2018

Investigating the Electrophysiology of Long-Term Priming in Spoken Word Recognition

Erin K. Bell
Cleveland State University

INVESTIGATING THE ELECTROPHYSIOLOGY OF LONG-TERM PRIMING IN SPOKEN WORD RECOGNITION

ERIN BELL

Bachelor of Science in Neuroscience

University of Mount Union

May 2016

Submitted in partial fulfillment of requirements for the degree

MASTERS OF ARTS

at the

CLEVELAND STATE UNIVERSITY

May 2018

We hereby approve this thesis

For

Erin Bell

Candidate for the Master of Arts degree

for the Department of

Psychology

And

CLEVELAND STATE UNIVERSITY'S

College of Graduate Studies by

Thesis Chairperson, Robert S. Hurley

Department & Date

Methodologist and Committee Member, Conor T. McLennan

Department & Date

Committee Member, Ilya Yaroslavsky

Department & Date

_________May 3, 2018__________

Student's Date of Defense

INVESTIGATING THE ELECTROPHYSIOLOGY OF LONG-TERM PRIMING IN SPOKEN WORD RECOGNITION

ERIN BELL

ABSTRACT

When participants are listening to the same words spoken by different talkers, two types of priming are possible: repetition priming and talker-specific priming. Repetition priming refers to the exposure of a stimulus improving responses to a subsequent exposure. Talker-specific priming refers to the exposure of words spoken by same talkers improving responses relative to those same words spoken by different talkers. There are conflicting theories regarding whether talker-specific priming should be observed. Abstract representational theories suggest that episodic details (e.g., talker identity) are not stored in the mental lexicon, while episodic theories of the lexicon posit that lexical representations include episodic details. According to the time-course hypothesis, the mental lexicon includes both types of representations, and abstract representations are accessed earlier than episodic representations. In the present experiment, long-term priming in spoken word recognition was tested using a technique that is particularly well-suited for answering questions about timing: event-related potentials (ERPs). Participants heard words spoken by two different talkers in each of two separate blocks. Stimuli in the second block consisted of three different priming conditions, which are described in relation to what participants heard in the first block: new, unprimed, words (control), repeated words spoken by the same talker (match), and repeated words spoken by different talkers (mismatch). Evidence for long-term repetition priming was obtained in reaction times and accuracy. Electrophysiological evidence of repetition priming was obtained in low frequency words. Talker-specific priming effects were observed in accuracy, with more accurate responses in the match condition than in the mismatch condition, consistent with episodic representational theories. However, there was no evidence of talker-specific priming in the ERP data, which, when considered alone, is consistent with abstract representational theories. The current results provide the first physiological evidence (ERPs) of long-term repetition priming in spoken word recognition, setting the stage for future empirical investigations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Spoken word recognition is complicated by a wide range of variability among talkers, including factors such as speaking rates, timbre, and prosody (Abercrombie, 1967; Pisoni, 1997), referred to as indexical variability. Given this variability, there are at least two different types of priming within spoken word recognition: *repetition priming* and *talker-specific priming.* Repetition priming occurs when stimuli are processed more efficiently (e.g., faster, more accurate, or both) after repeated exposures. Responding more efficiently to words spoken by the same talker compared to those same words being spoken by different talkers, is known as talker-specific priming.

There is an active debate within the field of spoken word recognition regarding whether or not indexical variability is stored in the mental lexicon, the mind's repository of lexical representations. *Episodic representational theories* state that episodic details (e.g., talker identity, speaking rate) underlying indexical variability of spoken words are stored in the lexicon (e.g. Campeanu, Craik & Alain, 2013; Craik & Kirsner, 1974; Palmeri, Goldinger & Pisoni, 1993). *Abstract representational theories* argue the opposite: that episodic details are discarded when listening to spoken words, and that episodic information is not stored in the mental lexicon (Gaskell & Marslen-Wilson, 1997; Luce, Goldinger, Auer & Vitevitch, 2000; McClelland & Elman, 1986; Pisoni, 1993).

Empirical tests of these competing theories have investigated participants' processing of variability. Participants in one study engaged in an auditory perceptual identification task, during which they were either exposed to stimuli spoken by a single talker or to stimuli from multiple talkers. When asked to identify words that were embedded within noise through headphones, identification was less accurate when there were multiple talkers compared to when there was only a single talker (Mullennix, Pisoni & Martin, 1989). While the presence of this talker variability effect appears to be evidence for episodic representational theories, the authors claimed that the phenomenon could still be explained by an abstract theory, as it may have arisen through the attenuation of auditory normalization. According to this abstract interpretation, indexical variability is treated as perceptual "noise" and is stripped away when recognizing spoken words (Nearey, 1989). Consequently, Mullenix, Pisoni, and Martin (1989) suggest that additional information brought on by multiple talkers adds to the complexity of perceptual processing, attenuating the normalization processes, and ultimately producing detrimental effects on spoken word recognition. This normalization account makes it difficult to disentangle predictions from abstract and episodic theories of the lexicon, unless a long-term repetition-priming paradigm is used, which more directly addresses questions about the nature of lexical representations.

The Time-Course Hypothesis

In an attempt to reconcile abstract and episodic representational theories, McLennan and Luce (2005) proposed the *time-course hypothesis*, which is consistent with a hybrid account of the mental lexicon. According to the time-course hypothesis, both abstract and episodic representations are stored in the mental lexicon, but when listening to a spoken word abstract representations are accessed earlier and episodic representations are accessed later. According to time-course hypothesis, abstract representations are more likely to affect spoken word recognition when processing is relatively fast and easy and episodic representations are more likely to affect spoken word recognition when processing is relatively slow and difficult. Thus, repetition priming should be observed both when processing is fast and slow, but talker-specific priming is more likely to be observed when processing is relatively slow.

McLennan and Luce (2005) directly tested the time-course hypothesis in three separate experiments, each employing a *long-term priming* paradigm, where a stimulus is repeated in a separate block of trials at least a few minutes later. The first experiment involved *lexical decision tasks*, during which participants judged whether words are real English words (e.g. "cat") or *nonwords* with no meaning (e.g. "caz"). Nonwords were divided into those that were more or less word-like, making the discrimination between real words and nonwords either relatively difficult or relatively easy, respectively. Results were consistent with the time-course hypothesis: talker-specific priming effects (faster reaction times to words spoken by the same talker than to those same words spoken by different talkers) only emerged in the difficult version of the lexical decision task in which participants' reaction times were slower.

Results from the two additional experiments further supported the time-course hypothesis, one with intra-talker variations in speaking rate (slow versus fast), and one with a delayed-response shadowing paradigm, where participants were asked to repeat (shadow) the stimulus words either immediately after hearing the word (speeded), or following a delay after a cue was presented. When presented with variations in speaking rate, talker-specific priming effects again only emerged in the difficult version of the lexical decision task. Additionally, when presented with variations in talker identity within speeded and delayed response shadowing tasks, only when processing was probed relatively late (from the delayed-response paradigm) did talker-specific priming effects emerge (M^c^Lennan & Luce, 2005).

The central claim of the time-course hypothesis focuses on how the perception of spoken words unfolds over time. Yet, descriptions and discussions about "time" maintain more general terms: relatively early versus relatively late. A physiological measure that provides a more specific time stamp of underlying neural activity could enrich the hypothesis. Consequently, event-related potentials (ERPs), which provide temporally precise information, were collected in the current study.

ERPs and Language Studies

The contrasting abstract and episodic theories, as well as hybrid models, which are consistent with the time-course hypothesis, have provided a spring-board for debate and empirical investigations within the field of spoken word recognition. Numerous studies have been conducted, providing evidence for one theory over the others. However, no study to date has been conducted with a physiological measure that is sensitive to time.

As mentioned above, ERPs index neural processing in real time with excellent temporal resolution. As spoken word recognition unfolds over time, this physiological measure provides a unique and novel approach to testing and contrasting theoretical concepts such as abstract and episodic representations, and the time-course hypothesis.

Several ERP components may be relevant for examining repetition priming, talker-specific priming, and the time-course of spoken word recognition. Spoken word recognition is thought to occur initially at *sublexical levels* of processing, followed by *lexical levels,* and different ERP components are thought to be sensitive to each. Sublexical processing essentially involves the sound-based components that are combined into whole words, such as phonemes. For example when hearing the word "cat", processing that word at a sublexical level would include phonemic processing (e.g., /k/, /@/, and /t/). In contrast, processing at the lexical level involves recognition of the whole word (e.g., /k@t/)[1]. ERP components thought to reflect sublexical versus lexical processing are described below.

