Acoustic Cues Used by Learners of English

by

Danica Reid

B.A., Simon Fraser University, 2015

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Linguistics
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Fall 2018

Approval

Name: **Danica Reid**

Degree: **Master of Arts**

Title: **Acoustic Cues Used by Learners of English**

Examining Committee: **Chair:** **Chung-hye Han**
Professor

Ashley Farris-Trimble
Senior Supervisor
Assistant Professor

Henny Yeung
Supervisor
Assistant Professor

Molly Babel
External Examiner
Associate Professor
Department of Linguistics
The University of British Columbia

Date Defended/Approved: December 3, 2018

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Second language learners must acquire the ability to use word boundary cues to segment continuous speech into meaningful words. Previous studies have used two types of *s*+stop clusters to test second language English speakers on their ability to segment fluent English speech: cross-boundary clusters (*thi**s** **t**able*) where allophonic aspiration is present and word-initial clusters (*this **st**able*) where allophonic aspiration is absent. These studies suggested that first language segmentation strategies influence second language segmentation. The goal of this study was to test real-time processing of these cluster types by second language learners from one language where cue adaptation was possible (Mandarin Chinese) and one where a new cue would have to be learned (French). Results did not support the idea that first language segmentation strategies influence second language segmentation, but found that both language groups had high accuracy of identification despite showing uncertainty in real-time processing.

Keywords: English as a second language; eye-tracking; online processing; speech segmentation; word boundary cues

Acknowledgements

First and foremost I would like to thank my thesis supervisor Ashley for picking me out of the first Phonology class she taught at SFU and pulling me into the world of research. Thank you for all of your help and guidance over the past few years. You have taught me so much and for that I will be forever grateful.

To the members of the Phonological Processing Lab, both past and present, I would like to thank every single one of you for helping me make our windowless room into the warm and welcoming place it is today. I don’t know where I would be without your continued help and support.

Finally, I wouldn’t have gotten through this without the love and support of my family and friends. Thank you for listening to my complaints, for forcing me out of my house, for the endless words of encouragements, and the kindness you have given me over the past few years. You all mean the world to me.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Learners of a second language must acquire the ability to use acoustic and phonotactic information to segment long streams of continuous speech into meaningful words. In the process of learning to identify word boundaries in their second or additional language, learners have two options: adapt a strategy from their native language into their second language, or learn a new strategy based on the input from their second language. Regardless of the strategy that a learner uses, they will eventually need to be able to identify a combination of word boundary cues in order to process the speech input. The word boundary cues which are the focus of this paper are fine-grained acoustic cues including phoneme duration and voice onset time (Barry, 1981; Lehiste, 1960; Nakatani & Dukes, 1977). These fine-grained cues are particularly useful in identifying word boundaries in phrases where more than one boundary location is possible. For example, when hearing the phrase [ðɪsteɪbəl], two interpretations are possible depending on the cues which are present: “this table” and “this stable.” For a native English speaker, cues such as vowel and consonant duration and voice onset time (VOT) provide sufficient information to identify the word boundary in an example such as the one given above. A second language learner of English attempting to adapt a segmentation strategy from their first language may not use the relevant cues in the same way that they are used in English which could lead to erroneous segmentation; learning to segment speech by forming strategies based on the English input will take time and may also lead to erroneous segmentation.

The goal of this thesis is to examine the segmentation of word boundaries in two groups of second language English speakers who differ in available first language segmentation strategies. Specifically the project aims to determine how the presence or absence of different cues in a learner’s first language affect the accuracy and time-course of segmentation in English. Second language learners of English are expected to try to use word boundary cues from their first language when segmenting English speech. While this method of adaptation may initially work for learners from some language backgrounds which have similar cues to those found in English, learners from other language backgrounds should quickly identify that this approach is not workable and will begin to learn the new segmentation strategies needed for English. In this thesis

English learners from a language where adaptation of cues would be possible (Mandarin Chinese) and from a language where learning of new cues was necessary (French) were tested in order to answer how language background affects speech segmentation.

1.1. English Segmentation Strategies

There are two main categories of cues that are used by speakers of a language in speech segmentation: lexical and sub-lexical. Lexical cues are knowledge-driven and are comprised of higher-level linguistic processes such as syntactic structure, semantic plausibility, and pragmatics. These cues arise from explicit word knowledge. Sub-lexical cues are signal-driven and provide a listener with lower-level linguistic information from the phonetic output of a sequence. Use of these signal-driven cues reflects the ability of speakers to use implicit linguistic knowledge in speech processing. For example, the use of phonotactic constraints in speech segmentation requires learned co-occurrence restrictions that speakers are not consciously aware of. In the segmentation of normal speech, listeners have been found to weight the lexical cues in the speech signal over the sub-lexical cues except in cases where the signal is degraded in some way (Mattys & Melhorn, 2007; Mattys et al., 2005; Sanders & Neville, 2000, 2003). The design of the present experiment negates any potential influence of lexical information on speech segmentation and these cues will not be discussed here. The following section outlines four main types of sub-lexical cues which are used in the segmentation of English speech.

One of the sub-lexical cues that is useful in speech segmentation is based on English phonotactics. Phonotactics constrain the sounds that can be found adjacently in a language and vary cross-linguistically. These phonotactic constraints combine with the ability to learn which sounds are more likely to occur together to provide listeners with a probability that two sounds will occur adjacent to one another within a word or across a boundary. Phonotactic constraints and the transitional probabilities that result from them have previously been found to be used by infants as young as 8 months old (Aslin, Saffran, & Newport, 1998; Mattys & Jusczyk, 2001; Saffran, Aslin, & Newport, 1996), school aged children (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), and adults (McQueen, 1998; Mirman, Magnuson, Estes, & Dixon, 2008; Saffran, Newport, & Aslin, 1996; Saffran et al., 1997) to identify word boundaries and learn non-words in strings of nonsense speech. In English, both single segments and clusters provide word boundary cues (Berko Gleason & Bernstein Ratner, 2013). For example, the velar nasal [ŋ] and the consonant cluster [lp] are never found word-initially. When processing continuous speech, listeners can use these phonotactic constraints to determine the location of a syllable and possibly word boundary.

A second sublexical word boundary cue used by native speakers of English is the location and predictability of metrical stress. The majority of content words in English begin with a strong syllable and follow the typical trochaic (strong-weak) stress pattern (Cutler & Carter, 1987). While previous research has shown that metrical stress is most useful in regular speech segmentation when the signal is degraded in some way (Mattys & Melhorn, 2007; Mattys et al., 2005), stress has been shown to be an important cue in word spotting and identification (Cutler & Butterfield, 1992; Cutler & Norris, 1988; Mattys, 2000; McQueen, Norris, & Cutler, 1994). For example, when a listener is presented with speech in noise, they can use the onset of a strong syllable to identify a potential word boundary location and help in parsing the speech stream into the intended message.

The acoustic cues involved in speech segmentation can be phonetic, including pitch, formant structure, intensity, and durational properties of speech, or allophonic. The main phonetic cue of interest is duration. Vowel, consonant, and total syllable durations have been found to influence speech segmentation in English (Bion, Benavides-Varela, & Nespor, 2011; Christophe, Gout, Peperkamp, & Morgan, 2003; Fry, 1958; Klatt, 1976; Lehiste, 1960; Nakatani & Schaffer, 1978; Redford & Randall, 2005; Turk & Shattuck-Hufnagel, 2000). In consonants, durational properties such as closure duration, VOT, and frication are used in the identification of segmentation points between two consonants. Durational properties of vowels are typically associated with the stress on a syllable which can provide listeners with information about word boundaries based on the typical English stress patterns discussed above.

Allophonic word boundary cues are the result of highly predictable phonological processes that occur in specific environments created by word boundaries (Church, 1987; Lehiste, 1960). Several acoustic cues are used in the segmentation of speech by native English speakers including glottal stop and/or laryngeal voicing at the onset of a vowel-initial word, and velarization of [l] when found in the coda of a syllable (Nakatani & Dukes, 1977). Allophonic aspiration has been found to be a particularly strong cue in speech segmentation despite being the property of a syllable rather than a word (Wells, 1990). This is perhaps due to the prevalence of strong initial syllables in content words and the typical trochaic stress pattern in English discussed above (Cutler & Carter, 1987). As such, aspirated voiceless stops are often found in word-initial position. For example, in the phonetic stream [thɪstʰeɪbl], the aspiration on the [tʰ] would indicate to a listener that there is a boundary preceding this sound and the listener would be able to parse the sequence into “this table.”

