

## THE DYNAMIC ROLE OF SUBPHONEMIC CUES

THE DYNAMIC ROLE OF SUBPHONEMIC CUES IN SPEECH PERCEPTION:  
INVESTIGATING COARTICULATORY PROCESSING ACROSS SOUND CLASSES

By JESSICA ARBOUR, B.A., B.A. (Hons.)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the  
Requirements for the Degree Master of Science

McMaster University © Copyright by Jessica Arbour, August 2012

McMaster University MASTER OF SCIENCE (2012) Hamilton, Ontario (Cognitive  
Science of Language)

TITLE: The Dynamic Role of Subphonemic Cues in Speech Perception: Investigating  
Coarticulatory Processing Across Sound Classes

AUTHOR: Jessica Arbour, B.A., B.A. (Hons.) (McMaster University)

SUPERVISORS: Dr. Tae-Jin Yoon, Dr. Anna L. Moro and Dr. John F. Connolly

NUMBER OF PAGES: viii, 65

## **Abstract**

Neural responses to anticipatory coarticulatory cues were investigated across systematically varying phonological conditions. Congruent or incongruent subphonemic information was placed between an initial consonant and a vowel in a consonant-vowel-consonant (CVC) spoken word (Archibald & Joanisse, 2011). Due to physical and temporal differences across sound classes, the objective was to investigate whether coarticulatory information would be processed differently across controlled manipulations of onset (fricative vs. stop) and vowel type (height vs. backness). Event-related potentials (ERPs) were recorded during a *printed-word/spoken-word matching paradigm*, in which participants indicated whether a visual prime stimulus and a spoken word matched/mismatched. The “*Phonological Mapping Negativity*” (PMN) component provides strong evidence that the use of coarticulatory information in speech recognition varies in strength and timing as a function of onset type (fricative vs. stop) and vowel height (high vs. low). Coarticulatory cues were more readily perceived in spoken word beginning with fricatives than with stops. Similarly, subphonemic variations were more easily detected in low vowels than in high vowels. Observed perceptual and temporal differences are interpreted to reflect variations in subphonemic and phonological processing.

## **Acknowledgments**

This work has been supported by McMaster Arts Research Board (ARB) (awarded to T.J Yoon, A. Moro and J.F. Connolly), Natural Sciences and Engineering Research Council (NSERC) (awarded to T.J Yoon and J.F. Connolly) and the Dean Faculty of Humanities. I would like to thank my supervisory committee, Dr. Tae-Jin Yoon, Dr. Anna Moro and Dr. John F. Connolly, for their guidance and support during the completion of this thesis.

I would also like to extend a special thank you to Dr. Elisabet Service and Dr. Victor Kuperman for their assistance with statistics, and to Janice Lam for helping me test participants. Thank you also to Richard Mah, for his technical help with editing the ERP figures.

## Table of Contents

Descriptive Note.....	ii
Abstract.....	iii
Acknowledgments.....	iv
Table of Contents.....	v
Lists of Figures, Tables and Appendices.....	vi
Declaration of Academic Achievement.....	viii
1.0 Introduction.....	1
1.1 Coarticulation in production.....	3
1.2 Coarticulation in perception.....	4
1.3 Electrophysiological responses to coarticulation.....	8
1.4 Models of speech perception.....	12
1.5 Acoustic characteristics.....	14
1.5.1 Consonants.....	15
1.5.2 Vowels.....	17
2.0 The Present Study.....	19
3.0 Methods.....	22
3.1 Participants.....	22
3.2 Stimuli and experimental conditions.....	22
3.3 Procedure.....	27
3.4 Electrophysiological recording.....	29
4.0 Results.....	32
4.1 Behavioural results.....	32
4.2 Electrophysiological results.....	33
4.2.1 N100.....	33
4.2.2 P200.....	34
4.2.3 PMN.....	36
5.0 Discussion.....	42
5.1 N100 and P200.....	42
5.2 PMN.....	44
5.2.1 Coarticulatory congruency.....	44
5.2.2 Onset type.....	48
5.2.3 Vowel type.....	49
6.0 Conclusions.....	52
References.....	55
Appendix.....	59

## List of Figures, Tables and Appendices

Figure 1: List organized into minimal sets by onset type.....	23
Table 1: Examples of stimuli used for splicing .....	25
Figure 2: Examples of stimuli containing congruent and incongruent coarticulatory cues .....	26
Table 2: Summary of conditions.....	27
Figure 3: Summary of one trial in the printed-word/spoken-word matching paradigm .....	29
Figure 4: Electrode summary .....	30
Table 3: Mean accuracy (SD) of behavioural responses for each condition.....	32
Figure 5: N100 component - Mean amplitude of electrophysiological responses ( $\mu\text{V}$ )...33	
Figure 6: Grand averaged ERP waveforms from all 9 participants at three electrode sites (FZ, CZ, PZ). Grand averages display electrophysiological responses to congruency of coarticulation (congruent (blue) versus incongruent (red)) and onset type (fricatives versus stops) .....	34
Figure 7: P200 component - Mean amplitude of electrophysiological responses ( $\mu\text{V}$ ) across levels of congruency (congruent versus incongruent) and vowel height (high versus low) .....	35
Figure 8: Grand averaged ERP waveforms from all 9 participants at three electrode sites (FZ, CZ, PZ). Grand averages display electrophysiological responses to congruency of coarticulation (congruent (blue) versus incongruent (red)) and vowel type (high versus low).....	36
Figure 9: Mean amplitude ( $\mu\text{V}$ ) of PMN across levels of congruency (congruent versus incongruent) and vowel height (high versus low) for fricative and stop onsets .....	38
Figure 10: Grand averaged ERP waveforms from all 9 participants at three electrode sites (FZ, CZ, PZ). Grand averages display electrophysiological responses across four condition types; congruent coarticulation (blue), incongruent coarticulation with fricative onsets (red), incongruent coarticulation with stop onsets (green) and unrelated incongruent initial phoneme (yellow).....	40
Table 4: Summary of ANOVA Results.....	41
Table 5: Summary of ANOVA Results – Benchmark PMNs.....	41

Appendix 1: Summary of conditions and trial counts.....	59
Appendix 2: List of subconditions and examples of stimuli.....	60
Appendix 3: Participant demographic information and screening form.....	61
Appendix 4: Letter of information.....	62
Appendix 5: Participant debriefing form.....	64
Appendix 6: Consent form.....	65



## **Declaration of Academic Achievement**

The original research question was initiated by my supervisory committee, Dr. Tae-Jin Yoon, Dr. Anna Moro and Dr. John F. Connolly. Together, the committee and I designed the experimental paradigm and created the stimuli list. Speech tokens were recorded by Dr. Anna Moro and a splicing program was generated by Dr. Tae-Jin Yoon. I assisted with all stages of preparing the experiment; including manipulating the speech tokens and overseeing the computer programming (coded by Alex Beaverstone).

Janice Lam and I tested participants to obtain event-related potential (ERP) data. I prepared, averaged and analyzed the collected data. Dr. Tae-Jin Yoon, Dr. Elisabet Service and Dr. Victor Kuperman offered guidance with statistical analyses. Dr. John Connolly provided insights into event-related potentials (ERPs). Finally, Richard Mah assisted with edits which were performed on Figures 6, 8 and 10. I created all figures, graphs and tables included in this manuscript.

## 1. Introduction

A central area of investigation in speech recognition research is how the perceptual system recognizes spoken words despite phonetic variability contained within the speech signal (Gow & McMurray, 2007; Smits, 2001). Speech sounds are often significantly altered through various systematic phonological modifications, which produce subtle and continuous changes to the acoustic signal (Gow & McMurray, 2007; Gaskell, 2001; Gow, 2001). Despite this lack of invariability however, listeners are capable of successfully recognizing spoken words by mapping acoustic information onto speech sounds (Tobin, Whan Cho, Jennett & Magnuson, 2010; McQueen & Cutler, 1999; Tjaden & Sussman, 2006). One question that has been considerably debated in the literature is the function of phonetic detail in speech recognition. Increasing empirical evidence has supported the view that subcategorical cues (*within*-category variations) may provide a “rich” source of information which is valuable to listeners (Gow & McMurray, 2007). In fact, a range of studies have provided evidence that phonetic detail is preserved, and that this information assists listeners during speech recognition (Dahan, Magnuson, Tanenhaus & Hogan, 2001; Joanisse & Archibald, 2011; Gow & McMurray, 2007; McQueen, Norris & Cutler, 1999; Smits, 2001).

Although empirical support suggests the preservation of subphonemic information in speech recognition, the *extent* to which these perceptual cues influence lexical activation is unknown. One source of phonetic variability in speech production is coarticulation. This is a process in which speech sounds are modified as a result of

articulatory overlap of individual speech sounds (Hardcastle & Tjaden, 2011). For example, coarticulation is exemplified in the production of the words “*key*” and “*cool*”. The initial /k/ segment in each is produced in different areas of the vocal tract due to coarticulatory effects stemming from the subsequent vowels (/k/ in the former is articulated further forward on the palate, and the /k/ in the latter is produced further back). Archibald and Joanisse (2011) were the first researchers to have examined neural components reflecting phonological processing in response to congruent/incongruent coarticulatory cues during spoken word recognition. Their neurophysiological findings offer strong evidence that coarticulatory cues are perceived in real-time, and are integrated continuously during speech recognition. However, their stimuli consisted of an uneven distribution of word onset types and vowel types given restrictions associated with using imageable items (e.g. methodology limits possible stimuli). Given that this imbalance consisted of sound classes that differed in both temporal and physical properties, it was unclear whether coarticulatory information is processed differently in some speech sounds compared to others. In the present study, phonologically controlled stimuli were examined in order to investigate whether coarticulatory effects reported in Archibald and Joanisse (2011) would be observed across different sound classes. More interestingly, this study was also conducted to examine whether these manipulations would evoke neurophysiological differences across different phonological contexts. It may be the case that due to the nature of particular speech sounds, some classes transmit perceptually greater coarticulatory cues while these effects are weaker among others (Archibald & Joanisse, 2011; Tobin, Whan Cho, Jennett & Magnuson, 2010). Thus, as an

extension to Archibald and Joanisse (2011), neural responses to matching and mismatching anticipatory coarticulatory cues were examined, with the objective of investigating the perceptual strength and time-course of coarticulatory information across select sound classes through the use of systematically varying phonological conditions.

### ***1.1 Coarticulation in production***

Coarticulation is a process in which communicative gestures instantaneously transition from one speech configuration to another (Hardcastle & Tjaden, 2011). Phonetic variability introduced by coarticulatory processes can result in changes to both spatial and temporal properties of phonological segments (Parush & Ostry, 1983).

Spatial/configurational variation is the result of modifications made to a ‘targeted’ place of articulation, in order to ease integration of neighbouring phonological segments (Kühnert & Nolan, 1999). Temporal changes may be associated with the time in which articulatory gestures start and stop in relation to one another (e.g. by shortening or lengthening the duration of a gestural property) (Kühnert & Nolan, 1999; Flemming, 1997). Direction and extent are elements that are frequently used to describe coarticulatory processes (Hardcastle & Tjaden, 2011). Coarticulatory effects can be *anticipatory*, meaning that speech sounds are influenced by upcoming segments (right to left), or *carryover/perservative*, referring to the impact a current segment has on a following speech sound (left to right) (Parush & Ostry, 1983; Hardcastle & Tjaden, 2011). Articulation of phonological segments can be constrained by several upcoming and/or preceding phonemes, resulting in coarticulatory effects between both adjacent

(Nittrouer & Whalen, 1990) and nonadjacent phonemes (Martin & Bunnell, 1981, 1982; Parush & Ostry, 1993). Moreover, coarticulatory processes have been shown to extend across long ranges. Empirical findings have demonstrated that the effects of coarticulation can occur across syllables (Recasens, 2002; Martin & Bunnell, 1980, 1981) and even across word boundaries (Tobin, Whan Cho, Jennett & Magnuson, 2010). These results draw attention to the considerable impact that coarticulatory effects have on the production of speech, and further highlights the importance of examining how this inherent variability influences the perceptual system.

### ***1.2 Coarticulation in perception***

As suggested in Tjaden and Sussman (2006), a number of sources have provided evidence suggesting the perceptual relevancy of anticipatory coarticulatory information. First, a wide range of behavioural investigations have shown listeners are capable of perceiving fine-grained, coarticulatory cues contained within the speech signal. Nittrouer and Whalen (1990) for example, observed that listeners were capable of identifying a subsequent vowel contained within a CVC sequence, when only provided with a sliced portion of an initial fricative sound. Thus, anticipatory information introduced by coarticulatory effects is beneficial to the perceptual system, as it permits rapid detection of upcoming phonemes. Second, experimental investigations have also used anomalous coarticulatory information to demonstrate listener's ability to detect such cues (Tjaden & Sussman, 2006). For instance, Martin and Bunnell (1981, 1982) utilized cross-spliced final syllables contained within CVCV sequences to create congruent/incongruent

coarticulatory information. Spoken sentences (e.g. “I say /CV1zV2/”) were crossed by taking the appropriate syllable (e.g. zV2) from one sentence and splicing it with an inappropriate counterpart syllable from another sentence. Thus, formant transitions in the previous context provided misleading information about the identity of upcoming phonemes. Increased reaction times were observed in an identification task in which subjects were to report the identity of the second vowel, when the first vowel contained incongruent (versus congruent) coarticulatory cues. This result was interpreted to reflect the fact that incongruent coarticulatory cues were actively processed by listeners, and that anomalous information mislead the perceptual system about later arising speech segments (Martin & Bunnell, 1982; Tjaden & Sussman, 2006). Martin and Bunnell (1980) concluded that coarticulatory cues could inhibit or facilitate the perception of upcoming targets, depending on whether a previous context contains congruent or incongruent anticipatory information. These experiments support the view that the perceptual system is tuned to fine-grained acoustic changes produced by coarticulatory processes. Moreover, coarticulatory information provides an important source of information that is influential in yielding proficient speech recognition (Tjaden & Sussman, 2006).