N200. The N200 component, typically occurring 200-270 ms post stimulus presentation is related to sublexical phonological processing, before semantic meaning can be established (Connolly, Phillips, Stewart, & Brake, 1992, Huang, Yang, Zhang & Guo, 2014; Lee, Harkrider & Hedrick, 2012; van den Brink, Brown, & Hagoort, 2001). The N200 amplitude is not only thought to reflect sublexical competition and encoding, but also to be sensitive to phonologically unexpected mismatches in speech stimuli (van den Brink, Brown, & Hagoort, 2001; Hunter, 2013; Lee, Harkrider & Hedrick, 2012).

N400. The N400 component is thought to reflect the process of *lexical access*, where the meaning of a word is retrieved from a set of activated word representations within the lexicon (Chwilla, Brown, & Hagoort, 1995; Kutas & Hillyard, 1980, 1984;

[1] Klattese is being used here to indicate the sublexical and lexical levels of processing. Klattese is a computer-friendly system of phonetic transcription (Vitevich & Luce, 2004).

Kutas & Iragui, 1998). Semantically unrelated targets elicit larger N400 components than semantically related targets, demonstrating the N400 to be a key component in semantic processing (Holcomb, 1993). N400 effects have also been noted in paradigms with violations in expectation (Magne et al., 2007; Olichney et al., 2006), where stimuli that violate semantic expectations (such as non-words) generate a larger N400 response. Utilizing a *short-term* priming paradigm, repetition priming effects have been found as early as 100-200 ms post stimulus onset, and lasting as late as through the P350 and N400 time windows (Dufour, et. al., 2017). Additionally, talker-specific priming effects have been observed within the N400 time window for low frequency words. This pattern of ERPs, suggests that spoken word recognition primarily relies on abstract representations, with the occurrence of talker-specific priming in later stages of processing: a pattern resembling that of the time-course hypothesis.

The Current Study

The purpose of the current study was to examine the time course of spoken word recognition using physiological measures in a *long-term* priming paradigm. By examining accuracy, reaction times, and ERPs in response to new, unprimed words, primed words spoken by the same talker (match), and primed words spoken by a different talker (mismatch), the results of the current study were expected to find support for one of three competing theories (abstract representational theories, episodic representational theories, a hybrid representational theory, as described in the time-course hypothesis), and to provide novel data and interpretations with respect to long-term priming within this context.

Predictions. According to the *time-course* hypothesis (M^c^Lennan & Luce, 2005), two sets of predictions were outlined for the behavioral outcomes (accuracy and reaction times). For accuracy, it was predicted that responses to primed words spoken by the same talker (match condition) would be significantly more accurate than responses to new, unprimed words (control condition), demonstrating repetition priming. It was additionally anticipated that responses to primed words spoken by the same talker (match condition) would be significantly more accurate than responses to primed words spoken by a different talker (mismatch condition), demonstrating talker-specific priming. The second set of behavioral predictions was based on reaction times. Repetition priming was expected to be characterized by faster reaction times for the match than for the control condition, while a talker-specific priming effect was anticipated through faster reaction times for the match than for the mismatch conditions.

The critical factor for the current study was in the analyses of the ERPs, as they carried the potential to supplement our understanding of previous behavioral findings. Talker-specific priming effects could emerge at either early or late ERP time-windows, yielding four possible outcomes (Figure 2).

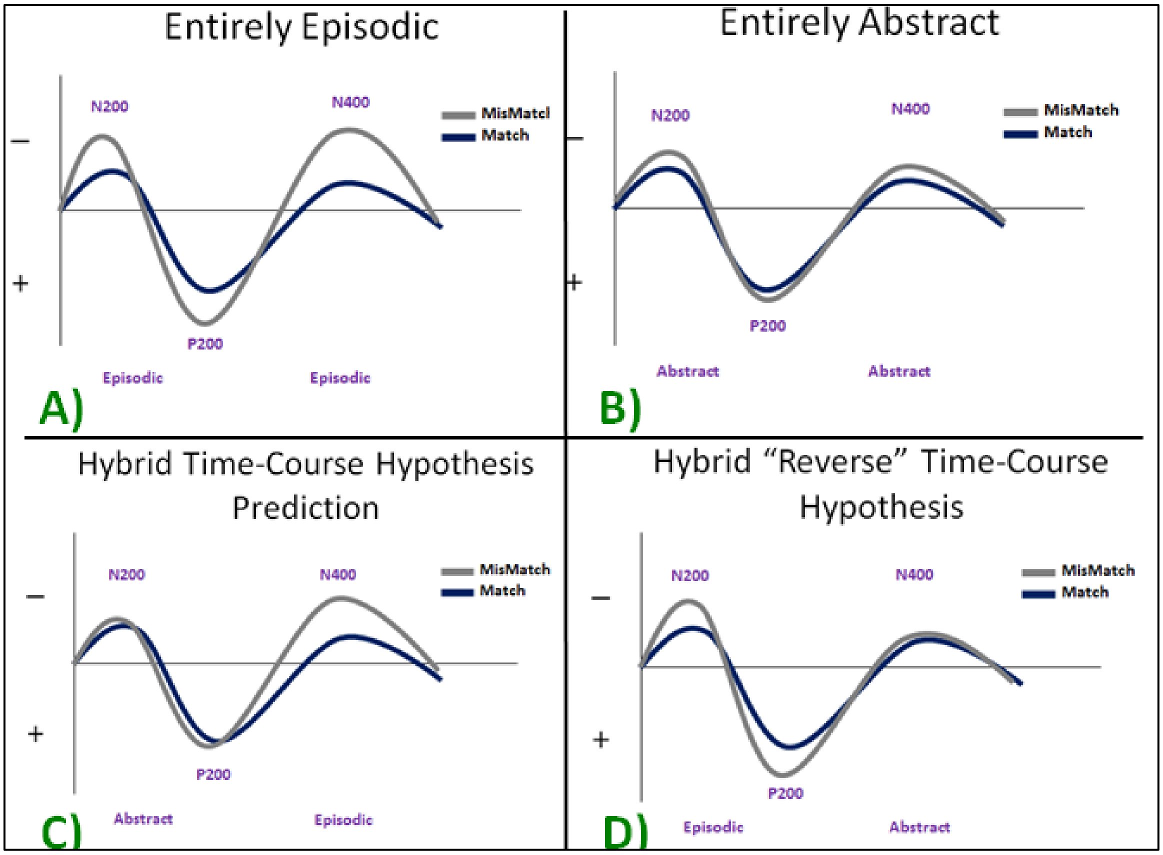


Figure 1. Possible ERP outcomes.

One possible ERP waveform outcome would show amplitude effects consistent with talker-specific priming (differences between the match and mismatch conditions) at both the earlier N200 and the later N400 time windows (Figure 1A). In discovering amplitude differences across the entire epoch, additional evidence for the episodic representational theories would be obtained, indicating that for the duration of spoken word recognition, listeners were processing episodic details. Another possible ERP waveform outcome would demonstrate no amplitude effects across the epoch (Figure 1B). In contrast with the previous possibility, this would indicate that at no point in the processing of spoken word recognition (at least within the time windows measured), was there any processing of episodic details, providing more concrete evidence in support of abstract representational theories.

Predictions from the time-course hypothesis are shown in Figure 1C; no significant talker differences in N200 amplitudes, but rather amplitude differences emerging in the later N400 time windows (Figure 1C). The differences in amplitudes in the later time window in direct correspondence with slower reaction times would represent the access of episodic representations, while the lack of differences in the earlier time window in direct correspondence with quicker reaction times would represent the access of abstract representations—as predicted by the time-course hypothesis.

One final possible outcome was considered, namely one that is directly opposite of the time-course hypothesis predictions: a waveform divergence at the earlier N200 time window with no amplitude effect at the later N400 window (Figure 1D). This pattern would constitute a "*reverse time-course hypothesis*" in which episodic representations come into play earlier than abstract representations. This particular finding could possibly suggest that processing of mismatched words is potentially delayed within the auditory network, and/or the language network, during a sublexical stage of processing (during the N200 time window), creating a bottleneck which subsequently pushes back lexico-semantic processing, and ultimately prolongs reaction times. The possibility of two "reverse" outcomes further demonstrates the utility of the ERP technique: although the time-course and reverse time-course hypotheses made equivalent predictions for reaction times, their predictions for N200 and N400 waveforms are dissociated.