Allophonic aspiration and phonetic consonant duration are the two main cues to the segmentation of English *s*C clusters. *s*C clusters are sequences of consonants where the first element is a sibilant sound like /s/ or /ʃ/ and the second element is a glide, liquid, nasal, or stop. Examples of words beginning with *s*C clusters include ‘stop’, ‘snail’, and ‘swim’. This thesis will focus on the segmentation of *s*+stop clusters found in English, /sp, st, sk/. When native English speakers are segmenting utterances that contain *s*+stop clusters, the presence or absence of aspiration on the stop and the duration of the sibilant element provide cues as to the location of a word boundary (Christie, 1974; Klatt, 1975; Lehiste, 1960). There are two classifications of *s*+stop clusters that are distinguished specifically by these two cues: cross-boundary and word-initial clusters. In word-initial clusters the word boundary precedes both segments of the cluster. They are cued by a longer /s/ and an absence of aspiration following the stop because the environment for allophonic aspiration is not present. In cross-boundary clusters the word boundary falls between the two elements of the cluster where the /s/ is the offset of one word and the stop is the onset of the following word. Because the stop is found in word-initial position in these clusters, the environment for allophonic aspiration is met and the word boundary is marked by the presence of aspiration and a short /s/-duration. Figure 1 shows approximate durations of these two boundary cues compared for cross-boundary and word-initial *s*+stop clusters.

Use of these cues can be found in the parsing of a phrase like [ðɪsteɪbəl], where both a cross-boundary or word-initial *s*+stop cluster would create acceptable phrases in English. The cross-boundary interpretation “this table” would be cued by a relatively short /s/-duration and the presence of allophonic aspiration on the voiceless stop /t/. The word-initial interpretation of “this stable” would have the inverse cues, a relatively long /s/-duration and the absence of aspiration.

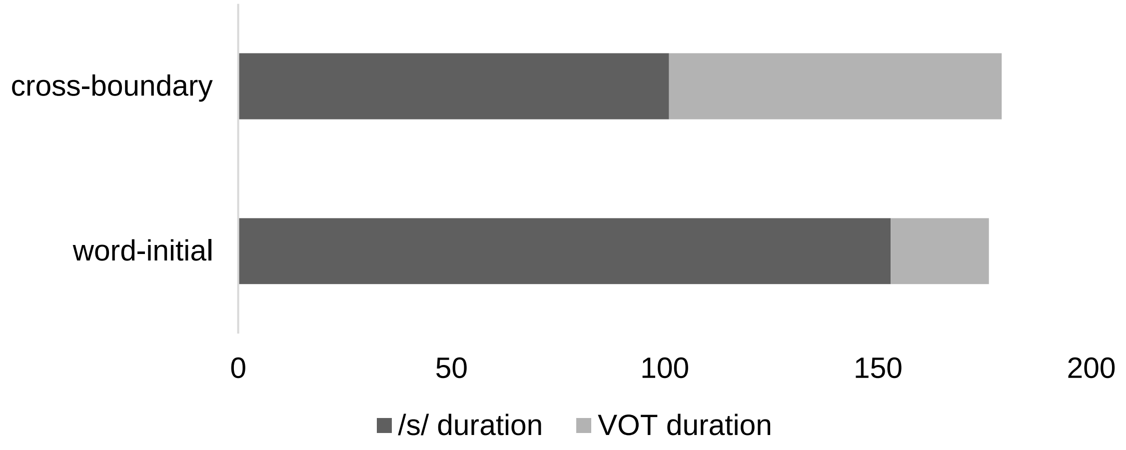

Figure 1. **Approximate durations of word boundary cues to illustrate how the /s/-duration and VOT duration might differ in the two cluster types**

1.2. Second Language Segmentation Strategies

Learning how to segment speech in a second language is heavily influenced by the segmentation strategies of a person's first language (Carroll, 2004). Previous research has shown that in second language acquisition, learners will initially try to adapt the sub-lexical segmentation cues such as rhythm (Cutler, 2002; Cutler, Mehler, Norris, & Segui, 1986, 1992; Goetry & Kolinsky, 2000; Sanders, Neville, & Woldorff, 2002) and phonotactic constraints (Weber & Cutler, 2006) from their first language into their second language. As mentioned in the previous section, English segmentation is guided by the position of stress in a word; the typical stress pattern of English is strong-weak. Cutler and her colleagues have shown that in a language like French, segmentation is motivated at the level of the syllable, and a native French speaker who has learned English as a second language will continue to use a syllable-based segmentation strategy in English speech processing even in instances where the speech signal does not support the use of such a strategy. While rhythmic cues have been fairly well studied, only a few researchers have looked at how second language listeners use allophonic word boundary cues in speech segmentation.

The previous body of literature on the segmentation of *s*+stop clusters by second language speakers of English has found that the acquisition of new word boundary cues that are necessary for speech segmentation is difficult compared to the adaptation of cues from a speaker's first language into their second language. Research on English learners who natively speak languages without systematic aspiration, like Spanish (Altenberg, 2005) and French (Shoemaker, 2014), showed that the rate of correct identification of word boundaries cued by the presence or absence of allophonic aspiration was low, but significantly above chance (58.5% for Spanish speakers and 58.3%-60.9% for French speakers, depending on level of English). English learners who natively spoke Japanese, which has weakly aspirated word-initial stops, had better identification accuracy (73.1%; Ito & Strange, 2009).

Together the results of these studies suggested that the presence or absence of highly predictable cues, like aspiration, were weighted higher than other cues to word boundary and provided the most relevant information to the listeners. However, only the native Spanish speakers showed a significant difference in accuracy between cluster types, with the presence of aspiration in cross-boundary clusters leading to a higher level of accuracy than the absence of aspiration in word-initial clusters. Results from the native Spanish speakers showed that when aspiration was absent, the participants had below chance accuracy (44.1%) compared to when aspiration was present (73.0%). The native French speakers tested in a first-year English class did show a trend toward higher accuracy when aspiration was present (61.3%) compared to when it was absent (55.3%), but this difference was not significant and the trend was not found with a class of more experienced English language students. The Japanese speakers tested showed no such difference between the two aspiration conditions.

The major finding of these three studies was that having a phonological process that cues the location of a word boundary in one's native language provides a clear advantage for discrimination and identification of the same cue in an additional language. For English learners from languages like Japanese, which has an allophonic process where word-initial stops are weakly aspirated, it is possible to adapt the boundary cue from the first language into an additional language. This ability to transfer and adapt a native cue rather than learn a new one clearly provided the Japanese learners of English with an advantage in word boundary identification. With English learners from languages like Spanish and French, with no systematic aspiration, their only option was to acquire the English aspiration cue and, until understanding the environment for allophonic aspiration, try to use strategies from their first language. For these learners, the presence but not the absence of allophonic aspiration provided stronger evidence for word boundary and led to higher accuracy in identification. The reason for stronger evidence coming from the presence of aspiration is likely due to its novelty for the English learners from French and Spanish backgrounds. For these learners, the default voiceless stop is unaspirated, as such, the presence of aspiration in cross-boundary clusters is a noticeable pattern and can be used more effectively than the lack of aspiration on word-initial clusters. Despite identifying the /s/-duration in these clusters as a potential cue for word boundary identification, none of the previous studies provided any discussion of how the /s/-duration cue could have influenced word boundary identifications for the three speaker groups. More research on the acquisition and/or adaptation of word boundary cues is necessary to identify the usefulness of the "secondary" cues of things like /s/-duration.

1.3. Models of Second Language Acquisition

Several models of second language acquisition have been developed to explain how a first and second language interact during learning. The Perceptual Assimilation Model (PAM; Best & Tyler, 2007), the Speech Learning Model (SLM; Flege, 1995), and Brown's (1998) feature-geometry based model of acquisition are all examples of models that attempt to explain how the phonetics and phonology of a person's first language influence their perception of sounds in their second language. One of the overarching ideas behind these models is that when listening to an unfamiliar, non-native sound, second language listeners are likely to assimilate that sound to the most articulatorily similar sound in their native language. When a non-native sound is too different from any native phoneme, a new phoneme category can be developed by the learner based on the quality and amount of input available for the non-native phoneme. The way these non-native sounds are assimilated then has implications for the discrimination ability expected for sound contrasts which include these non-native sounds. The main goal of all three of these models is to show that the acquisition of second language contrasts is dependent on a learner's language experience.

