McGurk rate than either of the Chinese speaker groups. However no Group by Language interaction occurred. Experiments 2 and 3 investigated the auditory and visual lexicon influence on the McGurk effect. The results were only partially consistent with the hypothesis. Finally, Experiments 2 and 3 tested the McGurk and lexicon interaction using electrophysiological measures. However, the prediction was only supported when the

stimuli were in Chinese. The unsupported predictions are discussed below.



**Figure 13.** The ERPs of Experiment 3 over the left centro-parietal region. (channel 66 equivalent to CP3 in a 10-10 system). The dark line represents the brainwave of the standard condition; the gray line represents the brainwave of the McGurk condition. During the time period around 88 to 340 ms (indicated by the arrow) the McGurk condition had more negative amplitude than the standard condition.



**Figure 14.** The ERPs of Experiment 3 over the right parietal region. (channel 170 equivalent to P8 in a 10-10 system). The dark line represents the brainwave of the standard condition; the gray line represents the brainwave of the McGurk condition. During the time period around 88 to 340 ms (indicated by the arrow) the McGurk condition generated a smaller negative amplitude than the standard condition.

## **CHAPTER V**

### **Discussion**

In Experiment 1, the McGurk effect was studied using audio /ba/ dubbed onto a video /ga/ in three experimental groups. The results confirmed that Visual-Auditory input would affect the accuracy of the McGurk condition compared to the Audio-Only condition. This study also confirmed that the McGurk effect had a negative impact on the reaction time compared to the Audio-Only condition, so that participants took a longer time to make a choice in the McGurk condition. In the analysis of ERP data, the McGurk effect generated a smaller negative N1 amplitude than the standard condition only, but only over the left hemisphere. This result was not reported by previous researches, but all investigators either used a different paradigm or used markedly different stimuli than the current study. The left hemisphere difference was in line with previous findings that the left hemisphere played an important role in the McGurk effect (Saint-Amour, De Sanctis et al. 2007). A result from Experiment 3 indicated that the McGurk effect also generated a larger negative peak amplitude for the period from 88 to 340 ms over the left hemisphere than the standard condition. However, it produced a smaller negative amplitude over the right hemisphere during the same time period. This finding is consistent with earlier reports (Lebib, Papo et al. 2004). Since the McGurk effect and speech perception involve language learning and reproduction, researchers believe that it might not happen until later in the brain responses. However the present results indicate

that the audiovisual integration occurs at a very early stage of time, during the first 50 ms as indicated by the ERP findings.

No overall reaction time or ERP differences were found between the different language groups. However, in terms of the McGurk effect, the group of American English speakers did generate a weaker McGurk effect than the group of Chinese who had less than three years residency in the United States. This finding contradicts earlier reports in the literature (Sekiyama and Burnham 2008). The discrepancy may be related to the fact that the American participants in the current study could have had more knowledge about the experiment because of their college major (i.e., Psychology) and their required courses (e.g., a broad survey course such as Introductory Psychology or more advanced courses in Cognitive Psychology or Language Processes). The majority of the American English speakers were recruited from the Department of Psychological and Sciences. They likely acquired prior information about the McGurk effect, so in the experiment they might have intentionally attended more to the auditory stimuli while trying not to be misled by the visual stimuli. Alternatively, the less experienced language users may have perceived more intently on both auditory and visual cues making them more accessible to the effect.

Semantic information was introduced in Experiments 2 and 3. The results indicated that American participants generated a smaller McGurk effect than the two Chinese groups. The same situation existed for both the Chinese and English stimulus conditions. Chinese stimuli generated a smaller McGurk effect than the English stimuli. The analysis results of the Chinese participants were consistent with previous findings, in which those exposed to a non-native language produced a larger McGurk effect. This is

consistent with Sekiyama and Burnham (2008). However, the result of the American group was inconsistent with this finding. The discrepancy was likely due to the fact that Chinese stimuli were prepared and presented by a non-Chinese speaker who possessed different articulation skills than the Chinese listeners in this study (foreigner effect). It is recommended for a future study to utilize a native Chinese speaker as an actor for the video portion of the stimulus set so that the foreigner effect could be controlled.

Differences between the two Chinese groups only occurred in their language related assessment (PPVT), but not in any other measurements in the experiments. It was suggested that it might be because that the visual dependency of the two Chinese groups reached a ceiling effect and therefore was not easily detectable by electrophysiological measurements. The following sections will systematically address the support for hypotheses tested and the implications of such support as well as potential reasons for the failure to find support for other hypotheses.

## **5.1 Experiment 1**

This experiment sought to understand the blending of audio and visual percepts within the McGurk Paradigm using both behavioral and event-related potential assessment techniques. In this cross-cultural study, we examined the McGurk effect in native English speakers and native Chinese speakers (high and low exposure to an English speaking environment). The McGurk effect was studied in Experiment 1. The results from tests of the six hypotheses are discussed below.

### **5.1.1 Hypothesis 1**

For Hypothesis 1 we expected that the combined incongruent Audiovisual (AV) input (McGurk effect) would impair the accuracy of speech perception relative to Audio-Only input when subjects were instructed to respond to the correct answer (McGurk and MacDonald 1976; MacDonald and McGurk 1978; Sekiyama and Burnham 2008).

Hypothesis 1 was supported. Performance accuracy of the McGurk condition was significantly lower than that for the Audio-Only condition. This finding was consistent with previous literature (McGurk and MacDonald 1976; MacDonald and McGurk 1978; Sekiyama and Burnham 2008), and indicated that the present study induced the McGurk effect in the tested population.

### **5.1.2 Hypothesis 2**

For Hypothesis 2 we hypothesized that the McGurk effect would have a negative impact on the reaction time (RT) relative to the Audio-Only condition and the audiovisual condition (van Wassenhove, Grant et al. 2005; Sekiyama and Burnham 2008).

