Cortical Processing of Phonetic and Emotional Information in Speech: A Cross-Modal Priming Study

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Erin Kathleen Diamond

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Dr. Yang Zhang

May 2015

Acknowledgements

Funding for this project provided by the Brain Imaging Research Project Award from the College of Liberal Arts. Many thanks to my advisor and mentor, Dr. Yang Zhang, for his patient guidance and invaluable support from the first day I stepped into his lab.

Dedication

This thesis is dedicated to the faculty of the Department of Speech-Language-Hearing Sciences and the administrators in the College of Liberal Arts at the University of Minnesota who saw the researcher in me and encouraged me to keep asking questions.

Abstract

The present study utilized a cross-modal priming paradigm to investigate dimensional information processing in speech. Primes were facial expressions that varied in two dimensions: affect (happy, neutral, or angry) and mouth shape (corresponding to either /a/ or /i/ vowels). Targets were CVC words that varied by prosody and vowel identity. In both the phonetic and prosodic conditions, adult participants responded to congruence or incongruence of the visual-auditory stimuli. Behavioral results showed a congruency effect in percent correct and reaction time measures. Two ERP responses, the N400 and late positive response, were identified for the effect with systematic between-condition differences. Localization and time-frequency analyses indicated different cortical networks for selective processing of phonetic and emotional information in the words. Overall, the results suggest that cortical processing of phonetic and emotional information involves distinct neural systems, which has important implications for further investigation of language processing deficits in clinical populations.

Table of Contents

List of Figures

Chapter 1: Introduction

This chapter provides a literature review motivating the current thesis project on phonetic and prosodic processing in the brain. It begins with an overview of the elements of speech, which is then further broken down into the elements of interest for this investigation: phonetics and prosody. Following is a brief discussion of priming paradigms and their use in behavioral and event-related potential research. The event-related potential components that are particularly relevant to this study (the N400 and late positive response) are then reviewed. The fifth and sixth sections review the importance of localization and cortical rhythm analyses for evaluating phonetic and prosodic processing. The final section presents research questions for the current investigation of phonetic and prosodic processing in an affective priming paradigm.

1.1 Elements of speech

Simply defined, speech is the acoustic representation of language. But researchers' efforts to reduce speech to its essential elements have not yielded a single clear description due to the co-existence of multiple informational dimensions in the speech signal. The seemingly concise dichotomies of *linguistic* and *paralinguistic* or *segmental* and *suprasegmental* do not adequately describe the complexity of the speech signal.

The linguistic content of speech is universally viewed as an arbitrary language code that can be broken down into vowels, consonants, or their attributes (Fox, 2000). The term “segmental” describes these discrete units, the most basic of which is the phoneme. Laver and Hutcheson (1972) propose that paralinguistic information encompasses all other elements of speech, including non-linguistic and non-verbal features (both vocal and non-vocal). The related term “suprasegmental” describes the phonological structure superimposed on the segments (i.e., the use of stress on a syllable), which can also carry language-specific features.

The word *prosody* comes from the Greek *prosodia* meaning “sung to music.” Describing prosody in the context of the whole speech signal presents a challenge. Let us first consider prosody in the context of paralinguistic information. Laver and Hutcheson (1972) propose that the non-verbal, vocal paralinguistic features can be described as the speaker’s tone of voice. However, Fox (2000) points out that prosodic features of linguistic significance also express tone of voice. Crystal (1969) suggests that prosodic features consist of variations in pitch, loudness, duration and silence whereas paralinguistic features, while vocal, are identified independently of these characteristics. Reflecting on these comparisons, it should come as no surprise that there is some conceptual overlap in the definitions between the terms *paralinguistic* and *prosodic* (Fox, 2000; Roach, Stibbard, Osborne, Arnfield & Setter, 1998).

While we can use the phonetic characteristics of length, accent, intonation, and tone to distinguish prosodic from phonetic features, these descriptors alone are insufficient. Prosody is too complex for its features to simply be superimposed on speech segments as a collection of suprasegmental characteristics. The acoustic cues of pitch frequency, duration, and intensity have been described as the “quasi-independent” basic parameters of prosody (Fox, 2000). This is to say that these acoustic cues are often mutually dependent, despite the possibility of manipulating them independently of one another.

Manipulating these cues allows speakers to communicate emotion through prosody. In the current experiment, we were particularly interested in the emotional aspect of prosody as opposed to its phonetic characteristics. Understanding the brain mechanisms that govern the proper use of these affective cues along with the expression of linguistic content is important for theories on the neural representations of language as well as practical applications such as intervention for individuals with communication difficulties in terms of affective speech comprehension/production. Brain imaging studies have investigated the timing, localization, and cortical rhythm characteristics related to processing the phonetic and prosodic elements of speech, which will be further discussed in the upcoming sections.

1.2 Priming paradigms

One common experimental task in psycholinguistic research is the use of priming. *Priming* refers to the facility with which a participant processes a stimulus (target) based on an earlier experience with a specific stimulus (prime). Researchers measure the accuracy rate and reaction times associated with different types of targets to make inferences about memory organization and the cognitive processes involved. In an early priming experiment, Segal (1966) measured activation of a primed word based on its appearance in a subsequent association task. Priming has often been investigated through an identification task using degraded stimuli, where the participant has studied only some of the stimuli in a prior session (Roediger & McDermott, 1993, as cited in Greene, Eaton & LaShell, 2001). Reduced reaction times in response to semantically or conceptually congruent prime-target pairs (i.e., pet-cat, furniture-chair) have generally been seen to reflect spreading-activation theories for memory (Collins & Loftus, 1975).

Priming paradigms have been employed in behavioral and event-related potential studies to evaluate the intersection between pairs of stimuli both within and across modalities. In these paradigms, primes and targets are combined to investigate the effects of their congruency or incongruency in a specified aspect. Greene et al. (2001) explored the effect of within-modal and cross-modal priming on spatio-temporal event processing. They observed an effect for visual priming of auditory targets, but not the reverse. The authors therefore proposed that a visual event provided specific information that could facilitate processing of an auditory target, whereas the auditory primes were more vague and had the potential to correspond to a wide range of visuals.

1.2.1 Affective priming

An adjective evaluation task created by Fazio, Sanbonmatsu, Powell, and Kardes (1986) was an early model of the affective priming. In that experiment, affectively related primes facilitated an evaluative decision of adjective targets, as demonstrated by shorter latencies preceding the adjective evaluation ("good" or "bad"). This has been described as "automatic attitude activation" (for a review, see Fazio, 2001).

Affective priming paradigms since that time have extended beyond the traditional visual word prime-target pairs to explore interactions between stimulus domains and modalities. Picture primes and written word targets have been used to investigate the neural mechanisms at play during cross-domain visual affective priming paradigms (Zhang, Lawson, Guo & Jiang, 2006; Zhang, Li, Gold & Jiang, 2010).

Other prime-target pairs have spanned two modalities, combining stimuli such as affective sentence prosody and written words (Schirmer, Kotz & Friederici, 2002) or musical stimuli and written words (Goerlich, Witteman, Schiller, Van Heuven, Aleman & Martens, 2012). Facial expressions have been used as targets in cross-modal priming paradigms with sentence primes (Czerwon, Hohlfeld, Wiese & Werheid, 2013) and musical primes (Lense, Gordon, Key & Dykens, 2014). Other cross-modal affective priming paradigms have used facial expressions as primes for emotional words (Schirmer & Kotz, 2006) and musical stimuli (Kamiyama, Abla, Iwanaga & Okanoya, 2013).

1.2.2 Stimulus onset asynchrony

When investigating the neural mechanisms at play in priming paradigms, it is important to consider the influence of the stimulus onset asynchrony (SOA). The SOA, or the duration from the onset of the prime to the onset of the target, can impact brain responses to affective priming. Hermans, De Houwer, and Eelen (2001) found an affective priming effect for an SOA of 300 milliseconds (ms), but not 1000 ms. Based on this research, Goerlich et al. (2012) employed a constant SOA of 200 ms in their investigation of cross-modal affective priming in music and speech.

Zhang et al. (2010) investigated the effect of short SOAs in their affective priming experiment with picture primes and word targets. Congruent and incongruent picture-word pairs were presented with stimulus onset asynchronies of 150 ms and 250 ms. Results suggest that emotional picture primes affect target word processing, even at very short SOAs.

1.3 Behavioral and event-related potential measures

In order to determine the facilitative effect of primes on targets, a variety of behavioral and brain measures have been utilized. As mentioned previously, differences in reaction time in response to congruent vs. incongruent pairs are thought to reflect increased or decreased facilitation of target processing. In conjunction with this measure, event-related potentials (ERPs) are especially valuable for investigating the timing of neural components underlying both phonetic and prosodic processing. Of particular interest to this experiment are the N400 and the late positive response (late positive component [LPC] or late positive potential [LPP]).

1.3.1 N400 component

Event-related potential research is a noninvasive method of measuring neural responses to specific events. It provides high temporal resolution suitable for investigating brain responses to acoustic and linguistic processing at the millisecond level. A wide research base has established specific ERP components that occur in response to specific linguistic violations (e.g. semantic or syntactic violations). One such response is the N400 component, a negativity occurring approximately 400 ms after a violation of meaning (Kutas & Hillyard, 1980). Traditionally, this component has been studied in the context of semantic expectancy violations in sentences (for a review, see Kutas & Federmeier, 2011). The N400 response has also been observed in semantic congruency studies that pair speech with visual stimuli involving gestures (Kelly, Creigh & Bartolotti, 2010; Kelly, Ward, Creigh & Bartolotti, 2007; Özyürek, Willems, Kita & Hagoort, 2007; Stevens & Zhang, 2013).

Researchers have addressed the role of emotion in speech using a variety of approaches and have identified different components based on those paradigms (for summary, see Appendix A). Recent studies have observed the N400 component in the context of emotional valence congruency violations. Increased N400 amplitude responses have been observed in response to visual affective incongruence of prime-target word pairs (Czerwon et al., 2013) and picture-word pairs (Bostanov & Kotchoubey, 2004; Zhang et al., 2010). Kamiyama et al. (2013) report the presence of the N400 component in response to mismatched stimuli in an affective priming paradigm in which a happy or sad musical stimulus is preceded by a congruent or incongruent facial expression. In the prosodic context, Schirmer and Kotz (2003) identified an N400 component following incongruent conditions of an auditory emotional Stroop task where participants judged the valence of an adverb or the emotional prosody with which it was spoken. Similarly, greater N400 responses occurred in experiments where the valence of a word was incongruent (as opposed to congruent) with the prosody of a preceding sentence (Schirmer et al., 2002; Schirmer, Kotz & Friederici, 2005).

In their variant of a cross-domain affective priming paradigm, Aguado, Dieguez-Risco, Méndez-Bértolo, Pozo, and Hinojosa (2013) observed a reverse priming effect wherein the N400 response was larger in response to congruent compared to incongruent stimuli (facial primes and emotional word targets). Paulmann and Pell (2010) identified the same response in a facial affective decision task. However, reverse priming occurred only for the short primes (200 ms), not for the medium-length primes (400 ms). For these, a classic N400 response was observed.

Zhang et al. (2010) identified both an N400 and a late positive response in their cross-domain affective priming experiment. The N400 response occurred for incongruent picture-word prime-target pairs at a stimulus onset asynchrony (SOA) of 150 ms. In contrast, the late positive potential occurred in response to prime-target incongruency at an SOA of 250 ms. The next section will discuss the late positive response in greater detail.