The downfall of these models of second language acquisition lies in their failure to address the effect of phonological context on the acquisition of certain contrasts. Sounds involved in highly predictable allophonic contrasts can only be identified based on the surrounding environment. PAM mentions the influence of context by stating that context does change how the non-native sounds are assimilated into first language phoneme categories. SLM, on the other hand, relies on the phonetic characteristics of sound to develop second language production targets but does not reference the necessary acquisition of second language allophonic processes. There have been few studies which have looked at the acquisition of second language allophones (Shea & Curtin, 2010, 2011) but the current study will provide a source of information which can help to identify the way in which the properties of a learner’s first language influences the perception and production of a second language allophonic contrast.

1.4. Word Recognition

Word recognition relies on the ability of a listener to identify word boundaries by using phonotactic and acoustic cues throughout the processing of a speech stream. When hearing the onset of an auditory input, there is immediate lexical activation and all potential candidates are activated in parallel. As more of the signal is received, the candidate set is updated incrementally and new parts of the input will identify the non-viable candidates. During this process, the candidates compete with one another for activation. The more frequent or better matched words will inhibit activation of the less favoured words until a single candidate can be identified (Allopenna, Magnuson, & Tanenhaus, 1998; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Tanenhaus, Magnuson, Dahan, & Chambers, 2000).

Models of continuous word recognition such as the Cohort model (Marslen-Wilson & Welsh, 1978), the Shortlist model (McQueen, Cutler, Briscoe, & Norris, 1995; Norris, McQueen, & Cutler, 1995), and the Good Start model (Gow & Gordon, 1995) describe the ways in which the process of lexical competition occurs. An example of the continuous activation of word candidates following these models are seen in Table 1 and Table 2 using the input /ðɪsteɪbəl/. Each cell in these tables shows words that might be activated upon hearing that segment; words that are eliminated because they no longer match the input are struck through. In Table 1 below, the phonetic input is marked by a longer /s/-duration and a lack of aspiration of the /t/. Listeners would begin by activating words that start with the voiced interdental fricative [ð], then limit the candidate set upon hearing the high front lax vowel [ɪ]. These two sounds alone suggest to the listener that the ideal candidate based on the phonetic input is “this.” The presence of the [s] in the input will confirm the ideal candidate activation of “this” and will begin activation of words beginning with the voiceless alveolar fricative; the duration of the /s/ will make the activation of /s/-initial words stronger. Upon hearing the unaspirated [t] in the input, activation will continue with /s/ initial words because the activation process is sensitive to the phonetic environment cues which indicate that there is no word boundary preceding the [t]. This process of activation continues using the presented input to identify the ideal candidate “stable” and the listener will recognize the phrase as “this stable.”

Table 1. Process of word activation for the sequence 'this stable' which contains a word-initial s+stop cluster

input	**[ð]**	**[ɪ]**	**[sː]**		**[t]**	**[eɪ]**	**[b]**	**[ə]**	**[l]**
candidate	that then these they this the thus …	this thither ~~that~~ ~~they~~ ~~the~~ ~~thus~~ …	this	sand soap sign stop …	stop stable stand stool stick ~~sand~~ ~~sign~~ …	stable stage stadium stake ~~stop~~ ~~stand~~ ~~stool~~ …	stable stabilize ~~stage~~ ~~stadium~~ ~~stake~~ …	stable stabilize …	stable

In Table 2, the phonetic input is marked by a short /s/, in comparison to the input in Table 1, and an aspirated [t^h]. The activation of “this” in this input will continue in the same way as previously described. The [s] in the input will again begin to activate words beginning with /s/, however, the presence of aspiration on the following [t] will penalize the activation of these /s/-initial words because aspiration of voiceless stops in English is associated with onset position and is not present in word-initial *s*+stop clusters. Activation of word candidates which begin with aspirated [t] will continue and the candidate set will become more limited with each consecutive sound until the word “table” and phrase “this table” are recognized by the listener.

Table 2. Process of word activation for the sequence 'this table' which contains a cross-boundary s+stop cluster

input	**[ð]**	**[ɪ]**	**[s]**		**[t^h]**	**[eɪ]**	**[b]**	**[ə]**	**[l]**
candidate	that then these they this the thus …	this thither ~~that~~ ~~they~~ ~~the~~ ~~thus~~ …	this	sand soap sign stop …	tap tooth table tarp ~~sand~~ ~~soap~~ ~~stop~~ …	table tail take tame ~~tap~~ ~~tooth~~ ~~tarp~~ …	table tabor tablespoon ~~tail~~ ~~take~~ ~~tame~~ …	table tablespoon …	table

The main difference between the activation of these two phonetic inputs results from the two cues relevant for distinguishing English cross-boundary and word-initial *s*+stop clusters. The duration of the /s/ does indicate by itself where the boundary may be located, and in each of the two inputs, listeners will confirm the activation of “this” and activate words that begin with [s]. The presence and absence of aspiration is the more relevant cue. The absence of aspiration indicates that there is no boundary preceding the [t] and activation of [s]-initial words will continue unimpeded. However, if aspiration is present, it indicates that there is a syllable or word boundary preceding the [t]. Since there are few, if any, words in English which begin with [ðɪst^h], and based on the typical strong-weak stress pattern in English, listeners should penalize activation of [s]-initial words and instead activate [t]-initial words. Because listeners should be able to identify “this” as the ideal candidate early in the input, their following activations are limited to syntactic categories which are permitted to follow “this.”

1.5. Word Recognition by Non-Native Speakers

Non-native speakers of English follow the same general processes of lexical activation upon hearing speech input. As sounds are heard, candidates are activated in parallel, and these candidates compete until the more frequent or better input match inhibits the less favored words resulting in a single ideal candidate. Eye-tracking research has provided evidence that non-native and bilingual speakers simultaneously activate word candidates within- and between-languages (Blumenfeld & Marian, 2007, 2013; Ju & Luce, 2004; Marian & Spivey, 2003a, 2003b; Marian, Spivey, & Hirsch, 2003; Weber & Cutler, 2004). In these studies, when participants are presented with an auditory target word, there is a target image, a within-language competitor, and a between-language competitor. For example, given the target word *‘van’* the images on the screen for a bilingual French-English participant might be a *van*, a *vacuum*, a *cow* (*‘vache’* is the French word for cow), and a completely unrelated item like a *swing*. When hearing the target *‘van’*, listeners would be expected to fixate on the target item, the within-language competitor, and the between-language competitor until it becomes clear, based on the input signal, that the ideal candidate is the van. This simultaneous activation of both languages causes non-native speakers of English to have slower overall processing of the speech signal as they are forced to suppress the lexical activation of words in their native language in favor of the non-native candidates.

Several models of bilingual activation have been proposed to detail orthographic word recognition (BIA and BIA+; Dijkstra & van Heuven, 2002) and auditory word recognition (Grosjean, 1997; Shook & Marian, 2013). The models of orthographic word recognition illustrate similar processes compared to the models of spoken word recognition but do not consider the communicative contexts in which spoken word recognition occurs. The general process of activation for bilinguals in auditory word recognition proceeds in much the same way as it would for native monolingual speakers of a language. The models suggest that upon hearing the first sound in a sequence the listener will activate words in both languages that have this same onset. As the speech input progresses, activation of words from both spoken languages will be influenced by different factors depending on the model of activation which is being used. In the Bilingual Language Interaction Network for Comprehension of Speech (BLINCS; Shook & Marian, 2013), semantic and phonological competitors of the active words feed back information which increases activation of some competing words and decreases activation of others; this reflects the influence of top-down and bottom-up information respectively. In the Bilingual Model of Lexical Access (BIMOLA; Grosjean, 1997) factors of bilingual activation relate to the language mode (monolingual or bilingual) being used, in addition to the top-down and bottom-up information that is prominent in BLINCS. Despite their differences, these models of bilingual spoken word recognition demonstrate that continuous activation progresses in roughly the same way for monolingual and bilingual language listeners.

1.6. Current Study

Previous studies that have investigated the ability of non-native English speakers to identify the correct member of a phrase pair based on word boundary cues have focused on the systems of aspiration that are found in the participants' native languages. The influence of consonant duration cues, as well as languages with phonemically contrastive systems of aspiration, have been disregarded thus far leaving a gap in the research. The current study aims to fill this gap by investigating how speakers use the phonological properties of their first language to process speech input in their second language. More specifically I ask how a phonemic contrast not used for word boundary identification is adapted into English by native speakers of Mandarin Chinese and how an unknown contrast is learned as an English word boundary cue by native speakers of

French. I will be looking at how these two cases are reflected in real-time processing in addition to accuracy of identification to further extend this field of research. In addition to investigating how language background influences processing in a second language, I am also interested in how second language word activation occurs more generally.