Hypothesis 2 also was supported, with The McGurk condition produced a larger RT than the Audio-Only and the Audiovisual conditions; the Audio-Only condition induced a larger RT than the audiovisual conditions. The results were consistent with previous studies (van Wassenhove, Grant et al. 2005; Sekiyama and Burnham 2008), suggesting that the incongruent information slowed down information processing.

### 5.1.3 Hypothesis 3

For Hypothesis 3 we hypothesized that the McGurk effect would generate a larger amplitude than the standard condition for the N1 (peak amplitude occurring between 50 ms to 150 ms), P2 (peak amplitude occurring between 150 ms to 250 ms), N300 (peak amplitude occurring between 250 ms to 350 ms) and a later component (a slow wave occurring between 350 to 600 ms) recorded at frontal, central and parietal regions (Colin, Radeau et al. 2002; Lebib, Papo et al. 2004; van Wassenhove, Grant et al. 2005; Saint-Amour, De Sanctis et al. 2007; Stekelenburg and Vroomen 2007).

The first discrepancy between our results and the hypothesis was that the N1 peaked at 56 ms instead of near 100 ms. The shortened N1 latencies were probably due to the short internal stimulus interval (ISI) and the possible stimulus onset asynchrony (SOA) between the visual and audio stimuli (Gomes, Sussman et al. 1999; Gilley, Sharma et al. 2005; Sussman, Steinschneider et al. 2008). In the current study the ISI varied from 1.9 to 2.4 seconds. In previous studies using MMN paradigm the ISI tended to be very short. For example, in Colin and colleagues' study (Colin, Radeau et al. 2002) they presented stimuli at approximately every second. However, in another study that did not use the MMN paradigm and required a button press, the ISI varied from 500 to 1500 ms, and the N1 occurred around 100 ms after auditory stimulus onset (van Wassenhove, Grant et al. 2005).

In the present experiment, N1 varies across hemispheres and conditions. In a previous study (Colin, Radeau et al. 2002), the P1-N1-P2 peak complex was more negative during the McGurk condition than the standard condition and no hemisphere effect was reported. Saint-Amour, De Sanctis et al. (2007) reported that the McGurk

effect was associated with activities in the left hemisphere, especially the left STG but the differentiation did not occur until 175 ms following auditory stimulus onset. However, the results from previous studies could not be directly compared to that of the current study, because both of the previous studies used a MMN paradigm in which one condition occurred with much greater frequency than another while the current study used an equal number of trials across conditions.

Studies that did use the equal probability paradigm reported much less negativity (van Wassenhove, Grant et al. 2005). Lebib, Papo et al. (2004) used a similar design but found that the McGurk condition generated a negativity around 300 ms compared to the standard condition over frontal, central and parietal areas. The current study did not duplicate their findings, an inconsistency that might be due to the fact that Lebib and Papo used French vowels, which because of their acoustic properties might elicit very different brain responses than the English consonants used in the current study.

Although our hypothesis was driven by previous studies, these studies were not equivalent in design to the current study. However, even though the current result did not support the hypothesis, it still provided evidence for differential processing as indexed by ERPs of the McGurk effect from the standard conditions.

#### **5.1.4 Hypothesis 4**

For Hypothesis 4 we expected that the degree of the McGurk effect would be largest in the group of native English speakers, followed by the Chinese participants who have been in United States of America more than three years. The effect was expected to be smallest in Chinese participants who have been in the United States of America less

than three years (Sekiyama 1997; Sekiyama and Burnham 2008).

Contrary to the expectations, the results indicated that American group (Group 3) had a significant lower McGurk Rate than the Chinese Group who had been in the United States of America for less than three years and there was no significant difference of McGurk rate between the two Chinese groups or between American group and the Chinese group who had been in the United States of America for longer than three years (Group 2).

As illustrated in Figure 6, most marked effect of the McGurk rate was in Group 1, following Group 2 and 3. This order was opposite to that hypothesized for this study. Sekiyama (1997) found that the longer the Chinese participants had been in Japan, the larger the McGurk effect. In a later study (Sekiyama and Burnham 2008) reported that English speaking children produced a higher McGurk rate than the Japanese children for English syllables. Thus, the current study results were inconsistent with previous work, a result perhaps due to the fact that the subjects from Group 3 were mostly recruited from a Department of Psychological and Brain Sciences. Given their exposure to a Psychology curriculum, it is possible that the students knew something about the McGurk effect and consequently paid more attention to the auditory stimuli during testing. Even though previous study suggested that the McGurk effect would occur even when the effect was known prior to the test, it was possible that the McGurk rate was still depressed by prior knowledge. For example, because the participants were instructed to detect what they heard, students with previous knowledge would try looking away from the screen in order to avoid being misled by the visual stimuli. Future studies should include additional measurements such as eye tracking in order to verify that participants were attending to

the screen during the experiment. In contrast to the American participants, none of the Chinese participants knew about the effect prior to the study. Further work is needed to resolve this issue.

The results also suggest that even though the Group 1 and Group 2 Chinese participants had different durations of residence in an English-speaking country, their behavioral performance on the McGurk effect did not differ. There was no behavioral separation of the Chinese participants who had been in American for longer than three years versus those who resided there for less than three years. Sekiyama & Burnham (2008) noted that an English-speaking environment increased the McGurk rate more for older children than for younger ones, but a Japanese-speaking environment did not have the same effect on native Japanese speakers. The discrepancy between previous findings and our study could result from several reasons. Sekiyama & Burnham (2008) used four different noise levels in contrast to the current study. They noted that a noisy environment decreased the McGurk rate and enlarged the differences between different groups. In the current study, however, participants were tested only in a quiet setting. Thus the difference between conditions might not have been ideal to find the differences noted by Sekiyama and Burnham. Another contributing factor to the lack of effects could be the environment. Even though Group 2 Chinese had lived in the United States of America longer, it is still common for Chinese speakers to spend more of their time within more exclusively Chinese speaking communities, thereby possibly negating or reducing any English language exposure effects that might otherwise have distinguished between Chinese participants who had been in the United States of America for more vs. less time.