1.3.2 Late positive response

A late positive response has been observed in affective priming experiments and is generally seen to reflect increased attention to unexpected targets (see summary in Appendix A). Mismatched affective face pairs (e.g. angry-happy) revealed a parietally distributed late positive potential (LPP) between 500–600 ms after presentation of the target face (Werheid, Alpay, Jentzsch & Sommer, 2005). The authors proposed the late positivity was a response to enhanced arousal resulting from the sudden change as opposed to the emotional valence (see Schupp, Cuthbert, Bradley, Cacioppo, Ito & Lang, 2000). Zhang et al. (2010) observed an increased LPP in response to affectively incongruent words between 550–700 ms.

This positive-going deflection is often discussed as a possible variant of the P300 component, reflecting updating working memory (Hajcak, Dunning & Foti, 2009; also see Donchin & Coles, 1988). The P300 is known to be a neurocognitive index of novelty detection and attentional capture, and its amplitude is strongly dependent on the stimulus context and task demands (Nie, Zhang, & Nelson, 2014). Because Zhang et al. (2010) also observed an N400 component in their experiment, they propose that the LPP may in fact be a P300 component reflective of updating working memory. They further suggest that the late positive potential may be indicative of conscious processing of the picture stimuli, evidenced by the lack of a late positive potential in both a word-word paradigm (Zhang et al., 2006) and a subliminal affective priming paradigm with facial stimuli (Li, Zinbarg, Boehm & Paller, 2008).

Other studies have identified a late positive response in experiments of emotion or prosody. The late positivity was identified in various tasks during which the participant was consciously attending to some characteristic of the stimuli, such as congruency (Aguado et al., 2013; Chen, Zhao, Jiang & Yang, 2011; Kamiyama et al., 2013), sound intensity deviation (Chen et al., 2011), or level of arousal (Paulmann, Bleichner & Kotz, 2013).

1.4 Localization

Localization of these processes is another area of investigation. Various methodologies have been implemented to study the regional as well as hemispheric distribution of phonetic and prosodic processing. It is generally accepted that language is primarily localized in the left hemisphere. Emotion processing has greater right hemisphere involvement.

Buchanan et al. (2000) used functional magnetic resonance imaging (fMRI) to investigate the localization of phonemic and prosodic processing. They observed left hemisphere lateralization in the frontal lobe and auditory cortex during phonemic tasks and right hemisphere lateralization in the same regions for the emotion detection task (Buchanan et al.). In another fMRI study, Grandjean et al. (2005) found enhanced responses in the superior temporal sulcus (STS) for angry compared to neutral stimuli presented in a dichotic listening paradigm. Although activation was observed bilaterally, there was a general increase in right STS activity when participants attended to the left ear and no corresponding increase in the left STS when participants attended to the right ear. The authors assert that these effects are attributable to the difference in prosody and could not be accounted for by changes in acoustic cues (i.e. frequency and amplitude).

Schirmer & Kotz (2006) challenge the idea that vocal emotion is uniquely a right hemisphere process. Rather, they propose that prosodic processing is a multi-step process with differential involvement of both hemispheres. Recent investigations into the lateralization of prosody support this model (Iredale, Rushby, McDonald, Dimoska-Di Marco & Swift, 2013; Paulmann et al., 2013; Witteman et al., 2014).

The first step of the working model is the early processing of acoustic information, demonstrated by N1 and mismatch negativity (MMN) responses in the secondary auditory cortex (Schirmer & Kotz, 2006). These are both ERP components that occur in response to changes in frequency or intensity; the N1 is a negativity occurring ~100 ms after stimulus onset, and the MMN has been observed in oddball paradigms 100–200 ms after presentation of the deviant stimulus (N1: Engelien, Schulz, Ross, Arolt & Pantev, 2000; MMN: Tse, Tien & Penney, 2006).

Oddball paradigms with standard and deviant emotional vocalizations also demonstrate an MMN effect, this one occurring at approximately 200 ms (Schirmer, 2004). Schirmer & Kotz (2006) take this as evidence that the listener uses acoustic cues to determine the emotional significance. Based on existing research, the authors infer that the right anterior superior temporal sulcus may be activated for the emotional MMN effect (Grandjean et al., 2005).

In the final stage of the current model, the listener uses higher order thinking to apply emotional meaning. This is the process most directly explored by experiments in which the task is to label emotions, such as those described in the previous section of this chapter. The right inferior frontal gyrus (IFG) is implicated here for evaluative judgments, while the left IFG is active for semantic judgments of words spoken with incongruous (compared to congruous) prosody.

1.5 Cortical rhythms

A relatively new area of analysis in ERP research is the application of time-frequency analysis to examine degree of trial-by-trial coherence in cortical rhythms that may give rise to the salient components in the averaged ERP waveforms (Luck, 2014). There is a growing body of literature on the cortical rhythms that mediate phonetic and prosodic processing in an audiovisual priming paradigm. The different cortical rhythms are considered to reflect resonant neural networks that code and transfer information across brain regions to support various sensory, motor and cognitive processing. Researchers have investigated how frequency bands are modulated in response to different auditory and visual cues (for a review of EEG coherence, see Weiss & Mueller, 2003). Of particular interest in the current investigation are the following oscillations: delta (1–4 Hz), theta (4–8 Hz), beta (12–30 Hz) and gamma (>30 Hz). Senkowski, Schneider, Foxe, and Engel (2008) reviewed the implications of these cortical oscillations on cross-modal sensory integration.

A subset of the studies reviewed by Senkowski et al. (2008) investigated gamma band activity in audiovisual semantic priming paradigms. Increased gamma oscillations were observed in response to incongruent audiovisual information in an object recognition task where the participant determined which animal they saw (visual) or heard (auditory) depending on the experimental block (Yuval-Greenberg & Deouell, 2007). Schneider, Debener, Oostenveld, and Engel (2008) performed a similar study using an audiovisual priming paradigm in which participants categorized only the auditory target. Their results demonstrated increased gamma oscillations in response to auditory targets that were incongruent with visual primes.

Theta band power has been tied to responses of arousal (Basar, Basar-Eroglu, Karakas & Schürmann, 1999). It may also increase in response to semantic violations (Hald, Bastiaansen & Hagoort, 2006) and working memory of verbal stimuli (Klimesch, et al., 2006; Scheeringa, Petersson, Oostenveld, Norris, Hagoort & Bastiaansen, 2009; Summerfield & Mangels, 2005). Furthermore, theta activity has been tied to visual presentation of emotional faces to adults (Balconi & Pozzoli, 2009; Knyazev, Slobodskoj-Plusnin & Bocharov, 2009) as well as affective or formant-exaggerated speech presented to infants (Orekhova, Stroganova, Posikera & Elam, 2006; Radicevic, Vujovic, Jelicic & Sovilj, 2008; Santesso, Schmidt & Trainor, 2007; Zhang et al., 2011).

Delta modulations have also been shown to increase in response to linguistic and speech processing (Giraud & Poeppel, 2012; Radicevic et al., 2008; Scheeringa, et al., 2009). Similarly, delta band power may be more synchronized in response to emotional faces compared to neutral ones (Knyazev et al., 2009). Zhang et al. (2011) also found increased delta synchronization in infants when they were presented with formant-exaggerated vowels.

Investigations of beta power changes in response to emotional stimuli represent a relatively small subset of the cortical rhythm research. However, it is generally observed that beta oscillations occur in response specifically to the visual information of emotional facial stimuli (Okazaki, Kaneko, Yumoto & Arima, 2008, as cited in Balconi & Pozzoli, 2009). Güntekin and Basar (2007) found a notably stronger increase in beta in response to angry compared to happy facial stimuli. Further investigation into this phenomenon revealed increased power for both beta and gamma bands in response to unpleasant pictures presented in a block design, but no effect on either frequency band in a random design (Güntekin & Tülay, 2014).

1.6 Research questions

In this event-related potential study, we explored phonetic and prosodic processing using two visual priming conditions to examine behavioral and neural responses associated with the identification of phonetic mismatch and prosodic mismatch. In line with the priming literature, we hypothesized that participants would take more time to react to incongruent audiovisual stimuli than to congruent audiovisual stimuli, regardless of condition (prosodic vs. phonetic). We anticipated that behavioral accuracy for detecting incongruent audiovisual pairs would be dependent on dimensional information (prosodic vs. phonetic) with potential dimensional interaction effects. We further predicted that participants would exhibit both an N400 response and a late positive response to incongruent audiovisual stimuli in both conditions. In addition to the conventional ERP waveform analysis, we applied a source localization method to test whether different cortical regions were involved in generating the N400 and late positive responses for the two conditions (prosodic vs. phonetic). We were also interested in determining whether different cortical rhythms mediated the generation of the N400 and late positive responses in the two conditions. The results would collectively provide a better understanding of the brain mechanisms underlying the processing of prosodic and phonetic information in spoken language.

Chapter 2: Methods

2.1 Participants

Twelve right-handed adults (6 male, 6 female) between the ages of 18 and 24 participated in the experiment. All participants were native English speakers with no history of speech, language, or hearing impairment. Prior to the start of the experiment, each participant read and signed an informed consent form (see Appendix B). Participants were compensated ten dollars per hour for their participation.

2.2 Stimuli

The stimuli included both visual primes and auditory targets. The visual primes were four photographs of a male face showing a happy or an angry expression with a mouth shape that was representative of either an /ɑ/ or an /i/ vowel (Appendix C). The same male model produced the four auditory targets. These were consonant-vowel-consonant (CVC) sequences /bɑb/ ("bob") and /bib/ ("beeb") produced with happy or angry prosody.

2.3 Procedure

During the EEG recording session, participants were seated in a comfortable chair in a soundproof booth (ETS-Lindgren Acoustic Systems). Participants were fitted with a stretchable 64-channel Waveguard cap, and continuous EEG data were recorded using the Advanced Neuro Technology system. The Ag/AgCl electrodes were arranged to match the standard International 10-20 Montage System and intermediate locations, with the ground electrode located at the AFz electrode. The bandpass filter for EEG recording was for the 0.016-200 Hz range, and the sampling rate was 512 Hz. Impedances for the individual electrodes were kept at or below 5 kΩ.

Visuals were presented in the center of the screen against a green background (Appendix C). Each visual prime was presented for 400 ms before the onset of the target auditory stimulus whose duration was 295 ms (Figure 1). There were 160 trials per block. Auditory stimuli were presented at 60 dB sensation level (Rao, Zhang & Miller, 2010). The presentation order for the phonetic and prosodic blocks was counterbalanced across participants.

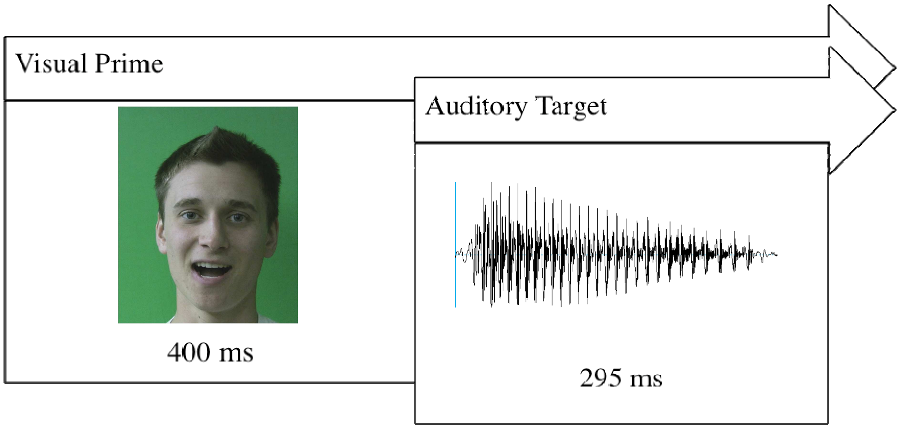

Figure 1. Visual schematic of affective priming paradigm. The visual prime was presented for 400 ms before the onset of the auditory target, whose duration was 295 ms.

