Cross-Lingual Diphthong Perception: A Simultaneous EEG/fMRI Investigation

David Olonzo Sorensen

Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Social and Behavioral Sciences Commons

BYU ScholarsArchive Citation
https://scholarsarchive.byu.edu/etd/7602

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
Cross-Lingual Diphthong Perception: A Simultaneous EEG/fMRI Investigation

David Olonzo Sorensen

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

Master of Science

David L. McPherson, Chair
Christopher B. Kirwan
Michael Robert Stark

Neuroscience Center
Brigham Young University

Copyright © 2018 David Olonzo Sorensen
All Rights Reserved
ABSTRACT

Cross-Lingual Diphthong Perception: A Simultaneous EEG/fMRI Investigation

David Olonzo Sorensen
Neuroscience Center, BYU
Master of Science

Previous research indicates that humans develop a phonological library in infancy. As humans grow into adulthood, their phonological library becomes well established. Upon encountering phonemes from a new language, humans process these phonemes by comparison to their native phonological library. Event-related potentials (ERP), specifically the mismatch negativity, have been shown to indicate that this process of comparing non-native phonemes to our native phonological library is not improved through learning the new language as an adult. An alternative explanation may be that there is an underlying change in the neural generators as the non-native phonemes are learned, but that this change is not reflected in the ERP. The current study seeks to examine this hypothesis through the simultaneous collection of ERP and blood-oxygen-level-dependent functional MRI (fMRI) data. The findings of the ERP and fMRI data are inconclusive. The study also explores the processing of diphthongs, a category of phonemes rarely tested before, through both behavioral and neuroimaging methods. The study presents behavioral data demonstrating that non-native diphthongs are processed based upon the separate elements of the phonemes, rather than as complete units.

Keywords: diphthong, simultaneous EEG/fMRI, mismatch negativity, phoneme perception
ACKNOWLEDGEMENTS

I would like to thank Richard Harris for help recording the stimuli used throughout the experiment. I also express gratitude to the BYU MRI Research Facility and staff that made collection of simultaneous EEG and functional MRI data possible. Finally, I thank my graduate committee, especially my chair, David McPherson, who has served as a mentor to me throughout this experience.
# TABLE OF CONTENTS

List of Figures ......................................................................................................................... v

Introduction .............................................................................................................................. 1

- Perceptual Assimilation Model ......................................................................................... 2
- Mismatch Negativity ............................................................................................................ 3
- Simultaneous EEG/fMRI ....................................................................................................... 4
- Khmer: A Unique Language .............................................................................................. 5

Methods .................................................................................................................................... 6

- Behavioral Discrimination Study ....................................................................................... 7
- Neuroimaging Study .............................................................................................................. 10

Results .................................................................................................................................... 14

- Stimuli ................................................................................................................................. 14
- Behavioral Discrimination Study ....................................................................................... 15
- Neuroimaging Study .............................................................................................................. 17

Discussion .............................................................................................................................. 22

- Behavioral Discrimination Study ....................................................................................... 22
- Neuroimaging Study .............................................................................................................. 23
- Conclusion .......................................................................................................................... 26

References .............................................................................................................................. 28

Appendix A ................................................................................................................................ 33

Appendix B ................................................................................................................................ 35
LIST OF FIGURES

Figure 1 ........................................................................................................................................... 8
Figure 2 ......................................................................................................................................... 16
Figure 3 ......................................................................................................................................... 17
Figure 4 ......................................................................................................................................... 19
Figure 5 ......................................................................................................................................... 20
Figure 6 ......................................................................................................................................... 21
Figure 7 ......................................................................................................................................... 24
Speech processing is vital to everyday life. Our everyday interactions with others depend on our ability to comprehend speech. The brain needs to be able to correctly determine meaning from the often incomplete and noisy speech stimulus produced by the speaker. While numerous stages of speech processing occur, at its most basic level, speech processing involves extracting and categorizing phonemes, the basic sounds of speech, from the continuous stream of sounds produced. Errors in phonemic categorization by the recipient can cause speech to be completely misunderstood, as is the case with mondegreens (e.g. “Jose, can you see” as the first line of “The Star-Spangled Banner”).

Phonemic categorization is also integral to speech production (Flege, MacKay, & Meador, 1999; Sebastian-Galles & Baus, 2005). Production of speech is reliant upon the phonemes that comprise the phonological representation of words. When the phonological representation is imperfect—in other words, the phonemes are not categorized correctly—errors occur in speech production that can affect meaning. Poor phonemic categorization by non-native speakers can, in part, explain foreign accents that cause native listeners difficulty in comprehending non-native speakers (Sebastian-Galles & Baus, 2005). Thus, phoneme perception and categorization are vital processes of our language system. Yet the neural substrates underlying this phonemic system are not well understood, particularly in the case of multilingual individuals, who must be able to process varying phoneme sets from multiple languages. This thesis reviews the existing research regarding individuals’ ability to perceive phonemes of non-native languages, and presents the findings of an exploratory study aimed at finding neural activity and substrates responsible for this perception.
Perceptual Assimilation Model

The perceptual assimilation model (PAM) is a working model that describes and makes predictions about how listeners perceive sounds of a non-native language. The PAM is derived from a direct realist view that supposes speech stimuli are encoded as speech gestures, not sounds, but the difference is insignificant as it applies to the proposed research. As described by Best (1995), the PAM predicts that unfamiliar, non-native sounds are compared to the learned set of native sounds. Resulting from this comparison is an assignment of the non-native sound as an exemplar of a native speech phonemes (with varying degrees of goodness), an uncategorized speech phoneme, or a non-speech sound. This categorization is dependent upon the available native phonology. That is, native phonological categories either include the non-native sound, exclude the sound but have surrounding categories suggesting the non-native sound as possible, or exclude the sound and have no similar sounds to suggest such a sound is possible for speech. The PAM thus describes six types of non-native speech contrasts between two non-native phonemes:

1. Two-Category Assimilation: non-native sounds are both categorized as exemplars of distinct native phonemes.

2. Category-Goodness Difference: non-native sounds are both categorized as exemplars of the same native phoneme. However, they differ in the goodness of fit to the ideal exemplar of the phoneme.