### **5.1.5 Hypothesis 5**

Hypothesis 5 predicted that the impact on the reaction time for the McGurk condition compared with the Audio-Only condition and the congruent audiovisual condition would be in the same order as described for the McGurk effect in Hypothesis 4 (Sekiyama 1997; Sekiyama and Burnham 2008).

Our results indicated that the McGurk effect did impact the reaction time in different conditions, with the longest reaction time in the McGurk condition, followed by a median reaction time in the Audio-Only condition and the shortest reaction time in the Audiovisual condition. However membership in the different groups did not differ between the conditions. This result was inconsistent with Sekiyama and Burnham's study (2008), in which they found that the RT difference between the Audiovisual condition and the McGurk condition was larger for English speaking participants versus Japanese speaking participants. They used 24 subjects in each group; but in the current study there were only 12 subjects in each group, with an observed power of .240. Therefore, the power of this study was not sufficient to detect the RT time difference between groups.

### **5.1.6 Hypothesis 6**

For Hypothesis 6 we expected that the McGurk markers found in Hypothesis 4 would be the largest in American English speakers group, followed by the Chinese group who has been in the United States of America for longer than three years. The Chinese group who has been in the United States of America for less than three years would have the lowest peak amplitudes in these markers. (Sekiyama 1997; van Wassenhove, Grant et al. 2005; Sekiyama and Burnham 2008). This was not the case. The possible reason for

not finding significant differences in the Post-hoc tests might come from the low McGurk rate in some of the participants. As discovered in the test of hypothesis 1.4, the American English speaking group had a much lower McGurk rate than the Chinese groups. From Figure 6, it is obvious that the average McGurk rate for the American English speakers was approximately 0.3, which means the McGurk effect was only generated in 3 of every 10 McGurk stimuli for the American Group. This low rate and the consequent mixing of both McGurk and non-McGurk effects in the McGurk condition likely contributed to the lack of the expected effect.

To further examine the weak McGurk effect, subjects were further categorized in three additional groups: weak McGurk (McGurk rate < 0.25), median McGurk effect (0.35 < McGurk rate < 0.50) and strong McGurk (McGurk rate > .75). The distribution of the different levels of the McGurk effect in experimental groups is illustrated in Table 8.

**Table 8**

The distribution of the different levels of the McGurk effects in three experimental groups

Number of subjects	Weak McGurk	Median McGurk	Strong McGurk
Group 1	2	0	10
Group 2	2	1	9
Group 3	7	2	3

Previous studies all reported a higher prevalence of the McGurk effect. For example, McGurk's study (McGurk and MacDonald 1976) indicated that 98% of the time

an audio /b/ dubbed with visual /g/ was perceived as /d/ by the American English speakers; in another study (Saint-Amour, De Sanctis et al. 2007), one out of 12 American English speakers were excluded because they did not perceive the effect. It is interesting that even though the majority of the American English speakers experienced the McGurk effect in the previous literature, large individual differences still occurred in the present study. No one to our knowledge has investigated the reason that why some people do not produce the illusion. It might have something to do with the linguistic boundary mentioned in the introduction. Future studies study further such individual differences. from the preent study, it is clear that there is no guarantee that all participants will detect the McGurk effect. Consequently, in any future study, subjects should be screened to determine whether they generate the effect. By utilizing a more homogeneous sample researchers might be able to examine the phenomenon more closely. Unfortunately, in the present study, too few subjects made up some groups, precluding our ability to reanalyze the data for only strong responders.

Previous studies have suggested that the McGurk effect generates a different brainwave than the standard condition in various ERP components on frontal, central and parietal regions (Colin, Radeau et al. 2002; Lebib, Papo et al. 2004; van Wassenhove, Grant et al. 2005; Saint-Amour, De Sanctis et al. 2007). In the current study, a hemisphere difference as early as 50 ms was found in the McGurk condition compared to that of the standard condition. Even though the McGurk effect involves high functionality and complicated processing, the brain appears to be triggered at a very early stage of processing. In addition, the present study used the direction comparison between conditions in an equal probability paradigm and found the results similar to those

reported that used an oddball paradigm. Both approaches confirmed the early activation of the brain during the McGurk effect on a higher functionality level. The condition related group difference was not found. It might be due to the fact that the American group intentionally avoided being misled by the visual illusion because they had acquired previous knowledge about the effect. The two Chinese groups might have reached their ceiling in visual dependency, so that visual dependency can no longer reflect their English proficiency. The results definitely require further investigation, given the small number of subjects in the present American English speaker sample who generated strong McGurk effects.

## **5.2 Experiment 2 and 3**

Experiments 2 and 3, investigated the McGurk effect on different language and lexical properties. The same cross-cultural groups of participants were tested. The results of four hypotheses are discussed below.

### **5.2.1 Hypothesis 1**

For Hypothesis 1 we hypothesized that for the Chinese LANGUAGE, the Chinese Groups (Group 1 and Group 2) would produce a higher McGurk rate than the American Group (Group 3); In the English LANGUAGE, the American Group (Group 3) was expected to generate the highest McGurk rate, followed by the Chinese Group who had been in the United States of America for longer than three years (Group 2), with the Chinese Group who had been in the United States of America for less than three years (Group 1) having the smallest degree of the McGurk effects (Sekiyama 1997; Sekiyama

and Burnham 2008).

The results included a significant GROUP effect, with Group 3 producing a lower McGurk rate than either Group 1 or Group 2. A main effect of LANGUAGE was also found, indicating when the language was Chinese lower ratio of McGurk effect occurred than when the language was English (Figure 9). However no GROUP by LANGUAGE interaction was found. The results were in contrary with the results from Sekiyama and Tohkura's study (Sekiyama and Tohkura 1993), in which American English speakers generated a larger McGurk effect than Japanese participants. As mentioned above, the low McGurk rate in the American participants might have occurred because the American participants were recruited from a Department of Psychological and Brain Sciences where they might have been exposed to information about the effect and consequently paid more attention to the auditory stimuli.