In the prosodic condition, participants were instructed to evaluate a match or mismatch between the emotion of the face and the emotion of the voice. In the phonetic condition, participants were instructed to evaluate a match or mismatch between the articulation and the auditory word target (Figure 2). They indicated their responses (match vs. mismatch) by pressing the left or right arrow key on a keyboard. In the phonetic block, participants were instructed to evaluate a match or mismatch between the vowel of the word and the mouth shape. Again, they indicated a match or a mismatch by pressing the left or right arrow key.

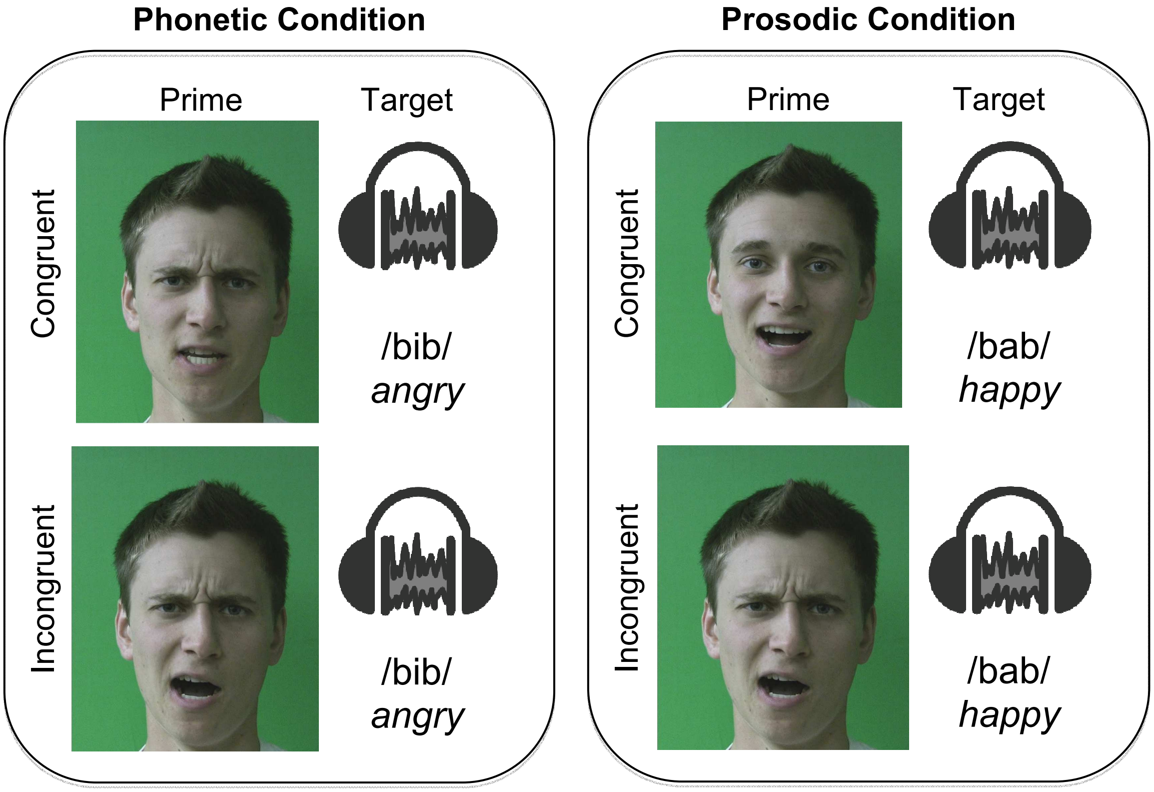

Figure 2. Congruent and incongruent stimuli. These are examples of congruent and incongruent stimuli for phonetic and prosodic conditions.

2.4 Data analysis

2.4.1 Behavioral data analysis

Behavioral responses were analyzed for percent correct accuracy and mean reaction time. Analysis of percent accuracy accounted for all possible response categories (hits, correct rejections, misses, and false alarms). The pairings of a face and a voice were classified as either congruent or incongruent. A "correct" response was agreement ("yes") with a congruent pairing or disagreement ("no") with an incongruent pairing.

Mean reaction times were calculated for each subject for each of the four conditions (phonetic congruent, phonetic incongruent, prosodic congruent, prosodic incongruent). Mixed repeated-measure ANOVA tests evaluated two main factors and their interaction: congruency (congruent vs. incongruent) and condition (phonetic vs. prosodic). Post-hoc tests were also conducted to further investigate interaction effects.

2.4.2 ERP waveform analysis

ERP averaging was performed offline in BESA (Version 6.0, MEGIS Software, GmbH, Germany). Artifact correction parameters were set at 100.0µV for horizontal electrooculogram (HEOG) and 150.0µV for the vertical electrooculogram (VEOG) to minimize the effects of eye drift and blink, respectively. After the artifact correction was applied, the raw EEG data were bandpassed at 0.5 – 40 Hz. The ERP epoch length was 1500 milliseconds, including a pre-stimulus baseline of 100 milliseconds. The automatic artifact scanning tool in BESA was applied to detect noisy signals. The automatic artifact rejection criterion was set at plus or minus 50μV. Additionally, trials where the difference between two adjacent sample points exceeded 75 μV were excluded from analysis.

To improve the signal-to-noise ratio of the data, nine electrode regions were defined for analysis (Figure 1). They were organized from anterior to posterior and left to right. The left anterior (LA) region included channel sites F7, F5, F3, FT7, FC5, and FC3, the middle anterior (MA) region included F1, FZ, F2, FCZ, FC1, and FC, and the right anterior (RA) region included F8, F6, F4, FT8, FC6, FC4. The left central (LC) region included channel sites T7, C5, C3, TP7, CP5, and CP3, the middle central (MC) region included C1, CZ, C2, CP1, CP2, and CPZ, and the right central (RC) region included T8, C6, C4, TP8, CP6, and CP4. The left posterior (LP) included channel sites P7, P5, P3, PO7, PO3, and O1, the middle posterior (MP) region included P1, PZ, P2, POZ, and OZ, and the right posterior (RP) region included P8, P6, P4, PO8, PO4, and O2. Similar channel groupings were used in previous studies (Chen et al., 2011; Schneider et al., 2008; Stevens & Zhang, 2014; Zhang et al., 2011).

Based on visual inspection and evidence from previous literature, two time windows were selected for analysis: an early time window from 250 – 450 ms (N400 component, Aguado et al., 2013; Kamiyama et al., 2013; Kotz & Paulmann, 2011) and a late time window from 700 – 1000 ms (late positive response, Chen et al., 2011; Paulmann, Bleichner & Kotz, 2013; Kotz & Paulmann, 2011). These latencies were measured based on the onset of the auditory stimulus, which occurred at 400 ms. However, the baseline for the epoch was calculated based on the 100 ms preceding the onset of the visual prime to prevent interference of visual processing activities. Repeated measures ANOVA tests were performed for these two peaks of interest. Within-subject factors were condition (phonetic and prosodic), laterality (left, middle, and right), and site (anterior, central, and posterior) (Figure 3).

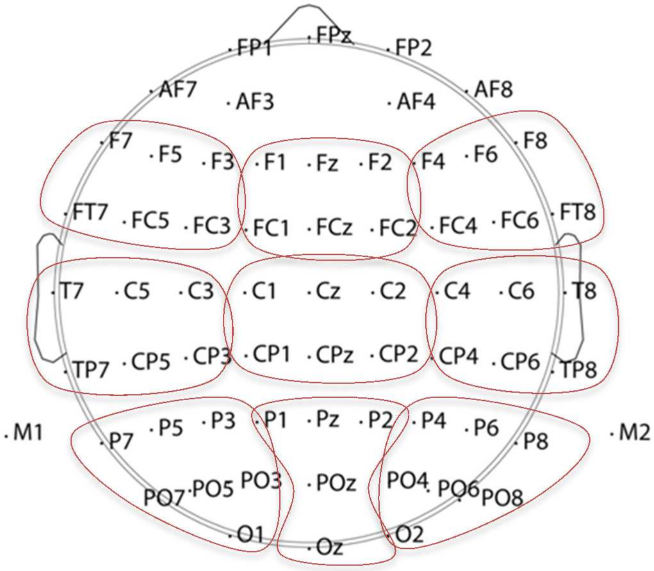

Figure 3. Electrode grouping. Electrode channels were grouped into nine regions for statistical analysis: left anterior (F7, F5, F3, FT7, FC5, FC3), middle anterior (F1, FZ, F2, FCZ, FC1, FC), right anterior (F8, F6, F4, FT8, FC6, FC4), left central (T7, C5, C3, TP7, CP5, CP3), middle central (C1, CZ, C2, CP1, CP2, CPZ), right central (T8, C6, C4, TP8, CP6, CP4), left posterior (P7, P5, P3, PO7, PO3, O1), middle posterior (P1, PZ, P2, POZ, OZ), right posterior (P8, P6, P4, PO8, PO4, O2) (Chen et al., 2011; Schneider et al., 2008; Stevens & Zhang, 2014; Zhang et al., 2011).

2.4.3 Source localization analysis

Source localization analysis was performed using the minimum norm estimation (MNE) in BESA software (Zhang et al., 2011). MNE analysis approximated the current source space, using the smallest norm to explain the measured ERP signals. These were reference-free estimates of cortical current activities. MNE was implemented in the process outlined below:

1. The electrode montage was calculated by using the standard positions for the WaveGuard EEG cap relative to the standard head model (Boundary Element Model) in BESA.
2. Depth weighting and spatio-temporal weighting were adopted to avoid bias towards superficial sources and improve the focality and reliability of the source activities.
3. The total activity at each source location (750 dipole locations in the left hemisphere and 750 in the right hemisphere) was calculated as the root mean square of the dipole source activities. These solutions were then projected to the standard realistic brain model in BESA. The current source data for the prefixed locations at all latencies were further analyzed for temporal and spatial interpretations.
4. The total MNE activities in each hemisphere were added at each time point. A two-tailed z-test relative to baseline mean and variance was applied to the MNE differences between the two stimuli at each sample point.
5. Regional contributions to the total MNE activities were examined using standard anatomical boundaries in the Talairach space for each region of interest (ROI) in the brain space.

2.4.3 Time frequency analysis

Time frequency analysis was also performed for the nine regions of interest (ROIs): left anterior (LA), middle anterior (MA), right anterior (RA), left central (LC), middle central (MC), right central (RC), left posterior (LP), middle posterior (MP), and right posterior RP). Inter-trial coherence in terms of phase locking values in delta (1–4 Hz), theta (4–8 Hz), beta (13–30 Hz) and gamma (>30 Hz) frequency bands was computed for each subject in each of the two conditions (phonetic vs. prosodic) with the open source EEGLAB package (Delorme & Makeig, 2004). The inter-trial coherence measure is an estimate of mean normalized phase across trials, which can range from 0 (indicating random phase coherence or complete lack of synchronization) to 1 (indicating perfect phase synchrony across trials). The inter-trial coherence data (also referred to as phase locking values) were averaged across the frequencies within the range of each frequency. The peak phase locking values corresponding to the N400 and late positive response components in their respective windows were identified for each frequency band for each listening condition on an individual basis. For each experimental condition (prosodic vs. phonetic), a direct comparison of the time-frequency analysis data with a false discovery rate (FDR) (Benjamini & Hochberg, 1995) corrected p-value threshold of 0.01 was conducted to determine what frequency bands mediated the N400 and late positive responses.