3. Single-Category Assimilation: non-native sounds are both categorized as exemplars of the same native phoneme with the same variance from the ideal.

4. Two Uncategorizable: non-native sounds are both categorized as phonological sounds not belonging to native phonological categories.
5. Uncategorized versus Categorized: one non-native speech sound is categorized as an exemplar of a native phoneme. The other is categorized as not belonging to native phonology, but still an acceptable speech sound.

6. Nonassimilable: non-native sounds are both categorized as non-speech sounds.

These different non-native speech contrasts have different levels of discrimination. Single-category assimilation contrasts have the worst discrimination, as both non-native phonemes are perceived as being equidistant from the same native phoneme. Two-category assimilation contrasts likely produce the best discrimination, as the two non-native phonemes are matched to contrastive native phonemes (Best, 1995).

Mismatch Negativity

To examine these claims of the PAM at the level of brain activity, the current study utilized the mismatch negativity (MMN). The MMN is an auditory event-related potential (ERP). It is most commonly evoked by a passive oddball paradigm. In a passive oddball paradigm, subjects engage in a task while auditory stimuli not relevant to the task are presented. These background stimuli consist of standard stimuli, which are presented more often, and deviant stimuli, which are presented intermittently. The MMN is evoked when the deviant stimulus is perceived. It appears between 150 to 250 milliseconds following the presentation of the deviant stimulus. The MMN has been described as a pre-attentive response to the change in stimulus, and can be evoked by tonal stimuli as well as phonemic stimuli.

The MMN has been used frequently to study speech processing in bilingualism. Because phonemes are processed categorically, phonemic contrasts elicit the MMN. This has been shown for monophthong vowel contrasts (Diaz, Baus, Escera, Costa, & Sebastian-Galles, 2008; Ylinen, Shestakova, Huotilainen, Alku, & Naatanen, 2006), consonant-vowel contrasts (Bomba, Choly,
& Pang, 2011), and even lexical tone (Kaan, Wayland, Bao, & Barkley, 2007). Thus, paired with the assumptions of the PAM model, the MMN should have different responses for the different non-native contrasts. The MMN should be best elicited by Two-category Assimilation vowel contrasts, which are predicted to have excellent discrimination. Such differences have been demonstrated for Two-category Assimilation and Uncategorized-Categorized Assimilation (Grimaldi et al., 2014). Other categories of non-native contrasts have yet to be tested in the same way. Even more importantly, the locus of differential neural activity producing the decreased MMN in poorer contrasts has yet to be described.

**Simultaneous EEG/fMRI**

There are two commonly utilized techniques to examine the activity of the whole brain in humans. The first, electroencephalography or EEG, measures the electrical activity of the brain. Recording electrodes are applied to the scalp, and measure the electrical potential at different locations. These potentials change as neuronal activity, which is electrical in nature, occurs in the brain. Small but stable variations in this activity in response to certain stimulus types results in ERPs, such as the MMN discussed earlier. Changes in electrical potential are conducted quickly, allowing near-millisecond resolution of relevant brain activity. EEG is limited, however, as identification of brain regions responsible is speculative. One procedure commonly used to improve localization of EEG activity is the digitization of electrodes using a commercial 3D magnetic digitizer, such as the Polhemus FastTrak (Law & Nunez, 1991). This procedure is even more critical when using a realistic head model for source localization (Spinelli, Andino, Lantz, Seeck, & Michel, 2000).

The second technique, functional magnetic resonance imaging or fMRI, gives a precise picture of the brain, providing a clear identification of the location of activity. Functional
magnetic resonance imaging is based on measuring changes in blood flow to various brain regions. Because neurons are oxygen dependent, as brain activity in a region increases, localized vessels allow increased blood flow to the active area. This increased blood flow due to neuronal activity is detectable (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001); however, it occurs far more slowly than the communication of neuronal signals. Thus, fMRI provides millimeter resolution on the location of brain activity but poor temporal resolution.

One technique currently being explored is a merger of these two most common techniques. Because their strengths and weaknesses complement each other, doing them simultaneously gives us insight into both where and when brain activity occurs (Mulert et al., 2004). Much of the work that has been done using simultaneous collection of fMRI and EEG data has dealt with seizures (Krakow et al., 1999). Seizures create a strong enough level of activity to be reliably processed for both methods under the constraints of joint collection. The field is expanding, however, to look at cognitive responses to simple stimuli, which produce more mild activity such as observed in ERPs (Otzenberger, Gounot, & Foucher, 2005; Scarff et al., 2004). The research presented aimed to explore the capability of this combined technique to provide further insight into the neural substrates that generate the MMN.

Khmer: A Unique Language

For non-native speech contrasts, the current research study utilized phonemes from the Khmer, also known as Cambodian, language. Khmer represents a unique contrast to English, the native language for the bilingual speakers used in this study. Much previous work has been done with bilinguals for which both languages come from the same language family (Flege et al., 1999; Sebastian-Galles, Rodriguez-Fornells, de Diego-Balaguer, & Diaz, 2006; Sebastian-Galles et al., 2012; Weber et al., 2013). While this makes sense, as most functional bilinguals learn
related languages due to proximity, contrasts between languages of different families should also be examined.