The effect of native language could have played a role as well. In the Chinese groups, they produced a lower McGurk effect for their native language and a greater effect for their non-native language. This finding was consistent with Sekiyama's studies (Sekiyama and Tohkura 1993; Sekiyama and Burnham 2008), in which non-native language generated a higher visual influence in the McGurk conditions.

In the American group, participants produced a greater McGurk effect for their native language than for their non-native language. This finding was opposite to those reported in previous studies. So far there was no theoretical reason that could explain the difference. Further studies will need to be conducted to confirm these findings.

No significant difference was found between the two Chinese groups, who had different length of English language experience. This finding was contrary to that

reported in Sekiyama's study (Sekiyama 1997) in which the length of residence in Japan of Chinese participants was correlated with a higher McGurk effect in Japanese. Additional correlations were assessed between the years of residence and the average visual influence of each McGurk stimulus for the subjects in the two Chinese groups. However no significant correlation was found. Combined with the result from Experiment 1, Hypothesis 4, the study failed to support previous findings that language experience modulates the McGurk effect on behavioral responses.

### **5.2.2 Hypothesis 2**

In Hypothesis 2 we expected that the mixed stimuli that produced a fused perception of a word would induce a higher McGurk rate than the mixed stimuli that produced a fused perception of non-words. The word/non-word effect was expected to affect Chinese groups in the Chinese LANGUAGE and all three groups in the English LANGUAGE with the same order of degrees as in Hypothesis 1: highest in the American Group (Group 3), next highest in the Chinese Group who had been in America for longer than three years (Group 2), and lowest in the Chinese Group who had been in the United States of America for less than three years (Group 1) (Brancazio and Brancazio 2004).

The results included a significant auditory lexical influence by language interaction, indicating that the auditory lexical influence decreased the McGurk rate when the stimulus language was English (the word condition had a lower McGurk rate than the nonword condition); in contrast, it increased the McGurk rate when the stimulus language was Chinese (the word condition had a higher McGurk rate than the nonword condition) (Figure 7). The negative influence on English was consistent with previous work

(Brancazio and Brancazio 2004). It suggested that when the auditory information was a word, participants would be more likely to choose the auditory perception as the final response, thereby diminishing the McGurk effect.

However, the positive influence when the stimulus language was Chinese was contrary to the findings of previous studies. This discrepancy could be due to the different acoustic properties of the stimuli. Four sets of lexical combinations were manipulated in the present study: (1) both auditory and visual stimuli were words; (2) the auditory stimulus was a word and visual stimulus was a nonword; (3) the auditory stimulus was a nonword and the visual stimulus was a word; and (4) both auditory and visual were used as nonwords. Only one set of words/nonwords was used for each lexical combination for each language. Consequently, the lexical properties were confounded with the acoustic properties of the sounds. In future work, using multiple sets of words for every lexical combination would control for this factor.

The results included a significant language by visual lexical influence interaction, indicating that when the stimulus language was Chinese, the visual influence had a positive influence on the McGurk rate. This means that when the visual lexical condition was a word, participants were more likely to choose the visual stimulus as their response. This finding is consistent with earlier findings reported by Brancazio et al (2004). However, the effect was not supposed to occur for American participants because they had no previous knowledge of Chinese. Such a result supports the notion that the lexical properties were once again confounded with the acoustic properties of the sound.

When the stimulus was English, the visual-lexical influence was not significant. Thus, when the video stimulus was a word, participants were not more likely to rely on

the visual modality. This result was inconsistent with Brancazio et al (2004). Further study including more sets of stimuli should be conducted to further clarify these issues.

### **5.2.3 Hypothesis 3**

For Hypothesis 3 we hypothesized that if the auditory source of the McGurk condition was a word (such as /bench/) and the visual source generated a nonword (such as /dench/), then the McGurk nonword condition would elicit a more negative peak around 400 ms (N400 with the peak latency during 350 to 450 ms) than the audiovisual word condition (such as audiovisual /bench/). This should only have happen in the Chinese groups (Group 1 and 2) when the language was Chinese (Kutas and Hillyard 1980; Kutas and Hillyard 1984; Friederici 2002). The hypothesis was supported.

Factor analysis indicated that Factor 4 encompassed variability closest to N400 peak with the effective time ranging from 292 to 380 ms and with the highest peak at 338 ms. The latency shift was very likely due to the offset issue discussed in session 5.1.3. As illustrated in Figure 10, if the latency of the first peak was adjusted from 56 ms to 100 ms, Factor 4 would be adjusted to 382 ms, and that was within the latency range of N400.

Analysis of the CONDITION by HEMISPHERE by GROUP interaction indicated that the Chinese group generated a more negative N400 response to the McGurk condition than to the standard condition for all the electrode sites over the left hemisphere, but not over the right hemisphere. However no condition difference was found in American group over either hemisphere.

Previous studies indicated that N400 was more prominent for nonwords than words (Bentin, McCarthy et al. 1985; Holcomb 1988; Kounios and Holcomb 1994), but

no one had studied the perception of word/nonword under the McGurk condition. In the present study, the Chinese participants perceived the stimulus as a nonword even though the auditory stimulus was a word (/bie3/) and the video stimulus a nonword (/die3/). The Chinese participants generated more negative N400 amplitudes in response to the McGurk condition than in response to the standard condition. This finding supports the notion that the visual stimulus influenced the auditory processing on the perceptual, cognitive and neuro-electrophysiological levels in the Chinese participants.

#### **5.2.4 Hypothesis 4**

Hypothesis 4 we predicted that if the auditory source of the McGurk condition was a word (such as /bench/) and the visual source generated a nonword (such as /dench/), then the McGurk nonword condition would produce a more negative N400 peak amplitude (with the peak latency during 350 to 450 ms) than the audiovisual word condition (such as audiovisual /bench/). This effect was expected to occur across all three groups when the language was English, but the degree of the negativity was expected to be the largest in the American group (Group 3), followed by the Chinese group who have been in the United States of America for more than three years (Group 2); and the smallest in the Chinese group (Group 1) who had been in the United States of America for less than three years (Kutas and Hillyard 1980; Kutas and Hillyard 1984; Sekiyama and Tohkura 1993; Sekiyama 1997; Friederici 2002; Sekiyama and Burnham 2008).