2.4.5 Statistical analysis

All statistical analyses were completed in Systat 10. A repeated measures analysis of variance (ANOVA), with $\alpha = 0.05$, was conducted to examine the statistical significance of listening condition (phonetic vs. prosodic) on N400 and late positive response latencies and amplitudes recorded at the selected regions of interest. The repeated-measures ANOVA was also applied in evaluating behavioral and neural responses in the phonetic and prosodic conditions. Post-hoc t-tests were also conducted to further determine how the different factors contributed to the significant interaction effects in the ANOVA tests.

Chapter 3: Results

3.1 Behavioral results

The percent correct showed a congruency effect (*F*(1,11) = 5.79, $p < 0.05$**)** and there was a significant interaction between congruency and condition (*F*(1, 11) = 8.16, $p < 0.05$**)** (Figure 4). Post-hoc t-tests revealed that the prosodic condition (but not the phonetic condition) showed a significant accuracy difference between the congruent and incongruent trials ($p < 0.05$).

The reaction times showed a congruency effect (*F*(1,11) = 28.67, $p < 0.01$). Participants took more time to respond to incongruent stimuli than congruent stimuli, regardless of condition. Post-hoc t-tests revealed that both the prosodic condition and the phonetic condition showed a congruency effect in reaction time ($p < 0.05$).

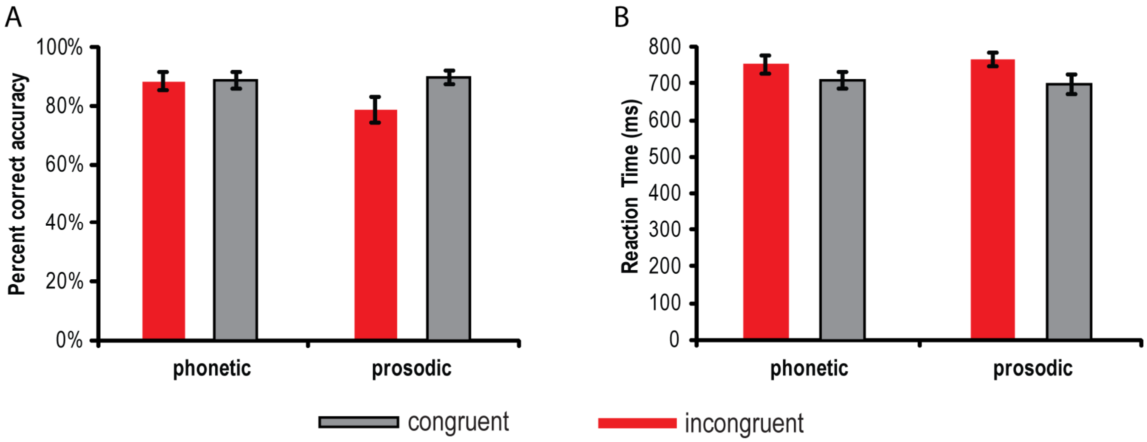

Figure 4. Percent correct accuracy (A) and reaction times (B). Presented for phonetic and prosodic conditions.

3.2 ERP results

3.2.1 N400

A repeated measure ANOVA in the early time window (250 – 450 ms after onset of the auditory target, 650 – 850 ms in figures) revealed a significant main effect for listening condition ($F(1,11) = 7.95, p < 0.05$) (Figures 5-7). Congruent trials were subtracted from incongruent trials to create the subtracted waveform seen in Figure 7A. The subtracted waveforms show clear N400 peaks across all regions of interest. These peaks are followed by a late positivity. The phonetic condition elicited stronger N400 activity for the incongruent stimuli than the prosodic condition. There was also a significant main effect for laterality (left, middle, right) ($F(1,11) = 7.17, p < 0.01$). Furthermore, there was a condition × laterality interaction ($F(1,11) = 4.33, p < 0.05$), indicating that the stronger N400 activity for incongruent stimuli in the phonetic condition was dependent on lateral location (left, middle, right). There was also a condition × site × laterality effect ($F(1,11) = 4.04, p < 0.01$) for the N400 component.

Post-hoc t-tests revealed that the N400 in the phonetic condition was left dominant in the central and posterior sites ($p < 0.05$). No significant laterality effect was found for the N400 in the prosodic condition. The topographic distribution (Figure 7B) for the N400 component confirms left hemisphere dominance for the phonetic condition, whereas the distribution appears to be more bilateral for the prosodic condition.

3.2.2 Late positive response

A repeated measure ANOVA in the late time window (700 – 1000 ms after onset of the auditory target, 1100 – 1400 ms in figures) revealed significant effects for electrode site (anterior, central, posterior) ($F(1,11) = 7.49$, $p < 0.01$) and laterality (left, middle, right) ($F(1,11) = 19.03$, $p < 0.01$). In contrast to the N400 component results, there was a significant interaction effect for condition × site ($F(1,11) = 3.89$, $p < 0.05$). Finally, there was a condition × site × laterality effect ($F(1,11) = 3.10$, $p < 0.05$).

Post-hoc t-tests showed anterior vs. posterior significant differences in both hemispheres for both the prosodic and phonetic conditions ($p < 0.05$). But there was no significant laterality effect in any of the anterior, central, or posterior sites in either condition. The topographic distribution (Figure 7) for the late positive response appears relatively similar between the two conditions, with the exception of a greater frontal negativity in the phonetic condition. However, there is not a clear hemispheric pattern like that seen for the N400 component.

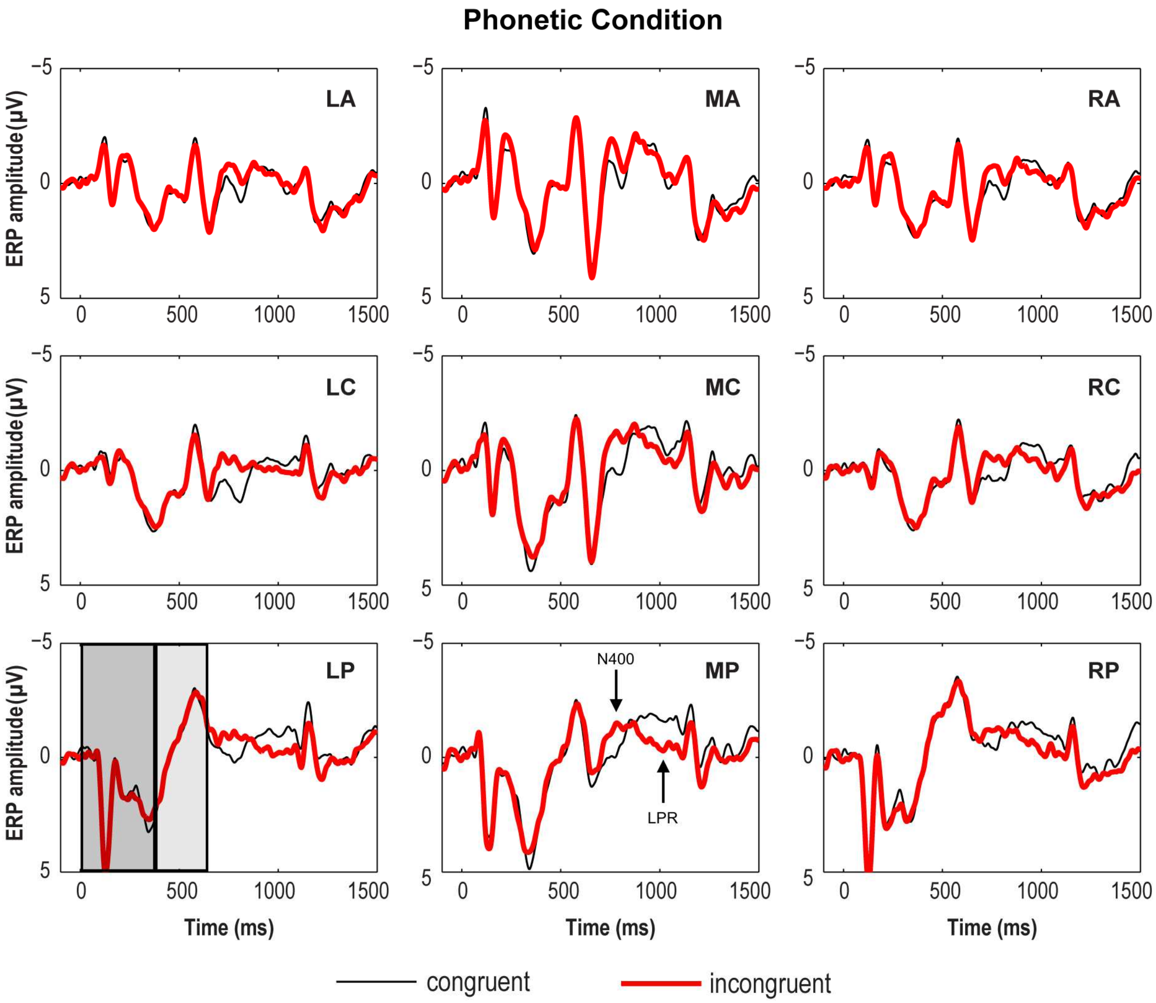

Figure 5. ERP average waveforms for phonetic condition. ERP average waveforms are presented for nine regions comparing responses to congruent and incongruent stimuli in the phonetic condition. Regions are organized by laterality and site: left anterior (LA), left central (LC), left posterior (LP), middle anterior (MA), middle central (MC), middle posterior (MP), right anterior (RA), right central (RC), and right posterior (RP). In the waveform plot for LP, the dark gray bar shows the length of the visual prime and the light gray bar shows the presentation of the auditory target. The N400 and late positive response (LPR) latencies were measured from the onset of the auditory target at 400 ms. These are denoted with arrows on the MP waveform.