The previous studies motivating the current research used forced-choice tasks to test groups of English learners on their ability to accurately discriminate and identify English cross-boundary and word-initial *s*+stop clusters in phrases where both cluster types could be used to create valid utterances (Altenberg, 2005; Ito & Strange, 2009; Shoemaker, 2014). The phrases used in prior studies included “Lou spills” vs “loose pills”, “keep sparking” vs “keeps parking”, and “cook struck” vs “cooks truck.” The stimuli were selected from a larger list of potential phrases based on naturalness ratings (Altenberg, 2005), and were used in all three experiments. Naturally produced stimuli were used in all three experiments to provide participants with all available acoustic-phonetic word boundary cues. Measurements of the recordings used in the experiments showed that there was a significant difference between the aspiration in cross-boundary and word-initial clusters. A significant difference in the /s/-duration between the two cluster types was also identified in the recordings used in Ito and Strange (2009) and Shoemaker (2014); /s/-duration was not reported in Altenberg (2005).

The use of a forced choice task in the previous studies only allowed for an investigation of the accuracy of identification, making it impossible to draw any conclusions about the sublexical cues that are being used by second language English speakers from different language backgrounds, nor is it possible to identify when within the speech signal listeners become aware of these cues. The current experiment uses eye-tracking in the visual world paradigm (Allopenna et al., 1998; Huettig, Rommers, & Meyer, 2011) to investigate the segmentation abilities of native Mandarin and native French speakers of English. The visual world paradigm allows for a measure of real-time language processing. The tasks are simple and do not require overt metalinguistic judgments, making them quite natural for participants. In the visual world paradigm, participants are seated in front of a screen displaying a given number of images and their eye movements around the display are recorded as they hear spoken instructions or descriptions of items. It is possible to infer real-time processing because participants unconsciously make multiple eye movements to the different objects on the screen every second. The visual world paradigm has been used to investigate phonetic all the way to prosodic processing (Huettig et al., 2011) and is sensitive to processing in a way that other more overt methodologies, such as the forced choice task used in previous studies, are not.

Using eye tracking in the visual world paradigm in this experiment allows for measures of both identification accuracy and timing of activation. During the task, participants are required to make selections based on the instructions or descriptions that they hear, making it possible to identify correct responses. This allows for a comparison to the results of previous studies on segmentation, which used allophonic aspiration as a word boundary cue. By tracking the eye movements of participants as they hear the stimuli, inferences can be made about lexical activation processes. The current experiment uses a four-item visual world paradigm where each display contained two phonetically related and two phonetically unrelated items. As activation of lexical items occurs simultaneously, when hearing one of the two phonetically related items, fixations to both items are expected to remain similar until the point at which there is enough phonetic information to identify the target word. At this time, fixations to the competitors will be suppressed and the proportion of fixations to the target will be the highest.

Based on the results of previous studies, there are clear expectations regarding identification accuracy by speakers of languages with access to different types of acoustic cues. For native Mandarin speakers who have a native aspiration contrast, using the allophonic aspiration cue in the segmentation of English should be easier and lead to higher levels of accuracy. For second language English speakers from a language like French, which has no distinction in aspiration of voiceless stops, identification accuracy is expected to be lower because the French speakers will not have any basis for using this information in word recognition

Since previous studies have not looked at real-time processing of word boundary cues in second language learners, specific predictions are harder to make. As discussed in the previous section, second language and bilingual speakers can have overall slower activation in real-time spoken word recognition tasks because they activate both between- and within- the languages that they speak. Slower activation has also been found to result from overall uncertainty of interpretations (Farris-Trimble, McMurray, Cigrand, & Tomblin, 2014; McMurray, Samelson, Lee, & Tomblin, 2010). As a result, compared to a group of native English speakers who serve as a baseline, participants in this study are also expected to show slower activation of target words because there may be more possible competitors. This slower fixation to target items is also expected to influence reaction times in identification. Between the two languages of interest there are also expected processing differences. Native Mandarin speakers are expected to have faster activation and identification of English target words than native French speakers because they have more experience using an aspiration contrast for word recognition. On account of the native French speakers not having an aspiration contrast to adapt into English, they are predicted to have slower overall processing and reaction times than both a native English baseline and native Mandarin speakers.

Chapter 2. Experiment 1

2.1. Language of Interest

Mandarin has a two-way system of contrastive aspiration with both voiceless aspirated and voiceless unaspirated stops that can occur syllable initially (Duanmu, 2000; H. Lin, 2001; Y.-H. Lin, 2007). The contrastive nature of the allophonic process implies that, unlike English, aspiration is not used as a word boundary cue in Mandarin. Relevant minimal pairs that illustrate this contrastive system are shown in Table 3.

Table 3. Minimal pairs showing contrastive VOT in Mandarin

	Voiceless aspirated		**Voiceless unaspirated**	
Labial	[pʰa]$_{51}$ ‘to fear’	82 ms	[pa]$_{51}$ ‘father’	14 ms
Dental	[tʰi]$_{35}$ ‘to lift’	81 ms	[ti]$_{35}$ ‘flute’	16 ms
Velar	[kʰou]$_{214}$ ‘mouth’	92 ms	[kou]$_{214}$ ‘dog’	27 ms

Average duration values from Chao & Chen (2008)

These aspirated and unaspirated stops found in Mandarin fall into the same categories of those found in the cross-boundary and word-initial *s*+stop clusters found in English, long lag (60-100ms) and short lag (0-25ms) respectively. However, in an experiment comparing Mandarin and English voiceless stops, Chao and Chen (2008) found that while the two languages share the same space along the VOT continuum, the VOT values between the two languages did show significant differences in that Mandarin aspirated stops have a longer VOT than English aspirated stops. Unlike English, the phonotactic constraints of Mandarin do not allow for any type of word-initial clusters or any cross-boundary *s*+stop clusters (Duanmu, 2000; H. Lin, 2001). These differences between the two languages could make it more difficult for Mandarin speakers to segment the two *s*+stop cluster types found in English.

2.2. Hypotheses

Based on the results of previous studies regarding identification accuracy of English phrases that are differentiated by the presence or absence of allophonic aspiration (Altenberg, 2005; Ito & Strange, 2009; Shoemaker, 2014), I expect that native Mandarin speakers will show accuracy levels similar to those found for native speakers of Japanese. I make this prediction because the presence of an aspiration difference in

Japanese made it possible for the native Japanese speakers to adapt to the English allophonic aspiration process found at word boundaries. The native Mandarin speakers have a system of aspiration that can be adapted for use in English speech segmentation rather than having to learn a new cue entirely.

With regards to the measures of real-time processing, native Mandarin speakers are expected to be slower to activate lexical items than native English speakers. This expectation is based on the results of previous research which has shown that activation of within- and between-language competitors and uncertainty of interpretations slow overall processing (Blumenfeld & Marian, 2007, 2013; Farris-Trimble et al., 2014; Ju & Luce, 2004; Marian & Spivey, 2003a, 2003b; Marian et al., 2003; McMurray et al., 2010; Weber & Cutler, 2004). This hypothesis would be supported in the data by slower fixations to target items and slower suppressions of competitor items. The presence of the aspiration contrast in Mandarin should give the native Mandarin speakers enough information to fully activate the target word regardless of cluster type. Despite this, the native Mandarin speakers are expected to have lower overall levels of activation due to the possible uncertainty of interpretations (Farris-Trimble et al., 2014; McMurray et al., 2010). Support for this hypothesis would be found by measuring maximum fixations to target items.

2.3. Participants

A total of fifty-three participants were recruited at Simon Fraser University and in the Metro Vancouver area: twenty-six native English speakers and twenty-seven native Mandarin speakers. I defined “native speaker” of a language as a speaker who had learned the language from birth and continued to use it in their daily lives; simultaneous Mandarin-English bilinguals were not recruited for the study. Additionally, native English participants run in the study were required to not speak Mandarin as an additional language; no other restrictions were put on additionally spoken languages. All participants for this study were between the ages of 19-45 following the age range used in previous studies (Altenberg, 2005; Ito & Strange, 2009; Shoemaker, 2014). Demographic information was collected from each participant regarding the age that participants began learning English, years of English learning, age of arrival in North America, and length of residence in North America. Alongside the demographic information, participant proficiency was assessed (Yeung & Lin, 2018). This test required participants to listen to a passage and circle incorrectly pronounced words on a response sheet. A total of twenty words in the passage were pronounced by the recorded speaker with the incorrect stress pattern; if the pattern should be strong-weak it was pronounced weak-strong, and vice versa. This test replaced subjective measures of proficiency. All participants were required to have normal or corrected-to-normal vision as well as normal hearing to ensure that all participants could complete the eye-tracking portion of the study. Participants received $10 or Linguistics course credits for their participation.