Khmer presents a good contrast to English. Unlike many other Asian languages, Khmer is atonal. Thus, differences in perception of lexical tone (Kaan et al., 2007) did not act as a possible confounding variable. The Khmer phonology, as described by Huffman (1970), contains thirteen vocalic diphthongs, while American English is described with 5 vocalic diphthongs (Menn, 2011). Diphthongs represent phonemes that are difficult for non-native speakers to perceive and to produce. The complexity involved in gliding between vowel positions may present a unique challenge to perception. Additionally, many differences that are allophonic in English, such as /p/ and /ph/, are contrastive in Khmer. A full comparison of Khmer and English phonetic structure can be found in Appendix A.

Another advantage that Khmer has over other languages commonly utilized in second language research is its relative obscurity. When using naïve participants, it can be difficult to find individuals who have truly never been exposed to the second language before. Most American English speakers have been exposed to Spanish before, due to the relative frequency of Spanish-speakers in America. Relatively few have heard spoken Khmer before, due to the scarcity of Khmer speakers in America.

Methods

The study was conducted to address whether there are detectable differences in nonnative phoneme processing for bilingual individuals as compared to monolingual individuals. The research conducted consisted of two separate phases: a behavioral discrimination study to determine the ability of native English speakers to discriminate contrastive Khmer diphthongs,
and a neuroimaging study to examine differential neural processes underlying perceptual differences across groups with varying levels of experience with Khmer.

**Behavioral Discrimination Study**

**Stimuli.** Stimuli were recorded by a native Khmer speaker (female, age 31 years) from Phnom Penh who was living in the United States for four years and 3 months at the time of recording. The speaker used Khmer daily in her employment and to communicate with family in Cambodia.

Each of the four diphthongs tested (/ao/, /aa/, /aw/ and /iw/) were recorded in an isolated t__ context. The diphthongs /aw/ and /iw/ cannot have a word final consonant, so only a word initial consonant was included. The t__ context forms Khmer words with each of the four diphthongs. No other words or syllables were recorded. The words were recorded by having the speaker read a line of printed text on a paper suspended in front of the speaker. Each line of text consisted of one instance of each of the four words. The order of the four words was randomized in each line. Because of known speaker effects (i.e. rising on the first word and falling on the final word of each line), only words in the second and third position on each line were considered valid tokens. In total, nine lines were generated with unique order of the second and third position words, as can be seen in Appendix B. The speaker was instructed to read through all nine lines twice during recording. On the first reading, the speaker paused after each word on the line with a longer pause at the end of the line. On the second reading, the speaker read the line at a normal pace, with pauses only at the end of the line.

The word tokens in the second and third position on each line were isolated using Adobe Audition (Adobe Audition CC 2015.1; Adobe Systems, Inc.) and saved as separate files with 100 ms of the surrounding recording both before onset and after offset of the word, for a total of
200 ms of recording surrounding the word token. Within each word token, the time position when the formants reached a steady state was noted. The 100 ms preceding the steady state time point was silenced to remove the consonant, and the beginning of the file was trimmed to maintain 100 ms of fade in time before the vowel (Figure 1). A fade-out was applied to the final 100 ms of each file. All 18 tokens were normalized for loudness using the ITU standard. The two tokens that were judged by the author as most representative of each diphthong were then selected. Each of these selected tokens was edited to a length of 600 ms by removing glottal pulses. An equal number of glottal pulses were removed from each element of the diphthong to
maintain balance between the two elements of the diphthong. This resulted in 8 total stimuli, two each of /ao/, /aə/, /aw/, and /ɨw/.

**Subjects.** Subjects consisted of 18 young adult (age 18-35) monolingual American English speakers. For the purposes of this study and the neuroimaging study, individuals not conversant in a second language and at least 5 years removed from formal instruction in a second language were considered monolingual.

**Behavioral Discrimination Task.** An AX discrimination task, also called a same-different task, was used to determine the diphthong pair to which monolingual English speakers have the poorest discrimination. The task consisted of pairs of diphthongs presented with a 500 ms interstimulus interval between the two vowels in the pair. The participants were asked to respond after each pair whether the two diphthongs in the pair were the same vowel or different vowels. All stimulus presentation was done using ePrime 3 software (Psychology Software Tools) at about 50 db HL through a Grason Stadtler GSI 61 Clinical Audiometer to EarTone 3A insert earphones. Trials were separated by 1500 ms (ITI), during which time participant responses were collected. Responses were collected using a RB-540 series response pad (Cedrus Corporation, California). Participants pressed a red button on the right side of the pad to indicate the vowels were different, or a blue button on the left side of the pad to indicate the vowels were the same. No other buttons on the pad were functional during the task. A fixation cross was displayed on a monitor placed approximately 30 cm away from the participant through the duration of the task.

A total of 192 trials were presented. Every contrasting pair was presented a total of 16 times with 8 times in each configuration (e.g. 8 /ao/-/aə/ trials and 8 /aə/-/ao/ trials, for 16 total trials contrasting the /ao/ and /aə/ diphthongs). Presentations were balanced so as to provide an
equal number of "same" and "different" pairs, resulting in 24 trials for the "same" pairing of each diphthong. The task was designed such that one token of each of the four diphthongs was assigned initial position, and the other token was assigned to final position in the pair.

**Behavioral Discrimination Analysis.** The accuracy and response times recorded in the behavioral discrimination task were analyzed for each participant and each pair of vowels, including “same” pairs. The accuracy analysis utilized Signal Detection Theory to produce $d'$ discriminability scores for each contrastive vowel pair. Correct responses to “different” pairs were labeled hits, and incorrect responses to “same” pairs of either of the two contrastive vowels in the pair were labeled false alarms. The hit and false alarm rates were calculated for each contrastive pair, and $d'$ scores calculated from these rates. In the case of 1.0 hit rate or 0.0 false alarm rate, the rate was adjusted by subtracting or adding $1/(2N)$ for the purpose of calculating a $d'$ score where $N$ was the total number of “different” or “same” pairs, respectively. Mean response times (RT) to each contrastive pair were also calculated. The individual $d'$ scores and mean response times were submitted to repeated measures ANOVA, with the vowel pair as the within-subjects factor (Macmillan, 2002).