More recently, the importance of coarticulatory information in speech perception has supported the view that coarticulatory cues influences lexical activation (McQueen, Norris & Cutler, 1999; Gow & McMurray, 2007; Dahan, Magnuson, Tanenhaus & Hogan 2001; Tobin, Whan Cho, Jennett & Magnuson 2010; Archibald & Joanisse, 2011; Hawkins, 2003). As discussed in Archibald and Joanisse (2011), subphonemic modifications could produce gradient (rather than discrete) changes to the acoustic signal,

providing listeners with cues that contain several sources of information that are advantageous to processing (Gow & McMurray, 2007). Archibald and Joanisse (2011) present two lines of evidence demonstrating such gradient effects in speech production. First, acoustic analysis by Gow (2001) revealed that coronals involved in labial assimilation showed spectral formant properties of both an unmodified coronal and an underlying labial phoneme involved in the assimilatory processes (e.g. green part). These findings suggest that the place of articulation of the initial and the subsequent segment sound may be “simultaneously encoded” into the assimilated speech sound, and this could potentially provide listeners with “perceptually enriching” subcategorical information (Gow, 2001). Archibald and Joanisse (2011) also highlight the fact that gradient effects have been observed by Parush and Ostry (1993) who reported that physiological variations (e.g. position of articulators) during the production of consonants vary according to surrounding vocalic context. Ultrasound recordings of VCV sequences showed that a consonant was produced with a narrower pharynx when the initial and/or final vowel was a low vowel. Greater medial movements from the consonant to the vowel were also observed among low vowels, whereas this amplitude was smaller for high vowels (Parush & Ostry, 1993). Due to the graded nature of acoustic and physiological variations, Gow and McMurray (2007) have hypothesized that subcategorical information is beneficial during speech recognition and may provide more “robust” word recognition. Based on this hypothesis, subphonemic cues may offer quicker disambiguating information, since coarticulatory effects facilitate recognition of upcoming phonological segments (Gow & McMurray, 2007; Joanisse & Archibald, 2011). Likewise, sensitivity

to phonetic variation may also prevent the perceptual system from selecting an erroneous candidate, before disambiguating information is conveyed in the acoustic signal (Gow & McMurray, 2007; Joanisse & Archibald, 2011). Evidence for the influence of coarticulatory cues impacting lexical activation was provided by McQueen, Norris & Culter (1999), who observed that misleading subcategorical cues contained within spoken utterances affected mean lexical-decision latencies. Subjects responded significantly faster to words containing congruent coarticulatory information, than to words containing incongruent subphonemic cues resulting from cross-splicing.

Although a range of behavioural evidence supports the view that coarticulation plays a role in speech perception, these measures alone are restricted in their capability of observing the time-course of spoken-language processing (Gow & McMurray, 2007). Investigating coarticulatory cues in speech recognition requires experimental paradigms that can capture the “temporal dynamics” of lexical activation. Eye-tracking methodologies have offered insights into real-time spoken language processing. In fact, several visual processing experiments have supported the hypothesis that subcategorical cues are continuously integrated, and that they assist with spoken word recognition (Tobin, Whan Cho, Jennett & Magnuson, 2010; Dahan, Magnuson, Tanenhaus & Hogan, 2001; McMurray, Clayards, Tanenhaus, & Aslin, 2008). In a *visual world paradigm* for example, Dahan and colleagues (2001) cross-spliced final consonants in word pairs (for example, *net* and *neck*) to obtain matching and mismatching coarticulatory cues contained within spoken words. Eye-movements demonstrated listeners’ sensitivity to subtle coarticulatory information, and revealed that coarticulatory congruency had strong



influence on lexical competition and activation. In cases in which mismatching subphonemic cues matched a competitor word, increased latency of fixation to the target word picture was observed. Furthermore, Dahan and colleagues (2001) also reported a divergence in fixation proportions across matching/mismatching coarticulation at approximately 600 milliseconds (ms) after the onset of the spoken word. These results were interpreted to reflect processing of coarticulatory information contained within the vowel, in which case a competitor word and target word competed for activation in instances of incongruent coarticulation. Thus, visual processing data provide further evidence that the perceptual system integrates coarticulatory information as spoken words become available, and that these cues interact dynamically during lexical activation. Nevertheless, eye-tracking methodology does not supply information reflecting distinct stages of speech perception, providing no insight into the particular cognitive mechanisms that are involved in coarticulatory processing.

### ***1.3 Electrophysiological responses to coarticulation***

A recent electrophysiological study (Archibald and Joanisse, 2011) has also offered support for Gow and McMurray's (2007) continuous integration hypothesis. Event-related potentials (ERPs), a measure of brain activity across time, provide an invaluable source of information for studying phonetic variation. ERPs offer fine-grained, temporal measure specific to discrete sensory and cognitive processes involved in linguistic processing (Connolly, Phillips, Stewart & Brake, 1992; Desroches, Newman & Joanisse, 2008). Due to their temporal accuracy and their links to "distinct" stages of

spoken word processing, ERPs are beneficial for examining coarticulatory cues in speech recognition. In fact, specific ERP responses have been associated with certain levels of processing, including prelexical (e.g. acoustic/phonological processing) and lexical stages of word recognition. For example, an early ERP component (the N100) is a negative-going waveform that peaks approximately 100ms post-stimulus. The N100 is elicited with a frontocentral distribution on the scalp and is associated with sensory processes, including elements such as the volume/intensity of an auditory stimulus or the brightness of a visual stimulus (Steinhauer & Connolly, 2005). Following the N100, is a distinct positive-going P200 waveform which peaks at around 200ms post-stimulus. Although less is known with regards to the dissociated P200, it is most prominently distributed along midline of the scalp (Martin, Tremblay & Stapells, 2007). In previous work (Tremblay, Kraus, McGee, Ponton & Otis, 2001), larger P200 amplitudes have been associated with changes in neural plasticity linked to “training”, during which time subjects are repetitively presented with an auditory stimulus. Collectively, the N100-P200 “complex” reflects sound detection in the auditory cortex (e.g. the onset/offset or modifications made to an acoustic signal), and is not associated with the discrimination of speech sounds (Martin, Tremblay & Stapells, 2007).

The *Phonological Mapping Negativity (PMN)* (Steinhauer & Connolly, 2005; Newman & Connolly, 2009) is a negative-going waveform that peaks between 250-350ms post-stimulus, and is characterized by a fronto-central scalp distribution (Connolly, Service, D’Arcy, Kujala & Alho, 2001). The PMN component reflects a

process in which a “phonological template” (created from a previous word prime) is compared against an incoming speech signal (Connolly & Phillips, 1994; Connolly, Service, D’Arcy, Kujala & Kimmo 2001; Newman & Connolly, 2009). This electrophysiological response is greater when phonological expectations are violated by a detected inconsistency between expected and heard phonemes (Connolly & Phillips, 1994; Connolly, Service, D’Arcy, Kujala & Alho, 2001; Newman & Connolly, 2009). The PMN is evoked to both spoken words and non-words, providing evidence that it does not involve any processing of lexical/semantic information (Connolly, Service, D’Arcy, Kujala & Alho, 2001; Newman, Connolly, Service & McIvor, 2003). As a result, this response has been interpreted to reflect an early and distinct neural mechanism representing a fundamental phonological stage of prelexical speech recognition (Connolly & Phillips, 1994; Connolly, Service, D’Arcy, Kujala & Alho, 2001; Newman, Connolly, Service & McIvor, 2003; Newman & Connolly, 2009). Furthermore, Newman et al. (2003) observed no neural differences to violations that ranged in phonological similarity. Thus, despite whether the mismatch involved just an initial phoneme or several subsequent phonemes, both instances elicited a PMN of equal amplitudes. As a result, this finding suggests that the PMN is an all-or-nothing component (Newman et al., 2003).

While previous studies have demonstrated the PMN’s sensitivity to deviations between phonemes of spoken words/nonwords and phonological expectations (Connolly, Service, D’Arcy, Kujala & Alho, 2001; Newman, Connolly, Service & McIvor, 2003), Archibald and Joanisse (2011) have also shown that the PMN is elicited to subphonemic

mismatches. In an electrophysiological experiment, Archibald and Joanisse (2011) used a *picture/spoken word matching* paradigm, in which subjects were shown a visual prime followed by the presentation of an auditory stimulus. Word pairs were spliced to create congruent and incongruent cues (placed between the onset of the word and the following vowel). For example, the initial segment /f/ in the word “feed” was spliced onto the word “food” (and vice versa) to create incongruent coarticulation. Neural responses were recorded as subjects listened to the auditory stimuli containing matching or mismatching subphonemic information. It was hypothesized that if coarticulatory effects were simply disregarded as noise, then fine-grained changes introduced by coarticulation would be associated with modulations of the N100. As Archibald and Joanisse (2011) described, previous work on the PMN component had assumed that the PMN was not affected by subcategorical information. In contrast with this assumption, neural results showed that both phonemic and coarticulatory mismatches resulted in increased amplitudes of PMNs. This finding confirmed that the PMN is sensitive to both *within-* and *between-*category variability. Alternatively, when ERPs were time-locked to the vowel, mismatches were detected earlier for subphonemic information and later for a mismatching vowel context. This supported the view that subphonemic cues are conserved and processed online, demonstrating that the human brain responds to subtle acoustic-phonetic changes as they become present in the acoustic signal. Archibald and Joanisse (2011) concluded that prelexical processing is beyond merely phonemic categorization, and that processing of subtle subphonemic information should be considered during the phonological stage of speech recognition.

#### ***1.4 Models of speech perception***

A range of investigations have proposed that spoken language is processed online (information contained in the acoustic signal is incorporated as it conveyed), giving rise to competition for lexical activation between words that are “phonologically related” (Desroches, Newman, & Joanisse, 2008; Dahan, Magnuson, Tanenhaus & Hogan, 2001). Several theoretical models of speech perception have tried to account for this competition, although the ways in which speech recognition is achieved has been a matter of debate. For example, in the TRACE model (local connectionist) (McClelland & Elman, 1986), lexical activation occurs at any time during processing through lateral inhibition, established by a match between the incoming speech signal and competitors (Dahan, Magnuson, Tanenhaus & Hogan, 2001; Joanisse & Archibald, 2011). Alternatively, in the Cohort model (Marslen-Wilson & Tyler, 1980) disambiguating information progressively restricts a set of competing candidates until only one item is selected (Desroches, Newman & Joanisse, 2008). Although research supports the claim that listeners are tuned to subphonemic information contained within the acoustic signal, theoretical speech recognition models have seldom included phonetic variability into their accounts (Gow & McMurray, 2007). Nevertheless, research investigating the PMN component has revealed that phonemic and subphonemic cues may influence an early, prelexical stage of speech recognition (Newman & Connolly, 2009; Newman et al., 2003; Archibald & Joanisse, 2011). In fact, the PMN is thought to reflect a process (prior to building cohorts or lexical candidates), in which phonological expectations interact with an incoming signal (Newman & Connolly, 2009; Newman et al., 2003).

Like TRACE, the MERGE model (Norris, McQueen & Cutler, 2000) suggests that lexical competition occurs through lateral inhibitions. MERGE is a strictly feedforward (bottom-up) model, in which phonological information is processed without lexical “feedback” (Newman et al., 2003). Newman et al. (2003) have proposed that the PMN is consistent with the MERGE model of speech recognition. Previous studies (Newman et al., 2003; Connolly, Service, D’Arcy, Kujala & Alho, 2001) have hypothesized that larger PMN amplitudes, elicited to phonological violations, could reflect the fact that greater work is required to generate “new hypotheses” from an incoming speech signal. It has been suggested (Newman et al., 2003) that the choice of lexical candidates is influenced by phonological expectations built from the preceding word prime (top-down). However, consistent with the MERGE model, this “phonological template” will not affect bottom-up processing from the speech signal, but instead influences “decisions at the higher level” (Newman et al., 2003). Therefore, it is hypothesized that the PMN represents a “merging” of both top-down (phonological expectations) and bottom-up (acoustic input) during a prelexical stage of speech recognition (Newman et al., 2003).

Given that recent work (Archibald & Joanisse, 2011) has suggested that the previous view of the PMN was too narrow, it is necessary to expand the role of this component to subcategorical cues and to examine these effects across various phonological contexts. For example, it may be the case that some coarticulatory mismatches are less salient across some sounds classes and therefore, elicit weaker PMNs in response to minimal phonological processing. It is important to further investigate

variation introduced by coarticulatory processes in order to examine the dynamic role of these cues across sound classes and to develop accurate models of spoken language recognition.