The data from a total of eleven participants were excluded from the analyses for this experiment for a variety of reasons. Four participants were removed because their self-reported language background did not fit the language background criteria that were required for the experiment (three participants run as native English speakers and one participant run as a native Mandarin speaker were removed based on this criterion). Seven other participants were excluded from analyses because they were at least two standard deviations from the mean for their language group in three key measures used in the eye-tracking analyses (one native English and two native Mandarin speakers were at least two standard deviations below the average identification accuracy for their respective groups; two native Mandarin speakers had reaction times that were at least two standard deviations slower than the average for their group; one native English and one native Mandarin speaker made too few fixations to the target). This left twenty-one native English speakers and twenty native Mandarin speakers in the analyses of perception. Demographic information collected from the remaining participants is provided in Table 4. A comparison of accuracy on the proficiency test for the two language groups showed that the native English speakers (mean accuracy: 80%) were significantly more accurate than the native Mandarin speakers (mean accuracy: 45%; *t*(39) = 5.6, *p* < .001).

Table 4. Experiment 1 demographic information

	Native English	**Native Mandarin**
Sample Size	21	20
Mean Proficiency	80% (45-95%)	45% (10-95%)
Mean age began learning		7.15 (4-15)
Mean years of learning		14.6 (7-18)
Mean age of arrival		16.45 (4-21)
Mean years of residence		5.38 (1-15)

Demographic information presented as mean (range)

2.4. Stimuli

In developing the word list for this experiment, minimal pairs of nouns were selected for each word-initial voiceless stop in English, /p/, /t/, /k/. These minimal pairs differed in onset with one member beginning with an *s*+stop cluster and the other beginning with an aspirated voiceless stop. A total of thirty minimal pairs were selected, ten for each word-initial voiceless stop. In addition to these thirty minimal pairs, sixty filler items were selected – two per minimal pair – to create four-item experimental sets. Filler items were selected so that no word began with a voiceless oral stop or sibilant consonant. Experimental sets were created such that the filler items in the sets were not phonetically or semantically related to the experimental items. All 120 words used in the experiment were a maximum of two syllables in length with trochaic stress. Examples of these sets are provided in Table 5 with a full list found in Appendix A.

Table 5. Examples of experimental sets

/p/ vs /sp/		/t/ vs /st/		/k/ vs /sk/	
park	spark	tack	stack	cone	scone
ladder	mug	bell	hand	leaf	hanger
pine	spine	table	stable	key	ski
watch	lobster	mitten	bike	lemon	door

One of the main concerns in selecting the words for this experiment was word frequency, as Ito and Strange (2009) found that the participants' familiarity with words biased their responses. The SUBTLEX$_{US}$ database (Brysbaert & New, 2009), a collection of words from the American English subtitles of over 8,000 films, was used to compare word frequencies of each item. Comparisons of word frequencies for the cluster-initial items compared to the singleton-initial items showed that for each place of articulation there was no significant difference in word frequency as tested using a paired two sample t-test: /p/ vs /sp/ $t(9) < |1|$; for /t/ vs /st/ $t(9) < |1|$; for /k/ vs /sk/ $t(9) < |1|$. Additionally, word frequency of filler items overall compared to the frequency of all experimental items was not significant: $t(118) < |1|$.

2.4.1. Auditory Stimuli

The auditory stimuli were recorded by an adult male speaker of Canadian English. Stimuli were digitally recorded in Audacity at 44,100 Hz in a quiet room. The selected pairs were recorded in the frame “click on this ______” to create the environment of a cross-boundary *s*+stop cluster, but not cause inconsistencies with the frame between word types. By recording the word-initial *s*+stop clusters in the same frame as the cross-boundary clusters the potential acoustic information available to participants was increased as the /s/-duration in word-initial clusters was lengthened with the addition of the final /s/ in “this.” The filler items were recorded in the frame “click on a ______”. By using the two separate frames, I expected participants to learn over the course of the experiment which items were targeted by which sentence frames. This allowed participants to more quickly rule out filler items making it possible to determine more precisely how the relevant word boundary cues were being used.

Noise reduction was performed on all recordings using Audacity’s Noise Reduction effect which uses spectral noise gating to reduce static background noise in an audio file based on a selected quiet sound segment. In total, six tokens of each word were recorded and the best four tokens were selected for use in the experiment ensuring there were no noticeable differences in pronunciation. The four selected tokens of each word were extracted from the noise reduced audio and amplitude was normalized using the *normalize* function in Praat (Boersma & Weenink, 2016). Silence was added to the beginning of each file to ensure that any sound created by a mouse click in the experiment would not mask the onset of the stimuli.

Each of the four tokens of the experimental pairs were manually annotated in Praat to allow for measures of the duration of the vowel /ɪ/ in “this”, the duration of the /s/, and the VOT. The previous research motivating this experiment also measured the closure duration preceding the voiceless stop in cross-boundary and word-initial clusters as a possible word boundary cue, but their analyses suggested that closure duration was not a distinctive cue for distinguishing cross-boundary and word-initial *s*+stop clusters (Ito & Strange, 2009). I thus chose to exclude this measure in the following analyses. Using the available functions in Praat, the durational properties of these three relevant segments were extracted. Figure 2 shows average durations for each cluster type and each place of articulation. Univariate ANOVAs were run on the durational measures with cluster type (cross-boundary or word-initial *s*+stop clusters) and place of articulation of the voiceless stop in the word pairs (/p/, /t/, or /k/) as fixed effects. The ANOVA run on the vowel duration showed no effect of cluster type ($F(1,54) < 1$), a significant effect of place of articulation ($F(2,54) = 6.3$, $p = .004$), and a significant interaction between cluster type and place of articulation ($F(2,54) = 5.2$, $p = .009$). To investigate the effect of place of articulation on vowel duration I ran t-tests comparing the individual places of articulation for each cluster type. T-tests comparing place of articulation for the cross-boundary clusters revealed a significant difference between vowel duration between /t/ and /k/ ($t(18) = 3.6$, $p = .002$). This difference is such that the /ɪ/-duration preceding cross-boundary clusters with an alveolar stop /t/ is significantly longer than in clusters with the velar stop /k/. Additional t-tests comparing the places of articulation for word-initial clusters showed a significant difference in vowel duration between /sp/ and /st/ ($t(18) = 3.6$, $p = .002$) and between /sp/ and /sk/ ($t(18) = 3.7$, $p = .002$). In the word-initial stimuli the difference was such that clusters with the labial stop /p/ had a shorter preceding /ɪ/-duration than clusters with alveolar or velar stops. These results suggest that vowel duration may cue participants to the place of articulation of the stop in the cluster, but because there is no main effect of cluster type, there is not a concern for the preceding vowel duration in “this” to cue participants to the potential location of a word boundary.

The analysis of /s/-duration showed the expected differences based on the boundary cues found in English. There was a significant effect of cluster type ($F(1,54) = 411.8$, $p < .001$), where /s/-duration was longer in cross-boundary than word-initial clusters, and of place of articulation of the voiceless stop ($F(2,54) = 13.5$, $p < .001$). T-tests comparing the individual places of articulation for cross-boundary clusters showed significant differences between /p/ and /t/ ($t(18) = 4.9$, $p < .001$) and between /p/ and /k/ ($t(18) = 4.1$, $p = .001$). These differences were such that /s/-duration preceding labial stops was shorter than that preceding alveolar and velar stops. Comparisons of word-initial clusters showed a significant difference between /sp/ and /st/ ($t(18) = 2.8$, $p = .012$) and a marginally significant difference between /sp/ and /sk/ ($t(18) = 1.7$, $p = .099$. These differences were such that the /s/-duration in labial *s*+stop clusters was shorter than in alveolar clusters and was slightly shorter than in velar clusters. It is not clear why the duration of the sibilant /s/ changed significantly with the different places of articulation, but the main difference expected between the two types of stimuli was found which suggests that /s/-duration will be an available cue to the participants of this experiment.