The Factor analysis failed to locate any component with a peak latency that overlapped the 400 ms period post stimulus onset. However, analysis of a component that occurred from 88 to 340 ms identified a significant Condition by Hemisphere interaction.

Further analyses indicated that ERPs elicited during the McGurk condition generated a larger negative amplitude than the standard condition over the left hemisphere. However over the right hemisphere, the standard condition produced a larger negative amplitude than the McGurk condition.

The factor analysis did not identify any ERP component that occurred around 400 ms. This latency difference could be due to several reasons. First, /dench/ might have been perceived as a word instead of a nonword. For example, some participants mentioned that /dench/ could be a word. Some dictionaries even define 'Dench' as a name. Names occur in everyday life just like words, so if the participants perceived 'Dench' as a name, their brainwave would not generate a N400 nonword characteristic peak. As a consequence, the negativity found between 88 and 340 ms could be due to difference in perception of the McGurk and the standard condition. If this alternative explanation is correct, these results would be in line with the results from Experiment 1, Hypothesis 3 where only N1 was modulated by the McGurk effect. In that case, the McGurk condition produced a smaller N1 amplitude over the left hemisphere. In Experiment 3, hypothesis 4, we found that McGurk condition generated a smaller negative amplitude on the time region from 88 to 340 ms. This finding was consistent with a previous study (Lebib, Papo et al. 2004), in which the McGurk effect generated a larger negative amplitude than the McGurk condition centered around 300 ms over frontal, central and parietal scalp regions. This effect did not occur in Experiment 1, perhaps due to the fact that the McGurk effect was relatively weak, especially for Group 3. The inclusion of non-McGurk trials could have reduced some condition differences. In Experiment 3, the McGurk effect was fairly strong for the ewnm condition. Given this

finding, it should be possible to isolate the participants with a strong McGurk effect, and include only the trials on which they demonstrated the McGurk effect.

No GROUP related effects were significant in the analysis of Factor 2 during the 88 to 340 ms period. Throughout all the experiments, the group effect was not significant across a number of experiments (Experiment 1, hypothesis 4, 5 and 6; Experiment 2 and 3, hypothesis 2 and 4). The only group effect found occurred when comparing both Chinese groups with the American group for Experiments 2 and 3, hypotheses 1 and 3). None of the analyses were able to differentiate between the two Chinese groups, a finding inconsistent with prior research. In previous behavioral studies, the McGurk effect correlated with a longer residence in a foreign country (Sekiyama 1997); an English-speaking environment also generated a larger McGurk effect than a non-English speaking environment (Sekiyama and Burnham 2008).

The significantly different PPVT scores between the two Chinese groups indicated a difference in their English proficiency. However PPVT did not correlate with the magnitude of the McGurk effect. Sekiyama (1997) previously noted that native Chinese speakers who had lived in Japan for 4 months to 6 years exhibited a significant positive correlation with the McGurk effect and the length of residence in Japan. In the current study, the length of residency in the United States of the native Chinese speakers ranged from 6 months to 10 years, but no such correlation was found. One possible contributing factor for this discrepancy could be variations in the English proficiency levels between the Chinese groups in the two studies. In Sekiyama's study, even though the native Chinese speakers learned English since they were 12 years old, many of them claimed that they did not speak English (Sekiyama 1997). In the current study, all the

Chinese participants started learning English no later than 12 years of age, and most of them passed the TOEFL and GRE exams, indicating a reasonably high level of English proficiency. Thus, it is likely that the experiment reached a ceiling effect for the McGurk effect in both groups of Chinese participants, especially given that the McGurk effect found in the experiment was higher in the native Chinese speakers than the native English speakers. Further research is needed to investigate this phenomenon in native Chinese speakers with low English proficiency.

### **5.3 Summary and Conclusions**

The first major finding of the present study was that a hemisphere difference was observed as early as 50 ms during the McGurk condition compared with the standard condition. With speech perception and McGurk effect thought to demand high level (e.g., cortical) functionality, it was expected that the integration would occur later in time. However the present results indicate that the brain is triggered very early in the process. In addition, the study was conducted using a direct comparison between conditions and it generated the similar results from previous studies using the oddball paradigm, thereby strongly indicating that this early process is not an artifact of the paradigm being utilized.

The second major finding of the present study was evident in the interaction between the McGurk effect and the lexicon properties leveling an analysis of the ERP data. Although previous studies provided behavioral evidence, the present study is the first one to provide the electrophysiological evidence of this interaction.

The hemisphere difference found in Experiment one had an opposite pattern in the left and right hemispheres. Although no dipole or source analyses were conducted, such a finding is consistent with the presence of a unilateral dipole. Further study with source

analysis should elucidate the location of the dipole, so that the results can be compared with the those from previously conducted fMRI studies.

The present study confirmed that if a word instead of a nonword was presented in the auditory channel (auditory lexical influence), the McGurk rate was decreased in the English language. However, it also indicated the auditory lexical influence had an opposite effect in the Chinese language users, a finding inconsistent with previous reports. This unexpected result might be due to different acoustic properties of the stimuli. The study found that the visual lexical influence had a positive impact on the Chinese stimulus for both groups. The finding in the Chinese groups was consistent with previous research, but the finding in the American group was not expected. It could again be related with the physical difference instead of the lexical nature of the sounds.

N400 was detected in the Chinese stimulus conditions. It was found to differ between the nonword and word condition in the Chinese groups, but not in the American group. The perceived nonword generated a characteristic larger N400 than the audio word condition. The effect occurred during the 292 to 380 ms period over most of the electrode sites and scalp regions sampled. This ERP finding was consistent with previous reports that the McGurk effect interacted with cognitive processing. Importantly, this ERP study provided the first electrophysiological evidence for such an interaction. The manipulation of N400 in the English stimulus condition was not successful, perhaps because participants might have interpreted the designated nonword as a word (e.g. 'Dench') in some situations.