### *1.5 Acoustic characteristics*

Results provided by Archibald and Joanisse (2011) offer evidence coarticulatory information is detected in real-time and influences spoken word recognition, however their stimuli lacked an even distribution of word onset types due to the use of imageable stimuli. Only items that could be represented by pictures were selected as stimuli within the experiment, limiting researchers' ability to control phonological variables (e.g. including onset and nuclei type). For example, most of auditory stimuli began with a fricative (18 out of 30 pairs) while others onsets consisted of an affricate (3 out of 30) or a nasal (9 out of 30). Since these sound classes consist of variations in temporal and physical properties, it is unknown whether coarticulatory cues stemming from particular consonant classes are processed differently than others. Similarly, an imbalanced distribution of spliced vowel (nuclei) pairs generated uncertainty about whether a subsequent vocalic context also influences the perceptibility of coarticulatory information. For instance, most stimulus pairs consisted of /u/- /i/ distribution (13 out of 30 pairs), while others included /ɑ/ - /æ/ (5 out of 30 pairs), /ɑ/ - /i/ (5 out of 30 pairs), /u/ - /æ/ (6 out of 30 pairs) and /o/ - /æ/ (1 out of 30 pairs). In a previous investigation (Tobin et al., 2010) listeners' sensitivity to anticipatory long-range coarticulatory mismatches were examined through the use of systematically varying consonant onsets. Auditory stimuli were created by cross-splicing the final words in phrases such as “pick up a pole”

with “pick up a pail”. Their stimuli were designed to vary in “high” and “low” resistance to coarticulatory effects. For example, segments with low coarticulatory resistance included phonemes such as /p/ and /f/, and high resistant segments such as /t/ or /s/. Resistance was based on the prediction that coarticulatory effects were likely to spread across phonemes that do not require strong constraints on the position of the tongue. They (Tobin et al., 2010) reported that listeners were slowest to fixate the target picture in the low resistance coarticulatory mismatch condition (high resistance coarticulatory mismatch approached significance); suggesting that coarticulatory effects are stronger in some phonological environments and more subtle in others. However, Tobin and colleagues (2010) noted that manipulations were “relaxed” in some instances to permit enough speech tokens. It may be the case that certain sound classes permit maximal configuration of the oral cavity in anticipation of a following vowel, and this produces greater detectable coarticulatory effects in some sound classes compared to others. In fact, Yeni-Komshian & Soli (1981) have examined the perception of FV (Fricative-Vowel) syllable coarticulation, and have suggested that coarticulatory effects do not occur consistently across classes of speech sounds. Thus, configurational differences in Archibald and Joanisse (2011) may have had an impact on the processing of coarticulatory cues, and therefore, may have elicited variations/modulations in ERP responses.

**1.5.1 Consonants.** Consonant phonemes vary considerably in their acoustic properties. For example, certain consonants produce more noise than others during production.



Likewise, certain consonants are generated with complete closure of the vocal tract (e.g. stops) whereas others involve a narrow constriction (e.g. fricatives) (Kent & Read, 1992). The following discussion focuses on a description of stop and fricative consonants, reflecting variations in these class differences stemming from differences in manner of articulation (the ways in which speech organs are used to produce sounds). As a result, differences in acoustic cues may influence the detection (e.g. strength, timing) of coarticulatory information during speech perception.

Stop consonants are recognized as the class of sounds with the most articulatory “constrictions/obstructions” of airflow (Raphael, 2008). Stops are produced with momentary blockage in the vocal tract, followed by the release of air pressure built from the constriction (Kent & Read, 1992; Raphael, 2008). As a result, one of the most noticeable characteristics of a stop consonant in a spectrographic analysis is a “spike”, indicating a burst of acoustic intensity produced by the rapid release of airflow (Raphael, 2008). The burst of stops (5-40ms) are known to be one of the briefest events in the production of speech (Kent & Read, 1992). Word initially, voiceless stops are produced with a brief aspiration (25-100ms) that occurs between the burst and the voicing of a subsequent vowel. Consonant-Vowel (CV) formant transitions from an initial stop to a following vowel are associated with articulatory shaping during the transition from C to V, and are significant cues utilized in speech recognition. This is well known from perceptual experiments which have demonstrated that listeners can distinguish between consonants /b/, /d/, /g/, based on the second formant (Liberman, Safford Harris, Hoffman & Griffith, 1957). Formant transitions from a stop to a vowel (or vice versa) take place in

approximately 50ms, during which time frequencies change from stop “values” to those of a vowel (Kent & Read, 1992). These brief transitions (from a stop to a vowel) are associated with the quick articulatory movements characterized by stops (Kent & Read, 1992).

Fricatives are produced by channeling airflow through a narrow constriction, which results in turbulent or aperiodic sound. In contrast to stop consonants, fricatives are produced with a gradual onset and a greater duration in “rise time” acting as a cue to manner of articulation (Raphael, 2008). As a result, the produced “frication” is typically longer in duration compared to other aperiodic segments, including elements such as the stop bursts or the aspiration in a stop release (Raphael, 2008). Since these consonant classes are different in terms of manner of articulation, coarticulatory information contained within fricatives and stops may be perceived differently.

**1.5.2 Vowels.** Unlike consonants, vowels are known to have the greatest unobstructed airflow. It is well known that the primary acoustic cues to the perception of vowels are the formants (or vocal tract resonances), which vary according to the place (or point) of articulation. This section is devoted to describing differences across vowel height (high and low) and vowel backness (front and back) for reason that the perception of subtle subcategorical cues may be influenced by the type of vowel involved in coarticulatory process.

Formant patterns are utilized to distinguish a vowel and to determine the relation between acoustic and perceptual boundaries. Despite some exceptions, an overall rule is

that F1 differs with the height of the tongue (high-low) and F2 differs with “advancements” of the tongue (front-back) (Kent & Read, 1992). Consequently, a high F1 is associated with low vowels and a low F1 is linked to high vowels (Kent & Read, 1992). Moreover, back vowels consist of a low F2 (little variation between F2-F1), while front vowels are known to have a greater F2 (bigger divergence in F2-F1) (Kent & Read, 1992). In examining VCV sequences by means of an ultrasound which measured coarticulation of pharyngeal wall movements, Parush and Ostry (1993) found that the articulatory position during the consonant was influenced by vowel height. Researchers found greater amplitude in movement of the lateral pharyngeal wall for low back vowels compared to high vowel in the CV transition. When the first vowel was fixed, the consonant was produced with a narrower pharynx when the final vowel was /a/, compared to when it was preceded by a /u/ vowel. Therefore, they observed that the magnitude of the CV movement involved with /u/ was “significantly less” than that of /a/. Furthermore, movements from the consonant towards the low vowel (/a/) began earlier than movements towards the high vowel (/u/). Parush and Ostry (1993) concluded by suggesting that the “position of the articulators” and the “magnitude/direction” of the gestures should be examined when accounting for phonological context. Thus, the perceptual strength of coarticulatory information may also vary as a function of the nature of the coarticulated vowel (e.g. place of articulation), due to factors such as the extent of the transition from the initial consonant to the subsequent vowel and the constraints placed on the articulators. To investigate the effects of coarticulation across acoustic properties, it is necessary to strictly manipulate phonological conditions.

## **2. The Present Study**

The purpose of this experiment was to investigate whether Archibald and Joanisse's (2011) findings could be extended using a set of systematically varying phonological conditions. Examination of these fine details across differing sound classes could provide additional evidence supporting the hypothesis that the perceptual system continuously integrates coarticulatory cues online, and that this information influences speech recognition. Moreover, controlled manipulations were also examined to determine whether neurophysiological differences would be observed across sound classes. It is unknown whether coarticulatory cues are processed similarly or asymmetrically given temporal, physical and physiological differences underlying these classes. If such variations are obtained, this knowledge could provide enhanced insights into the perceptual strength/extent and the temporal dynamics of subphonemic effects in during speech recognition.

To achieve these objectives, word onsets and vowel types were strictly manipulated and balanced through the use of minimal sets. The aim of this experiment was to have greater control over these phonological manipulations, even though fewer sounds were used compared to Archibald and Joanisse (2011). As a result, matching and mismatching coarticulatory cues were placed between an initial consonant (fricative or stop) and a subsequent vowel (e.g. high front, high back, low front, and low back) of a CVC spoken word. Codas remained consistent within minimal sets to reduce any regressive coarticulatory effects on the preceding phonemes.

In order to achieve these strict criteria, a printed word was used to create a phonological expectation. Event-related potentials (ERPs) were recorded as subjects participated in a *printed-word/spoken-word matching* paradigm, in which a visual prime was used to create a phonological expectation for a subsequently heard auditory word. This experimental paradigm is not subject to the imbalanced stimuli problem that arose from the *picture/spoken-word matching paradigm* conducted in earlier experiments (Joanisse & Archibald, 2011, Desroches, Newman & Joanisse, 2009). In conditions that contained subphonemic mismatches, formant transitions in the consonant provided misleading coarticulatory information about the subsequent vowel. Speech tokens were created using a splicing technique, in which words such as ‘feed’ /f<sup>i</sup>id/ and ‘food’ /f<sup>w</sup>ud/ were spliced to create /[f<sup>i</sup>]eed/ (congruent) and /[f<sup>w</sup>]eed/ (incongruent). Three experimental conditions were designed (Examples provided in Table 2): 1) coarticulatory/phonemic/orthographic match condition (congruent coarticulation, ‘feed’ – /f<sup>i</sup>id/), 2) coarticulatory mismatch/phonemic and orthographic match (incongruent coarticulation, ‘feed’ – /f<sup>w</sup>id/) and 3) coarticulatory match/phonemic mismatch/orthographic mismatch (unrelated condition, ‘seat’ – /s<sup>i</sup>it/).

Consistent with the findings from Archibald and Joanisse (2011), it was hypothesized that larger PMNs would be elicited in cases of incongruent (versus congruent) coarticulation given that these subtle misleading cues violated listeners’ phonological expectations. This result would strengthen Archibald and Joanisse (2011) findings, providing evidence that coarticulatory cues are perceived and impact spoken

word recognition in real time, during a prelexical stage. Furthermore, neurophysiological differences (e.g. gradations) were also anticipated as a function of the nature of the sound classes. First, coarticulatory variations were expected to evoke ERP differences across onset types (stops versus fricatives). Specifically, subphonemic information may be processed faster in stops than in fricatives given differences in temporal properties across these sounds classes. Second, ERP differences were also anticipated in the perception of coarticulatory cues across vowel place of articulation (height versus backness). Subphonemic information may be greater detectable in instances in which vowels have reduced articulatory constraints (e.g. low, open vowels). Moreover, stronger coarticulatory effects may be generated with low vowels given that the tongue is positioned as far as possible from the roof of the mouth, and this transition produces larger amplitude of movement from the initial consonant to a low vowel.

### **3. Methods**

#### ***3.1 Participants***

Fourteen-participants (8 female, 6 male) were recruited through advertisements. Data from five participants were excluded from the analysis due to poor quality ERP recordings or failure to complete the experiment. Recordings from nine-individuals (4 female, 5 male) were used for analysis. Participants ranged in age from 18 to 31 years, with a mean of 22.56 ( $SD = 5.02$ ). Participants were undergraduate students or university graduates at McMaster University. All participants were typically developed individuals, and no reported neurological or other relevant medical histories (See Appendix 3; Screening form). Furthermore, all participants were native speakers of English, some of whom had limited experience with other languages (5 monolingual, and 4 with limited experience). Participation in the study was voluntary and financially compensated. Participants provided signed informed consent prior to the experiment (See Appendix 6; Consent Form), which was authorized by McMaster University's Research Ethics Board. See Appendix 4 and 5 for letter of information and debriefing forms.

#### ***3.2 Stimuli and experimental conditions***

Stimuli consisted of 30 CVC words, organized into minimal sets (See Figure 1). Each minimal set was composed of a grouping of words with different meanings that deviated by only one phoneme (the vowel). The onset of each word began with a fricative /f, s, ʃ, h/ or a stop /p, t, k/. Four corner vowels, /i, u, æ, α/, were used for maximal contrast. One of the following consonants, /p, t, d, k, l/, served as the coda for each of the

four words consisting of a minimal set., The objective of using the same (coda) phoneme within sets was to reduce any potential effects of regressive coarticulation. Given the strict criteria of implementing phonologically controlled stimuli, the minimal set for fricative /f/ was divided into two sets of three words (as can be seen in Figure 1) because a combination of four words was not possible.

**Figure 1.** List of stimuli organized into minimal sets.

**Fricatives: [f], [s], [ʃ], [h]**

[f]

Feed	Food
[i]	[u]
Fad	
[æ]	[ɑ]

Feet	
[i]	[u]
Fat	Fought
[æ]	[ɑ]

[s]

Seat	Suit
[i]	[u]
Sat	Sought
[æ]	[ɑ]

[ʃ]

Sheet	Shoot
[i]	[u]
Shat	Shot
[æ]	[ɑ]

[h]

Heat	Hoot
[i]	[u]
Hat	Hot
[æ]	[ɑ]



**Stops – [p], [t], [k]**

[p]

Peal	Pool
[i]	[u]
Pal	Paul
[æ]	[ɑ]

[t]

Teak	Toque
[i]	[u]
Tack	Tock
[æ]	[ɑ]

[k]

Keep	Coop
[i]	[u]
Cap	Cop
[æ]	[ɑ]

The list of stimuli was not restricted to the use of imageable words so that stricter criteria could be implemented to create more systematically varying phonological conditions. Words were presented visually (white letters on a black background) for 1500s ms and were followed immediately by the target word. The spoken target word had an average duration of 516.59ms.