Finally, analysis of VOT also showed the expected differences based on the word boundary cues of English. There was a significant effect of cluster type such that VOT was longer in the cross-boundary clusters where the environment for allophonic aspiration was met ($F(1,54) = 502.5$, $p < .001$). There was also a significant effect of place of articulation found ($F(2,54) = 9.3$, $p < .001$). T-tests were also run to compare VOT for each place of articulation. For the cross-boundary clusters there was a significant difference in VOT between /p/ and /k/ ($t(18) = 2.4$, $p = .030$) and between /t/ and /k/ ($t(18) = 2.2$, $p = .040$). In the word-initial clusters, VOT was significantly different between all places of articulation: /sp/ and /st/ ($t(18) = 4.2$, $p = .001$), /sp/ and /sk/ ($t(18) = 4.9$, $p < .001$), /st/ and /sk/ ($t(18) = 2.3$, $p = .035$). The differences found in cross-boundary and word-initial *s*+stop clusters showed that the VOT duration in velar stops is consistently longer than labial and alveolar stops. These differences in VOT between the three relevant places of articulation match the findings of previous research (Cho & Ladefoged, 1999; Klatt, 1975) that suggest that the VOT for /k/ is longer than /p/ and /t/ for both articulatory and perceptual reasons.

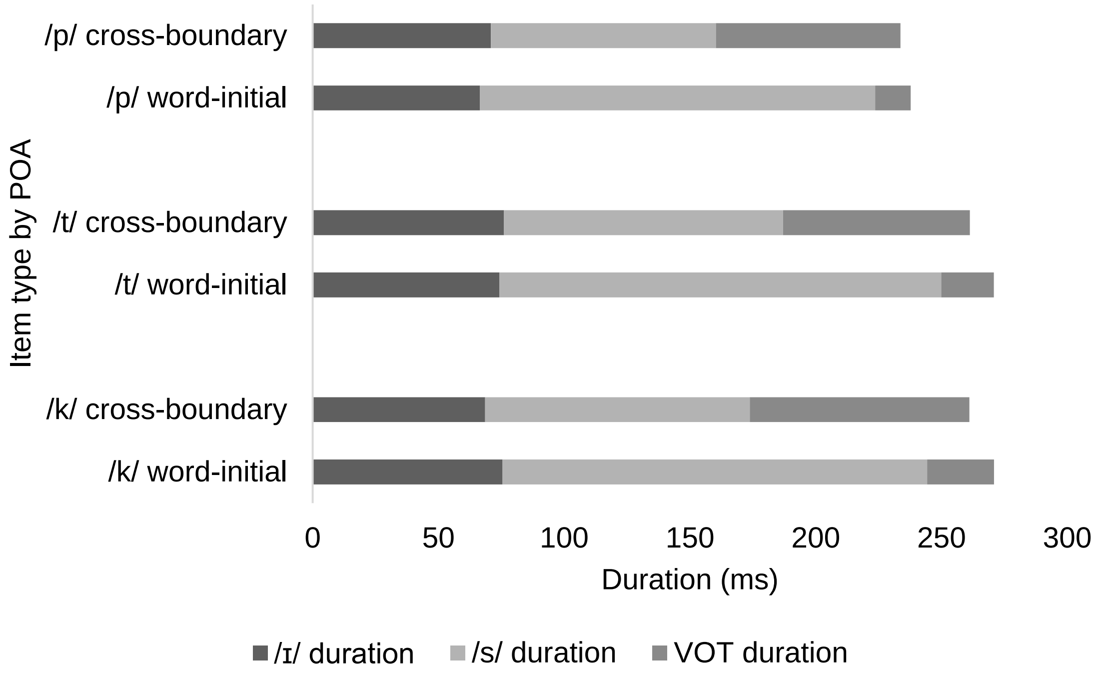

Figure 2. **Durational measures for auditory stimuli separated by cluster type and place of articulation**

2.4.2. Visual Stimuli

The majority of the visual stimuli used in this experiment were taken from a database of images developed in the Mechanisms of Audio-Visual Categorization Lab at the University of Iowa and maintained in the Phonological Processing Lab at SFU. Images found in this database are selected by a committee, approved by someone with extensive experience using the visual world paradigm, and are edited for clarity, typicality, and salience. If an image could not be found in the database, clipart.com was used to provide several image options and the same processes used in the creation and maintenance of the database were used in the development of these images. In each experimental set, in addition to ensuring that the filler items were not phonologically related to the experimental items, we made sure that the pictures were all clearly different from one another and had roughly equivalent visual salience.

2.5. Production Task

One portion of the study had participants complete a production task which served two main purposes. The first was to familiarize participants with the word-picture pairings used throughout the perception stage of the experiment. The second was to collect acoustic data which could be analyzed to determine whether the second language learners of English were producing the relevant word boundary cues. In this familiarization task, participants were shown screens that contained a picture with the associated word and a short sentence underneath. For example, they would see a picture of a key, and beneath it would be the word “KEY” and the sentence “Look at this KEY.” Participants were asked to read the word followed by the sentence for each item in the experiment and were recorded while completing this task to allow for analyses on production.

In the production task all 120 items (60 experimental and 60 filler) were pseudo-randomized in a PowerPoint presentation; word pairs in the randomization were never consecutive. Each participant was presented with these items in the same order. Participants were digitally recorded in Audacity using a Blue Yeti microphone at 44,100 Hz. If a participant was unsure of the pronunciation of a word they could ask the experimenter, and if a participant mispronounced a word the experimenter would correct the participant’s pronunciation.

2.6. Visual World Paradigm Experimental Design

In the visual world paradigm perception task each of the 60 experimental and 60 filler items were presented four times, each time with a different auditory token, for a total of 240 experimental and 240 filler trials (480 trials in total). The experiment was presented in 15 blocks of 32 trials with a break between each block. The trials were randomized individually for each participant using a MATLAB script ensuring order of presentation was not a factor in the results. In addition, the location of items on the screen were also randomized for each trial. The entirety of the experiment was presented on a 1280 x 1024 pixel monitor and each item on the screen was 50 pixels away from the edges of the screen. The experiment was built and presented using the Experiment Builder software developed by SR Research for use with the EyeLink 1000 eye tracker, which sampled participant fixations every 4 ms from the start of the trial until the response.

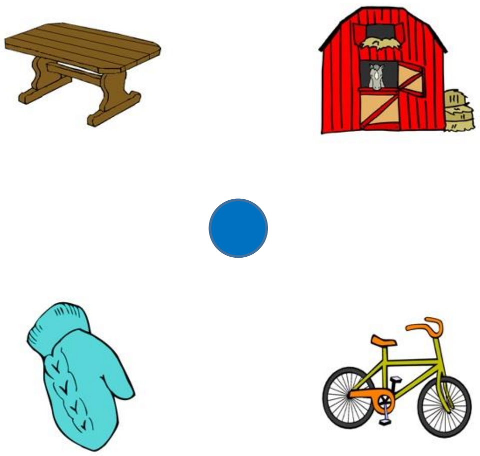

Figure 3. Sample visual world paradigm screen for the set table, stable, mitten, bike

2.7. Procedures

The experiment took approximately one hour to complete and was run in the Phonological Processing Lab at Simon Fraser University. After obtaining informed consent, participants were asked to fill out a language background questionnaire to collect the demographic information provided in Table 4 above. Following the completion of the language background questionnaire, participants were administered the proficiency test described in section 2.3. The next portion of the study was the production task which took around 10 minutes to complete. Finally, participants completed the visual world paradigm perception task, a sample of which can be seen in Figure 3, which took between 30 to 40 minutes to complete.

2.8. Results

The following section outlines the analyses conducted on the production and perception data collected during the experiment. Based on preliminary analyses of average group accuracy, three phrase pairs were removed from the analyses because the native Mandarin speakers performed below chance at identifying which member of the pair they heard: pout vs spout, tick vs stick, and cot vs scot.

2.8.1. Production Measures

In analyses of the production task, an additional three participants (one native English speaker and two native Mandarin speakers) were excluded because they paused between the /s/ in “this” and the /s/ in all *s*+stop cluster words so the full /s/-duration could be measured in the word-initial clusters. Since /s/-duration is an important cue in the processing of word-initial and cross-boundary *s*+stop clusters these speakers were removed from the production analyses. For nine participants (two native English speakers and seven native Mandarin speakers), tokens where the same pauses were found in production were excluded from analyses (range of excluded tokens: 1 to 10), however, these participants remained in the analyses. The remaining 18 native English and 11 native Mandarin speakers produced all tokens in the production task.