CHAPTER II

METHOD

Participants

A total of 51 undergraduate participants from Cleveland State University were recruited in exchange for research participation credit. Participants were native speakers of American English, right-handed, and reported no speech or hearing deficits. In addition, none reported having any forms of aphasia or other language impairments. Participants with overall mean reaction times above or below two standard deviations from the grand mean (two participants), or with accuracy scores below two standard deviations from the grand mean (six participants), were excluded from the analyses. An additional seven participants were excluded due to poor ERPs, characterized by the presence of too much noise. The final sample size consisted of 36 participants (27 female, 9 male).

Stimuli

A list of 120 words and 120 nonwords was generated and presented as auditory stimuli (appendices A and B). Each word and nonword was monosyllabic and conformed to a consonant-vowel-consonant (CVC) format, such as “cat.” Low frequency words were used in addition to high frequency words, as previous research has demonstrated that talker-specific priming effects are more likely to be observed in low frequency words (Dufour & Nguyen, 2014; Dufour, et. al., 2017). The nonwords were created by replacing the final consonants (codas) with codas from other real words (e.g., “bell” became “besh”).

Each word and nonword was recorded twice, once from a male talker and once from a female talker. Both talkers were native talkers of American English from within a 100-mile radius of Cleveland, Ohio, with clear diction and annunciation. The recorded sound clips were 600 milliseconds (ms) in duration.

Procedure

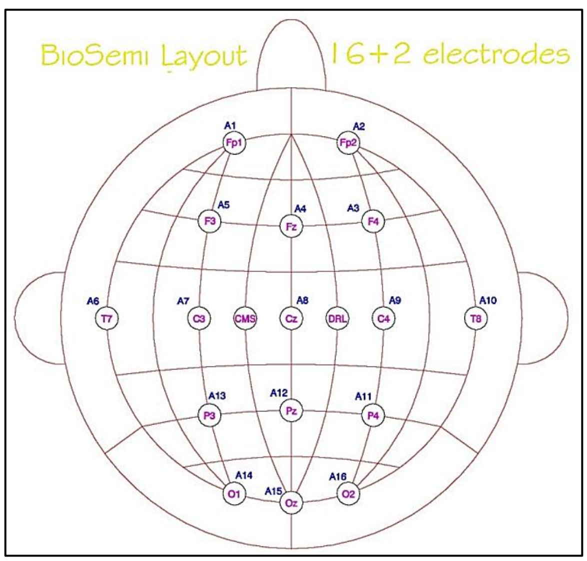


*Figure 2***.** Electrode Layout.

Once the participants had given signed consent, a 16 scalp electrode EEG cap was placed on their head. EEG was continuously recorded from 16 scalp electrodes (F1, F2, F3, F4, Fz, C3, C4, Cz, T7, T8, P3, P4, Pz, O1, O2, Oz) in an elastic cap through a BioSemi Active II amplifier system (Figure 2). Two additional electrodes were attached to the left and right mastoid bones as references.

Once the participants were set up and seated comfortably, they performed a difficult lexical decision task. As previous work demonstrated that a more difficult task elicited stronger talker-specific priming effects (M^cLennan & Luce, 2005), the nonwords were created to seem very similar to real words, increasing the difficulty of the task. Each participant was instructed to indicate whether the sound presented on each trial was a real English word or a nonword by pressing either the left or right mouse button with the index and middle fingers, respectively, of their right hand. The participants were instructed to make decisions as quickly and as accurately as possible. Instructions were presented to the participants both visually through a prompt on the computer screen, as well as verbally by the experimenter.

The first block consisted of 160 trials, with 80 words and 80 nonwords pseudo-randomly intermixed. Participants were then given a three-minute distractor test requiring participants to solve math equations. The purpose of the distractor test was to prevent any conscious effort to rehearse stimuli from the first block, and to provide sufficient time for long-term priming to occur. Following the distractor math task, participants were given a second block, which consisted of 120 words and 120 nonwords pseudo-randomly intermixed, yielding 240 trials. The same 80 words and 80 nonwords from the first block were repeated in the second target block (separate by between three and 20 minutes), creating an environment to test for effects of long-term priming. These priming trials were equally divided into 80 match condition trials (40 word & 40 nonword) where the clip was identical to the one heard during the first block (e.g., the same clip from a male speaker in the first and second blocks), and 80 mismatch condition trials where the same word or nonword was presented but recorded by a speaker of the opposite gender.

These repeated, or primed, trials were supplemented with 40 new, unprimed, words and 40 new, unprimed, nonwords (i.e., not present in the first block), creating an unprimed control condition. This within-participants design ensured that the same sound clips appeared in the match, mismatch, and unprimed control conditions, with specific words in each list counterbalanced across participants. Afterwards, participants were debriefed. Each experimental session lasted for approximately one hour.

CHAPTER III

DATA ANALYSIS

Overall, the means of the reaction times and accuracy scores to the nonwords were 967 ms and 89%, respectively. In all of the analyses reported below, we focused only on the real words, which were of primary theoretical interest in the current investigation. In addition, we only note effects that were in the expected direction, as the current study is an empirical test of a priori directional hypotheses (e.g., that words in the match condition and mismatch conditions should be responded to more efficiently, not less efficiently, than words in the control condition)[2].

Reaction Times and Accuracy

Stimuli to which > 30% of participants generated inaccurate responses (thus, close to the chance rate of 50% accuracy) were excluded from the analyses, leaving 110 words for analyses. After implementing the exclusionary criteria, there were an average of 32 trials per condition.

---

[2] Only one result was in an unexpected direction: within low frequency words, there was a talker-specific priming effect evident within the 250-325 time window ($F(1, 35) = 10.58$, $p < 0.01$, $\eta^2 = 0.23$). Despite a significant different at this particular time window, the conditions are not presented directionally as theoretically expected. In addition, this effect was not consistent across electrodes. Thus, this result is believed to be that of a type I error.

Each participant's individual accuracy score was calculated and any participants with scores below two standard deviations from the grand mean were excluded from the study all together (six participants). Grand means were then generated for each condition (match, mismatch, and control), both collapsed across high and low frequency words, as well as separately for low frequency words. For reaction times, only trials with accurate responses were used. Reaction times within each participant's individual set of response trials were winsorized at a cutoff level of two standard deviations above the mean. Mean reaction times were additionally calculated for each participant and scores two standard deviations above the grand mean were also excluded from the study (two participants). Grand means were generated for each condition for all words, and separately for low frequency words.

Two-tailed paired sample *t*-tests were used in order to directly evaluate both reaction times and accuracy in the comparisons of interest (i.e., the match and control conditions – repetition priming, and the match and mismatch conditions – talker-specific priming).

ERPs

EEG was recorded from 16 scalp electrodes using a BioSemi Active Two high-impedance amplifier (BioSemi Instrumentation, Amsterdam). EEG signals were acquired at a sampling rate of 512 Hz, and referenced to two averaged, offline mastoid electrodes. Data were epoched from -100 to 800 ms relative to stimulus presentation, and baseline corrected to the 100 ms pre-stimulus interval. An eyeblink-correction algorithm was implemented using EMSE (Source Signal Imaging, San Diego) to remove blinks from the EEG trace. Based on previous work demonstrating components, such as the

N200 (sublexical processing) (Huang, Yang, Zhang, & Guo, 2014; Lee, Harkrider & Hedrick, 2012; van den Brink, Brown & Hagoort, 2001) and the N400 (lexical processing) (Chwilla, Brown, & Hagoort, 1995; Kutas & Hillyard, 1980, 1984; Kutas & Iragui, 1998) were more likely to have repetition and talker-specific priming waveform divergences (Dufour, et. al., 2017), these two broad time windows were initially chosen for analysis a priori. These time windows were then adjusted and confirmed based on visual inspection of our data. Three time windows were thus identified and chosen for statistical analyses around the peak amplitude of three deflections: 150-200 ms, 250-325 ms, and 600-800 ms. ERPs were quantified as the average amplitude across each time window: average amplitude of each component was extracted and subjected to inferential analysis.