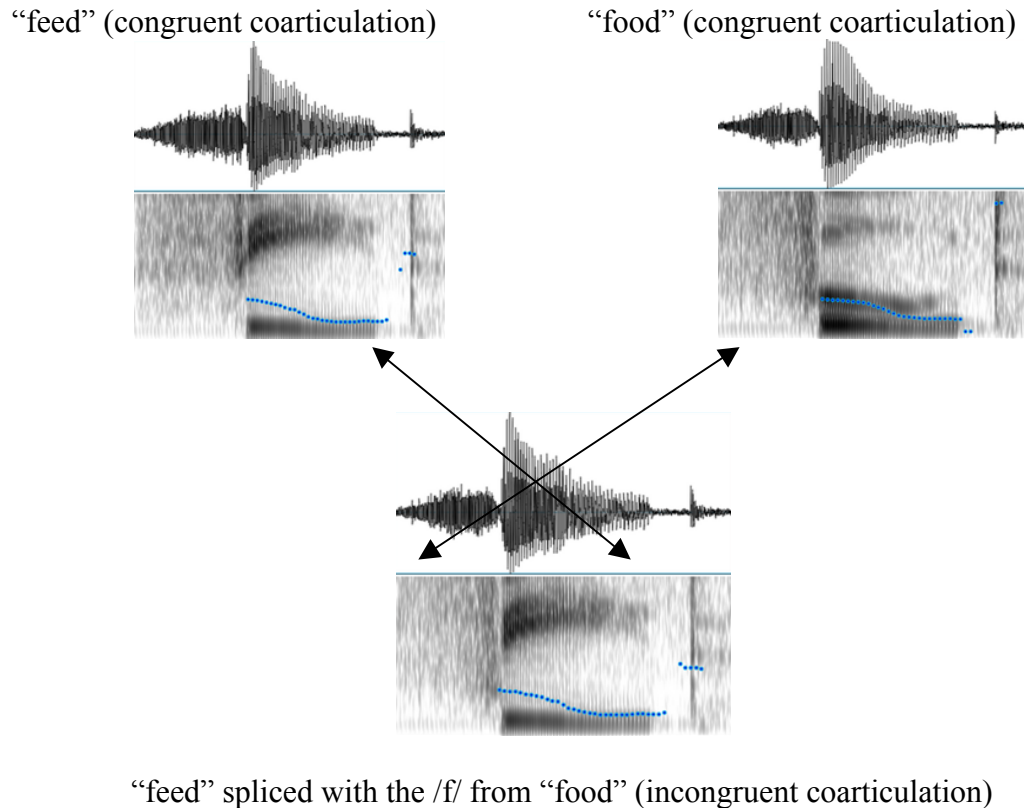
Speech tokens were recorded by a female speaker of Canadian English. The speaker spoke each target word three times. For consistency across recordings, the speaker spoke each stimulus item between the words “say” and “again” (e.g., “say (stimulus item) again”). Speech tokens were produced using M-Audio Microtrack II recorder (at a sampling rate of 44.1 K Hz and a quantization rate of 16 bit). Table 1 shows examples of stimuli used for splicing.

**Table 1.** Examples of stimuli used for splicing

	Stops		Fricatives	
High	Peal /pɪl/	Pool /pu:l/	Heat /hit/	Hoot /hu:t/
Low	Pal /pæl/	Paul /pa:l/	Hat /hæt/	Hot /hɒt/

A single speech token was chosen among the repetitions. Stimulus waveforms were viewed in Praat in order to identify the junction between the consonant and the vowel. Stimulus items were then zoomed in by 5 to 10 cycles and a zero crossing point (where the waveform crosses the horizontal axis reflecting the transition from the word's onset to the vowel) selected between the consonant and the vowel. Incongruent coarticulatory information was created by splicing stimuli with their associated minimal pairs (e.g. two words with separate meanings that differed by the vowel) using a custom made Praat script. The script was programmed to take the onset of one word and replace it with the onset of another word. For example, 'feed' /f<sup>i</sup>id/ and 'food' /f<sup>u</sup>ud/ were spliced to create [f<sup>i</sup>]eed (congruent) and [f<sup>u</sup>]eed (incongruent) (Whalen, 1991, McQueen, Norris & Cutler, 1999, Joanisse & Archibald, 2011). Figure 2 displays samples of spectrograms illustrating conditions of congruent and incongruent coarticulation within this experiment.

**Figure 2.** Examples of stimuli containing congruent and incongruent coarticulatory cues



From these splicing methods, three conditions were created (Table 2). In the coarticulatory/phonemic match condition (congruent), both coarticulatory information and initial spoken phoneme was consistent with the written word, (e.g., food – [f<sup>w</sup>]ood). In the coarticulatory mismatch/phonemic match (incongruent) condition, the written-word was paired with a spoken word that contained the correct initial phoneme but incongruent coarticulation information (e.g., food, [f<sup>d</sup>]ood). In coarticulatory match/phonemic mismatch (unrelated), coarticulatory information was congruent but the initial phoneme was incongruent (e.g., feet – [s<sup>i</sup>]eat). This final condition was created to

give participants an obvious mismatch response inasmuch as coarticulatory mismatches are very subtle.

**Table 2.** Summary of conditions

Condition	Visual	Auditory
coarticulatory match/phonemic match (congruent)	Food	[f <sup>w</sup> ]ood
coarticulatory mismatch/phonemic match (incongruent)	Food	[f <sup>d</sup> ]ood
coarticulatory match/phonemic mismatch (unrelated)	Feet	[s <sup>l</sup> ]eat

Thus, 218 tokens were presented to participants (88 congruent, 81 incongruent and 49 unrelated) per session, which amounted to 872 tokens (352 congruent, 324 incongruent and 196 unrelated) across a total of four experimental sessions. Stimuli were randomized within-participants. Of interest was the analysis of onset type (fricative or stop) and vowel type (high, low, front and back). A combination of these phonological conditions including onset type (fricative or stop) and vowel type (high, low, front and back), resulted in 8 sub-conditions for the congruent condition, and 8 sub-conditions for the incongruent. Five practice trials, selected at random from the last 18 items in the stimulus list, were randomized across participants. Practice trials were included in the ERP analysis. See Appendix 2 for full description of subconditions and trial counts.

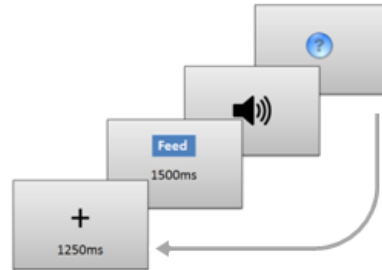
### **3.3 Procedure**

Auditory stimuli were delivered binaurally using earphones (Etymotic Research) and an amplifier (ARTcessories HeadAmp4). Both visual and auditory stimuli were

presented using Presentation (NeuroBehaviouralSystems Presentation 14.7). Participants saw visual prime stimuli (words – white letters against a black background) that were presented on a 24-inch Hewlett-Packard (HP) computer monitor position 1 m away. Participants were instructed to decide whether the visually presented prime word lexically matched or mismatched the following spoken word and to indicate their decision with a manual response – left button press for “match” and right button press for “mismatch.” Buttons responses were not balanced (e.g. left for “mismatch” and right for “match”) across or within-participants. Although lack of balancing could produce anticipation effects, this methodology would not influence neurophysiological data given previous research. Archibald & Joanisse (2011) reported no behavioural differences, despite observed neural variations.

A trial began with the presentation of a fixation cross, which appeared for 1250ms. The fixation cross was replaced by the visual word prime that was presented for 1500ms. Immediately following the word prime, a spoken word was played through the headphones while the word prime continued to be displayed. A subsequent trial would not begin until the participant decided on the match/mismatch relationship and this response was made 700 ms after the onset of the spoken word (See Figure 3).

**Figure 3.** Summary of one trial in the printed-word/spoken-word matching paradigm



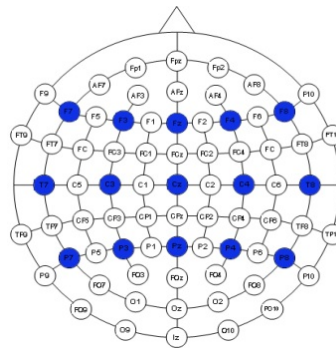
Five practice trials were given before beginning each test session. Subjects participated in a total of four test sessions, scheduled over two days of testing. Sessions one and two took place on the first day, and sessions three and four on the second day. Participants completed both testing days between a span of one to 11 days (three participants (33%) one day, two participants (22%) two days, one participant (11%) three days, one participant (11%) 7 days, one participant (11%) 9 days and one participant (11%) 11 days). The duration of one session was approximately 25 minutes in length. Each session was divided into 11 blocks, providing participants with a break of 15-seconds after a block of 21 tokens. Longer breaks were given between sessions to reduce effects of fatigue.

### ***3.4 Electrophysiological recording***

EEG was recorded (BioSemi Active-Two system) from 15 sites (F3, F7, C3, T7, P3, P7, PZ, FZ, F4, F8, CZ, C4, T8, P4, P8 according to the 10-20 system; see Figure 4) with Ag/AgCl electrodes using a 0.01 to 100 HZ bandpass, sampled at 512 Hz, digitally filtered offline with a 20 Hz low pass filter and referenced to the tip of the nose. Continuous recordings were made and epochs (200 ms prestimulus to 1000 ms

poststimulus) acquired at the onset of the spoken word target. Electrooculographic (EOG) activity was recorded from electrodes placed above and over the outer canthus of the left eye. The average duration of the initial stop onsets was 78.24ms ( $SD = 19.61$ ) and fricative onsets averaged 190.48ms ( $SD = 36.53$ ).

**Figure 4.** Electrode summary



The three ERP components of interest in this experiment were the N100, the P200 and the PMN. The N100 was identified as the most negative peak in the 99-219 ms latency window, the P200 as the most positive peak in the 219-319 ms latency window, and the PMN as the most negative peak in the 280-380 ms latency window. Bearing in mind the temporal properties known to characterize these electrophysiological components, latency windows for scoring were chosen based on visual inspection of the observed peaks in the ERP figures. Component amplitudes were scored as the mean integrated amplitude in a 50 ms window around the identified peak.

A repeated-measures analysis of variance (ANOVA) was conducted separately for each ERP component (N100, P200 and PMN). Each repeated-measures ANOVA was constructed with site with three levels (FZ, CZ, and PZ), onset type with two levels (fricative and stop), coarticulatory congruency with two levels (congruent and

incongruent), vowel height with two levels (high and low) and vowel backness with two levels (front and back). ERP amplitudes were measured in microvolts (uV) and compared across conditions. Furthermore, pairwise t-tests were conducted in instances of significant interactions. A summary of main effects and interactions for the electrophysiological data analysis are outlined in Table 4.



## 4. Results

### 4.1 Behavioural accuracy

Behavioural accuracy is displayed in Table 3. Participants' performance accuracy exceeded 96% in all conditions (See Table 3). It must be emphasized that responses indicating a “match” between the prime and target were scored as correct for conditions containing incongruent coarticulation, since visual and spoken items matched on a lexical level. Behavioural data were analyzed with a repeated-measures ANOVA with four factors: congruency (congruent and incongruent), onset (stop and fricative), and vowel height (high and low) and vowel backness (front and back). No effects were found indicating that behaviour was unaffected by any of the factors in this experiment.

**Table 3.** Mean accuracy (*SD*) of behavioural responses for each condition

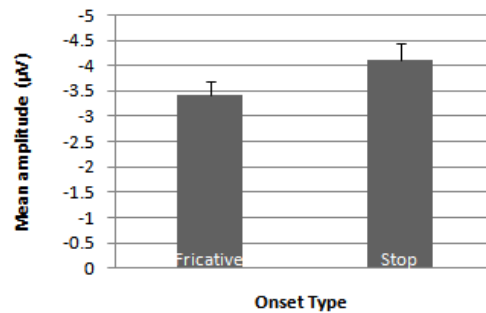
<i>Condition</i>	<i>Accuracy</i>
Coarticulatory Match, Phonemic Match, Stop, High (Congruent)	98.01 (.060)
Coarticulatory Match, Phonemic Match, Stop, Low (Congruent)	100.00 (0)
Coarticulatory Match, Phonemic Match, Stop, Back (Congruent)	99.65 (.007)
Coarticulatory Match, Phonemic Match, Stop, Front (Congruent)	99.90 (.003)
Coarticulatory Match, Phonemic Match, Fricative, High (Congruent)	99.55 (.007)
Coarticulatory Match, Phonemic Match, Fricative, Low (Congruent)	99.75 (.005)
Coarticulatory Match, Phonemic Match, Fricative, Back (Congruent)	99.40 (.009)
Coarticulatory Match, Phonemic Match, Fricative, Front (Congruent)	99.90 (.003)
Coarticulatory Mismatch, Phonemic Match, Stop, High (Incongruent)	96.45 (.068)
Coarticulatory Mismatch, Phonemic Match, Stop, Low (Incongruent)	97.39 (.065)
Coarticulatory Mismatch, Phonemic Match, Stop, Back (Incongruent)	96.31 (.068)
Coarticulatory Mismatch, Phonemic Match, Stop, Front (Incongruent)	97.53 (.070)
Coarticulatory Mismatch, Phonemic Match, Fricative, High (Incongruent)	98.48 (.014)
Coarticulatory Mismatch, Phonemic Match, Fricative, Low (Incongruent)	98.57 (.027)
Coarticulatory Mismatch, Phonemic Match, Fricative, Back (Incongruent)	98.82 (.015)
Coarticulatory Mismatch, Phonemic Match, Fricative, Front (Incongruent)	98.23 (.018)
Coarticulatory Match, Phonemic Mismatch (Unrelated)	98.80 (.006)

## 4.2 Electrophysiological results

**4.2.1 N100.** Analysis of N100 amplitudes revealed a main effect of onset type (stops and fricatives),  $F(1,8) = 6.49, p < .05, \eta^2 = .45$ , reflecting the significantly larger N100 in response to stop ( $M = -4.12, SD = .99$ ) than for fricative onsets ( $M = -3.41, SD = .84$ ) (Figure 5). The effect is also apparent in the waveform grand averages (Figure 6).