Recordings of participants underwent the same processing as the stimuli recordings. Noise reduction was done in Audacity, followed by segmentation and annotation of the /ɪ/ in “this”, the full /s/-duration, and the VOT of the voiceless stop. The durations of these annotated sections were extracted in Praat. Graphs of these measures by language and cluster type can be found in Figure 4, Figure 5, and Figure 6. Repeated measures ANOVAs were run on all three durational measures with language as the between-subjects factor. In terms of the vowel duration, the vowel was shorter in cross-boundary clusters than in word-initial clusters (*F*(1, 36) = 26.2, *p* < .001) and shorter for English speakers than Mandarin speakers (*F*(1, 36) = 5.5, *p* .024). For /s/-duration, the fricative was longer in word-initial clusters than cross-boundary clusters as was expected (*F*(1, 36) = 308.2, *p* < .001) and there was no difference in duration for native English and Mandarin speakers (*F*(1, 36) = 2.6, *p* = .118). In this measure there was a significant interaction between language and cluster type (*F*(1, 36) = 4.9, *p* = .033). Follow up t-tests showed that native English speakers had a marginally longer /s/-duration for word-initial clusters than native Mandarin speakers (*t*(36) = 2.0, *p* = .056) and that there was no difference in duration for cross-boundary clusters (*t*(36) < |1|. Finally, for VOT duration, the VOT was shorter in word-initial clusters than cross-boundary clusters (*F*(1, 36) = 522.6, *p* < .001) and marginally shorter for English speakers than Mandarin speakers (*F*(1, 36) = 4.1, *p* = .051). These results reflect the expected differences in duration for the /s/ and VOT for the two types of clusters. In cross-boundary clusters, participants had a shorter /s/-duration and a longer VOT duration, while in word-initial clusters participants had a longer /s/-duration and shorter VOT duration. The significant effect of cluster type for the /ɪ/-duration was not expected but the direction of the difference, shorter in cross-boundary than word-initial clusters, could be related to the longer /s/-duration found in word-initial clusters. The main effect of language that was found in all three measures could be related to speech rate or familiarity, but because the native Mandarin participants patterned the same way as the native English participants for the two cluster types, this effect did not drive production.

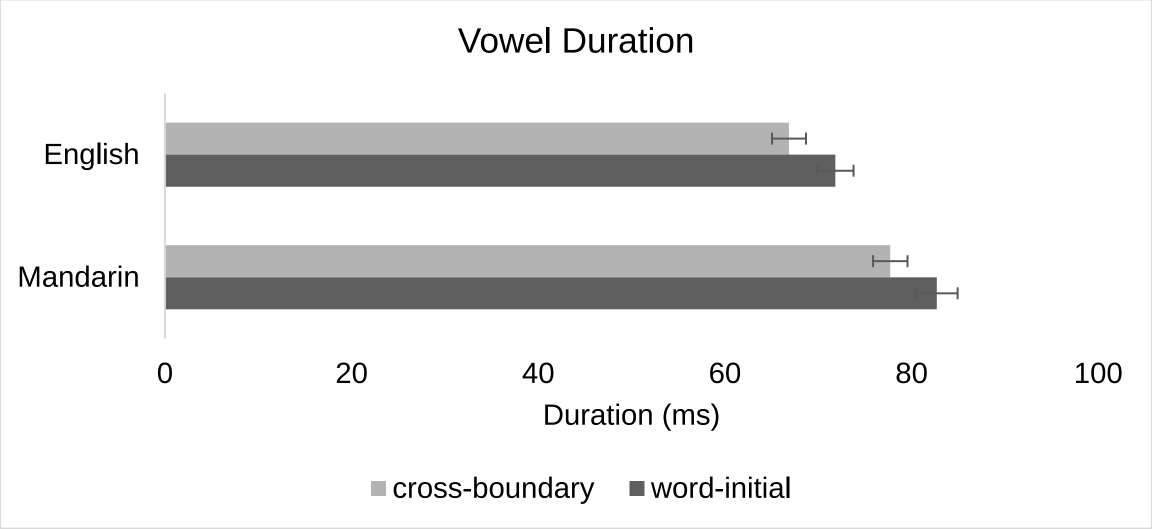

Figure 4. **Average /ɪ/ duration by group and cluster type**

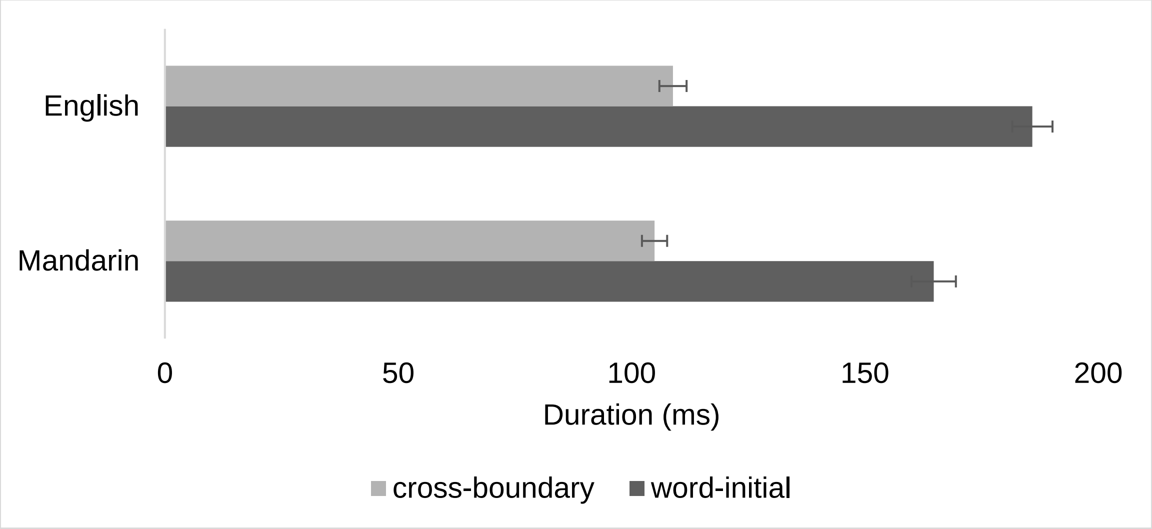

Figure 5. **Average /s/-duration by group and cluster type**

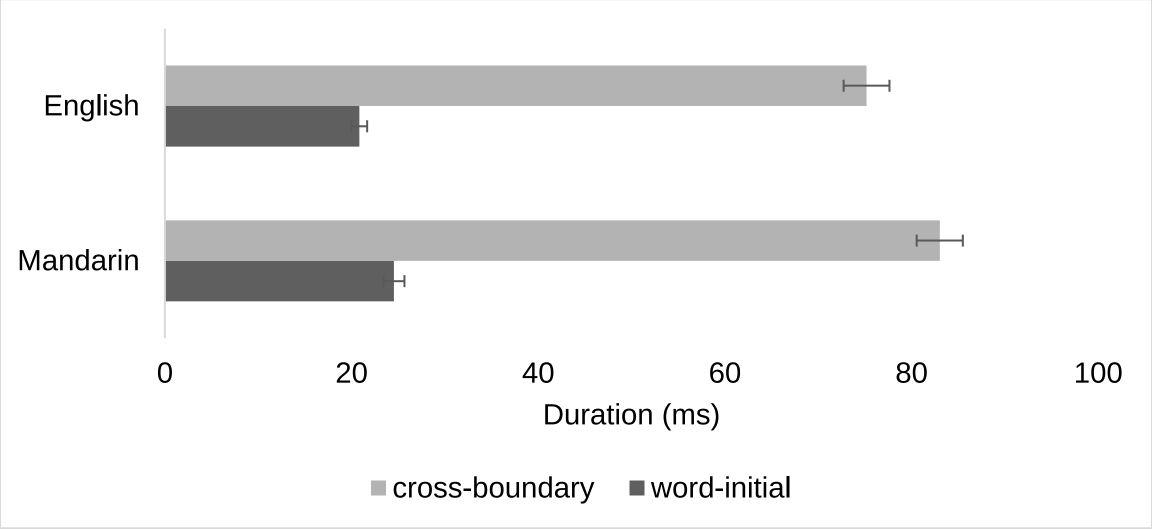

Figure 6. **Average VOT duration by group and cluster type**

2.8.2. Identification Measures

Overall accuracy and reaction times were calculated for each participant. Native English speakers showed ceiling effects with a mean accuracy of 99% (range 96-100%). Native Mandarin speakers were significantly less accurate, with a mean accuracy of 92% (range 82-99%; $t(39) = 6.0$, $p < .001$). Native English speakers also had faster reaction times (1707 ms; range 1535-1944 ms) than native Mandarin participants (2072 ms; range 1674-2588 ms; $t(39) = 6.1$, $p < .001$). Plots of accuracy and reaction time are found in Figure 7 and Figure 8 respectively. Accuracy did not differ based on cluster type (word-initial or cross-boundary) for either of the two language groups (English: $t(20) <$

|1|; Mandarin: *t*(19) < |1|). However, for native English speakers, reaction times to cross-boundary clusters were faster than word-initial clusters (*t*(20) = 2.4, *p* = .027). This difference of reaction time for cluster type was not found in the native Mandarin participants (*t*(19) = 1.5, *p* = .147). This difference suggested that native English speakers may have been more reliant on the presence of aspiration in the cross-boundary clusters than the native Mandarin speakers.