Separate repeated measures ANOVAs were performed on ERP amplitudes in each of the selected time windows (150-200 ms, 250-325 ms, and 600-800 ms), with congruency (match, mismatch, control) and electrode location (C3, Cz, C4, P3, Pz, and P4) as within-group factors. Hypotheses were evaluated based on the main effect of congruency (match vs control, and match vs mismatch), while amplitude at each electrode location was treated as a repeated measurement rather than an effect of interest. The Greenhouse-Geisser (1959) correction was employed to adjust for non-independence of amplitudes at neighboring electrodes. Any participant who was excluded from the analyses of the behavioral data based on poor accuracy or slow reaction time was also excluded from the analyses of the physiological data.

CHAPTER IV

RESULTS

Behavioral Results

Paired sample *t*-tests of accuracy scores revealed a significant repetition priming effect between the match ($M \pm SD$ = 92±6) and control ($M \pm SD$ = 88±8) conditions ($t(35) = 3.60$, $p < .01$, $d = 0.57$). Accuracy on matching ($M \pm SD$ = 92±6) and mismatching trials ($M \pm SD$ = 90±7) did not significantly differ ($t(35) = 1.61$, $p = 0.12$), so there was no evidence of talker-specific priming.

Similarly to the accuracy results, paired samples *t*-tests of the reaction times demonstrated a significant repetition priming effect between the match ($M \pm SD$ = 880±86) and control ($M \pm SD$ = 909±94) conditions ($t(35) = -4.35$, $p < 0.01$, $d = 0.32$), and there was no significant talker-specific priming effect evident between the match ($M \pm SD$ = 880±86) and mismatch ($M \pm SD$ = 879±86) conditions ($t(35) = 0.11$, $p = 0.91$) (see Figure 3).

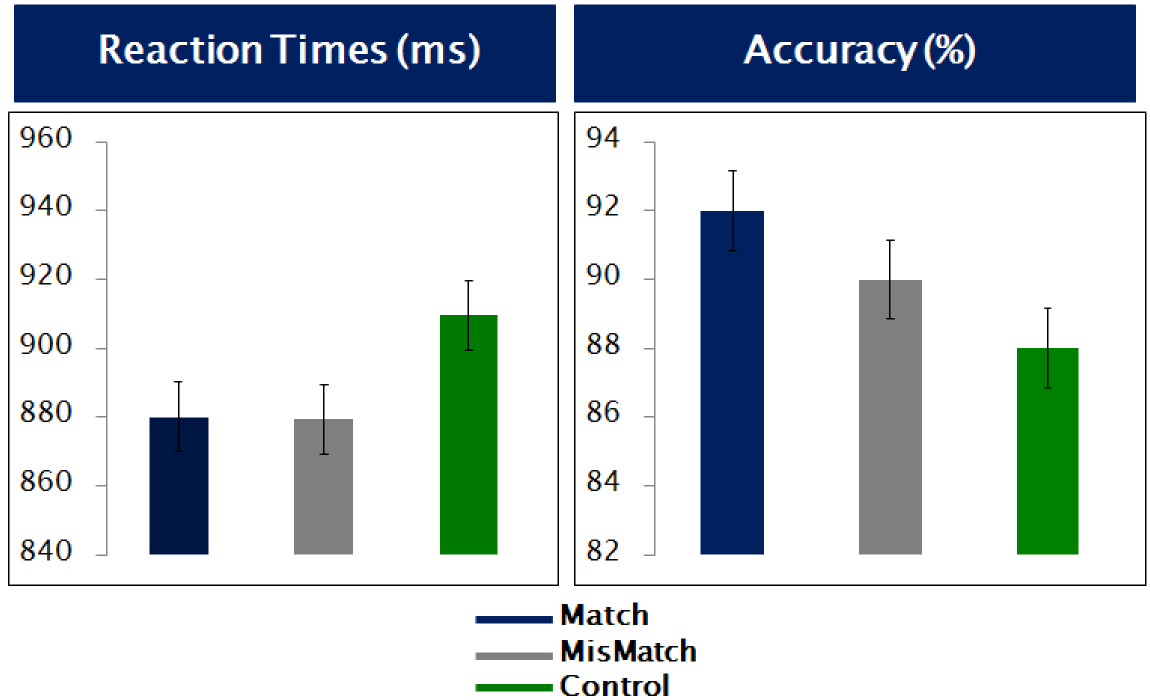


Figure 3. Behavioral Results. Bars indicate standard error.

ERP Results

There were no significant repetition priming effects (comparing the match and control amplitudes) in the 150-200 ms time window ($F(1, 35) < 0.01$, $p > .99$), the 250-325 ms time window ($F(1, 35) = 0.51$, $p = 0.48$), or the 600-800 ms time window ($F(1, 35) = 2.63$, $p = 0.11$) (see Figure 4). Additionally, there were no significant talker-specific priming effects (comparing the match and mismatch amplitudes) at the observed time windows: 150-200 ms ($F(1, 35) = 0.00$, $p > .99$), 250-325 ms ($F(1, 35) = 2.03$, $p = 0.16$), 600-800 ms ($F(1, 35) < 0.01$, $p > .99$) (see Figure 4).

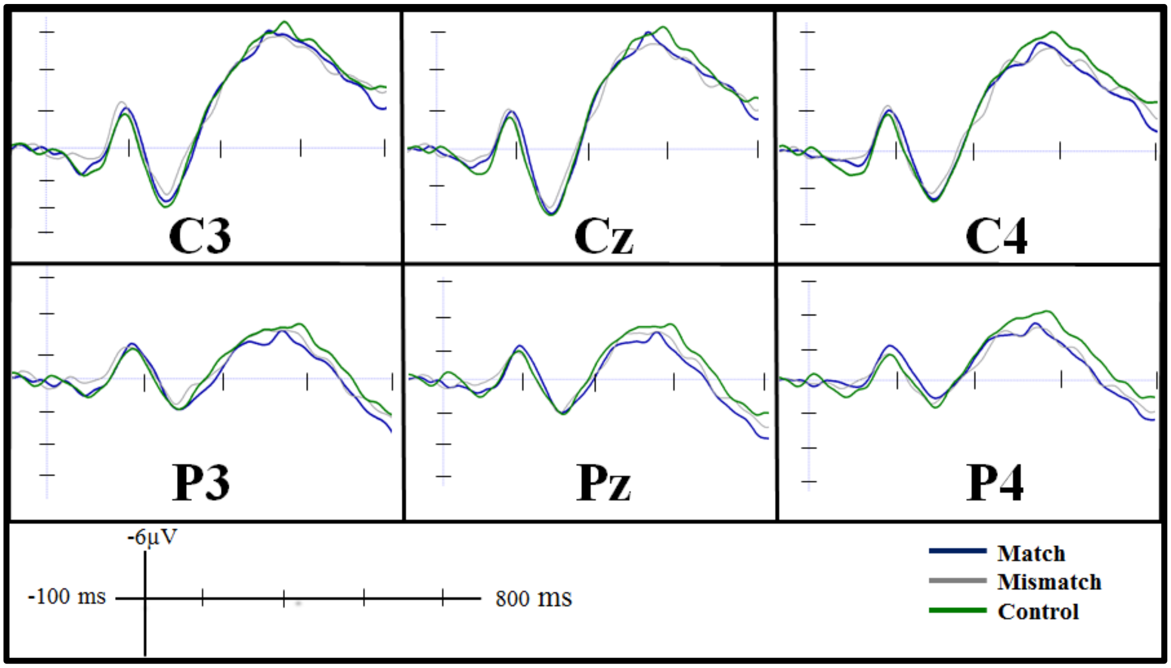


Figure 4. ERP Results.

Responses to Low Frequency Words

In comparing the behavioral results with those from the physiological data-, repetition priming effects were present within the reaction times and accuracy scores, but not within the ERPs. This could be an indication that the task was not sufficiently difficult. Previous work as demonstrated that talker-specific priming effects are more likely to be elicited from low frequency words (Dufour, et. al., 2017) and when a task is more difficult (McLennan & Luce, 2005). The unfamiliarity with low frequency words may have naturally increased the difficulty of the lexical decision task. In an attempt to boost any signal of a talker-specific priming effect, low frequency words (N = 55, $M \pm SD$ = 2.25±0.52) were separated from the high frequency words (N = 55, $M \pm SD$ = 3.43±0.38) for additional analyses

Behavioral Results. With regards to accuracy, paired samples *t*-tests of the low frequency words showed a significant repetition priming effect between the match

($M\pm SD$ = 90±7) and control ($M\pm SD$ = 83±11) conditions ($t(35) = 3.65$, $p < .01$, $d = 0.76$), as well as a talker-specific priming effect between the match ($M\pm SD$ = 90±7) and mismatch ($M\pm SD$ = 87±8) conditions ($t(35) = 2.45$, $p < 0.05$, $d = 0.40$).