**Neuroimaging Study**

**Subjects.** A total of 25 participants completed the neuroimaging study. Among the participants, 10 were monolingual English speakers, 10 were bilingual English-Khmer speakers who learned Khmer as adults, and 5 were native Khmer speakers.

**Simultaneous EEG/fMRI Task.** The simultaneous EEG/fMRI task consisted of a passive oddball paradigm modified for MRI data collection (Rusiniak et al., 2013). Based off the results of the behavioral discrimination task, the /ɑə/-/aw/ pair was chosen for use in the EEG/fMRI investigation as the most difficult contrast for naïve English listeners to discriminate.
The most difficult pair was chosen under the hypothesis that any improvements in processing non-native phonemes would be most prominent in the marginal perceptual space. /aa/ was designated as the standard stimulus, comprising 85% of all presentations, with /aw/ designated the oddball and comprising 15% of all presentations. No null trials were included. A session consisted of two blocks of 400 presentations for 800 total presentations (680 standard, 120 oddball). The stimulus onset asynchrony (SOA) was jittered about 1500 ms (maximum 1700 ms; minimum 1300 ms). During presentation, participants viewed a series of silenced cartoons. Participants were asked to ignore the sounds that they heard and pay attention only to the cartoons. In between blocks, a short break in recording took place, during which the cartoons continued playing but no auditory stimuli were presented. Each block took approximately 11 minutes.

The EEG was recorded using the MicroMagLink 64-channel cap from Compumedics Neuroscan. This cap includes electrodes for recording the EKG for reduction of the ballistocardiogram artifact. Electrode impedances were kept below 20 kΩ for the duration of the task by the use of an electroconductive gel (Signa Gel, Parker Laboratories) injected individually to each electrode well.

The fMRI data were collected using a Siemens 3T Trio scanner (Siemens Healthineers, Erlangan, Germany) using a 12-channel head coil. Two functional scans were done, one for each block of stimulus presentation. The EPI scan sequence consisted of 39 sagittal slices (slice thickness = 3 mm; field-of-view: 192 x 192 mm²; matrix size: 64 x 64) collected in interleaved order (TR = 2000 ms; TE = 28 ms; flip angle = 90°), with 320 volumes acquired in each block. Each functional scan session was preceded by 3 “dummy” scans to protect against T1 artifact. A localizer and T1 MP-RAGE structural scan (number of slices = 176; slice thickness = 1 mm;
Stimuli were presented using the ePrime 3.0 software (Psychology Software Tools). The video was displayed on a monitor placed behind the MRI scanner which participants were able to view through a mirror placed atop the head coil. Auditory stimuli were presented using Etymotics 30-A insert earphones. Prior to beginning the study, the sound level of the stimuli were calibrated to match the sound level presented in the sound booth previously described with the Audiometer set to 60 dB HL. The calibration resulted in the stimulus being presented within 1 dB of the value recorded in the sound booth for both RMS and peak amplitude.

**Event-related Potential Analysis.** The EEG data were processed and analyzed using the Curry 7 analysis package (Compumedics Neuroscan). Blocks recorded during the dummy scans before functional MRI data acquisition were excluded from artifact reduction and analysis. The artifact produced by the scanning sequence was removed by first applying a PCA-based artifact reduction (Niazy, Beckmann, Iannetti, Brady, & Smith, 2005), and then a subtraction-based artifact reduction built-in to the Curry 7 software to the time-locked scanner trigger (Allen, Josephs, & Turner, 2000). The ballistocardiogram artifact was identified in the EKG electrodes, and the PCA algorithm was applied to remove this artifact from the EEG channels (Ellingson et al., 2004). Finally, eye-blink detection and removal using vertical eye channels and a covariance algorithm built-in to the Curry 7 software.

The artifact-reduced data were then assigned to epochs from 200 ms before onset of auditory stimuli to 800 ms after onset and averaged according to stimulus type (standard or deviant). To balance the number of standard and deviant presentations used for the averaged epochs, only the standard stimuli that were presented 3 trials before a deviant stimulus were
included. The averaged epochs were corrected to a pre-onset baseline defined by the mean amplitude from -200 to 0 ms, and filtered to between 1 and 10 Hz. The averaged standard epoch was subtracted from the averaged deviant epoch to produce the difference waveform used to identify the MMN. The individual difference waveforms from each individual participant were averaged together to produce a group average waveform. The strongest negative deflection in the FZ electrode of the group average difference waveform between 300 and 450 ms post-stimulus onset was designated as the MMN. This time window was calculated by adding the divergence point of the diphthongs in the contrast, 150 ms post-onset, to the typical MMN window of 150-300 ms post change onset (Ou & Law, 2016). The mean amplitudes of the FZ, F3, F4, FC3, FC4, CZ, C3, and C4 electrodes for individual difference waveforms in a 50 ms time window centered on this group average peak were then acquired and subjected to repeated measures ANOVA with electrode as within-subjects factors, and group as between-subjects factors. These electrodes were chosen to be consistent with previous studies examining the MMN in response to non-native phonemes (Bomba et al., 2011; Grimaldi et al., 2014; Ylinen et al., 2006). The latencies of the maximum negative peak in the individual waveform falling within the 300-450 ms post-stimulus onset window for the MMN from the same group of electrodes was subjected to another repeated measures ANOVA, again with electrode as the within-subjects factor and group as the between-subjects factor.