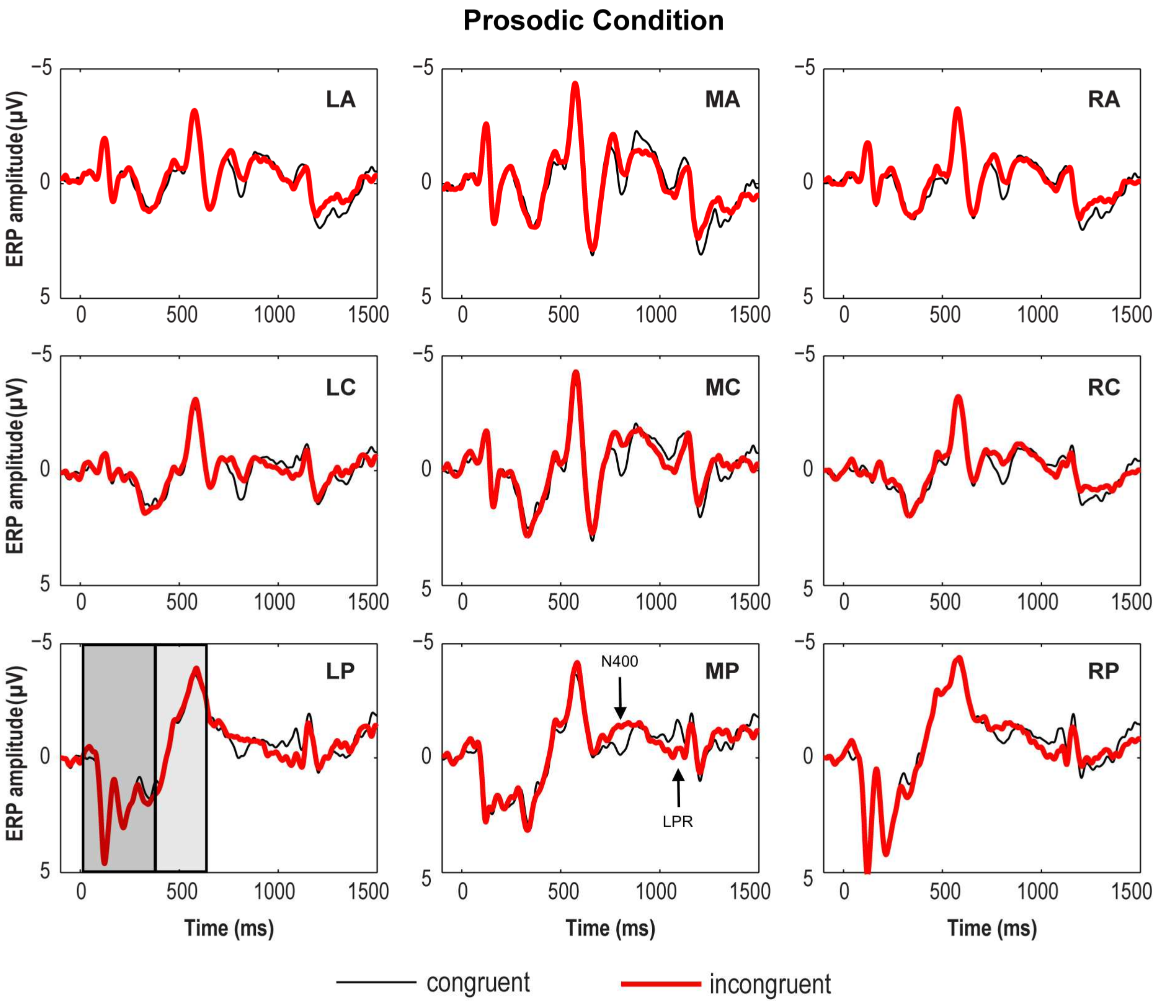

Figure 6. ERP average waveforms for prosodic condition. ERP average waveforms are presented for nine regions comparing responses to congruent and incongruent stimuli in the prosodic condition. Regions are organized by laterality and site: left anterior (LA), left central (LC), left posterior (LP), middle anterior (MA), middle central (MC), middle posterior (MP), right anterior (RA), right central (RC), and right posterior (RP). In the waveform plot for LP, the dark gray bar shows the length of the visual prime and the light gray bar shows the presentation of the auditory target. The N400 and late positive response (LPR) were measured from the onset of the auditory target at 400 ms. These are denoted with arrows on the MP waveform.

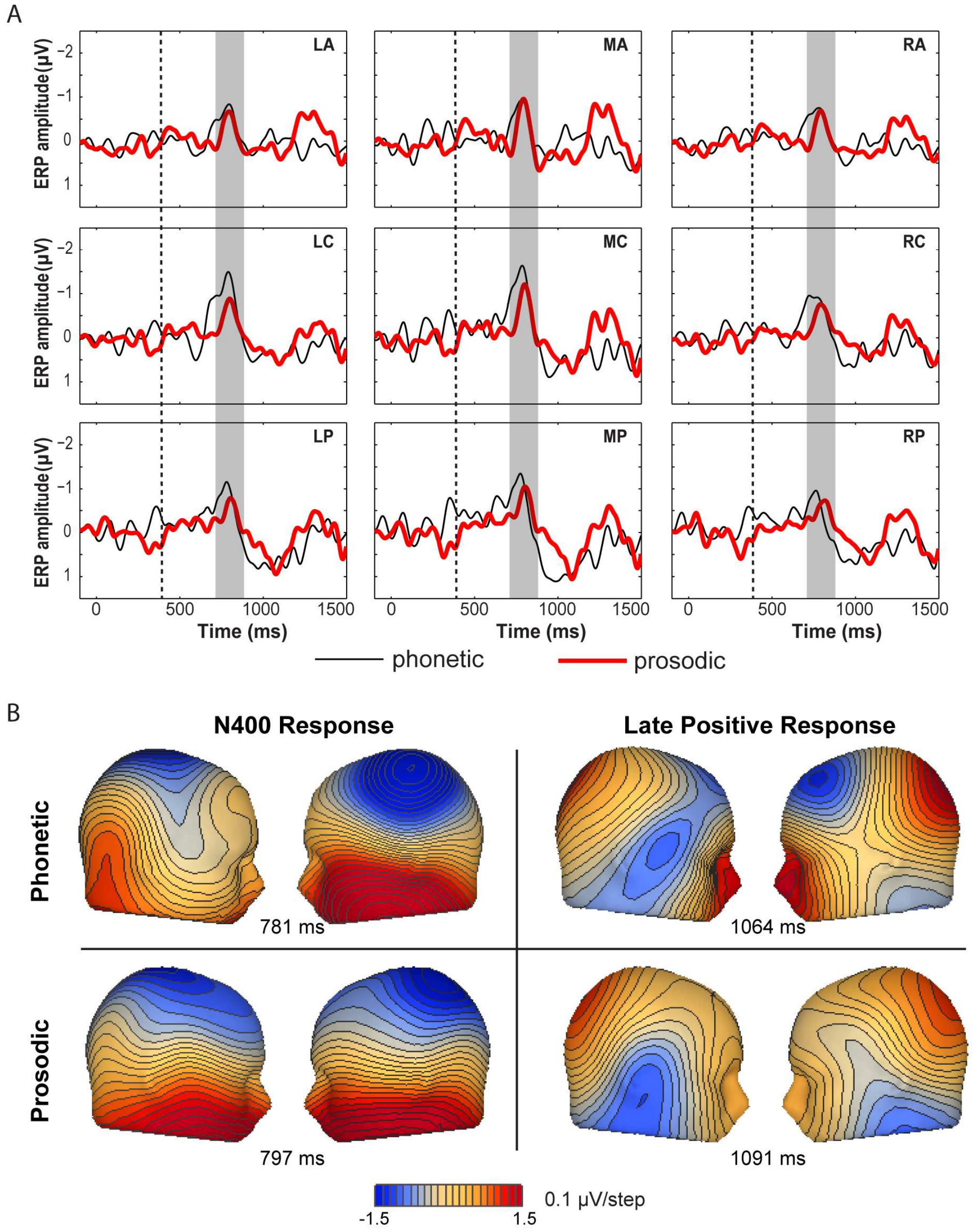

Figure 7. ERP difference waveforms and scalp topography. This figure shows ERP difference waveforms (incongruent - congruent). In the average waveforms (A), the gray bar highlights the N400 component. Scalp topography of the N400 and late positive responses for the two listening conditions is shown in (B).

3.2.3 Localization results

Source localization analyses provided a rough estimation of cortical activation patterns for the N400 and late positive responses in the current experiment (Figure 8). Total activity waveforms for the phonetic and prosodic conditions are not highly revealing because of the differential patterns of cortical activation. Regions where the minimum norm estimation analysis yielded a z score of 4 or greater ($p < 0.001$) are described below for each condition.

Source localization patterns for the N400 response show strong left hemisphere lateralization in the phonetic condition, with contributions from the superior temporal and inferior parietal regions as well as the primary motor cortex. In the prosodic condition, the N400 response shows a pattern of right hemisphere dominance with superior temporal and inferior parietal region activations.

The late positive response appears to have more distributed regions of activation for the phonetic condition. These regions generally include the parietal lobe in addition to the primary motor cortex, with possible contributions from the occipital region. Left hemisphere activations in the prosodic condition include parietal and occipital regions, while right hemisphere activity includes occipital, temporal, and inferior parietal regions.

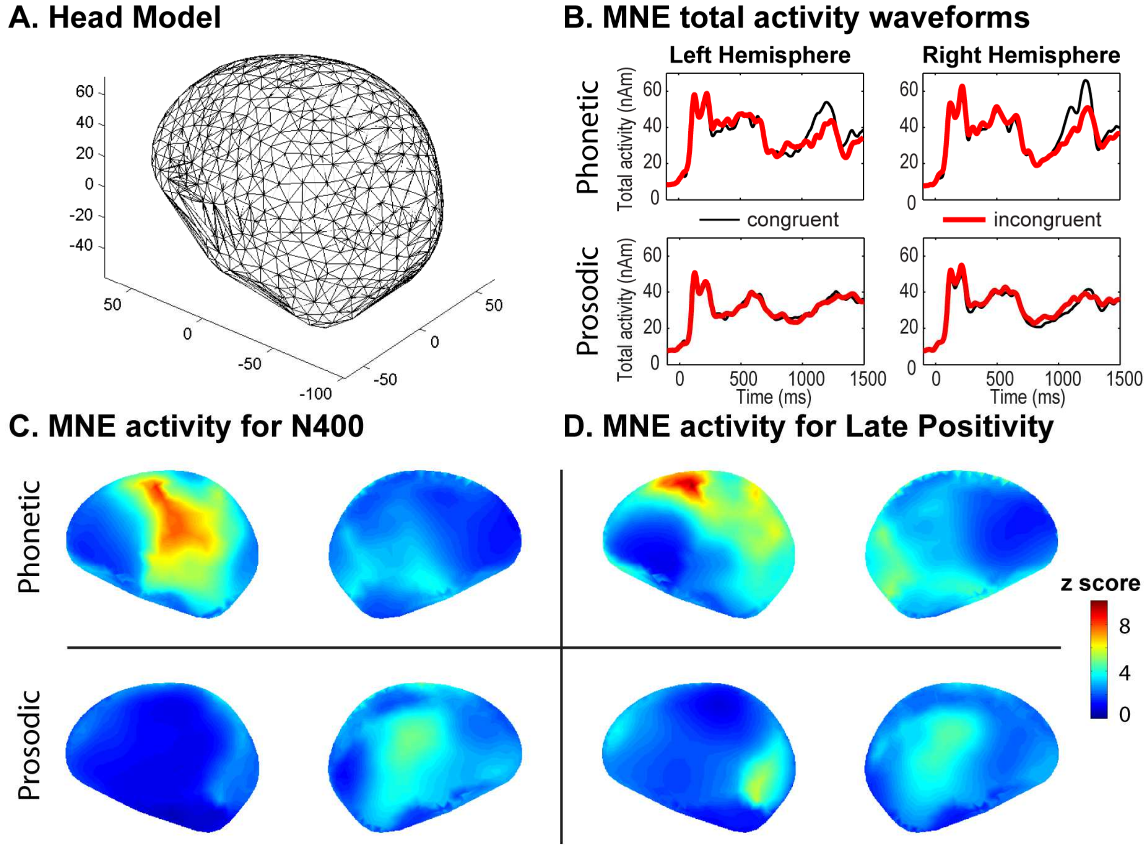

Figure 8. Minimum norm estimation activity. (A) Boundary Element Method (BEM) head model (B) total activity waveforms for phonetic and prosodic conditions in the left and right hemispheres (C) Minimum norm estimation (MNE) activity for the N400 response (D) MNE activity for the late positive response.