Within reaction times, there was a significant repetition priming effect between the match ($M\pm SD$ = 894±93) and control ($M\pm SD$ = 933±95) conditions ($t(35) = -5.81$, $p < 0.01$, $d = 0.40$), but there was no talker-specific priming effect between the match ($M\pm SD$ = 894±93) and mismatch ($M\pm SD$ = 902±90) conditions ($t(35) = -1.21$, $p = 0.24$) (see Figure 5).

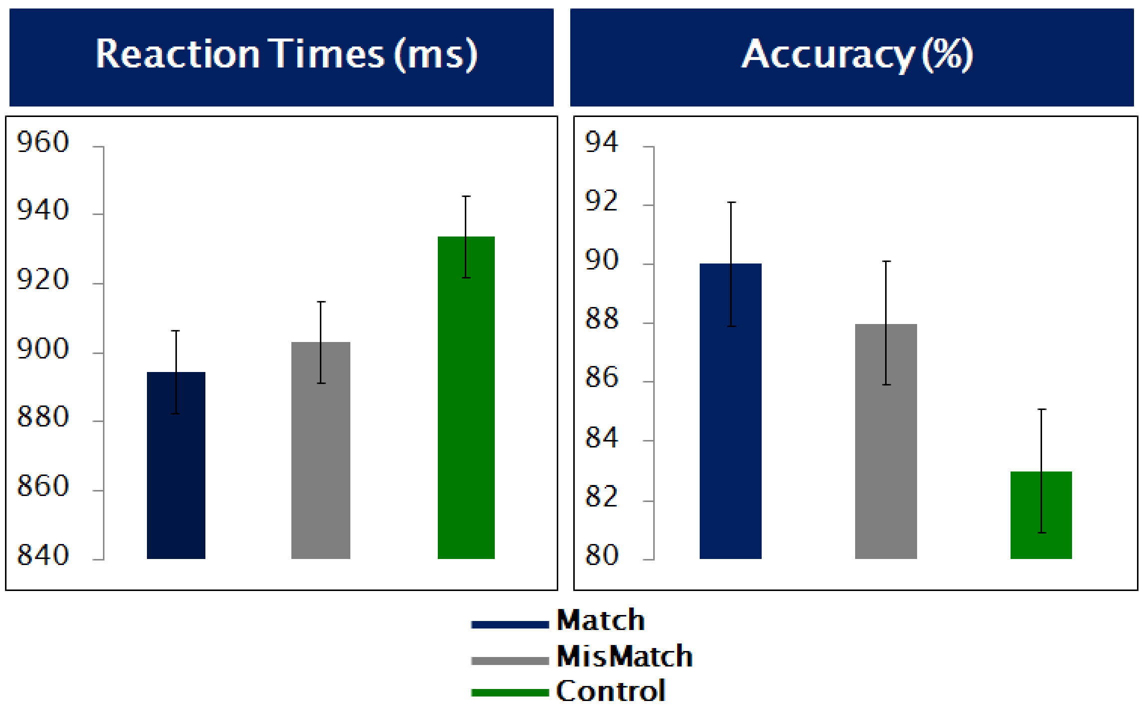


Figure 5. Behavioral Results – Low frequency words.

ERP Results. There was no significant repetition priming effect between the match and control conditions within the 150-200 ms time window for low frequency words ($F(1, 35) = 2.10$, $p = 0.16$). However, the repetition priming effect was significant both within the 250-325 ms time window ($F(1, 35) = 5.35$, $p < 0.05$, $\eta^2 = 0.13$); and within the 600-800 ms time window ($F(1, 35) = 7.79$, $p < 0.01$, $\eta^2 = 0.18$) (see Figure 5). There were no significant talker-specific priming effects within earliest or latest time windows analyzed: 150-200 ms ($F(1, 35) = 0.17$, $p = 0.68$); 600-800 ms ($F(1, 35) = 2.39$, $p = 0.13$) (see Figure 6, and footnote[2] regarding ERPs at the 250-325 time window).

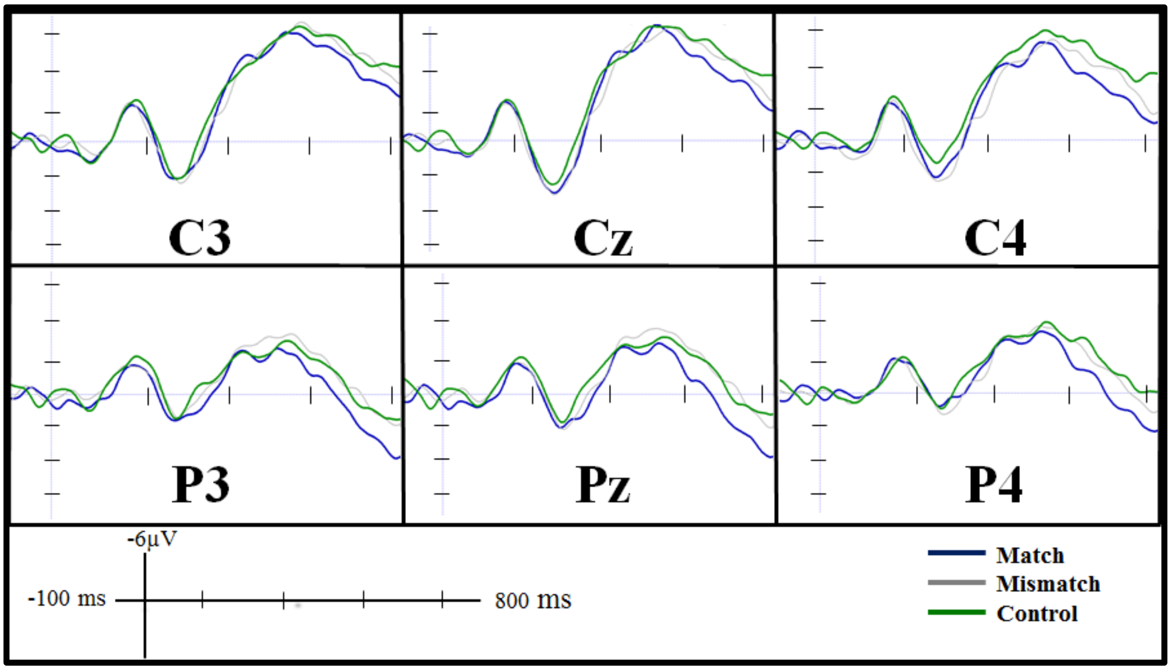


Figure 6. ERP Results – Low frequency words.

Table I. Reaction Times and Accuracy

	Match		Mismatch		Control	
	RT	A	RT	A	RT	A
All Words	880 ms	92%	879 ms	90%	909 ms	88%
Low Frequency Words	894 ms	90%	902 ms	87%	933 ms	83%
High Frequency Words	863 ms	94%	856 ms	93%	886 ms	90%