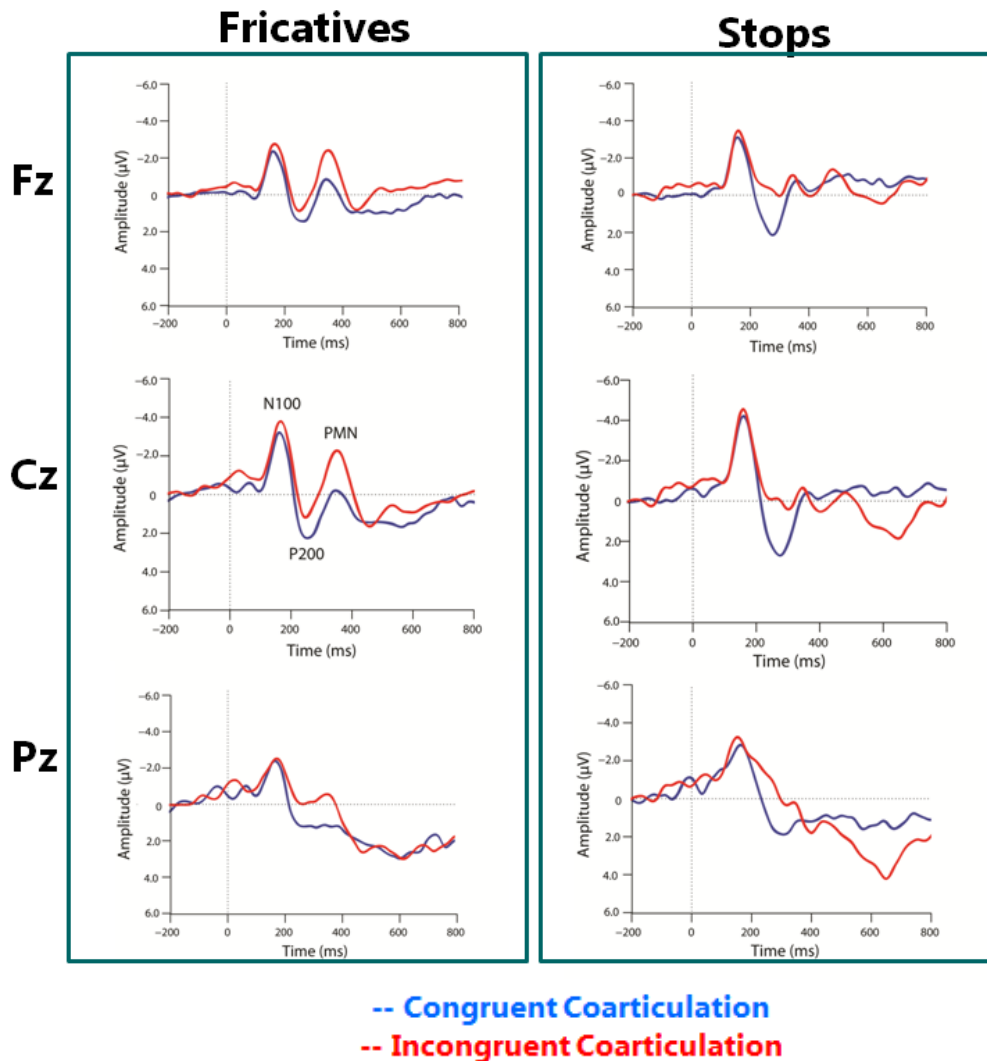
Although no significant distribution effect was noted, Figure 6 exhibits the N100's typical frontocentral distribution. No other main effects or interactions were observed.

**Figure 5.** N100 component - Mean amplitude of electrophysiological responses ( $\mu\text{V}$ )



\*NB: Error bar is 1 standard error.

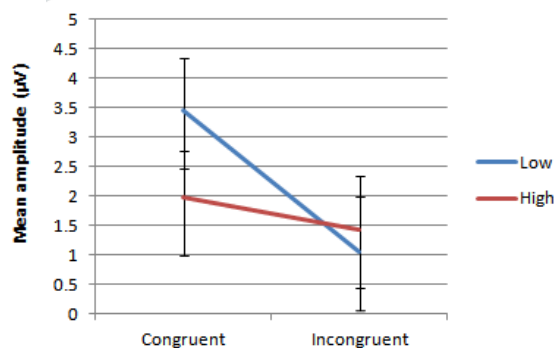
**Figure 6.** Grand averaged ERP waveforms from all 9 participants at three electrode sites (Fz, Cz, Pz). Grand averages display electrophysiological responses to congruency of coarticulation (congruent (blue) versus incongruent (red)) and onset type (fricatives versus stops).



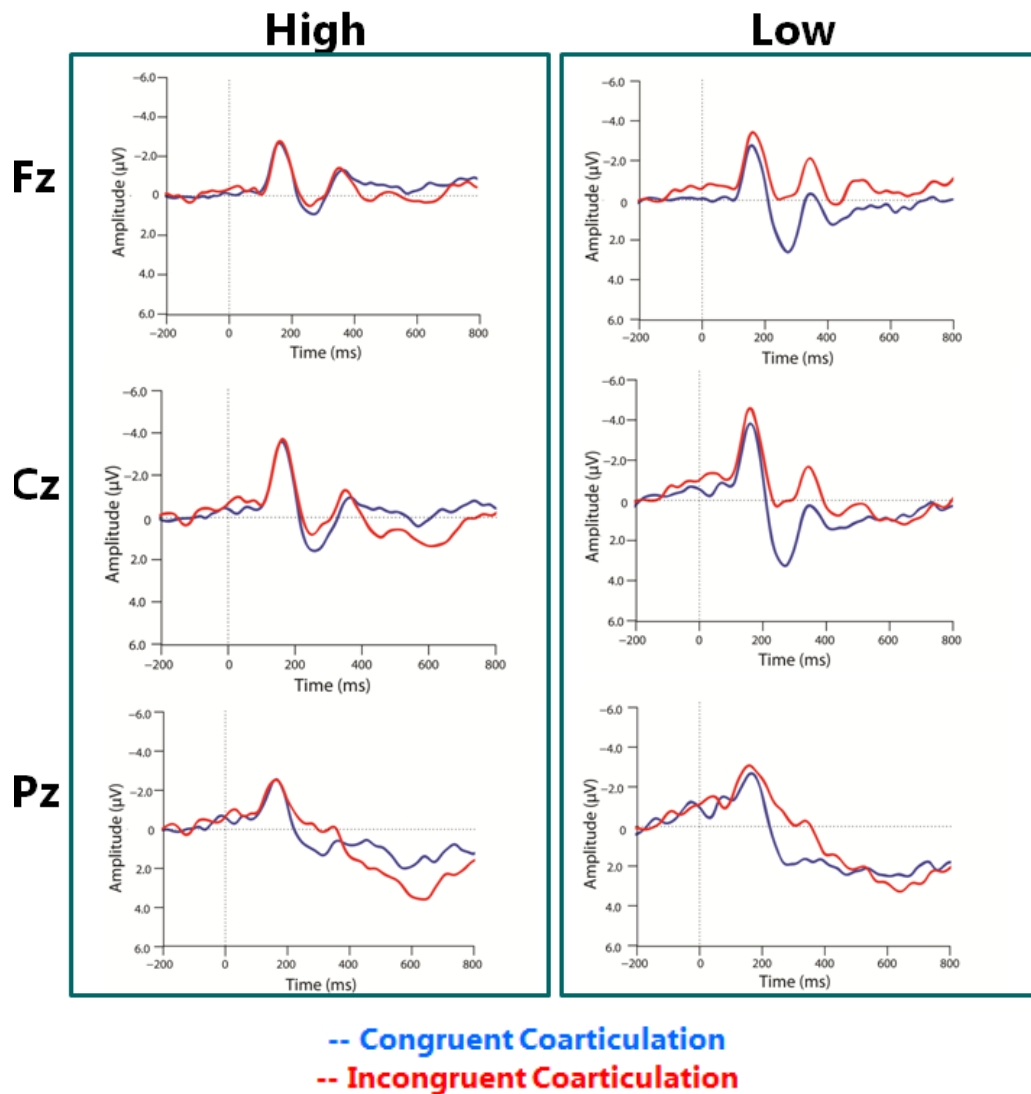
**4.2.2 P200.** Analyses of the P200 component revealed a significant main effect of congruency,  $F(1,8) = 17.79$ ,  $p < .01$ ,  $\eta^2 = .69$ , suggesting that the amplitude of the P200 is affected by coarticulatory miscues. Larger P200 responses were observed to congruent ( $M = 2.72$ ,  $SD = 2.46$ ) than incongruent coarticulation ( $M = 1.25$ ,  $SD = 2.69$ ). There was

also a significant interaction between coarticulatory congruency and vowel height,  $F(1,8) = 18.73, p < .01, \eta^2 = .70$ . Paired t-tests revealed that within spoken words containing low vowels, the P200 was significantly less positive to incongruent coarticulatory cues ( $M = 1.06, SD = 2.78$ ) than congruent ( $M = 3.46, SD = 2.64$ ),  $t(8) = 5.27, p = .001, r = .88$  (Fig. 8). No significant differences were found when comparing levels of coarticulatory congruency contained within spoken words with high vowels,  $t(8) = 1.51, p > .05$ . No significant differences were found for incongruent subphonemic information across high and low vowels,  $t(8) = -.91, p > .05$  (see Fig. 7). Lastly, Mauchley's test indicated that the assumption of sphericity had not been violated ( $\chi^2(2) = 2.16, p > .05$ ) and a significant interaction was obtained across electrode site and vowel backness,  $F(2,16) = 4.07, p < .05, \eta^2 = .34$ .

**Figure 7.** P200 component - Mean amplitude of electrophysiological responses ( $\mu V$ ) across levels of congruency (congruent versus incongruent) and vowel height (high versus low).



**Figure 8.** Grand averaged ERP waveforms from all 9 participants at three electrode sites (Fz, Cz, Pz). Grand averages display electrophysiological responses to congruency of coarticulation (congruent (blue) versus incongruent (red)) and vowel type (high versus low).



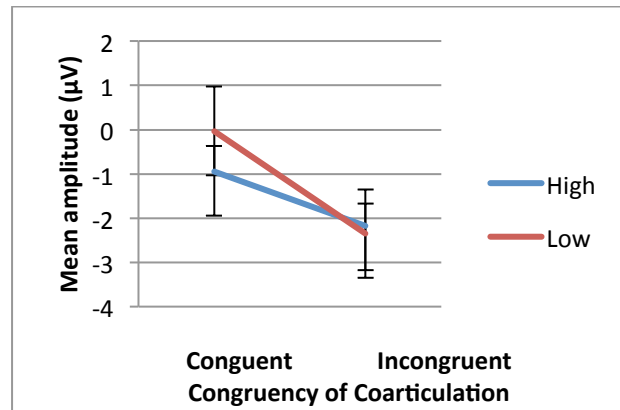
**4.2.3 PMN.** First, Mauchley’s test revealed that the assumption of sphericity had not been violated ( $\chi^2(2) = 2.49, p > .05$ ) and a significant 5-way interaction (site\*congruency\*onset type\*vowel height\*vowel backness) was observed,  $F(2,16) = 7.40, p < .01, \eta^2 = .48$ .