Neither behavioral nor electrophysiological differences of the two Chinese groups were found in Experiment 2 and 3. These findings contradict previous reports that an

English environment promotes audiovisual integration. This result also contradicted the previous findings that the McGurk effect could serve as an indicator of second language proficiency since no differences were noted between Chinese speakers who differed in their exposure to spoken English. It is recommended that future studies have a more delimited and a better defined period of language experience and document in a more standardized manner the level of linguistic expertise in the different language groups.

This study provides electrophysiological evidence for the interaction between perception and cognition. The results suggest that the McGurk effect modulated the ERP brainwave during both conditions. This study also provides some insights into the audiovisual integration occurring in the Chinese and English languages. The results of the study suggest that the Chinese language activated a lower degree of the McGurk effect than the English language. However, native Chinese speakers produced a higher degree of the McGurk effect than the native English speakers, on both Chinese and English. Even though the two Chinese groups differed in their English skills, they did not differ in any McGurk effect related tests. It was suspected that the McGurk effect would reach a ceiling after a certain level of second language proficiency. A changed neurological procedure in language learning may occur when the McGurk effect reaches the ceiling. As an additional control, future studies should be conducted on the native Chinese speakers with minimal English exposure to verify the hypothesis.

## REFERENCES

- Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809-823.
- Beauducel, A., & Debener, S. (2003). Misallocation of variance in event-related potentials: simulation studies on the effects of test power, topography, and baseline-to-peak versus principal component quantifications. *J Neurosci Methods*, *124*(1), 103-112.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalogr Clin Neurophysiol*, *60*(4), 343-355.
- Brancazio, L., & Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *J Exp Psychol Hum Percept Perform*, *30*(3), 445-463.
- Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport*, *10*(12), 2619-2623.
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593-596.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol*, *10*(11), 649-657.
- Chapman, R. M., & McCrary, J. W. (1995). EP component identification and measurement by principal components analysis. *Brain Cogn*, *27*(3), 288-310.
- Cienkowski, K. M., & Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear Hear*, *23*(5), 439-449.
- Colin, C., Radeau, M., Soquet, A., & Deltenre, P. (2004). Generalization of the generation of an MMN by illusory McGurk percepts: voiceless consonants. *Clin Neurophysiol*, *115*(9), 1989-2000.

- 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. *Clin Neurophysiol*, *113*(4), 495-506.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test-third edition*.
- Friederici, A. D. (1995). The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. *Brain Lang*, *50*(3), 259-281.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn Sci*, *6*(2), 78-84.
- Gentilucci, M., & Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. *Exp Brain Res*, *167*(1), 66-75.
- Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci*, *25*(20), 5004-5012.
- Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci*, *11*(5), 473-490.
- Gilley, P. M., Sharma, A., Dorman, M., & Martin, K. (2005). Developmental changes in refractoriness of the cortical auditory evoked potential. *Clin Neurophysiol*, *116*(3), 648-657.
- Giraud, A. L., & Truy, E. (2002). The contribution of visual areas to speech comprehension: a PET study in cochlear implants patients and normal-hearing subjects. *Neuropsychologia*, *40*(9), 1562-1569.
- Gomes, H., Sussman, E., Ritter, W., Kurtzberg, D., Cowan, N., & Vaughan, H. G., Jr. (1999). Electrophysiological evidence of developmental changes in the duration of auditory sensory memory. *Dev Psychol*, *35*(1), 294-302.
- Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philos Trans R Soc Lond B Biol Sci*, *363*(1493), 1055-1069.
- Hall, D. A., Fussell, C., & Summerfield, A. Q. (2005). Reading fluent speech from talking faces: typical brain networks and individual differences. *J Cogn Neurosci*, *17*(6), 939-953.
- Holcomb, P. J. (1988). Automatic and attentional processing: an event-related brain potential analysis of semantic priming. *Brain Lang*, *35*(1), 66-85.

- Kislyuk, D. S., Mottonen, R., & Sams, M. (2008). Visual Processing Affects the Neural Basis of Auditory Discrimination. *J Cogn Neurosci*.
- Kotz, S. A., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychol (Amst)*, 128(3), 514-527.
- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *J Exp Psychol Learn Mem Cogn*, 20(4), 804-823.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161-163.
- Lebib, R., Papo, D., Douiri, A., de Bode, S., Gillon Dowens, M., & Baudonniere, P. M. (2004). Modulations of 'late' event-related brain potentials in humans by dynamic audiovisual speech stimuli. *Neurosci Lett*, 372(1-2), 74-79.
- Lenneberg, E. H. (1967). The biological foundations of language.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36.
- Liu, Y., Shu, H., & Wei, J. (2006). Spoken word recognition in context: evidence from Chinese ERP analyses. *Brain Lang*, 96(1), 37-48.
- MacDonald, J., Andersen, S., & Bachmann, T. (2000). Hearing by eye: how much spatial degradation can be tolerated? *Perception*, 29(10), 1155-1168.
- MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Percept Psychophys*, 24(3), 253-257.
- Massaro, D. W., Cohen, M. M., & Smeele, P. M. (1995). Cross-linguistic comparisons in the integration of visual and auditory speech. *Mem Cognit*, 23(1), 113-131.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748.
- McPherson, W. B., Ackerman, P. T., Holcomb, P. J., & Dykman, R. A. (1998). Event-related brain potentials elicited during phonological processing differentiate subgroups of reading disabled adolescents. *Brain Lang*, 62(2), 163-185.