Note. RT = reaction times; A = Accuracy

All Spoken Words

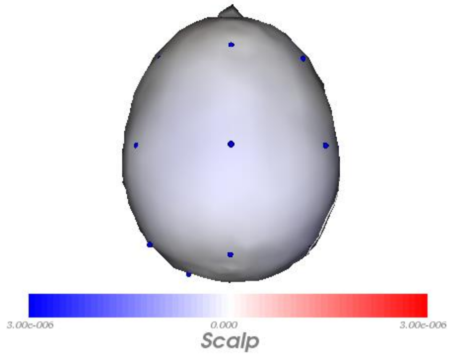
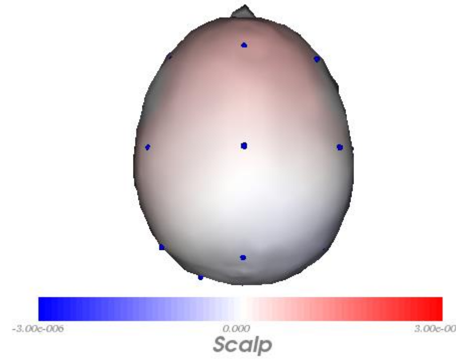
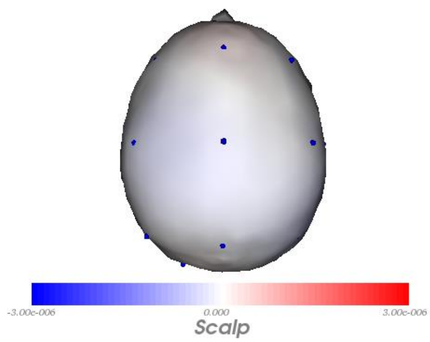
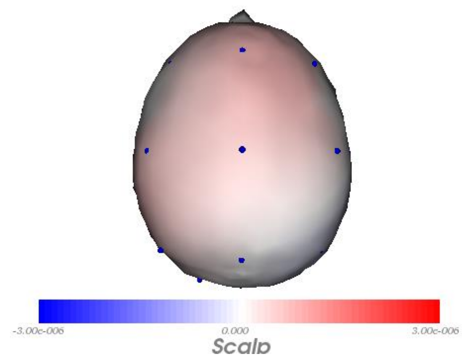
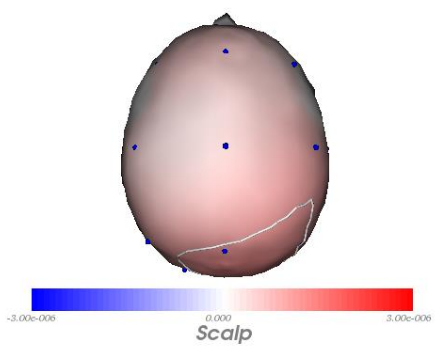
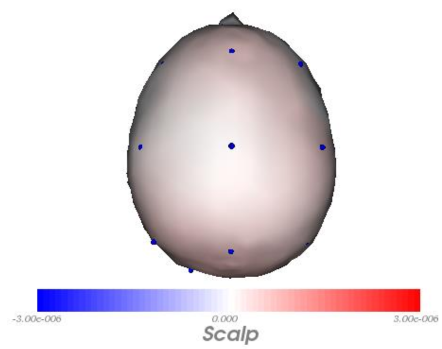
	Repetition Priming	Talker-specific priming
150-200 ms		
250-325 ms		
600-800 ms		

*Figure 7***.** Topographic Graphs—All Spoken Words**.** Topographic plots for repetition priming were calculated by subtracting the control from the match condition ERPs. Plots indexing talker-specific priming were calculated by subtracting the mismatch from the match condition ERPs.

Low Frequency

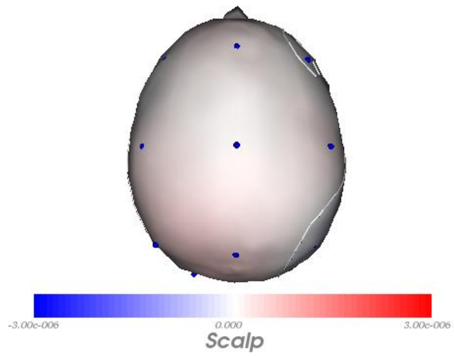
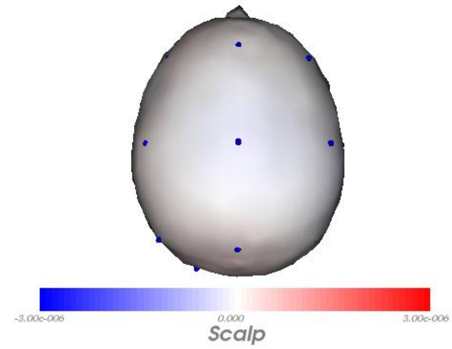
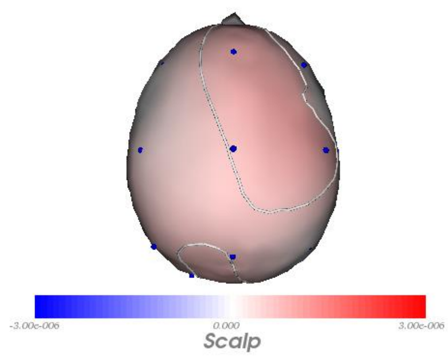
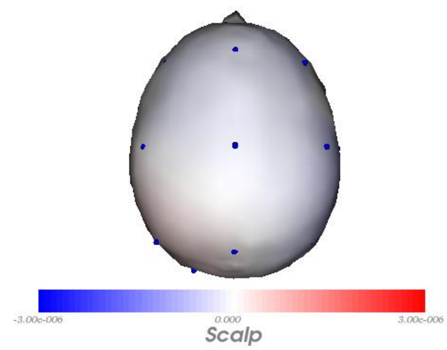
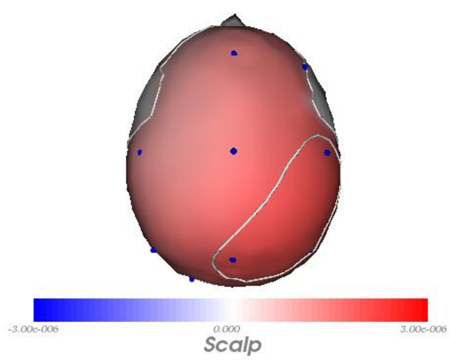
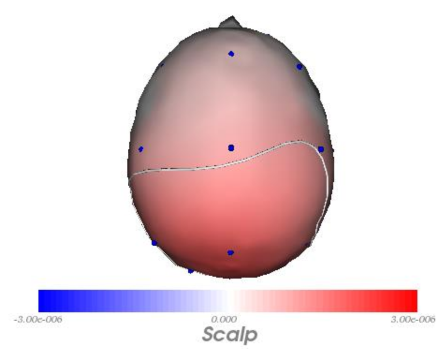
	Repetition Priming	Talker-Specific Priming
150-200 ms		
250-325 ms	* 	
600-800 ms	** 	

*Figure 8***.** Topographic Graphs – Low frequency words**.** Topographic plots for repetition priming were calculated by subtracting the control from the match condition ERPs. Plots indexing talker-specific priming were calculated by subtracting the mismatch from the match condition ERPs. * $p < 0.05$; ** $p < 0.01$

CHAPTER V

DISCUSSION

The goal of the present study was to address three competing theories of spoken word recognition (abstract representational theories, episodic representational theories, and a hybrid representational theory, as detailed in the time-course hypothesis) by examining behavioral and electrophysiological responses in a long-term repetition priming paradigm. By exploring accuracy, reaction times, and ERPs responses to a difficult auditory lexical decision task, we were expecting to find results supporting two concepts: *repetition priming* and *talker-specific priming*. We found behavioral and electrophysiological evidence for repetition priming (most evident for low frequency words), and behavioral evidence of talker-specific priming (in accuracy). We had additionally anticipated finding electrophysiological evidence supporting a hybrid model of spoken word recognition, the *time-course hypothesis* such that episodic representations would have been evident in later processing, while abstract representations would have been more dominant earlier in processing. While our ERP data did not reflect waveform patterns consistent with the time-course hypothesis, the presence of talker-specific priming (in accuracy) in a difficult lexical decision task with low frequency words did provide evidence consistent with a hybrid model.

An important and major finding within the current study is physiological evidence of long-term repetition priming. Obtaining long-term repetition priming within both reaction time and accuracy discrepancies is not as novel a finding as is that of long-term repetition priming through the ERPs. Previous physiological work to date has had important differences from the current study: including the use of full sentences (Dahaene-Lambertz, et. al. 2006) or single syllables (Belin & Zatorre, 2003), using a short-term priming paradigm (Dufour, et. al., 2017), and/or employing different physiological methods (Tang, Hamilton & Chang, 2017). To our knowledge, this study is the first to find evidence of long-term repetition (identity) priming in isolated spoken word recognition with ERPs.

While evidence of repetition priming within the ERPs of low frequency words is a new finding, the lack of repetition priming collapsed across high and low frequency words is important. High frequency words are those with which we are most familiar. The constant exposure to these words suggests that a priming effect of the high frequency words from one test block to another may not have been present. When collapsing both low and high frequency words into one analysis, the priming effect from the low frequency words was attenuated by the lack of priming effect from the high frequency words. Nevertheless, priming effects were obtained in the behavioral measures even when collapsing across high and low frequency words.