Second, the PMN proved to be sensitive to coarticulatory congruency,  $F(1, 8) = 13.09, p$

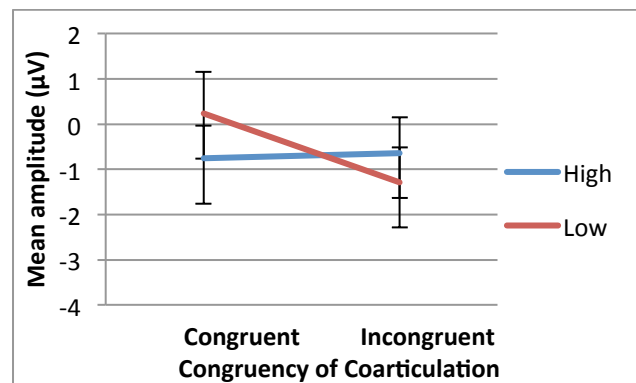
$< .01$ ,  $\eta^2 = .62$  exhibiting significantly larger amplitudes to coarticulatory miscues ( $M = -1.6$ ,  $SD = 1.84$ ) than coarticulatory congruencies ( $M = -.37$ ,  $SD = 1.62$ ). In addition, PMN differences were also observed in response to the nature of the onset (i.e., stops and fricatives) as reflected by the significant interaction between coarticulatory congruency and onset type,  $F(1, 8) = 5.91$ ,  $p < .05$ ,  $\eta^2 = .43$ , suggesting that congruency of coarticulation had different effects on mean amplitude of the PMN depending on which type of onset listeners heard (Figure 6). Paired t-test analysis revealed that there was a significant difference in mean amplitude across conditions of congruent ( $M = -.49$ ,  $SD = 1.62$ ) and incongruent coarticulation ( $M = -2.26$ ,  $SD = 1.94$ ) for fricative onsets,  $t(8) = 3.68$ ,  $p < .01$ ,  $r = .68$ . No significant difference was found for stops across congruency levels,  $t(8) = 2.23$ ,  $p > .05$ . A second interaction was also obtained in the temporal window of the PMN, revealing a significant effect of congruency of coarticulation and vowel height (high versus low – Figure 8),  $F(1, 8) = 8.28$ ,  $p < .05$ ,  $\eta^2 = .51$ . Further analysis revealed significant differences between congruent ( $M = .103$ ,  $SD = 1.79$ ) and incongruent subphonemic cues ( $M = -1.82$ ,  $SD = 2.01$ ) in spoken words containing low vowels,  $t(8) = 3.74$ ,  $p < .01$ ,  $r = .68$ . There was no significant difference in PMN responses for congruent/incongruent cues in words containing high vowels,  $t(8) = 1.91$ ,  $p > .05$  (see Figure 9).

**Figure 9.** Mean amplitude ( $\mu\text{V}$ ) of PMN across levels of congruency (congruent versus incongruent) and vowel height (high versus low) for fricative and stop onsets.

*Fricatives*



*Stops*

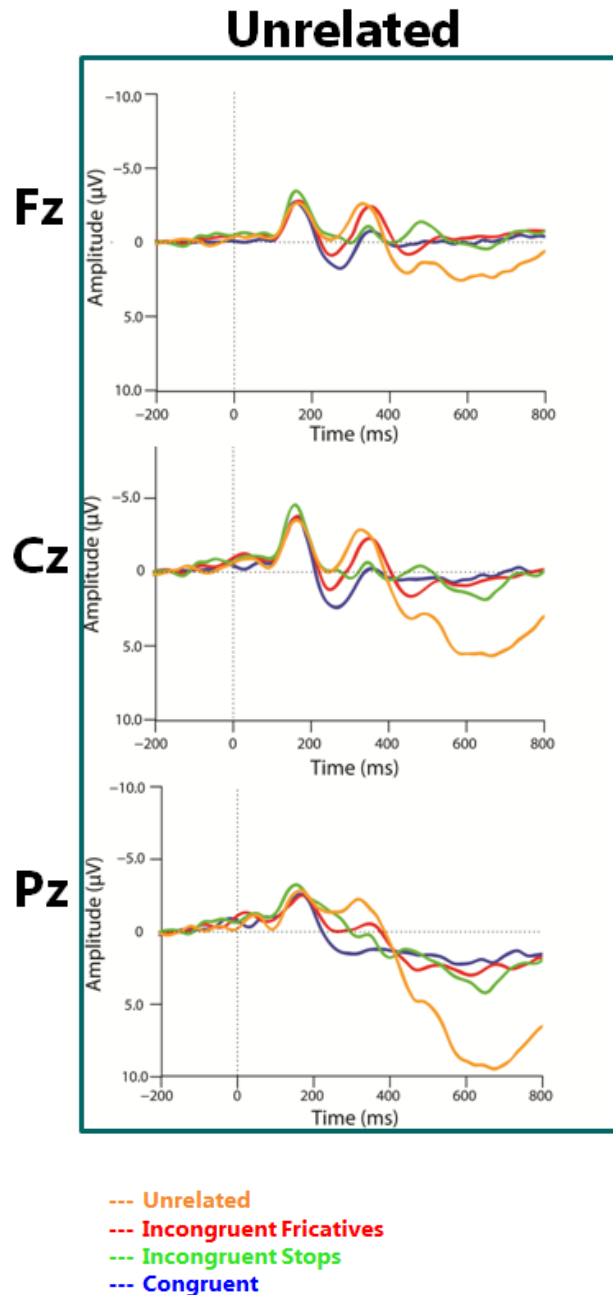


Lastly, a repeated-measures analysis of variance (ANOVA) was conducted as a benchmark for comparing the PMN against the unrelated initial phoneme. Previous analysis suggested that coarticulatory information was processed differently across fricative and stops onsets. Therefore, using conservative degrees of freedom (Greenhouse-Geisser, 1958), the repeated-measures ANOVA was constructed with 4 levels of conditions (incongruent coarticulation for fricatives, incongruent coarticulation

for stops, congruent coarticulation and the unrelated condition) and 3 levels of electrode sites (FZ, CZ and PZ). First, an interaction between condition type and electrode site approached significance using Greenhouse-Geisser correction,  $F(6,48) = 3.04$ ,  $p = 0.53$ ,  $\eta^2 = .28$ . Second, Mauchley's test indicated that the assumption of sphericity had not been violated ( $\chi^2(5) = 4.26$ ,  $p > .05$ ) and a main effect of condition was observed,  $F(3, 24) = 11.20$ ,  $p < .01$ ,  $\eta^2 = .58$  (Fig.10). Paired t-tests revealed a significant difference between congruent coarticulation ( $M = .07$ ,  $SD = 1.66$ ) and incongruent coarticulation for fricatives ( $M = -1.76$ ,  $SD = 1.76$ ) ( $t(8) = 4.09$ ,  $p < .01$ ,  $r = .69$ ), congruent coarticulation ( $M = .07$ ,  $SD = 1.66$ ) and unrelated ( $M = -2.82$ ,  $SD = 2.22$ ) ( $t(8) = 6.98$ ,  $p < .001$ ,  $r = .83$ ), and incongruent coarticulation for stops ( $M = -.79$ ,  $SD = 2.12$ ) and unrelated ( $M = -2.82$ ,  $SD = 2.22$ ) ( $t(8) = 4.06$ ,  $p < .01$ ,  $r = .76$ ). No significant differences were observed across congruent coarticulation and incongruent coarticulation for stops ( $t(8) = 1.93$ ,  $p > .05$ ), incongruent coarticulation for fricatives and incongruent coarticulation for stops ( $t(8) = -1.57$ ,  $p > .05$ ), or incongruent coarticulation for fricatives and the unrelated ( $t(8) = 1.56$ ,  $p > .05$ ). Results are summarized in Table 5.



**Figure 10.** Grand averaged ERP waveforms from all 9 participants at three electrode sites (Fz, Cz, Pz). Grand averages display electrophysiological responses across four condition types; congruent coarticulation (blue), incongruent coarticulation with fricative onsets (red), incongruent coarticulation with stop onsets (green) and unrelated incongruent initial phoneme (yellow).



**Table 4.** Summary of ANOVA Results\*

Component	Effect/Interaction	<i>df</i>	<i>F</i>	<i>p</i>	$\eta^2$
N100	Onset	1,8	6.49	<.034	.45
P200	Congruency	1,8	17.79	<.003	.69
	Congruency*Height	1,8	18.73	<.003	.70
	Site*Backness	2, 16	4.07	<.040	.34
PMN	Congruency	1,8	13.09	<.007	.62
	Congruency*Onset	1,8	5.91	<.041	.43
	Congruency*Height	1,8	8.28	<.021	.51
	Site*Congruency*Onset*Height*Backness	2, 16	7.40	<.005	.48

\*NB: Only significant main effects and interactions have been included in the summary table.

**Table 5.** Summary of ANOVA Results – Benchmark PMNs\*

Component	Effect/Interaction	<i>df</i>	<i>F</i>	<i>p</i>	$\eta^2$
PMN	Condition	3,24	11.20	<.001	.58

\*NB: Only significant main effects and interactions have been included in the summary table.

## **5. Discussion**

In this experiment, the temporal dynamics of neurophysiological responses to matching and mismatching anticipatory coarticulatory cues were examined across systematically controlled phonological conditions. Electrophysiological responses were recorded to spoken words containing systematically isolated onset types (stop and fricative) and vowel place of articulation (height and backness). As reported in Table 4 and 5, main effects and interactions were obtained across three components of interest; the N100, the P200 and the PMN. First, neural responses revealed significantly larger PMN amplitudes to subtle coarticulatory information that violated phonological expectations. These results are consistent with the findings reported in Archibald and Joanisse (2011). Unlike Archibald and Joanisse (2011), neurophysiological differences were also observed across systematically varying phonological conditions (onset and vowel type). The results of the present experiment reflect that the temporal and physical properties of particular sound classes produce more salient coarticulatory effects compared to others. The implications of these findings will now be discussed.

### ***5.1 N100 and P200***

N100 responses to word onsets were examined as a function of the initial phoneme (fricative versus stop). The observed negativity was consistent with distributional and temporal properties known to characterize the N100. Stop onsets evoked larger N100 responses than fricatives along the midline. These results suggest that differences in physical characteristics of sound classes modulate the N100. Larger

N100 amplitudes to stop onsets are interpreted as a sensory response associated with the abrupt onset (e.g. burst of acoustic energy) produced during a stop. Smaller N100 amplitudes may be linked to the gradual onset (e.g. greater duration in rise time) generated during the production of fricatives. Variations in N100 amplitudes associated with the onset type, is consistent with the suggestion that the N100 is an “exogenous” potential sensitive to the features of stimuli (e.g. intensity of onset). No significant differences were observed for congruency of coarticulation, suggesting that subphonemic variations were not simply processed as acoustic noise.

Nevertheless, there was a significant main effect of congruency of coarticulation in the temporal window of the P200, which may reflect the fact that listeners perceive subtle acoustic differences to some extent. Alternatively, this main effect may also reflect the “mechanical” influence of the adjacent PMN component. In fact, larger PMNs to incongruent (than congruent) coarticulation could have reduced the amplitude of the preceding P200 in instances of subphonemic mismatches. Similarly, a significant interaction of vowel height and congruency of coarticulation for the P200 component suggests that listeners detected acoustic subphonemic variations. Given these differences across congruency of coarticulation, the present results suggest that acoustic variations are more salient to listeners when presented with low vowels. This is likely due to the fact that information for subphonemic violations may be more prominent among low vowels due to the open configuration of the vocal tract, which could have maximized acoustic properties associated with coarticulatory differences

Visual inspection of the grand averaged waveforms shows an absence of the P200 in instances of stop onsets containing incongruent coarticulatory violations (Fig. 6). This result could indicate that the brain perceives subtle coarticulatory violations more rapidly in stops than in fricatives, and that a neural component independent of the PMN reveals this process. These data for the P200 might reasonably be interpreted as supporting arguments that coarticulatory differences may be less salient (but perceived more rapidly) in spoken words beginning with plosives, due to the short formant transition duration extending from the initial stop consonant to the subsequent vowel (Kent & Read, 1992). Previously, Newman, Connolly, Service & McIvor (2003) demonstrated latency and amplitude differences across the N100 and the P200 components, which they interpreted to be the result of variation of acoustic properties that were uncontrolled in their study. Therefore, investigating sensory responses (e.g. N100 – P200) associated with systematic manipulations of acoustic properties may reveal additional effects across other sound classes in future experiments (as suggested by Newman et al. (2003)).