- McPherson, W. B., Ackerman, P. T., Oglesby, D. M., & Dykman, R. A. (1996). Event-related brain potentials elicited by rhyming and non-rhyming pictures differentiate subgroups of reading disabled adolescents. *Integr Physiol Behav Sci*, 31(1), 3-17.
- McPherson, W. B., Newton, J. E., Ackerman, P., Oglesby, D. M., & Dykman, R. A. (1997). An event-related brain potential investigation of PTSD and PTSD symptoms in abused children. *Integr Physiol Behav Sci*, 32(1), 31-42.
- Molfese, D. L. (1978). Neuroelectrical correlates of categorical speech perception in adults. *Brain Lang*, 5(1), 25-35.
- Molfese, D. L., Alexandra, P., Fonaryova, K., & Macguire, M. J. (2005). Event-related potentials in speech perception. *The Handbook of Speech Perception*, .
- Molfese, D. L., & Molfese, V. J. (1988). Right-hemisphere responses from preschool children to temporal cues to speech and nonspeech materials: electrophysiological correlates. *Brain Lang*, 33(2), 245-259.
- Naatanen, R. (2003). Mismatch negativity: clinical research and possible applications. *Int J Psychophysiol*, 48(2), 179-188.
- Picton, T. W., & Taylor, M. J. (2007). Electrophysiological evaluation of human brain development. *Dev Neuropsychol*, 31(3), 249-278.
- Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci*, 18(6), 2188-2199.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu Rev Neurosci*, 27, 169-192.
- Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Percept Psychophys*, 59(3), 347-357.
- Saint-Amour, D., De Sanctis, P., 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(3), 587-597.
- Sallinen, M., Kaartinen, J., & Lyytinen, H. (1994). Is the appearance of mismatch negativity during stage 2 sleep related to the elicitation of K-complex? *Electroencephalogr Clin Neurophysiol*, 91(2), 140-148.
- Schiller, N. O., Horemans, I., Ganushchak, L., & Koester, D. (2008). Event-related brain potentials during the monitoring of speech errors. *Neuroimage*.

- Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: the McGurk effect in Chinese subjects. *Percept Psychophys*, 59(1), 73-80.
- Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Dev Sci*, 11(2), 306-320.
- Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, 21(4), 18.
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing Lips and Seeing Voices: How Cortical Areas Supporting Speech Production Mediate Audiovisual Speech Perception. *Cereb Cortex*.
- Stein, B. E., Meredith, M. A., & Wallace, M. T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Prog Brain Res*, 95, 79-90.
- Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci*, 19(12), 1964-1973.
- Sumbly, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, 26(2), 4.
- Sussman, E., Steinschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hear Res*, 236(1-2), 61-79.
- van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *J Cogn Neurosci*, 16(6), 1068-1084.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A*, 102(4), 1181-1186.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598-607.
- Wood, C. C., & McCarthy, G. (1984). Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. *Electroencephalogr Clin Neurophysiol*, 59(3), 249-260.

# APPENDIX---- Consent Form

UNIVERSITY of LOUISVILLE

Health Sciences Center  
UNIVERSITY OF LOUISVILLE  
BIRTH DEFECTS CENTER

BIRTH DEFECTS CENTER  
MOLFESE LAB

School of Dentistry  
University of Louisville  
Louisville, Kentucky 40292

Office: 502-852-8809  
502-852-8809  
Fax: 502-852-4702  
www.louisville.edu/hsc/  
birthdefectcenter/

IRB # 5-29-07 Valid Thru 5-28-09

## Subject Informed Consent Document

### ELECTROPHYSIOLOGICAL INVESTIGATION OF SPEECH PERCEPTION AND THE MCGURK EFFECT

IRB assigned number: 24 2,07  
Investigator(s) names: Dr. Dennis L. Molfeese and Jia Wu, M.A.  
Site(s) where study is to be conducted: University of Louisville  
Phone number for subjects to call for questions: (502) 852-2512

#### Introduction and Background Information

You are invited to participate in a research study. The study is being conducted by Dr. Dennis L. Molfeese and Jia Wu, M.A. The study is sponsored by the University of Louisville's Birth Defects Center. The study will take place at the University of Louisville. Approximately 60 subjects will be invited to participate.

#### Purpose

The purpose of this study is to investigate the brain processing of audiovisual speech perception in adult subjects. This study will compare brain activity of American and Chinese adults when processing congruent and incongruent speech stimuli under different language backgrounds. This information will help further research regarding the strategies of learning a new language.

#### Procedures

In this study, you will be asked to complete questionnaires that provide information on how well you perceive audio and visual information. You are free to decline to answer any particular questions that make you uncomfortable. Additionally, a full-scale IQ test will be administered to assess your English competence level and cognitive capability. Once this information has been collected, an electrode net will be placed on your head. This will involve first measuring your head to determine where the recording electrodes will be placed. Prior to placement, the net will be soaked in warm saline solution. Once the electrodes are in place, auditory and visual stimuli will be presented to you. The task will involve listening to spoken stimuli and watching lip movement and then deciding whether the stimuli presented follow a specified rule. The entire experiment will last approximately two hours.

#### Potential Risks

There are no foreseeable risks, although there may be unforeseen risks.

### **Benefits**

The possible benefits of this study include gathering a better understanding of the brain's involvement in speech perception in both the American and Chinese populations. The information collected may not benefit you directly. The information learned in this study may be helpful to others.

### **Compensation**

You will be compensated for your time, inconvenience, or expenses for your participation in this study. Upon completion of the study, you will be compensated \$20.

### **Confidentiality**

Total privacy cannot be guaranteed. Your privacy will be protected to the extent permitted by law. If the results from this study are published, your name will not be made public. While unlikely, the following may look at the study records:

The University of Louisville Institutional Review Board, Human Subjects Protection Program Office  
Office for Human Research Protections (OHRP),  
Office of Civil Rights and  
Research personnel affiliated with the collection of data

### **Voluntary Participation**

Taking part in this study is voluntary. You may choose not to take part at all. If you decide to be in this study you may stop taking part at any time. If you decide not to be in this study or if you stop taking part at any time, you will not lose any benefits for which you may qualify.

### **Research Subject's Rights, Questions, Concerns, and Complaints**

If you have any concerns or complaints about the study or the study staff, you have three options.