MRI Analysis. The MRI data were analyzed using the SPM 12 revision 6225 analysis package (Penny, Friston, Ashburner, Kiebel, & Nichols, 2011). After importing the DICOM files into the software, the images underwent slice time correction, realignment to the first image of a scan to correct for motion artifacts using a least squares approach and rigid body, co-registration to the structural image, normalization to MNI space, and smoothing using a Gaussian kernel of
8 mm FWHM. The functional data were then analyzed individually using a full factorial design with the 2 stimulus types as conditions. This analysis was then continued to a group level analysis for each of the three subject groups. The data were masked by the t-test positive effect of auditory stimulus presentation (both standard and deviant) over baseline with an uncorrected p-value of 0.05, and then the F-test comparing deviant and standard stimuli was analyzed—first with a family-wise error corrected p-value of 0.05, and then repeated with an uncorrected p-value of 0.05.

**Source Analysis.** The original structural images, co-registered and normalized structural image, and group analysis fMRI cluster image were imported into Curry 7 software for source analysis. The original structural image was used to determine electrode positions of the electrodes, in a process comparable to digitization (de Munck, van Houdt, Verdaasdonk, & Ossenblok, 2012; Koessler et al., 2007). The co-registered and normalized structural image was used to produce a BEM head model for analysis. The MMN peak in individual difference waveforms was used for sLORETA analysis using the electrode positions and BEM head model generated from the structural MRI data. The sLORETA analysis was repeated using the group-level fMRI clusters as a priori hypothesized locations weighted at 140% to compare differences with and without the fMRI data.

**Results**

**Stimuli**

Formant analysis was done using Praat v. 6.0.37 software (Boersma & Weenik, 2018) on the stimuli to determine F1 and F2 frequencies of the first and second elements of the diphthongs. Formants were analyzed at 20, 50, and 80% of the unedited vowel duration (Hillenbrand, Getty, Clark, & Wheeler, 1995). The results are summarized in Table 1. A repeated
measures ANOVA was performed with formant (F1 or F2) as the within-subjects factor and diphthong type as the between-subjects factor. This showed that the first element /a/ was not significantly different across the three diphthongs that included first element /a/ (/ao/, /aə/, and /aw/). It also revealed that the concluding vocalic /w/ element of the /aw/ and /iw/ stimuli were not significantly different.

Table 1

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Duration (ms)</th>
<th>F1 Frequency (Hz)</th>
<th>F2 Frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>20%</td>
<td>50%</td>
</tr>
<tr>
<td>/ao/a</td>
<td>456</td>
<td>907</td>
<td>829</td>
</tr>
<tr>
<td>/aə/b</td>
<td>450</td>
<td>905</td>
<td>812</td>
</tr>
<tr>
<td>/aw/a</td>
<td>425</td>
<td>922</td>
<td>617</td>
</tr>
<tr>
<td>/iw/b</td>
<td>402</td>
<td>681</td>
<td>492</td>
</tr>
</tbody>
</table>

a Means for these vowels were calculated from 10 recorded tokens
b Means for these vowels were calculated from 8 recorded tokens

Behavioral Discrimination Study

Mean response times (RT) to each type of different pair were calculated, and these mean response times are shown in Figure 2. The /iw/-/ao/ vowel pair had the shortest mean response time (920 ms), indicating the easiest discrimination. The vowel pair with the longest mean RT was /aə/-/aw/ at 1116 ms. Repeated measures ANOVA of the mean RTs found statistically
significant differences across vowel pairs, $F(5, 75) = 18.21, p < 0.001$. Post hoc tests with Tukey correction applied group the vowel pairs into four different sets, as indicated in Figure 2.

Discriminability ($d'$) scores were also analyzed using repeated measures ANOVA, showing significant differences across vowel pairs, $F(3.198, 47.967) = 31.21, p < 0.001$, Greenhouse-Geisser correction applied. $d'$ scores are shown in Figure 3. Post hoc tests with Tukey correction applied group the vowel pairs into three different sets, as indicated in the figure. The /aə/-/aw/ contrast had the lowest discriminability, with an estimated $d'$ of 0.995.

Since this contrast had both the highest mean RT and lowest $d'$, it was chosen for the neuroimaging study under the hypothesis that the most difficult contrast would illicit the largest differences between naïve and adult learner listeners.
Neuroimaging Study

Event-related potential results. The group averaged waveforms are displayed in Figure 4. The group average waveforms show that the native Khmer group had the most characteristic MMN waveform. The native Khmer MMN also had the shortest latency. The monolingual group also had a well-defined MMN waveform, but it appeared later than the native Khmer group. The bilingual group did not have a well-defined MMN waveform. There was a negative peak, but this was earlier than the defined window. The most negative deflection in the defined time window (300-450 ms) was selected as the MMN for analysis.

Averaged amplitudes of the MMN in the individual waveforms were subjected to repeated measures ANOVA. There were no significant differences between groups by the averaged amplitudes of the 8 electrodes used in the analysis, $F(2, 20) = 1.481; p = 0.251$. The latencies of the most negative peak in these electrodes in the time window used for averaging amplitudes was also subjected to repeated measures ANOVA. The F-test for group effects did
show a significant difference for latency, \(F(2, 20) = 6.30; p = 0.008\). Post hoc comparisons using Tukey correction for family wise error showed the native Khmer group was significantly earlier than the bilingual group (\(p = 0.008\); 95% confidence interval between 176 and 25 ms earlier). The native Khmer group was not significantly different from the monolingual group (\(p = 0.391\)). The monolingual group trended towards being earlier than the bilingual group, but this did not reach statistical significance (\(p = 0.08\)).