## 5.2 *PMN*

**5.2.1 *Coarticulatory congruency:*** In confirmation of previous findings, ERP data within the present study revealed the neurophysiological detection of fine-grained coarticulatory variations (Fig. 5 and 7). Despite any observed differences in behavioural responses, incongruent coarticulatory cues elicited larger PMN amplitudes than congruent. Visual inspection of the waveforms showed that the increased negativity was most prominent in the fronto-central electrode sites, a distribution commonly seen for the

PMN. Previous literature has suggested that larger PMN amplitudes reflect greater processing or termination of phonological matching, as a result of inconsistencies between expected and perceived phonemes (Connolly & Phillips, 1994, Newman, Connolly, Service & McIvor, 2003). As a result, larger PMNs to incongruent coarticulatory mismatches (despite a congruent initial phoneme) provide additional evidence that listeners are sensitive to subtle violations produced by subphonemic processes. In fact, even across systematically varying phonological conditions, the results of this study are comparable to the coarticulatory effects observed in Archibald and Joanisse (2011). Despite the fact that acoustically evoked potentials (EPs) could have been elicited by the spliced stimuli, variations in coarticulatory congruency did not evoke these types of sensory responses. Therefore, modulations of the PMN to coarticulatory congruency are interpreted in the phonological domain, and provide strong evidence that subphonemic information is phonologically processed by listeners in real time. The present neural data support Archibald and Joanisse (2011) and shows that the PMN component detects phonological processing of subtle coarticulatory variations. Together, these findings suggest that the previous view of the PMN was too narrow (e.g. *between* category variations), and that the phonological stage of speech recognition must be expanded to consider subphonemic processing. In line with previous literature (Newman, Connolly, Service & McIvor, 2003), larger PMN amplitudes were also evoked in cases in which coarticulatory cues were congruent but an initial consonant violated expectation (i.e., unrelated condition; see Fig. 10). While previous studies have demonstrated that the PMN is sensitive to phoneme level miscues (Newman, Connolly, Service & McIvor,

2003, Connolly & Phillips, 1994), the present findings offer additional support for Archibald and Joanisse's (2011) findings that fine-grained coarticulatory variations also evoke modulations of the PMN. These results may demonstrate that both *within* (e.g. subphonemic) and *between*-category (e.g. phonemic) mismatches are processed similarly. In this view, it is unknown whether the brain discriminates between variations in phoneme and subphonemic information. One interpretation is that the PMN is an all-or-none response to phoneme level violations, but that subphonemic miscues are detected to differing extents (e.g. variations in PMN) depending on phonological context. This effect may be explained by the gradient (rather than discrete) nature of subphonemic modifications. This perspective draws questions about whether the PMN in response to subphonemic anomalies is the same PMN elicited to phoneme level mismatches. Alternatively, these data could reflect strict subphonemic processing. In fact, Archibald and Joanisse (2011) concluded that the neural data they observed offered little evidence for a phonemic level of representation. This study, among others, lends credence to the argument that we need to reevaluate the notion of the phoneme from the perspective of perception. Future research in this area is essential in teasing apart levels of processing. Despite uncertainty concerning the smallest unit leading to the distinction of a phoneme, the present study supports the need to account for coarticulatory information in theoretical models of speech perception (Archibald and Joanisse, 2011). As first suggested by Archibald and Joanisse (2011), these results emphasize the need to reevaluate the prelexical stage of speech recognition, by considering "gestures and features" associated with subphonemic variability during phonological processing. Thus,

coarticulatory processes generate variations in articulatory movements (gestures) and the qualities or characteristics (features) that define subphonemic cues. As a result, neural data obtained in the present study highlights the importance of additionally exploring these fine-grained elements in future studies, in order to further refine speech recognition models.

Visual comparison of benchmark PMNs (Fig. 10) revealed a slightly earlier PMN in the unrelated (incongruent initial phoneme) than in conditions of coarticulatory mismatches (incongruent subphonemic cues). These results are consistent with the neural data reported by Archibald and Joanisse (2011), in which coarticulatory violations in word onsets evoked early modulations of the PMN, while phoneme level violations involving the subsequent vowel were elicited later. These temporal differences suggest that detection of acoustic-phonetic differences is linked to the presentation of these cues in the acoustic signal (Archibald & Joanisse, 2011), a result which is well captured by ERP methodology. Findings from the present study and from Archibald and Joanisse (2011), contribute to the emerging evidence that suggests that coarticulatory cues are processed in real-time. Likewise, the present results also offer additional evidence supporting Gow and McMurray's (2007) continuous cue integration hypothesis, which proposes that subcategorical information facilitates speech perception. Observed neurophysiological data offers strong evidence for the preservation and continuous integration of coarticulatory cues during prelexical processing. Therefore, the interpretations of these findings align with Archibald and Joanisse (2011), suggesting that



subphonemic cues assist the perceptual system in rapidly disambiguating targets during spoken word recognition.

**5.2.2. Onset type:** In contrast with Archibald and Joanisse (2011), stimuli in the present study were not restricted to the use of imageable words so that stricter criteria could be implemented to create more systematically varying phonological conditions. Differences in manner of articulation (fricative versus stop) across controlled manipulations of acoustic properties elicited modulations of the PMN. First, neurophysiological differences in response to the nature of the onset were observed. A significantly larger PMN was elicited when listeners heard incongruent (versus congruent) coarticulatory information in spoken targets consisting of fricative onsets. Stimuli consisting of initial stop consonants did not exhibit this same effect across congruency of coarticulation. This result suggests that fricatives produce more salient coarticulatory information regarding upcoming vowels than do stops, a result that may be explained by temporal and physical differences across these sound classes. Subphonemic cues may be more easily detectable within fricatives onsets due to the continuous property of this sound class. In fact, continuants (phonemes that are articulated with prolonged airflow) are usually followed by vowels comprising greater duration and this may have maximized coarticulatory effects between the initial consonant and the vowel (Archibald & Joanisse, 2011). As a result of these acoustic properties, listeners may have been provided with more prominent coarticulatory information for fricatives onsets than for stops. Alternatively, stop (plosive) onsets may have generated less noticeable subphonemic cues due to the closure phase, in which case airflow is momentarily blocked and minimal acoustic energy is

generated (Kent & Read, 1992). Visual inspection of the ERP waveform's morphology (Fig.6), has suggested that the detection of coarticulatory mismatches occurs earlier in the onsets of stops than fricatives (see discussion of P200); a result which has been interpreted to reflect durational differences between these sound classes. As aforementioned, these differences could be interpreted to be the result of shorter formant transitions extending from the initial stop consonant to the vowel, which could have permitted more rapid cue detection. Alternatively, temporal variations may have permitted perceptually greater vowel information (e.g. greater degree in overlap) in fricative onsets than stops due to the frication's longer duration. Therefore, due to these factors listeners may have detected subphonemic differences later in fricatives than in stops, a result which is well illustrated in the present neural data.

**5.2.3. Vowel type:** Similarly, PMN differences were observed across congruency of coarticulation and vowel place of articulation. Previous research (Archibald and Joanisse, 2011) did not control for the influence that the coda may have had on the preceding vowel. Thus, there was concern that differences in codas within stimulus pairs could have had differing effects on the vowel. As a result, the present study ensured that codas remained consistent within minimal sets. Similarly, an uneven distribution of vowels generated uncertainty about whether vocalic context may have differing neurophysiological effects. The present investigation employed a strict manipulation of the four corners of the vowel space, for maximal contrast within minimal sets. Observed neural data suggest that the detection of coarticulatory cues were distinct in the

dimension of high-low vowels, but generated comparable perceptual effects for front-and-backness. In fact, the PMN was significantly larger when listeners heard spoken words containing low vowel coarticulatory information that was incongruent (versus congruent). This effect was elicited irrespective of the initial consonant type (stop or fricative). No significant differences were observed across levels of vowel backness, suggesting that listeners processed coarticulatory information similarly across front and back vowels.

Overall, the present findings indicate that mismatches involving low vowel coarticulatory information were more readily detected. This result may be explained by several factors. First, spatial effects (e.g. position of articulators) associated with vowel place of articulation may contribute to the perceptual strength of coarticulatory effects in the acoustic signal. During the production of low vowels the tongue is positioned at a maximal distance from the roof of the mouth (e.g. open), whereas for high vowel it is in an elevated position (e.g., close). Articulation of low vowels may have permitted maximal shaping of the oral tract, producing greater anticipatory coarticulatory information. For example, it may be the case that an open configuration permits low vowels (as opposed to high vowels) to be more coarticulated with a previous consonant context. Second, transition from the initial consonant to a low vowel resulted in larger gesture movements, while transition from a consonant to a high vowel is smaller (Parush & Ostry, 1983, 1993). Thus, coarticulatory information may have also been more noticeable among low (versus high) vowels due to a greater magnitude of articulatory motion from the consonant to the vowel. These results support pulse-echoed ultrasound

research (Parush & Ostry, 1983, 1993), which reported smaller downward displacement from a stop to a high vowel /u/ and larger displacements towards a low vowel /a/. Parush and Ostry (1993) concluded that vowels were articulated differently across high and low vowels, but that back and front vowels followed similar “trajectories”.

## 6. Conclusion

ERP provide valuable source of information to study language processing. The present study was conducted in order to investigate the temporal integration of coarticulatory cues in speech recognition, and to obtain knowledge about whether these effects vary across phonological contexts. Consistent with Archibald and Joanisse (2011), the results indicate that listeners detect subtle coarticulatory cues in real time, as a strategy that is advantageous to spoken word recognition. These findings contribute to a range of studies that suggest that the perceptual system continuously integrates anticipatory information provided by phonetic variability, in order to predict information about the upcoming phonemes (Marin & Bunnell, 1982). However, the extent to which coarticulatory information is detected by listeners appears to vary according to phonological context.

Neurophysiological differences across sound classes have demonstrated that coarticulatory effects are perceptually more salient in some phonological environments and subtler in others, a result interpreted to reflect differences in temporal and physical properties. In particular, this work demonstrates that the brain responds differently to incongruent coarticulatory information among systematically controlled phonological conditions. Primarily, differences in onset type (fricative versus stop) and vowel height (high versus low) were observed, and are interpreted to be the result of the variations in subphonemic and phonological processing that underlie these sound classes.

Although it is evident that the brain responds to subtle coarticulatory cues, the behavioural results in this experiment make question whether listeners were fully aware of these differences. The task only required subjects to make a lexical (as opposed to phonological) decision between visual and auditory stimuli, and provided no insights into whether incorrect coarticulatory cues were fully identified. In fact, listeners may have been completely unaware of coarticulatory violations, despite the brain's clear ability to detect these differences. The present research also raises questions about whether the PMN is an all-or-none response (as previously suggested by Newman, Connolly, Service & McIvor, 2003). Newman, Connolly, Service & McIvor (2003) demonstrated that degree of phonological overlap evoked PMNs of similar amplitudes. However, given differences in amplitudes in response to subphonemic cues across varying phonological conditions, the PMN as a graded response should be re-evaluated. Furthermore, it would be interesting to explore additional neural differences found within the present study. For example, a late positive component (LPC) appears to be more prominent in stop onsets than in fricatives, which may represent increase working memory demands. Further interesting differences include an apparent negativity at about 400 ms (likely not the classic N400) in the case of incongruent coarticulation for stops, which may be related to “continued phonological processing” (discussed in Connolly, Service, D’Arcy, Kujala and Alho, 2000) thought to reflect “phonological working memory”.

In the future, it would be interesting to extend this line of research to additional sound classes, in order to further examine ERP responses to different acoustic properties and phonological contexts. Furthermore, of particular interest is the investigation of

coarticulatory processing in L2 (second language) learners. According to Beddor and her colleagues (2002), the perception of coarticulatory information may be dependent on “linguistic experience”, and sensitivity to coarticulatory cues may be influenced by “coarticulatory patterns” from a speaker’s native language (as discussed in Tjaden & Sussman, 2006). The current research may also be profitable to extend to clinical populations, such as individuals with dyslexia and specific language impairment. The prospective experiment is significant to offering real-time information advantageous to understanding fine-grained phonological processing in these populations, and to obtain information valuable to therapies.

## References

- Archibald, L. M., & Joanisse, M. F. (2011). Electrophysiological responses to coarticulatory and word level miscues. *Journal of Experimental Psychology: Human Perception and Performance*, 37(4), 1275-1291.
- Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30, 591-627.
- Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties to terminal words in sentences. *Brain and Language*, 43 1-18.
- Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6, 256-266.
- Connolly, J. F., Service, E., D'Arcy, R. C. N., Kujala, A., & Kimmo, A. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *Cognitive Neuroscience and Neuropsychology*, 12(2), 237-243.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16(5/6), 507-534.
- Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2008). Investigating the time course of spoken recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, 1-14.
- Flemming, E. (1997). Phonetic detail in phonology: Towards a unified account of assimilation and coarticulation. In K. Suzuki, & D. Elzinga (Eds.), *Proceedings of the 1995 southwestern workshop in optimality theory (SWOT)*. University of Arizona.
- Gaskell, M. G. (2001). Phonological variation and its consequences for the word recognition system. *Language and Cognitive Processes*, 16, 723-729.
- Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45, 133-159.



- Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. In J. S. Cole, & J. Hualdo (Eds.), *Papers in laboratory phonology 9* (9th ed., pp. 173-200)
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*, 373-405.
- Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesh, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47-B57.
- Hardcastle, W. J., & Tjaden, K. (2009). Coarticulation in speech impairment. In M. J. Ball, M. R. Perkins, N. Miller & S. Howard (Eds.), *Handbook of clinical linguistics* (pp. 506-524). Oxford, UK: Blackwell.
- Kent, R., & Read, C. (1992). The acoustic characteristics of vowel and diphthongs. *The acoustic analysis of speech* (pp. 87-104). San Diego: Singular Publishing Group.
- Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In W. J. Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 7-30). Cambridge: University Press.
- Liberman, A.M., Safford Harris, K., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358-368.
- Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *88*, 1-71.
- Martin, B. A., Tremblay, K. L., & Stapells, D. R. (2007). Principles and applications of cortical auditory evoked potentials. In J. Eggermont, M. Don & R. Burkard (Eds.), *Auditory evoked potentials: Basic principles and clinical applications* (pp. 482-507). Philadelphia: Lippincott, Williams & Wilkins.
- Martin, J. G., & Bunnell, T. (1980). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America*, *69*(2), 559-587.
- Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 473-488.

- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.
- McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064-1071.
- McQueen, J.M & Cutler, A. (1999). Cognitive Processes in Speech Perception. In J. Hardcastle, & J. Laver (Eds.), *The Handbook of Phonetic Sciences*. (pp.489 – 520). Oxford: Blackwell. McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decisions making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(5), 1363-1389.
- Newman, R. L., Connolly, J. F., Service, E., & McIvor, K. (2003). Influence of phonological expectation during a phoneme deletion task: Evidence from event related brain potentials. *Psychophysiology*, *40*, 640-647.
- Newman, R.L. & Connolly, .F. (2009). Electrophysiological markers of pre-lexical speech processing: evidence for bottom-up and top-down effects on spoken word processing. *Biological Psychology*, *80*(1), 114-121.
- Nittrouer, S., & Whalen, D. H. (1990). The perceptual effects of child-adult differences in fricative-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research*, *102*, 177-193.
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299-370.
- Parush, A., Ostry, D.J., Munhall, K.G (1983). A kinematic study of lingual coarticulation in VCV sequences. *Journal of Acoustical Society of America*, *74*(4), 1115-1125
- Parush, A., & Ostry, D. J. (1983). Similarities in the control of the speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 622-636.
- Parush, A., & Ostry, D. J. (1993). Lower pharyngeal wall coarticulation in VCV syllables. *Journal of the Acoustical Society of America*, *94*, 715-722.
- Raphael, L. J. (2008). Acoustic cues to the perception of segmental phonemes. In D. Pisoni, & R. Remez (Eds.), *The handbook of speech perception* (pp. 182-206). Oxford: Blackwell Publishing.

- Recasens, D. (2002). An EMA study of VCV coarticulatory direction. *Journal of Acoustical Society of America*, *111*(6), 2828-2841.
- Steinhauer, K., & Connolly, J. F. (2008). Event-related potentials in the study of language. In B. Stemmer, & H. Whitaker (Eds.), *Handbook of the cognitive neuroscience of language* (pp. 91-104). New York: Elsevier.
- Smits, R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1145-1162.
- Tjaden, K., & Sussman, J. (2006). Perception of coarticulatory information in normal speech and dysarthria. *Journal of Speech, Language, and Hearing Research*, *49*, 888-902.
- Tobin, S. J., Whan Cho, P., Jennett, P. M., & Magnuson, J. S. (2010). Effects of anticipatory coarticulation on lexical access. *Proceedings in COGSCI 2010*, 2200-2205.
- Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, *22*, 79-90.
- Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics*, *50*, 351-360.
- Yeni-Komshian, G., & Soli, S.D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*. *70*(4), 966-975.

## Appendix

### *Appendix 1: Summary of conditions and trial counts*

<i>Condition</i>	<i>Trial Count</i>	<i>Total Count</i>
Coarticulatory Match, Phonemic Match, Stop, High Front (Congruent)	10	88
Coarticulatory Match, Phonemic Match, Stop, High Back (Congruent)	9	
Coarticulatory Match, Phonemic Match, Stop, Low Front (Congruent)	9	
Coarticulatory Match, Phonemic Match, Stop, Low Back (Congruent)	10	
Coarticulatory Match, Phonemic Match, Fricative, High Front (Congruent)	14	
Coarticulatory Match, Phonemic Match, Fricative, High Back (Congruent)	12	
Coarticulatory Match, Phonemic Match, Fricative, Low Front (Congruent)	13	
Coarticulatory Match, Phonemic Match, Fricative, Low Back (Congruent)	11	
Coarticulatory Mismatch, Phonemic Match, Stop, High Front (Incongruent)	10	88
Coarticulatory Mismatch, Phonemic Match, Stop, High Back (Incongruent)	9	
Coarticulatory Mismatch, Phonemic Match, Stop, Low Front (Incongruent)	9	
Coarticulatory Mismatch, Phonemic Match, Stop, Low Back (Incongruent)	10	
Coarticulatory Mismatch, Phonemic Match, Fricative, High Front (Incongruent)	14	
Coarticulatory Mismatch, Phonemic Match, Fricative, High Back (Incongruent)	12	
Coarticulatory Mismatch, Phonemic Match, Fricative, Low Front (Incongruent)	13	
Coarticulatory Mismatch, Phonemic Match, Fricative, Low Back (Incongruent)	11	
Coarticulatory Match, Phonemic Mismatch (Unrelated)	42	42

*Appendix 2: List of subconditions and examples of stimuli*

<i>Condition</i>	<i>Visual</i>	<i>Auditory</i>	<i>Splicing?</i>
Coarticulatory Match, Phonemic Match, Stop, High Front (Congruent)	peal	[p <sup>l</sup> ]eal	None
Coarticulatory Mismatch, Phonemic Match, Stop, High Front (Incongruent)	pool	[p <sup>l</sup> ]ool	/p/ from <i>peal</i>
Coarticulatory Match, Phonemic Match, Stop, High Back (Congruent)	pool	[p <sup>w</sup> ]ool	None
Coarticulatory Mismatch, Phonemic Match, Stop, High Back (Incongruent)	peal	[p <sup>w</sup> ]eal	/p/ from <i>pool</i>
Coarticulatory Match, Phonemic Match, Stop, Low Front (Congruent)	pal	[p <sup>a</sup> ]al	None
Coarticulatory Mismatch, Phonemic Match, Stop, Low Front (Incongruent)	pool	[p <sup>a</sup> ]ool	/p/ from <i>pal</i>
Coarticulatory Match, Phonemic Match, Stop, Low Back (Congruent)	paul	[p <sup>a</sup> ]aul	None
Coarticulatory Mismatch, Phonemic Match, Stop, Low Back (Incongruent)	peal	[p <sup>a</sup> ]eal	/p/ from <i>paul</i>
Coarticulatory Match, Phonemic Match, Fricative, High Front (Congruent)	seat	[s <sup>l</sup> ]eat	None
Coarticulatory Mismatch, Phonemic Match, Fricative, High Front (Incongruent)	suit	[s <sup>l</sup> ]uit	/s/ from <i>seat</i>
Coarticulatory Match, Phonemic Match, Fricative, High Back (Congruent)	suit	[s <sup>w</sup> ]uit	None
Coarticulatory Mismatch, Phonemic Match, Fricative, High Back (Incongruent)	seat	[s <sup>w</sup> ]eat	/s/ from <i>suit</i>
Coarticulatory Match, Phonemic Match, Fricative, Low Front (Congruent)	sat	[s <sup>a</sup> ]at	None
Coarticulatory Mismatch, Phonemic Match, Fricative, Low Front (Incongruent)	suit	[s <sup>a</sup> ]uit	/s/ from <i>sat</i>
Coarticulatory Match, Phonemic Match, Fricative, Low Back (Congruent)	sought	[s <sup>a</sup> ]ought	None
Coarticulatory Mismatch, Phonemic Match, Fricative, Low Back (Incongruent)	seat	[s <sup>a</sup> ]eat	/s/ from <i>sought</i>

*Appendix 3: Participant demographic information and screening form*

Participant Demographic Information & Screening Form:  
**A Study of Speech Processing in the Brain**

Protocol: \_\_\_\_\_ Subject #: \_\_\_\_\_

**This information is strictly confidential and will not be shared with anyone.**

First name: \_\_\_\_\_ Last name: \_\_\_\_\_

Age: \_\_\_\_\_ Year of education: \_\_\_\_\_

First language: \_\_\_\_\_

Do you speak any other language(s): Yes No

If yes, please specify: \_\_\_\_\_

If yes, please specify when first learned: \_\_\_\_\_

Do your parents speak any other languages: Yes No

If yes, please specify: \_\_\_\_\_

Handedness: Right Left

Is your hearing normal? Yes No

Have you ever had any neurological or psychological problems? Yes No

Are you presently taking any medication? Yes No

If yes, which one(s): \_\_\_\_\_

Have you recently taken any medication? Yes No

Please rate your current state of alertness by circling a number below: (1 = extremely alert & 5 = extremely tired)

- 1 2 3 4 5 +

How many hours did you sleep last night? \_\_\_\_\_

*Appendix 4: Letter of information*

Letter of Information

**A Study of Speech Perception in the Brain**

Principle Investigator: Jessica Arbour  
McMaster Cognitive Science of Language  
McMaster University, Hamilton, Ontario, Canada  
spoken.language.lab@mcmaster.ca

**Purpose of the study**

The aim of this study is to examine brain responses while you listen to particular auditory stimuli. We will be recording your brain activity through event-related potential (ERP) responses by means of an electroencephalogram (EEG) while you listen to spoken words.

**What is an electroencephalogram (EEG)?**

An electroencephalogram (EEG) is used to collect event related potential (ERP) data whose source comes from within the human brain. ERPs are a measure of the brain's activation through the skull or scalp in response to perception and thought. Thus, utilizing the information ERPs supply, offers great insight into spoken word perception research.

**What will happen during the study?**

To guarantee quality recording of the EEG, we will start by asking you to fill out a questionnaire consisting of personal information. This is an essential part of the study because certain neurological conditions, hearing problems, head injuries as well as particular medications may interfere with the recordings that we are interested in examining. It is important that we are aware of these conditions because such factors could prevent you from participating in this study.

EEG recordings will be attained through the use of a traditional cap that will be placed on your head and secured by means of a chin-strap. This cap contains many sensors that are able to record your brain activity. In order to obtain good recordings, we must use an electrical conductive jelly-like substance called “electrolyte”, which we will insert into each of the electrode sites. Finally, we will insert a number of electrodes into each of the corresponding sites in the cap.

During this study, a written word will appear on the computer screen followed by a spoken-word presented to you through a set of headphones. You are to listen to the spoken-word and decide whether or not it matches the written word. As a result, you will be asked to make a decision (yes or no), by pressing a set of specific buttons on the keyboard. Before the task begins, you will be given detailed instructions on how you should proceed. The experiment should take approximately 1.5 hours to complete.

Please remember that there are no ‘right’ or ‘wrong’ answers within this study. Do not be concerned or upset if you are having difficulty with the task. Please also note that it will be

possible for you to communicate with the experimenter by intercom at any point in time throughout the experiment, if necessary.

**Potential Harms, Risks or Discomforts:**

There are no known harms or risks associated with the usage of the EEG technology we are employing. The present study however, does require you to sit still and remain focused for a considerable period of time. It also requires wearing a cap that has been well-secured to the head. There is a possibility that these factors could create minor discomfort for some participants.

**Potential Benefits:**

Beyond the financial compensation you have been offered, there will be no direct benefits for you. Your participation however, will help to contribute to further understanding the brain and how it processes spoken words. We hope that you find satisfaction in this knowledge.

**Compensation:**

Financial compensation is provided for participating within this experiment. You will be paid at a rate of \$10 per hour.

**Confidentiality:**

All personal information we attain will remain highly confidential. Within this study, you will be identified as a participant number. Any personal information you provide will only be accessible to the principal and co-investigators. This information will be strictly secured in a locked cabinet or on a password protected computer. No publication or discussion regarding this study will contain any personal and/or identifying information.

**Participation:**

Participation in this study is completely voluntary. You may choose to withdraw at any time without any consequences. If you withdraw, you will not need to give any explanation for the decision and your data will be destroyed immediately. You will also receive financial compensation for the amount of time you have spent in the laboratory.

**Information about the study and your participation as a study subject:**

You may obtain additional information about the results of the study by contacting the principle investigator. Contact information is provided on the first page.

This study has been approved by the McMaster Research Ethics Board. If you have any additional concerns or questions, you may contact:

McMaster Research Ethics Board Secretariat, Office of Research Services  
(905) 525-9140 Ext. 23142, [ethicsoffice@mcmaster.ca](mailto:ethicsoffice@mcmaster.ca)



*Appendix 5: Participant debriefing form*

Participant Debriefing Form  
**A Study of Speech Perception in the Brain**

Thank you for agreeing to participate in this study! The general purpose of this research is to gather event-related potential (ERP) data to investigate speech processing in the human brain. The human voice has evolved as a vehicle for conveying many different types of information. Human listeners have developed the ability to detect very small and subtle voice quality changes, and to interpret their function. In spite of the incredible abilities of the human voice and ear, our understanding of these abilities lags far behind. We are therefore interested in investigating how different sounds (phonetic) properties are utilized as cues in which human listeners rely on in order to decode words they are processing.

We invited participants who were typically-developed, native speakers of English with no history of medical and or neurological conditions. In this study, you were asked to take part in an odd-ball paradigm which involved looking at the spelling on a computer monitor, and pressing a button to identify the sounds you hear. The electroencephalograph (EEG) was employed during this task in order to measure the brain responses to perception and thought, while processing various words. The results from this study will provide insight on how humans process various sound properties when retrieving words. We hope that empirical data from this experiment will contribute to unanswered questions about human word recognition.

You may obtain additional information about the results of the study by contacting the principle investigator at [spoken.language.lab@mcmaster.ca](mailto:spoken.language.lab@mcmaster.ca). If you have any additional concerns or questions, you may contact the McMaster Research Ethics Board Secretariat at (905) 525-9140 Ext. 23142.

Thank you

*Appendix 6: Consent form*

Consent Form

**A Study of Speech Perception in the Brain**

I have read and understood the information letter about a study being conducted by Jessica Arbour of McMaster University. I have had the opportunity to ask questions about my involvement in this study and to receive any additional details I wanted to know about the study. I understand that I may withdraw from the study at any time.

---

Name of Participant

---

Signature

---

Date