The second aspect of our research explored talker-specific priming, an effect more likely to occur during a more difficult task and within low frequency words (Dufour, et. al., 2017; McLennan & Luce, 2005). To find this effect would have been consistent with episodic representational theories, suggesting that unique characteristics of a talker's voice are carried and stored within the mental lexicon. The accuracy data provide evidence for the *time-course hypothesis*. In a difficult task and within low frequency words, repeated words spoken by the same talker were responded to significantly more accurately than repeated words spoken by a different talker.

However, our physiological data, if taken alone (i.e., without consideration of the behavioral results), would support abstract representational theories of spoken word recognition (see Figure 2B)—given the lack of talker-specific priming. While contributing to the current body of work comparing and contrasting contradictory theories, our work did not elicit any talker-specific priming effects, despite findings suggesting that such effects would be more pronounced from low frequency words (Dufour & Nguyen, 2014; Dufour, et. al., 2017). Whether a lack of talker-specific priming within our data accurately reflects abstract representational theories, or that talker-specific priming was simply inaccessible through our measure, remains to be seen.

One possible explanation for not finding talker-specific priming within our study could be that these effects are not physiologically sustained over long intervals. Behavioral talker-specific priming effects have been noted within both short- and long-term priming paradigms. However, physiological evidence for talker-specific priming has been obtained within the short-term priming paradigm (Dufour, et. al, 2017), but not the long-term priming paradigm. Perhaps somewhere between short- and long-term priming, any physiologically effects of talker-specific priming decays. One avenue to pursue with continued research is a replication of the present study, utilizing a short-term priming paradigm, rather than long-term. Alternatively, by manipulating the time between prime and target exposures, a time line for talker-specific priming decay could potentially be established.

Another possible explanation for a lack of talker-specific priming within our data is that this type of effect occurs in areas of the brain where EEG coverage is limited, such as the sulcal depths of the neocortex or deep brain structures. Future studies employing other methods may help to localize talker-specific priming. From research using fMRI techniques, the *right anterior* superior temporal sulcus (STS) has been implicated as playing an important role in storing the representation of individual voices by demonstrating significantly less activation when syllables were spoken by a single voice than when spoken by different voices (Belin &Zatorre, 2003). This effect of talker variability has not only been found in the right STS, but also in the *left* STS, which has shown evidence of sentence priming, where there was less activation for repeated sentences, even when speakers varied (Dehaene-Lambertz, et. al., 2006). There seems to be an indication that the *left* STS contributes to the encoding of more abstract representations of spoken word, while the *right* STS may encode not only abstract representations, but also episodic representations related to the talker's voice.

In an electrocorticography study, intracranial electrodes in an area next to STS, the *left* superior temporal gyrus (STG), were sensitive to intonation, pitch contour, and talker, suggesting that talker-specific priming may be localized somewhere within the superior temporal lobe (Tang, Hamilton & Chang, 2017).

Another way to pursue this line of research would be to consider an analysis of nonwords[3]. When engaging in long-term repetition priming, the assumption is that spoken word recognition involves the access of lexical representations within the mental lexicon. Our data found more robust results for both repetition priming and talker-specific priming within low frequency words, substantiating this assumption. If the assumption is true, then an analysis of nonwords, which have no representation in long-term memory, should reveal no evidence of repetition priming. By establishing this dissociation between the priming for real words and the lack of priming for nonwords, we would have a better understanding about the contents and processing within the mental lexicon.

As mentioned previously, the accuracy results from the current study support the time-course hypothesis, as responses within a lexical decision task were more accurate when the talker remained unchanged from one block to the next, but only when the task was more difficult through the use of low frequency words. However, there is additional evidence supporting both abstract and episodic representational theories: the behavioral data demonstrated talker-specific priming, consistent with the latter, while the physiological data presented no talker-specific priming, consistent with the former. Utilizing a short-term rather than a long-term priming paradigm could potentially sway the current debate one direction or another. Additionally, attempts to localize talker-specific priming in the brain may provide stronger evidence supporting episodic representational theories. Other physiological measures could also be administered in order to better assess hybrid models, such as the time-course hypothesis. Our data

[3] An analysis of the nonwords is beyond what was necessary for the sake of this thesis project. However, this data was collected and there are plans to analyze it at a later time.

provide physiological evidence of long-term repetition priming, providing a springboard into a new line of research in spoken word recognition.

REFERENCES

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine

Belin, P. & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport*, *14*, 2105-2109

Campeanu, S., Craik, F. I. M. & Alain, C. (2013). Voice congruency facilitates word recognition. *PLoS One, 8*(2), e58778.

Chwilla, D., Brown, C. & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*, 274-285.

Connolly, J. F., Phillips, N. A., Stewart, S. H. & Brake, W. G. (1992). Event-related potential to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43,* 1-18.

Craik, F. I. M. & Kirsner, K. (1974). The effect of speaker's voice on word recognition, *The Quarterly Journal of Experimental Psychology*, *26*(2), 274-284.

Dehaene-Lambertz, G., et. al. (2006). Functional segregation of cortical language areas by sentence repetition. *Human Brain Mapping*, *27,* 360-371.

Desroches, A. S., Newman, R. L. & Joanisse, m. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influence of phonological similarity. *Journal of Cognitive Neuroscience*, *21,* 1893-1906.

Dufour, S., Bolger, D., Massol, S., Holcomb, P. J. & Grainger, J. (2017). On the locus of talker-specificity effects in spoken word recognition: An ERP study with dichotic priming. *Language, Cognition, and Neuroscience*, *32,* 1273-1289.

Dufour, S., Brunellière, A. & Frauenfelder, U. H. (2012). Tracking the time course of word frequency affects in auditory word recognition with event-related potential. *Cognitive Science, 34*, 489-507.

Dufour S. & Nguyen, N. (2014). Access to talker-specific representations is dependent on word frequency. *Journal of Cognitive Psychology, 26*(3), 256-262.

Friedrich, C. K., Kotz, S. A., Friederici, A. D. & Alter, K. (2004). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Journal of Cognitive Brain Research, 20,* 300-308.

Friedrich, C. K., Kotz, S. A., Friederici, A. D. & Gunter, T. C. (2004). ERPs reflect lexical identification in word fragment priming. *Journal of Cognitive Neuroscience, 16,* 541-552.

Friedrich, C. K., Schild, U. & Röder, B. (2009). Electrophysiological indices of word fragment priming allow characterizing neural stages of speech recognition. *Biological Psychology, 80,* 105-113.

Gaskell, M. G. & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*(5-6), 613-656.

Greenhouse, S. W. & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-111.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F. & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage, 30,* 1383-1400.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47-61.

Huang, X. H., Yang, J., Zhang, Q. & Guo, C. (2014). The time course of spoken word recognition in Mandarin Chinese: A unimodal ERP study. *Neuropsychologia, 63,* 165-174.

Hunter, C. R. (2013). Early effects of neighborhood density and phonotactic probability of spoken words on event-related potentials. *Brain & Language, 127*, 463-474.

Kutas, M. & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

Kutas, M. & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalogram Clinical Neurophysiology, 108*, 456-471.

Lee, J. Y., Harkrider, A. W. & Hedrick, M. S. (2012). Electrophysiological and behavioral measures of phonological processing of auditory nonsense V-CV-VCV stimuli. *Neuropsychologia, 50*, 666-673.

Luce, P. A., Goldinger, S. D., Auer, E. T. & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, *62*(3), 615-625.

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R. & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cerebral Cortex, 17,* 2659-2668.

McClelland, J. S. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.

McLennan, C. T. & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 306-321.

Mullennix, J. W., Pisoni, D. B. & Martin, C. S. (1989). Some effects of talker variability on spoken-word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America, 85*(5), 2088-2113.

Olichney, J. M., Iragui, V. J., Salmon, D. P., Riggins, B. R., Morris, S. K. & Kutas, M. (2006). Absent event-related potentials (ERP) word repetition effect in mild Alzheimer's disease. *Clinical Neurophysiology, 117*(6), 1319-1330.

Palmeri, T. J., Goldinger, S. D. & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(2), 309-328.

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate, and perceptual learning. *Speech Communication, 13*, 109-125.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9-32). San Diego, CA: Academic Press.