Functional MRI results. In the masked F-test comparing deviant and standard stimuli, no clusters reached statistical significance when correcting for family-wise error-rate. When the p-value threshold was adjusted to an uncorrected value of 0.05, the clusters displayed in Figure 5 were found. The native Khmer group had a cluster in the left middle temporal gyrus. The bilingual group had a cluster in the left superior temporal gyrus. The monolingual group had a prominent cluster in the right temporal lobe, as well as a cluster in the left temporal lobe.

Source analysis results. Individual sLORETA analysis was conducted for the MMN identified in the individual difference waveforms. This analysis was repeated using the respective group-level fMRI clusters identified at the \(p = 0.05\) level discussed above. The individual results were compared using the CDR ANOVA built into the Curry 7 suite. No group had consistent source analysis results, and no strengths were found in this analysis. sLORETA was also computed for the group averaged waveforms, and these results are shown in Figure 6.
Figure 4. The group averaged difference (black), oddball (gray), and standard (dotted) ERP waveforms. Arrows indicate the peak identified to establish the time window for analysis of individual MMN amplitudes. (a) Native Khmer group (n=5) (b) Bilingual group (n=10) (c) Monolingual group (n=8)
Figure 5. Suprathreshold fMRI voxels for all groups to the F-test comparing deviant and standard stimulus presentation with threshold p-value set to uncorrected 0.05. A) The native Khmer group (n = 5) had a cluster in the left middle temporal lobe. B) The bilingual group (n =10) had a cluster in the left superior temporal lobe. C and D) The monolingual group (n = 8) had a cluster in the right superior temporal lobe (C) and another cluster in the left superior temporal lobe (D). Images are shown in neurological orientation.
Figure 6. sLORETA results from group average difference waveforms. (a) Khmer group (b) Bilingual group (c) Monolingual group
Discussion

Behavioral Discrimination Study

The primary purpose in undertaking the behavioral discrimination study was to identify the contrast that was hardest for monolingual English speakers to discriminate. The results of the task revealed that the /aə/-/aw/ contrast was the most difficult for monolingual English speakers.

An interesting pattern emerges in looking at the grouping of the contrasts resulting from the post hoc tests. The d’ results grouped the pairs into three sets: /aə/-/aw/ and /aə/-/ao/; /ɨw/-/aw/, /ao/-/aw/, and /ɨw/-/aə/; and /ɨw/-/ao/. This result was initially surprising. The /ɨw/ diphthong was expected to be easily discriminated from the other three diphthongs because it has a different first element; only one contrast including /ɨw/ was significantly different than contrasts with a shared first-element. This three-tier grouping could be explained by the perceived number of shared first elements and shared second elements. Thus, the /ɨw/-/ao/ contrasts, with both the first and second elements perceived differently, is the most easily discriminable contrast. The /aə/-/aw/ and /aə/-/ao/ contrasts are perceived as having both shared first elements and shared second elements (with the second element of /aə/ being perceived as similar to both the second elements of /aw/ and /ao/). Finally, the middle tier consists of contrasts where one of the two elements is perceived as shared: the /ɨw/-/aw/ contrast shares a second element, the /ao/-/aw/ contrast shares a first element, and the /ɨw/-/aə/ contrast is perceived as sharing a second element. This explanation is also largely consistent with the reaction time data. Importantly, this implies that non-native diphthongs are assimilated under the PAM model broken down by individual elements of the diphthong rather than as a whole.

We can also use the results to make inferences regarding to which PAM category the contrast used for the neuroimaging study belongs. Because an identification task was not
performed, we cannot state with certainty to which English vowels these diphthongs are assimilated. It seems reasonable, however, to suggest that the most likely vowel for /ao/, /aə/, and /aw/ to assimilate to is the English /aʊ/ diphthong. The pattern of the results suggest this may be the case, and the low d’ scores would be consistent with a Single-Category assimilation contrast, in which the vowels in the contrast vary equally from the ideal of the category to which they are assimilated. The pattern of results is also consistent with a Two Uncategorizable contrast, where the /ao/, /aə/, and /aw/ phonemes are not assimilated to any English phoneme. This explanation seems less credible given the similarity of the Khmer phonemes to the English /aʊ/ phoneme, and the lack of another English phoneme that is similar to the Khmer phonemes. An identification task would need to be performed to state conclusively the type of contrast, but the discrimination task results together with the PAM concepts suggest the /aə/-/aw/ contrast is one of Single-Category assimilation.

**Neuroimaging Study**

**Event-related potential.** The ERP results from the neuroimaging study were largely inconclusive. While visual inspection suggested that the native group tended to have larger, more defined MMN peaks and shorter latencies, the statistical methods did not support this conclusion. This may in part be due large variance in MMN latency across individuals within the same group. This was particularly troublesome in the bilingual group. This large variance reduces the effectiveness of using the group average waveform to set the time window for averaging the MMN peak amplitude. Elements that appear to be the MMN in individual difference waveforms had no overlap with the time window identified by the group average waveform (Figure 7). The large variance may be explained by the difficulty of the contrast. The contrast may have been so
difficult that differences in individual auditory perception caused variance that was not seen in previous tests of phoneme discrimination (Diaz et al., 2008).

Another contributing factor may be the simultaneous collection of EEG and fMRI data. Previous studies have suggested that simultaneous data collection may affect the amplitude and latency of ERPs (Bregadze & Lavric, 2006; Chun, Peltier, Yoon, Manschreck, & Deldin, 2016). Most such studies have analyzed the P300 ERP element. The MMN has not been tested in this manner. It may be that the MMN, a subtler element than the P300, is more susceptible to such effects. Further feasibility studies should be undertaken to confirm that the MMN can be successfully produced in simultaneous EEG/fMRI conditions.