You may contact the principal investigator at (502) 852-2512.

If you have any questions about your rights as a study subject, questions, concerns, or complaints, you may call the Human Subjects Protection Program Office (HSPPPO) (502) 852-5188. You may discuss any questions about your rights as a subject, in secret, with a member of the Institutional Review Board (IRB) or the HSPPPO staff. The IRB is an independent committee composed of members of the University community, staff of the institutions, as well as lay members of the community not connected with these institutions. The IRB has reviewed this study.

If you want to speak to a person outside the University, you may call 1-877-852-1167. You will be given the chance to talk about any questions, concerns or complaints in secret. This is a 24 hour hot line answered by people who do not work at the University of Louisville.

---

Consent for the use of photography or video during the testing session and in our scholarly efforts:

\_\_\_\_\_ Yes, I do permit the use of photos and video to be used and for my face not to be masked.

\_\_\_\_\_ Yes, I do permit the use of photos and video to be used and for my face to be masked.

\_\_\_\_\_ No, I do not permit the use of photography or video recording to occur during the session.

---

This paper tells you what will happen during the study if you choose to take part. Your signature means that this study has been discussed with you, that your questions have been answered, and that you will take part in the study. This informed consent document is not a contract. You are not giving up any legal rights by signing this informed consent document. You will be given a signed copy of this paper to keep for your records.

---

Signature of Subject/Legal Representative	Date Signed
Signature of Person Explaining the Consent Form (if other than the Investigator)	Date Signed
Signature of Investigator	Date Signed

---

LIST OF INVESTIGATORS	PHONE NUMBERS
Dr. Dennis L. Molfese	(502) 852-2512
Jia Wu, M.A.	(502) 852-2512

262.07  
UNIVERSITY OF LOUISVILLE  
INSTITUTIONAL REVIEW BOARD  
Date App'd 5-29-07 Valid Thru 5-28-08

## **CURRICULUM VITA**

Name: Jia Wu

Address: 550 Whitney Ave. Apt 6, New Haven, CT, 06511

Email: [jia.wu@yale.edu](mailto:jia.wu@yale.edu)

Phone: 626-679-6247

### **EDUCATION & TRAINING**

B.S., Life Sciences  
Fudan University  
1997-2001

M.A., Experimental Psychology  
University of Louisville  
2002-2005

Ph.D., Experimental Psychology  
University of Louisville  
2005-2009

Post Doctoral Associate, Child Study Center  
Yale University School of Medicine  
2007-current

### **AWARDS**

Conference Travel Grant  
University of Louisville  
2006

New Student Fellowship  
University of Louisville  
2002

People's Scholarship of Fudan University  
Fudan University  
1997-2001

### **TEACHING EXPERIENCE**

Physiological Psychology & lab  
Hanover College

Spring 2007

Methods In Experimental Psychology  
University of Louisville  
Spring 2006 and Spring 2005

Quantity Methods In Psychology  
University of Louisville  
Fall 2005 and Fall 2004

## PRESENTATION AND PUBLICATION

- Crowley, M. J, Wu, J., Crutcher, C., Bailey, C. A., Lejuez, W. C., Mayes, L. C., (in press), Risk-Taking and the Feedback Negativity Response to Loss Among At-risk Adolescents, *Developmental Neuroscience*.
- Wu, J., Dykman, R., & Molfese, D. L. (June 2007). Sleep Restriction Decreased ERP Brainwave In Directional Stroop Task. Slide presentation accepted on the 21<sup>st</sup> Annual Meeting of the Associated Professional Sleep Societies, Minneapolis, Minnesota.
- Wu, J., Molnar, A., Wagner, M., Waford, R., Warren, C., & Molfese, D. L. (February, 2007). Sex differences in attention across ERP and near-infrared procedures on the same participants. Poster presented on the 35<sup>th</sup> Annual Conference of International Neuropsychological Society, Portland, Oregon.
- Wu, J., Yoder P.J., Molfese, D. L., Dykman, A.R., (October, 2006). Comparison of Speech-Perception in SLI and Normal Children, Slide presentation on the 36<sup>th</sup> Annual Conference of Society for Neuroscience (SFN), Atlanta, Georgia.
- Wu, J., Molnar, A., Wagner, M., Waford, R., Warren, C., & Molfese, D. L. (October, 2006). Sex differences in oddball task across ERP and Near-infrared methodology. Poster presented at Research!Louisville, Louisville, Kentucky.
- Molnar, A., Wagner, M., Wu, J., Waford, R., & Molfese, D. L. (October, 2006). Spatio-temporal characteristics of auditory oddball processing using ERP and near-infrared procedures. Poster presented at Research!Louisville, Louisville, Kentucky.
- Molfese, D.L., Wu, J., Wagner, M., Molnar, A., Wafford, R. (August 2006). The use of high-density array event-related potentials and near-infrared technologies to provide spatio-temporal insights into language processing. Paper Presented on Conference of the International Society of Oxygen (ISOTT-2006), Louisville, Kentucky
- Wu, J., Molfese, D.L., (April 2006). Removal of ERP Artifacts in Newborn Infants Using Independent Component Analysis, Paper presented at Annual Conference of Human

Development, Louisville, Kentucky

Molfese, D.L., Key, A.F., Peach, K., Weittig, C., Ratajczak, E., Ferguson, M., Straub S., & Wu, J., (October 2004). Changes In Brain Organization For Language In The First Two Years Of Life: Electrophysiological Evidence, Poster Presentation, Department of Psychological and Brand Sciences, Poster presented at Research!Louisville, Louisville, Kentucky

Molfese, D.L., Key, A.F., Peach, K., Ratajczak,E., Ferguson, M., Straub.S., &Wu, J., (October 2003). Sex-Related Memory Difference For Complex Tone Sounds In 1-Year-Old Infants Poster presentation, Department of Psychological and Brand Sciences, University of Louisville, Poster presented at Research!Louisville, Louisville, KY