3.2.4 Cortical rhythm results

Time-frequency analysis evaluated the contributions of delta (1–4 Hz), theta (4–8 Hz), beta (12–30 Hz) and gamma (>30 Hz) oscillations to ERP responses (Figure 9). In the phonetic condition, the lower frequency bands (delta, theta and beta rhythms) contributed to the N400 response and theta rhythm contributed to late positive response ($p < 0.01$). In the prosodic condition, the primary contributors to both N400 and late positive responses were beta and gamma rhythms. Also, theta band oscillations showed a significant difference between the congruent and incongruent trials ($p < 0.01$) at the late positivity window in the prosodic condition. Furthermore, there was significant early gamma activity before the onset of the auditory target for the prosodic condition ($p < 0.01$). A similar rest-state gamma rhythm, which was argued to reflect predictive coding of the following auditory target based on the visual information preceding the sound, was previously reported in an experiment investigating the McGurk effect (Keil, Muller, Ihssen & Weisz, 2012).

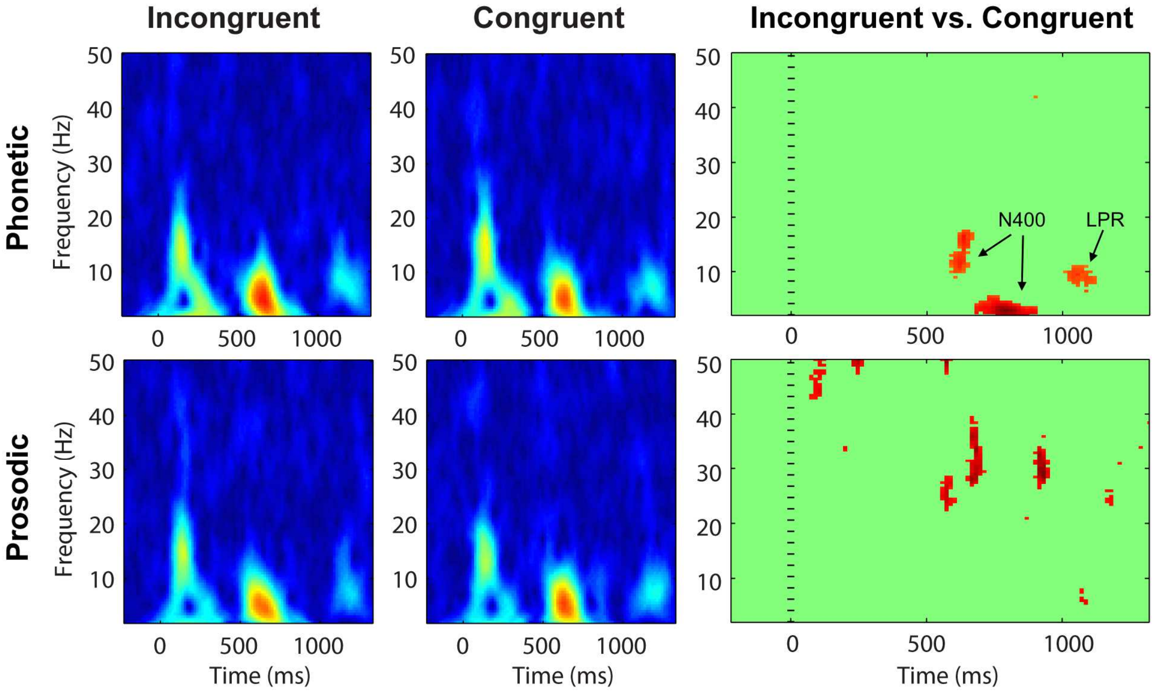

Figure 9. Cortical time-frequency analysis. Compares cortical oscillations for incongruent and congruent trials (and their difference) in phonetic and prosodic conditions.

Chapter 4: Discussion

4.1 Congruency effect in behavioral data

Overall, participants were more accurate at identifying congruent face and voice pairs than incongruent combinations. Furthermore, percent correct accuracy was dependent upon dimensional information. Participants were less accurate at identifying prosodic incongruency (78.56%) than prosodic congruency (89.54%) while percent correct accuracy for the phonetic condition was nearly identical for the incongruent and congruent conditions (88.13% and 88.45% respectively). In a recent experiment by Chen et al. (2011), participants were more accurate at identifying a prosodic match compared to a prosodic mismatch when attending to sound intensity deviation in a sentence. However, the same authors found no effect for accuracy in a congruency detection task using the same stimuli. Kamiyama et al. (2013) observed high accuracy (above 90%) for both congruent and incongruent face/music pairs. As different studies tested different informational dimensions, it is difficult to reach a simple consensus across the studies. Our data suggest that the phonetic judgment accuracy was not much affected whether the visual prime matched the auditory target or not, whereas prosodic judgment accuracy was significantly affected.

Results for reaction time demonstrated that participants took longer to respond to incongruent stimuli than congruent stimuli, irrespective of whether the condition was phonetic or prosodic. A similar effect has been observed in previous studies (Kamiyama et al., 2013; Stevens & Zhang, 2014). In the experiment by Kamiyama et al., participants with and without musical experience judged congruent face-music pairs more quickly than incongruent pairs. In a cross-language comparison, Stevens & Zhang identified an audiovisual congruency effect regardless of language background or inclusion of an independent gesture variable.

4.2 N400 audiovisual congruency effect

Consistent with our hypotheses, we observed both an N400 response and a late positive response to incongruent audiovisual stimuli. Our results were in line with the results of previous studies evaluating responses to various congruent and incongruent stimuli (Kamiyama et al., 2013; Schirmer & Kotz, 2013; Stevens & Zhang, 2014). In the current experiment, the phonetic condition elicited a larger N400 component than the prosodic condition. Furthermore, the N400 for the linguistic processing condition showed left hemisphere dominance, whereas the N400 for the prosodic processing condition did not (see further discussion in the localization results section).

Existing literature evaluating emotional processing spans a wide range of experimental approaches and identifies multiple ERP components. Among the studies that identified an N400 response, there was a mixture of priming and reverse priming effects (Aguado et al., 2013; Kamiyama et al., 2013; Paulmann & Pell, 2010). In a facial affective decision task presented in a priming paradigm, Paulmann and Pell observed both of these effects. They observed a normal effect for the medium-length prime of 400 ms (the length of the visual prime in the current experiment), whereas there was a greater response for congruent compared to incongruent stimuli when participants were presented with a short prime (200 ms). As a possible explanation for the reverse priming effect they observed in their experiment, Aguado et al. cites their use of stimuli with complex affective valence. The design differed from that of the current experiment in that the facial presentation was followed by the visual presentation of a word with emotional content rather than an auditory presentation of a non-word (absent of emotional content) spoken with different emotional prosody.

4.3 Late positive response congruency effect across conditions

Throughout the literature, researchers have posited a variety of explanations for the late positivity response. While there is not a clear consensus regarding its functional significance, the late positive response is generally characterized as reflecting increased attention to unexpected targets. This response has been identified in response to incongruency of affective face or word stimuli (Werheid et al., 2005; Zhang et al., 2010). Our findings in the current experiment support this explanation; we observed a late positive response in response to incongruent stimuli, regardless of condition. In contrast to these results, the late positive potential (LPP) identified by Aguado et al. (2013) was modulated by affective valence (positive vs. negative) of the target word, but not by congruency of the visual stimuli. The authors propose that the LPP reflected an evaluative priming effect.

Because the late positive response was observed following the N400 component, it is appropriate to entertain the possibility that the late response is in fact a variant of the P300 component, reflecting working memory updating (see Hajcak et al., 2009; Zhang et al., 2010). Conscious processing of visual stimuli might also contribute to the late response, as suggested by Zhang et al. to account for the lack of a late positivity in response to other priming paradigms (word-word: Zhang et al., 2006; subliminal affective priming: Li, Zinbarg, Boehm & Paller, 2008). The late positive response has been observed in experiments where the participant consciously attended to congruency of emotional or prosodic stimuli (Aguado et al., 2013; Chen et al., 2011; Kamiyama et al., 2013).

Witteman et al. (2014) observed an effect of the late positive potential (LPP) in an emotional task, but not the linguistic one. This effect was larger at posterior sites in the left hemisphere, but proximal sites in the right hemisphere. Our results show that the late positivity did not vary by condition in isolation, but interaction effects show an effect of condition dependent on site as well as an effect dependent on site and laterality. Effects for site and laterality were also observed independently of each other, indicating clear localization of the late positive response.

4.4 Cortical regions for the N400 and LPP responses

In the current experiment, participants were asked to attend to either the phonetic or the prosodic dimension of the same stimuli. Overall, our results indicate that different brain regions are recruited for each of these conditions for both the N400 and late positive response. This may reflect the underlying mechanisms of how the individuals selectively tune to one dimension of information (Rao, Zhang & Miller, 2010). For the N400 component, we observed left hemisphere lateralization in the phonetic condition but right hemisphere lateralization for the prosodic condition. In the same conditions, cortical activations for the late positive response were more broadly distributed.

It appears that there was some primary motor cortex involvement for both the N400 and late positive response. However, this pattern of activation was only seen in the phonetic condition. We speculate that the motor cortex involvement might partly be due to the participants indicating their response by pressing a key on a computer keyboard, which was more consistent in timing across trials for the phonetic condition than for the prosodic condition. It is important to note that source localization analysis represents a weak area in event-related potential research (Luck, 2014). But despite its imprecision, source localization analysis can help to determine which neural networks are implicated in a given task, particularly when compared to the oscillation rhythm patterns across the cortical network activations (see discussion in the following section).

4.5 Cortical rhythms mediating audiovisual congruency effects

The MNE analysis and time-frequency analysis results revealed cortical network activation patterns that varied across conditions. Delta, theta, and beta oscillations contributed to the N400 response in the phonetic condition. The late positive response in the same condition was modulated only by theta rhythms. In contrast, beta and gamma activity contributed to both the N400 and late positive response in the prosodic condition. There was also a significant difference in theta activity between incongruent and congruent conditions for the late positive response.

Theta oscillations have been observed in response to the visual presentation of emotional faces (Balconi & Pozzoli, 2009; Knyazev et al., 2009). Although both conditions involved the visual presentation of happy or angry faces, theta activity was only observed for both ERP responses in the phonetic condition and not for the prosodic condition. There was also increased theta activity for the difference between congruent and incongruent face/voice pairs in the late positive response.

Increased gamma activity is often seen in response to incongruent audiovisual information (for a review, see Senkowski et al., 2008). While cortical rhythms in the gamma frequency range were observed for incongruency in the prosodic condition, no such activity was present for incongruency in the phonetic condition in our study. Güntekin and Basar (2007) identified an increase in beta activity in response to angry compared to happy facial stimuli, but no such effect between emotions was observed in the current investigation.

These distinct patterns support our interpretation that different cortical regions are involved in mediating selective attention to phonetic and prosodic processes. How exactly these processes are related, however, remains unclear. Specific cortical sites may contribute to each ERP response, with different cortical rhythm patterns arising from these neural networks. Conversely, distinct cortical rhythms may be at play, leading to the recruitment of localized neural networks.

In addition to these cortical rhythm patterns for the N400 and late positive response, time frequency analysis yielded significant gamma activity preceding the onset of the auditory target in the prosodic condition. A similar predictive coding response has been observed during the resting state in experiments investigating the McGurk effect (Keil et al., 2012). This finding highlights the utility of time-frequency analysis in ERP research. In this case, time-frequency analysis revealed an apparent difference that was not visible in the waveform.

4.6 Limitations and future directions

One concern with our experimental design is the amount of lag time between the visual and auditory presentations. When Paulmann and Pell (2010) presented visual primes of different lengths, the resulting ERP patterns differed. While our results were consistent with the results of Paulmann and Pell's medium-length prime, we did not investigate the possible differences resulting from visual primes of different lengths.