Figure 7. Individual difference waveforms from the bilingual group showing variability of the MMN peak. Arrows indicate MMN peak, and the shaded region represents the time window for analysis from the group averaged difference waveform. (a) An example of the individual peak falling within the time window for analysis. (b) An example showing the individual peak occurring after the time window for analysis. (c) An example showing the individual peak occurring before the time window for analysis.
There is also some evidence of an earlier waveform that may be a MMN waveform occurring around 200 ms post stimulus onset, especially in the bilingual and monolingual groups (see Figure 3(b) and Figure 3(c)). This earlier waveform may be a MMN to accoustical variance present in the change from the /aə/ stimulus to the /aw/ stimulus. While the perception that the oddball stimulus is a different phoneme than the standard stimulus would not occur until the divergence point of 150 ms post-onset, there are acoustical differences that may be present at onset. The negative peak around 200 ms post stimulus onset could be a MMN waveform to this accoustic difference, and the waveform in the 300-450 ms time window the MMN waveform to the phonemic difference.

Functional MRI. The fMRI results were also inconclusive, though the voxels that tended toward significance were promising. One weakness of the current research is the small subject number (n = 5, 10, and 8 for native Khmer, bilingual, and monolingual English speakers, respectively). One study has suggested that even group sizes as large as 15 may be insufficient for fMRI data (Schafer et al., 2003). The subject number was chosen based off comparable studies using simultaneous EEG/fMRI data (Mulert et al., 2004; Scarff et al., 2004) and studies examining MMN differences across groups with varying second language experience (Bomba et al., 2011; Grimaldi et al., 2014; Ylinen et al., 2006). However, it seems likely that these group sizes were too small for this study, perhaps because the effect size of the MMN is smaller than the P300.

Examining those voxels that tended toward significance reveals an interesting result. While ERP research did not find differences between bilinguals and monolingual speakers to a non-native contrast (Grimaldi et al., 2014; Ylinen et al., 2006), there appear to be differences between these groups in the fMRI data. The voxels that tended toward significance in both the
native group and the bilingual group were in the left temporal lobe. The voxels that tended toward significance in the monolingual group were located in the right temporal lobe. These trends are consistent with the hypothesis that bilingual speakers process the nonnative contrast as a poorly defined speech contrast; whereas, monolingual speakers process the nonnative contrast based on the distinct auditory features. A more robust study with larger group sizes would need to be performed to confirm these results.

**Source Analysis.** The sLORETA results had high variance. No common areas were found within any of the groups. The group averaged sLORETA results were inconsistent with previous MMN findings, localizing the most of the activity to the frontal lobes. Because of the inconsistent sLORETA results and the lack of significant fMRI results, we cannot state whether the voxels that trended toward significance were directly related to the generation of the MMN. This, perhaps, highlights a weakness of simultaneous EEG and fMRI data collection: the difficulty of combining the results of the disparate methods meaningfully. While simultaneous data collection allows researchers to link EEG and fMRI data by showing they are responses to the same stimuli, the activity present in the ERP and the activity shown by fMRI BOLD response may not be caused by the same processes. It may take more advanced algorithms to successfully use fMRI data to inform the source of ERP components (Fabbiano, Vacca, Morello, & De Capua, 2013; Mangalathu-Arumana, Beardsley, & Liebenthal, 2012).

**Conclusion**

Because of the small group size, this study was underpowered in its ability to find neural generators for differential processing of difficult, non-native diphthong contrasts. There is a trend toward evidence of native and bilingual processing in the left temporal lobe and monolingual processing in the right temporal lobe, but more robust research is needed to confirm this finding.
The study did produce results that suggest that naïve listeners process nonnative diphthong contrasts by separately comparing the individual element components of each diphthong. This contrasts with native diphthongs, which are perceived as single phonemes and not as combinations of distinct phonemes.

Future studies should confirm the feasibility of collecting meaningful MMN data during simultaneous EEG/fMRI data collection. This should be done with pure tone stimuli to avoid confounding effects of language processing. After confirmation of the viability of recording the MMN in the MRI environment, large subject groups (n > 15) should be presented with a non-native diphthong contrast of middling difficulty to increase the contrast for fMRI detection and decrease the variance of the MMN. Decreasing the variance of the MMN should lead to improved source localization results, which would allow for successful integration of the fMRI data to the analysis.
References


doi:10.1006/nimg.2000.0599


Retrieved from http://www.praat.org


doi:10.1097/WNR.0b013e328347dada


doi:10.1016/j.ijpsycho.2006.01.010


doi:http://dx.doi.org/10.1016/j.neuroimage.2011.07.021


Appendix A

Comparison Between Khmer and American English Phonology

Table A1

Khmer Consonant Phonology

<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th>Dental</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stops:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>-p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-t</td>
<td>-c</td>
<td>-k</td>
<td>-q</td>
</tr>
<tr>
<td>Voiced</td>
<td>b</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Spirants:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>(f)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>s</td>
<td></td>
<td>-h</td>
<td></td>
</tr>
<tr>
<td><strong>Continuants:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>-m</td>
<td>-n</td>
<td>-ñ</td>
<td>-η</td>
<td></td>
</tr>
<tr>
<td>Semivocalic</td>
<td>-w</td>
<td></td>
<td>-y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td>-l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trilled</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes.** The transcription is the author’s own creation and does not match IPA, though it is very similar. The dash (-) symbol before a consonant indicates that it can appear in word initial position as well as final position. Adapted from *Cambodian system of writing and beginning reader with drills and glossary* (p. 6) by F. E. Huffman, 1970, New Haven: Yale University Press. Copyright 1970. In the public domain since 1975.