Tang, C., Hamilton, L. S. & Chang, E. F. (2017). Intonational speech prosody encoding in the human auditory cortex. *Science, 357*, 797-801.

van den Brink, D., Brown, C. M. & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13*(7), 967-985.

Vergara-Martinez, M. & Swaab, T. Y. (2012). Orthographic neighborhood effects as a function of word frequency: An event-related potential study. *Psychophysiology, 49*(9), 1277-1289.

Vitevitch, M. S. & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavioral Research Methods Instrumental Computation, 36*(3), 481-487.

APPENDIX A

Stimuli List of Real Words

bell	Jar	pad	yam
bile	jazz	pal	year
boat	jeep	pave	yearn
cab	jerk	pep	youth
cheese	jet	perk	zone
choice	juice	pill	zoom
chore	keg	pine	
church	king	pit	
curl	kiss	poll	
death	kit	pun	
dime	knit	rhyme	
dine	lace	rip	
ditch	lack	road	
dodge	lag	sag	
dog	lair	seal	
door	lamb	search	
faith	learn	seat	
feet	leech	shun	
file	leek	sing	
fill	liar	soak	
fin	line	soothe	
foal	load	tail	
foam	loathe	term	
foil	luck	theme	
gate	mead	tip	
gnome	meat	top	
gun	mess	tov	
gush	mice	town	
ham	mile	vase	
hat	moan	veil	
hood	mope	weave	
hook	mouth	wedge	
hoop	myth	weed	
hop	nag	weep	
hose	nail	wig	
hug	nap	wing	
hurl	niche	wipe	
jail	none	yacht	

APPENDIX B

Stimuli List of Nonwords (NW) with Klattese**

NW	Klattese	NW	Klattese	NW	Klattese	NW	Klattese
bidge	bIJ	hoong	huG	mouk	mWk	soog	sug
bipe	bYp	hov	hav	moze	moz	sov	scv
bup	b^p	jach	J@C	Muj	m^J	teeg	tig
buv	b^v	jaych	JeC	myb	mIb	terch	tRC
caj	k@J	jeeg	Jig	nach	n@C	thebe	Tib
cheem	Cim	jers	JRz	naich	neC	tith	tIT
churf	CRf	jev	JEv	newth	nuT	tush	t^S
desh	dES	joof	Juf	nidge	nIJ	vafe	vef
dij	dIJ	kack	k@k	niz	nIz	vage	veJ
dishe	dIS	keeb	kib	nup	n^p	veech	viC
dith	dIT	kej	kEJ	pab	p@b	verf	vRf
div	dIv	kib	kIb	pafe	pef	weej	wiJ
dobe	dob	kif	kIf	pash	p@S	weesh	wiS
doch	daC	kiv	kIv	perd	pRd	weeth	wiT
dofe	dof	laib	leb	pij	pIJ	wef	wEf
doov	duv	laj	l@J	pithe	pYD	wib	wIb
fid	fId	lav	l@v	poath	poT	wid	wId
fip	fIp	layshe	leS	pof	pof	wije	wYJ
fise	fYz	laz	l@z	pum	p^m	wis	wIs
foach	foC	learm	lRm	raig	reg	wiv	wIv
fom	fcm	leck	lEk	reiche	rYC	yaf	yef
foof	fuf	leeb	lib	riz	rIz	yath	ycT
fythe	fYT	lige	lYg	roje	roJ	yawp	ycp
gnoge	nog	loach	loC	roshe	roS	yeaf	yif
gouche	gWC	loaj	loJ	rothe	roD	yearp	yRp
gub	g^b	loash	loS	saj	s@J	zoke	zok
gud	g^d	lurve	lRv	saych	seC	zoog	zug
haab	heb	meag	mig	seeb	sib		
haav	hev	mep	mEp	seef	sif		
hadg	h@J	mife	mYf	shup	S^p		
hige	hYJ	moach	moC	soaf	sof		

**Klattese is a computer-friendly system of phonetic transcription (Vitevich & Luce, 2004).

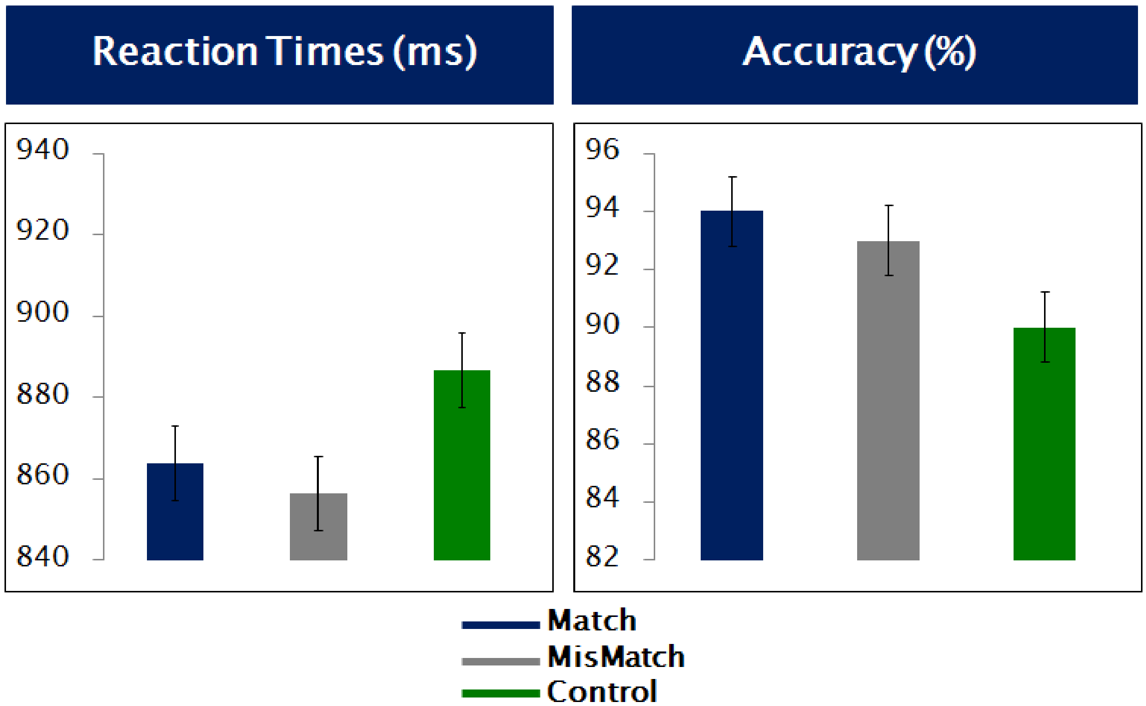
APPENDIX C

Low and High Frequency Words

Low Frequency		High Frequency	
bile	perk	bell	pal
chore	pill	boat	rip
curl	pine	cab	road
dime	pit	cheese	seal
dine	poll	choice	search
ditch	rhyme	church	seat
dodge	sag	death	sing
fin	theme	dog	tail
foal	veil	door	term
foam	weave	faith	tip
foil	wedge	feet	top
gnome	weed	file	town
gush	weep	Fill	vase
ham	wig	gate	wing
hoop	yacht	gun	wipe
hurl	yam	hat	year
jar	yearn	hood	youth
jazz	zoom	hook	zone
jeep		hop	
jet		hug	
knit		jail	
lace		jerk	
lag		juice	
lair		king	
lamb		kiss	
leech		lack	
leek		learn	
loathe		liar	
mead		line	
mice		load	
moan		luck	
mope		meat	
myth		mess	
nag		mile	
nap		mouth	
niche		nail	
pad		none	

APPENDIX D

Behavioral Data for High Frequency Words



Behavioral Results. With regards to accuracy, paired samples *t*-tests of the high frequency words showed a trending repetition priming effect between the match (*M*±*SD* = 94±7) and control (*M*±*SD* = 90±9) conditions (*t*(35) = 1.88, *p* = 0.069), yet no talker-specific priming effect between the match (*M*±*SD* = 94±7) and mismatch (*M*±*SD* = 93±9) conditions (*t*(35) = 0.44, *p* = 0.66).

Within reaction times, there was a trending repetition priming effect between the match (*M*±*SD* = 864±85) and control (*M*±*SD* = 887±98) conditions (*t*(35) = -2.02, *p* = 0.051), but there was no talker-specific priming effect between the match (*M*±*SD* = 864±85) and mismatch (*M*±*SD* = 856±87) conditions (*t*(35) = 0.89, *p* = 0.38).