Another factor to consider is the relationship between the selected vowel sounds and emotions. The /i/ vowel is produced with spread lips and lends itself more easily to a happy facial expression, while the /a/ vowel is taller and corresponds more closely to a large, angry facial expression. To combine the vowel sound and emotions that are inherently in conflict with each other (i.e. angry + /i/ and happy + /a/) required that the expression of happiness or anger be clearly displayed without sacrificing the appropriate mouth shape for the vowel. The model in the photographs compensated for this limitation by manipulating his brow to express happiness or anger.

Our experiment was designed in a way that required behavioral responses from the participants. It is worth considering whether a format that does not rely on an overt response might be effective and more suitable for testing individuals whose behavioral responses are limited because of impaired language or cognition. For example, the mismatch negativity (MMN) paradigm is a passive listening task, which would be more user-friendly to a wider range of participants.

MMN experiments investigating the pre-attentive responses to emotional prosody have identified differences in the ERP responses of men and women (Schirmer & Kotz, 2006; Fan, Tsu & Cheng, 2013). These and other experiments suggest that emotional prosody processing is dependent on gender (Schirmer et al., 2002; Schirmer et al., 2005). These differences have also been documented in gender comparisons of responses to facial expressions (Guntenkin & Basar, 2007). Future research should evaluate whether the ERP waveform, localization, and cortical rhythm findings presented here differ between men and women.

Conventional ERP research provides excellent time resolution but comparatively poor localization information. Minimum norm estimation can identify which broad regions at the cortical level are implicated in certain processes, but its resolution cannot rival that of more precise imaging techniques like fMRI. This limitation of the imaging method impacts our ability to make specific claims about the precise brain regions involved in the processing of incongruent audiovisual information in the phonetic and prosodic conditions. However, the localization results taken in tandem with the topographical potential distribution data and cortical rhythm activity can help elucidate the neural dynamics underlying phonetic and prosodic processing.

Time frequency analysis is still a relatively novel area of ERP research, but it is one that warrants further investigation. The dynamic information it provides beyond the level of the waveform is analogous to spectral analysis of speech. For example, consider the speech segments “ra” and “la” presented at the same intensity. At the waveform level, the phonemic contrast could be indistinguishable, but the spectrograms would easily reveal their characteristic formant transitions. Along the same line, time frequency analysis for the ERP data provides highly valuable information about the neural dynamics underlying the processing of the multidimensional speech signal (Zhang, 2008).

In typical development, children appear to implicitly and effortlessly tease apart the many informational dimensions that make up speech. These dimensions include linguistic cues, like phonetic and semantic information, as well as non-linguistic cues, such as emotion and speaker identity. However, research suggests that some clinical populations may exhibit difficulty with sorting these multiple sources of information (Gebauer, 2014; Marshall, Harcourt-Brown, Ramus & van der Lely, 2009; Paul, Auguestyn, Klin & Volkmar, 2005; Wang & Tsao, 2015), which would lead to developmental delays or disorders in language learning. The current investigation provides a baseline from normal adults for comparison to clinical populations, such as individuals with autism, language impairment, or aphasia. Future research should investigate how these individuals integrate audiovisual speech information and selectively attend to a single dimension of information while ignoring the others. If we can identify the neural markers that are tied to the clinical manifestation of language impairment, electrophysiological measures like those utilized in the current investigation may provide more objective diagnostic criteria for future clinical practice. However, further research efforts are necessary to establish a robust system that allows identification of reliable neural markers at the individual level.

Works Cited

Aguado, L., Dieguez-Risco, T., Méndez-Bértolo, C., Pozo, M. A., & Hinojosa, J. A. (2013). Priming effects on the N400 in the affective priming paradigm with facial expressions of emotion. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(2), 284-296.

Balconi, M., & Pozzoli, U. (2009). Arousal effect on emotional face comprehension: Frequency band changes in different time intervals. *Physiology & Behavior*, *97*(3), 455-462.

Başar, E., Başar-Eroğlu, C., Karakaş, S., & Schürmann, M. (1999). Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG?. *Neuroscience Letters*, *259*(3), 165-168.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 57, 289–300.

Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, *41*(2), 259-268.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jäncke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cognitive Brain Research*, *9*(3), 227-238.

Chen, X., Zhao, L., Jiang, A., & Yang, Y. (2011). Event-related potential correlates of the expectancy violation effect during emotional prosody processing. *Biological Psychology*, *86*(3), 158-167.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407.

David Crystal. (1969). *Prosodic systems and intonation in English* (Vol. 1). CUP Archive.

Czerwon, B., Hohlfeld, A., Wiese, H., & Werheid, K. (2013). Syntactic structural parallelisms influence processing of positive stimuli: Evidence from cross-modal ERP priming. *International Journal of Psychophysiology*, *87*(1), 28-34.

David, O., Kilner, J.M., & Friston, K.J. (2006). Mechanisms of evoked and induced responses in MEG/EEG. *Neuroimage, 31,* 1580-1591.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, *134*, 9–21.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating?. *Behavioral and Brain Sciences*, *11*(03), 357-374.

Engelien, A., Schulz, M., Ross, B., Arolt, V., & Pantev, C. (2000). A combined functional in vivo measure for primary and secondary auditory cortices. *Hearing Research*, *148*(1), 153-160.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, *15*(2), 115-141.

Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C., & Kardes, F.R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229–238.

Fox, A. (2000). *Prosodic features and prosodic structure: the phonology of suprasegmentals*. Oxford University Press.

Gebauer, L., Skewes, J., Hørlyck,L., & Vuust, P. (2014). Atypical perception of affective prosody in Autism Spectrum Disorder. *NeuroImage: Clinical, 6*, 370-378.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511-517.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8*(2), 145-146.

Greene, A. J., Easton, R. D., & LaShell, L. S. (2001). Visual–auditory events: cross-modal perceptual priming and recognition memory. *Consciousness and Cognition*, *10*(3), 425-435.

Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., & Martens, S. (2012). The nature of affective priming in music and speech. *Journal of Cognitive Neuroscience*, *24*(8), 1725-1741.

Güntekin, B., & Basar, E. (2007). Emotional face expressions are differentiated with brain oscillations. *International Journal of Psychophysiology*, *64*(1), 91-100.

Güntekin, B., & Tülay, E. (2014). Event related beta and gamma oscillatory responses during perception of affective pictures. *Brain Research*, *1577*, 45-56.

Hajcak, G., Dunning, J. P., & Foti, D. (2009). Motivated and controlled attention to emotion: time-course of the late positive potential. *Clinical Neurophysiology*, *120*(3), 505-510.

Hald, L. A., Bastiaansen, M. C., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, *96*(1), 90-105.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition & Emotion*, *15*(2), 143-165.

Iredale, J. M., Rushby, J. A., McDonald, S., Dimoska-Di Marco, A., & Swift, J. (2013). Emotion in voice matters: Neural correlates of emotional prosody perception. *International Journal of Psychophysiology*, *89*(3), 483-490.

Kamiyama, K. S., Abla, D., Iwanaga, K., & Okanoya, K. (2013). Interaction between musical emotion and facial expression as measured by event-related potentials. *Neuropsychologia*, *51*(3), 500-505.

Keil, J., Muller, N, Ihssen, N, & Weisz, N. (2012). On the variability of the McGurk Effect: Audiovisual integration depends on prestimulus brain states. *Cerebral Cortex, 22*(1), 221-231.

Kelly, S. D., Creigh, P., & Bartolotti, J. (2010). Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *Journal of Cognitive Neuroscience*, *22*(4), 683-694.

Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and language*, *101*(3), 222-233.

Klimesch, W., Hanslmayr, S., Sauseng, P., Gruber, W., Brozinsky, C. J., Kroll, N. E. A., ... & Doppelmayr, M. (2006). Oscillatory EEG correlates of episodic trace decay. *Cerebral Cortex*, *16*(2), 280-290.

Knyazev, G. G., Slobodskoj-Plusnin, J. Y., & Bocharov, A. V. (2009). Event-related delta and theta synchronization during explicit and implicit emotion processing. *Neuroscience*, *164*(4), 1588-1600.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205.

Laver, J., & Hutcheson, S. (Eds.). (1972). *Communication in face to face interaction: Selected readings*. Penguin books.

Lense, M. D., Gordon, R. L., Key, A. P., & Dykens, E. M. (2014). Neural correlates of cross-modal affective priming by music in Williams syndrome. *Social Cognitive and Affective Neuroscience*, *9*(4), 529-537.

Li, W., Zinbarg, R. E., Boehm, S. G., & Paller, K. A. (2008). Neural and behavioral evidence for affective priming from unconsciously perceived emotional facial expressions and the influence of trait anxiety. *Journal of Cognitive Neuroscience*, *20*(1), 95-107.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique* (2nd ed.). Cambridge, Massachusetts: The MIT Press.

Marshall., C.R., Harcourt-Brown, S., Ramus, F., & van der Lely, H. K. J. (2009). The link between prosody and language skills in children with specific language impairment (SLI) and/or dyslexia. *International Journal of Language and Communicative Disorders, 44* (4), 466-488.

Nie, Y., Zhang, Y., & Nelson, P. (2014). Auditory stream segregation using bandpass noises: evidence from event-related potentials. Frontiers in Neuroscience, 8, 277.

Orekhova, E. V., Stroganova, T. A., Posikera, I. N., & Elam, M. (2006). EEG theta rhythm in infants and preschool children. *Clinical Neurophysiology*,*117*(5), 1047-1062.

Özyürek, A., Willems, R., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605-616.

Paul, R., Auguestyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35*(2), 205-220.

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*.

Paulmann, S., & Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: how much is enough?. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 230-242.

Radicevic, Z., Vujovic, M., Jelicic, L., & Sovilj, M. (2008). Comparative findings of voice and speech: language processing at an early ontogenetic age in quantitative EEG mapping. *Experimental Brain Research*, *184*(4), 529-532.

Rao, A., Zhang, Y., & Miller, S. (2010). Selective listening of concurrent auditory stimuli: an event-related potential study. *Hearing Research*, *268*(1), 123-132.

Roach, P., Stibbard, R., Osborne, J., Arnfield, S., & Setter, J. (1998). Transcription of prosodic and paralinguistic features of emotional speech. *Journal of the International Phonetic Association*, *28*(1-2), 83-94.

Santesso, D. L., Schmidt, L. A., & Trainor, L. J. (2007). Frontal brain electrical activity (EEG) and heart rate in response to affective infant-directed (ID) speech in 9-month-old infants. *Brain and Cognition*, *65*(1), 14-21.

Scheeringa, R., Petersson, K. M., Oostenveld, R., Norris, D. G., Hagoort, P., & Bastiaansen, M. C. (2009). Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. *Neuroimage*, *44*(3), 1224-1238.

Schirmer, A. (2004). Timing speech: a review of lesion and neuroimaging findings. *Cognitive Brain Research*, *21*(2), 269-287.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*(2), 228-233.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research*, *24*(3), 442-452.

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, *15*(8), 1135-1148.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24-30.

Schneider, T. R., Debener, S., Oostenveld, R., & Engel, A. K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage*, *42*(3), 1244-1254.

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., & Lang, P. J. (2000). Affective picture processing: the late positive potential is modulated by motivational relevance. *Psychophysiology*, *37*(2), 257-261.

Segal, S. J. (1966). Priming compared to recall: Following multiple exposures and delay. *Psychological Reports*, *18*(2), 615-620.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends in Neurosciences*, *31*(8), 401-409.

Stevens, J., & Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: An event-related potential study. *Journal of Neurolinguistics*, *26*(1), 31-45.