<sup>a</sup> The consonant sounds /p t c k/ can also appear aspirated /pʰ tʰ cʰ kʰ/, which are contrastive when immediately followed by vowels. These aspirations are not included in the chart, however, as these aspirations are often analyzed as a consonant series: /ph/ /th/ /ch/ /kh/.

<sup>b</sup> The /f/ occurs only in loan words.

Table A2

English Consonant Phonology

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Interdental</th>
<th>Alveolar</th>
<th>Alveopalatal</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stop</strong></td>
<td>p b</td>
<td></td>
<td>t d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affricate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ʧ ʣ</td>
</tr>
<tr>
<td>Fricative</td>
<td>f v</td>
<td>Θ ɵ</td>
<td>s z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ʃ ʒ</td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semivowel</td>
<td>(w)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>j (w)</td>
</tr>
</tbody>
</table>

**Notes.** When two consonants appear in one cell, the first consonant is voiceless, and the second is voiced. Adapted from *Psycholinguistics: Introduction and applications* (p. 27) by L. Menn, 2011, San Diego: Plural Publishing, Inc. Copyright © 2011 Plural Publishing, Inc. All rights reserved. Used with permission.
Table A3

**Khmer Vowel Phonology**

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
<td>ɨ</td>
<td>ɨ̞</td>
<td>u</td>
</tr>
<tr>
<td><strong>Mid</strong></td>
<td>e</td>
<td>ə̞</td>
<td>o̞</td>
</tr>
<tr>
<td><strong>Lower Mid</strong></td>
<td>a</td>
<td>ɔ̞</td>
<td></td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>a</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Notes.* The transcription is the author’s own creation and does not match IPA, though it is very similar. Arrows indicate diphthongs, with the arrowhead the second vowel of the diphthong. Adapted from *Cambodian system of writing and beginning reader with drills and glossary* (p. 6) by F. E. Huffman, 1970, New Haven: Yale University Press. Copyright 1970. In the public domain since 1975.

Table A4

**English Stressed Vowel Phonology**

<table>
<thead>
<tr>
<th>i</th>
<th>heed, he, beat, heat</th>
<th>lowercase ɨ_a</th>
<th>ꞌ</th>
<th>Hudd, mud</th>
<th>wedge; turned ꞌ_a</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>hid, bid, hit, kid</td>
<td>small capital I</td>
<td>ꞔ</td>
<td>herd, hurt, bird, curd</td>
<td>reversed epsilon</td>
</tr>
<tr>
<td>eɪ</td>
<td>hayed, hay, bade, hate</td>
<td>lowercase e + ɨ</td>
<td>ꞌ</td>
<td>high, hide, bide, height</td>
<td>lowercase a + ɨ</td>
</tr>
<tr>
<td>e</td>
<td>head, bed</td>
<td>epsilon (Greek)</td>
<td>ꞔ</td>
<td>how, cow, cowed</td>
<td>lowercase a + o</td>
</tr>
<tr>
<td>æ</td>
<td>had, bad, hat</td>
<td>ash</td>
<td>ꞔ</td>
<td>(a) hoy, boy, Lloyd</td>
<td>open o + ɨ</td>
</tr>
<tr>
<td>o</td>
<td>hard, bard, hod, cod</td>
<td>script a</td>
<td>ꞔ</td>
<td>ear, beard</td>
<td>lowercase i + r</td>
</tr>
<tr>
<td>ɔ</td>
<td>haw, bawd, caw</td>
<td>open o</td>
<td>ꞔ</td>
<td>ear, beard</td>
<td>small capital I + r</td>
</tr>
<tr>
<td>ʊ</td>
<td>hoed, hoe, code</td>
<td>lowercase o + o</td>
<td>ꞔ</td>
<td>ber(ry), mer(ry)</td>
<td>epsilon + r</td>
</tr>
<tr>
<td>ʊ</td>
<td>hood</td>
<td>upsilon (Greek)</td>
<td>ꞔ</td>
<td>hare, bare, mare, Mar(y)</td>
<td>lowercase e + r</td>
</tr>
<tr>
<td>u</td>
<td>who, hoot, bood</td>
<td>lowercase u</td>
<td>ꞔ</td>
<td>hired, hire</td>
<td></td>
</tr>
</tbody>
</table>

*Notes.* The author notes in the text that the unstressed vowel /ə/ also occurs in English. Adapted from *Psycholinguistics: Introduction and applications* (p. 27) by L. Menn, 2011, San Diego: Plural Publishing, Inc. Copyright © 2011 Plural Publishing, Inc. All rights reserved. Used with permission.

a The 3rd and 6th columns are explanations of the IPA symbols used, and were included in the text to help the reader correctly identify the symbols.

b In the text, Menn (2011) notes that “most American English speakers say ear, beard, and words like them with a vowel sound somewhere between /i/ and /ɨ/” (p. 31).
Appendix B

Presentation of Words for Stimulus Recording

1. េី េី េី េី
    tiw  taw  tao  tao

2. េី េី េី េី
    taw  tao  tao  tiw

3. េី េី េី េី
    tao  tiw  taw  tao

4. េី េី េី េី
    tao  taw  tiw  tao

5. េី េី េី េី
    tiw  tao  taw  tao

6. េី េី េី េី
    tao  tao  tiw  taw

7. េី េី េី េី
    tao  taw  tao  tiw

8. េី េី េី េី
    tao  tiw  tao  taw

9. េី េី េី េី
    taw  tao  tao  tiw

Notes. The transcriptions were added for inclusion in the appendix to aid the reader. They were not presented to the speaker who recorded the stimuli. The speaker was instructed to read through the whole list twice, once with pauses between each word and again with pauses only after each line.