Stevens, J., & Zhang, Y. (2014). Brain mechanisms for processing co-speech gesture: A cross-language study of spatial demonstratives. *Journal of Neurolinguistics, 30*, 27-47.

Summerfield, C., & Mangels, J. A. (2005). Coherent theta-band EEG activity predicts item-context binding during encoding. *Neuroimage*, *24*(3), 692-703.

Tse, C. Y., Tien, K. R., & Penney, T. B. (2006). Event-related optical imaging reveals the temporal dynamics of right temporal and frontal cortex activation in pre-attentive change detection. *Neuroimage*, *29*(1), 314-320.

Wang, J.-E., & Tsao, F.-M. (2015). Emotional prosody perception and its association with pragmatic language in school-aged children with high-function autism. *Research in Developmental Disabilities, 37*, 162-170.

Weiss, S., & Mueller, H. M. (2003). The contribution of EEG coherence to the investigation of language. *Brain and Language*, *85*(2), 325-343.

Werheid, K., Alpay, G., Jentzsch, I., & Sommer, W. (2005). Priming emotional facial expressions as evidenced by event-related brain potentials. *International Journal of Psychophysiology*, *55*(2), 209-219.

Witteman, J., Goerlich-Dobre, K. S., Martens, S., Aleman, A., Van Heuven, V. J., & Schiller, N. O. (2014). The nature of hemispheric specialization for prosody perception. *Cognitive, Affective, & Behavioral Neuroscience*, 1-11.

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *The Journal of Neuroscience*, *27*(5), 1090-1096.

Zhang, Q., Lawson, A., Guo, C., & Jiang, Y. (2006). Electrophysiological correlates of visual affective priming. *Brain Research Bulletin*, *71*(1), 316-323.

Zhang, Q., Li, X., Gold, B. T., & Jiang, Y. (2010). Neural correlates of cross-domain affective priming. *Brain Research*, *1329*, 142-151.

Zhang, Y. (2008). Plurality and plasticity of neural representation for speech sounds. *The Journal of the Acoustical Society of America, 123*, 3880.

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., Tusler, E., & Carney, E. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Developmental Science*, *14*(3), 566-581.

Appendix A

Summary of ERP Components

Study	Task Design	ERP Component	Time	Rationale
Aguado, Dieguez-Risco, Méndez-Bértolo, Pozo & Hinojosa (2013)	Affective priming paradigm, double task procedure	N400	400 ms	Reverse priming effect; N400 was larger in response to congruent stimuli compared to incongruent stimuli
		Late positive potential (LPP)	700 ms	Evaluative priming effect
Chen, Zhao, Jiang & Yang (2011)	Participants attended to prosody match vs. mismatch	Early negative effect	150 – 250 ms	Occurred later than classic N1 component
		Positive effect	250 – 450 ms	Comparable to prosodic expectancy positivity (PEP); fronto-central location
		Late positive component (LPC)	450 – 900 ms	Location: mainly posterior
	Participants attended to separate task (visual probe)	Early negative effect	150 – 250 ms	Early negativity was been caused by the emotional prosody violation and not an acoustic change; anterior-central location
		Positive effect	250 – 400 ms	Integration process of expectancy violation

	Participants decided whether obvious sound intensity deviation occurred	Early negative effect	130 – 230 ms	Occurred earlier for environmental prosody deviations than spectrally rotated versions; anterior and central areas
		Positive effect	230 – 430 ms	Larger positivity for all mismatched stimuli compared to matched stimuli; all regions and hemispheres
		Late positive component (LPC)	430 – 900 ms	Match and Type effects over posterior regions
Goerlich, Witteman, Schiller, Van Heuven, Aleman & Martens (2012)	Affective categorization: judged valence of affective targets	N400	400 – 500 ms	No behavioral or ERP effects when participants categorized same affective targets based on non-affective characteristics; response competition plays a role
Iredale, Rushby, McDonald, Dimoska-Di Marco & Swift (2013)	Discrimination task: semantically neutral word pairs with congruent and incongruent prosody	N1	50 – 100 ms	Initial processing stage; greater response for emotional prosody than neutral prosody; larger amplitude at parietal sites than frontal sites
		P2	100 – 200 ms	Reflects differentiation of emotion; amplitude largest for happy in left hemisphere and largest for angry in right hemisphere
		N3	400 – 650 ms	Third processing stage (cognitive); localized in frontal cortex
Kamiyama, Abla, Iwanaga & Okanoya (2013)	Affective priming paradigm	N400	250 – 450 ms	Integrative processing of face-music pairs affected by congruency

Paulmann, Bleichner & Kotz (2013)	Participants were asked to rate arousal of pseudo-sentences or their own arousal (implicit vs. explicit instructions)	P200	170 – 230 ms	High arousing stimuli
		Late positivity component (LPC)	450 – 750 ms	High arousing stimuli
Paulmann, Jessen & Kotz (2012)	Emotional prosody judgment task	Prosodic expectancy positivity (PEP)	470 ms	Posterior electrode sites
	Linguistic prosody judgment task	Prosodic expectancy positivity (PEP)	620 ms	Anterior electrode sites
Paulmann & Pell (2010)	Priming paradigm; facial affective decision task	N400	440 – 540 ms	Normal priming effect for medium-length prime and reverse priming effect for short prime
Schirmer & Kotz (2003)	Auditory emotional word Stroop task	N400	350 – 650 ms	Larger negativity for emotionally incongruous words compared to congruous words; time course and scalp distribution similar to N400
Witteman, Goerlich-Dobre, Martens, Aleman, Van Heuven & Schiller (2014)	Dichotic detection task; emotional and linguistic tasks	Early negativity	100 – 140 ms	Left ear elicited stronger negativity than right ear in both hemispheres for emotional task; each ear elicited larger negativity in the contralateral hemisphere for linguistic task
		N2	180 – 320 ms	Larger response at frontocentral sites for emotional task; larger response at posterior sites for linguistic task
		Late positive potential (LPP)	350 – 900 ms	Effect present for emotional but not linguistic task; larger at posterior sites in the left hemisphere, but proximal sites in the right hemisphere

Zhang, Lawson, Guo & Jiang (2006)	Visual affective priming; participants identified target pleasantness	N400	480 – 680 ms	N400 sensitive to affective prime-target mismatches in the visual domain, no cross-modal differences
Zhang, Li, Gold & Jiang (2010)	Cross-domain visual affective priming paradigm; participants categorized target pleasantness	N400	350 – 450 ms	Incongruent affective stimuli; anterior scalp regions for stimulus onset asynchrony of 250 ms
		Late positive potential (LPP)	550 – 700 ms	Attentional resource allocation; across scalp regions for stimulus onset asynchrony (SOA) of 400 ms and across posterior scalp regions for positive targets for SOA of 250 ms

Appendix B
CONSENT FORM FOR ADULT PARTICIPANTS

Emotion Processing in Visual and Auditory Modalities: An Event-Related Potential Study

You are invited to participate in a research study titled "**Emotion Processing in Visual and Auditory Modalities**". This study is being conducted in the Department of Speech-Language-Hearing Sciences at the University of Minnesota. You were selected as a possible participant because you fit the profile we are interested in assessing, and have no medical history of hearing damage or brain injury. The target populations of the study are adults with no history of speech and hearing disorder or brain damage.

This form may contain words or language that are unfamiliar to you. Please ask the researcher if you would like something explained to you. We ask that you read this form and ask any questions that you may have before agreeing to be in the study.

The researchers in this project include Yang Zhang (Ph.D.) of the Department of Speech-Language-Hearing Sciences and Erin Diamond (Undergraduate Research Assistant) Department of Speech-Language-Hearing Sciences.

Background Information

The purpose of the study is to examine how the human brain processes emotion in visual and auditory modalities. We will take both behavioral and brain measures when you listen see a facial expression and/or hear a voice characterized by a specific emotion. For part of the experiment, you will be asked to identify the emotion presented as a facial expression on the screen or through a voice as one of three emotions. For the other part, you will be asked to identify whether or not the emotion presented in the face matches the emotion presented in the voice.

We hope that the results of this study will help us better understand how emotions are processed in the brain. Although the results of the tests will not be of direct benefit to you, they may yield information that will be helpful for the development of effective treatments for people who struggle with emotion processing, such as autistic individuals.

Procedures

If you agree to be in this study, we would ask you to do the following things:

In the recording session, you will sit in a comfortable chair in a sound-treated booth. A stretchable cap with electrodes sewn into it will be fit on your head much like a shower cap. The electrodes will touch the scalp on different spots to record electrical brain activities corresponding to those individual spots. The experimenter will put conductive gels on each electrode to automatically record your brain activities as you listen to sequences of sounds and watch the visually presented material. The set-up will take us about 15 minutes.

The study will consist of four different sections. In two of the parts, you will be asked to identify the emotion that you perceive, either by looking at a facial expression or by listening to a voice. In the other two parts, you will see a face and hear a voice simultaneously and be asked to choose whether the emotions of the face and voice are congruent or incongruent.

During the recording you will be asked to sit as still as possible and to relax your face and muscles as much as possible. You will be given multiple short breaks of about 2-3 minutes in between recording sessions. The experiment will take a total of two hours to complete. If you normally wear contact lenses, we suggest you wear glasses to minimize excessive blinking which can interfere with testing. After testing, your hair will be messy. A hair wash station with sanitized combs, shampoo, towels, and a hair dryer is available for you to use after the experiment is finished.

Throughout the experiments, we will be watching you from a video monitor in the control room via an intercom system. The monitoring system is necessary to ensure proper data collection and timely correction if we see any problematic data. The monitoring video is not recorded. You can stop the session at any time for any reason by simply telling the experimenter that you would like to stop.

Risks and benefits:

You may choose to end participation at anytime without negatively impacting your relationship with the University of Minnesota or the researchers.

This study follows the standard procedures in neurophysiological studies and there is no known risk. However, you may be bored of listening or watching the stimuli. Taking a short break may prevent the occurrence of boredom. Application of gel for recording brainwaves on the scalp is a standard procedure. The gel is made of non-toxic and non-allergenic materials and completely safe; it can be washed off easily using tap water. Accommodations will be made so that you can wash your hair and clean up after the experiment has been completed.

There is no direct benefit for participating in this study, although findings in this study will help us to better understand how emotions are processed in the human brain. This information can be helpful for developing treatment strategies for people with emotion processing problems.

Compensation:

You will receive compensation of $20.00 for your participation in this study. If you are unable or unwilling to complete the study, you will be compensated at a pro-rated rate of $10.00 per hour of participation.

Confidentiality:

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota. If you decide to participate, you are free to not answer any question or to withdraw at any time without affecting those relationships. In the case of withdrawal, the compensation fee will be based on the percentage of participation in the study.

Research Related Injury:

In the event that this research activity results in an injury, treatment will be available, including first aid, emergency treatment and follow-up care as needed. Care for such injuries will be billed in the ordinary manner, to you or your insurance company. If you think that you have suffered a research related injury let the study physicians know right away.

Contacts and Questions:

The researchers conducting this study are: Dr. Yang Zhang and Dr. Zhang's research assistants. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact them at 115 Shevlin Hall, 164 Pillsbury Drive, Minneapolis, MN 55455, Phone: (612) 624-3322, or email: zhang470@umn.edu, diamo057@umn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Statement of Experiment Consent:

I have read the above information. I have asked questions and have received answers. I consent to participate in the adult study.

Signature:____ Date: _____________

Signature of Investigator:__________________________________ Date: _____________

Appendix C

Visual Prime Stimuli