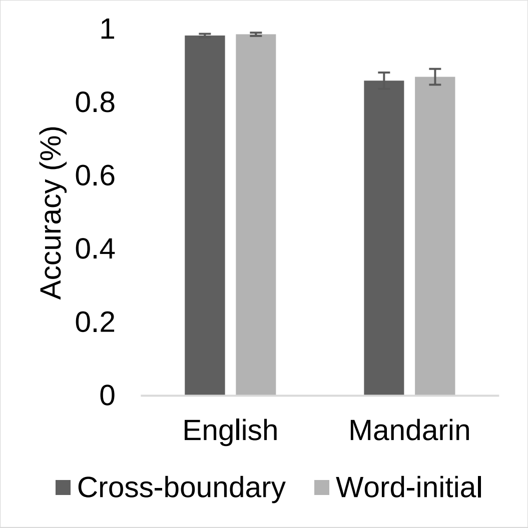

Figure 7. Accuracy by cluster type

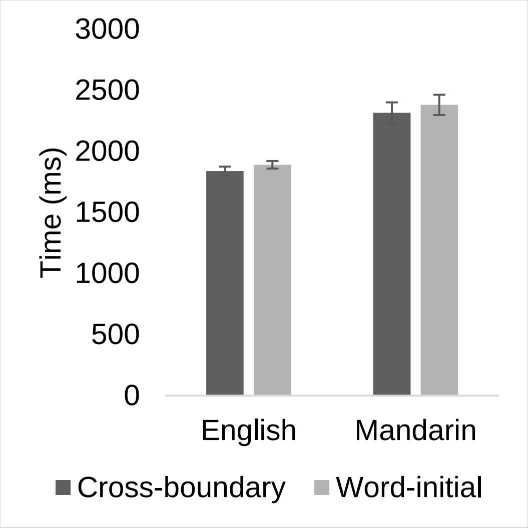

Figure 8. Reaction Time by cluster type

In addition to looking at the mean accuracy for each language group, response bias was analysed to determine if the errors made by native Mandarin speakers in the response task were biased toward one cluster type or the other compared to the native English speakers. Response bias was measured using z-scores of hit and false-alarm rates to calculate the beta value using a likelihood ratio[1]. For participants where hit and false-alarm proportions were 0 or 1, proportions were converted to 1/(2N) and 1-1/(2N) respectively, where *N* is the number of trials (Macmillan & Creelman, 2005). This conversion ensured that there would be no infinite values in the calculation of beta. A

---

[1] Application of the use of beta values in this analysis is unorthodox as a result of the number of choices presented to participants. Unlike typical signal detection theory applications, the participants of this study had four possible response options. However, in the analysed trials participants were expected to select one of two responses, either the target item or the competitor item, and ignore the filler items. Native English speakers selected a filler item in an experimental trial a total of 5 times (0.11% of the 4536 total experimental trials) and native Mandarin speakers selected a filler item in an experimental trial a total of 62 times (1.44% of the 4320 total experimental trials). Thus I felt the use of this response bias measure was justified in spite of the nature of the task.

comparison of beta for the two language groups found that there was no difference in response bias (mean beta for English speakers: 1.56; mean beta for Mandarin speakers 0.99; *t*(39) < |1|). When interpreting the level and direction of bias using beta, bias is centred around 1. In this analysis, values higher than 1 represent a bias toward selecting a cross-boundary cluster and values lower than 1 represent a bias toward selecting a word-initial cluster. An interesting observation based on the mean beta values for the two languages was that the mean beta value for native English speakers was further from the centred value than for native Mandarin speakers. This difference suggests that native English speakers were more biased toward selecting cross-boundary clusters, however, the bias resulted from a small number of errors which were distributed across participants (only four native English participants were completely accurate). The small number of errors made it difficult to make larger judgments about the representativeness of the bias for native English responses to cross-boundary and word-initial clusters in general.

2.8.3. Fixation Measures

In one approach to the analysis of eye-tracking data, measures of fixation curves are taken to identify differences in real-time processing. For each individual trial, participants do not make enough fixations to provide any indication of how they are processing an auditory input. However, when data are averaged across trials for target and competitor separately, smooth curves which show real-time processing emerge. Despite following different curves, target and competitor fixations typically increase together approximately 200 ms after the onset of acoustic information because it takes roughly 200 ms to plan and launch an eye-movement (Viviani, 1990). Figure 9 shows a plot of target fixations by language group and by cluster type. In this plot, time is given on the x-axis and the proportion of fixations is given on the y-axis. For this experiment, time was adjusted along the x-axis such that 0 ms represents the onset of the /s/ in “this” because this is the onset of the first word boundary cue; any fixations 200 ms or more after the adjusted 0 ms are expected to be related to the relevant word boundary cues. By adjusting the time it was also possible to ensure that the interpretations of the results were accounting for the differing durations of the sentence frame “click on _____” due to the variability of natural speech. Target fixations typically follow a logistic curve; fixations increase slowly at first, then exponentially, before levelling out. In this specific plot, target fixations increase slowly through the onset of the trial and start to increase rapidly after around 200 ms, levelling off around 1000 ms. Between the fixations for native English and native Mandarin participants, the English participants have a steeper slope and a higher proportion of fixations to target items.

Figure 10 shows a plot of competitor fixations by language group and by cluster type. As with the plot of target fixations, time is given on the x-axis with 0 ms being the onset of the /s/ in “this” and proportion of fixations is given on the y-axis. Competitor fixations typically follow a gaussian curve; they increase, peak, and then fall off as competitors are eliminated. In this plot, competitors increase after around 200 ms to their peak at around 500 ms and then decrease. The native English speakers in this plot show a clearer rise and fall of fixations around the peak, whereas the native Mandarin participants have a gradual rise to a peak followed by a plateau where competitor fixations do not fall off as much as would be expected for a typical gaussian curve.

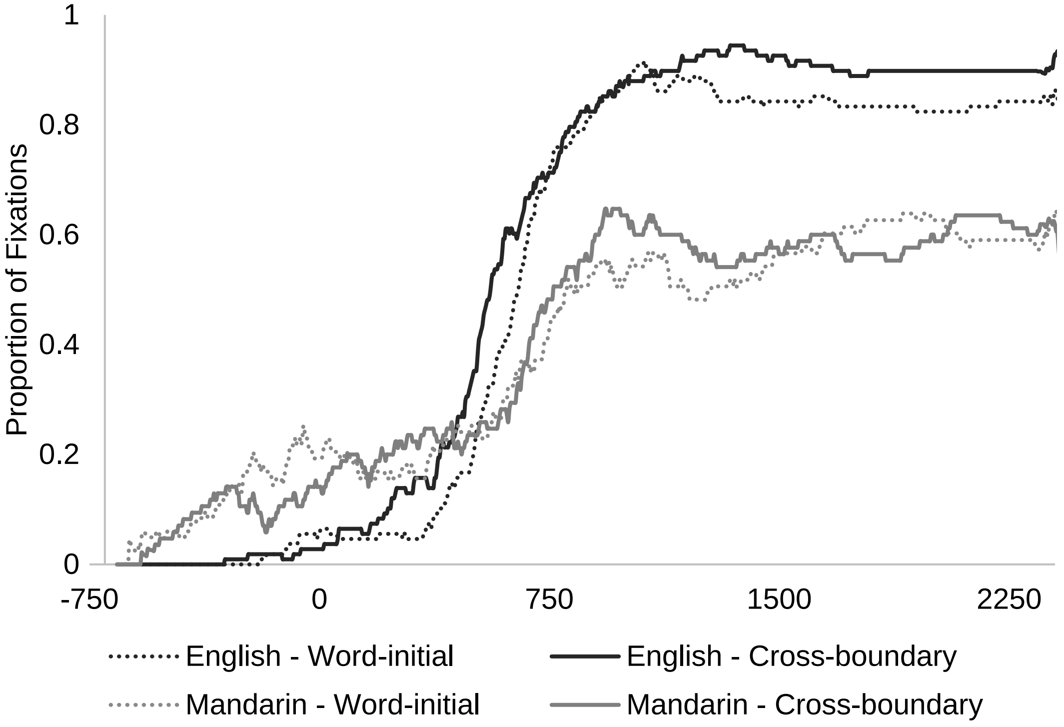

Figure 9. **Target fixations by language and cluster type**

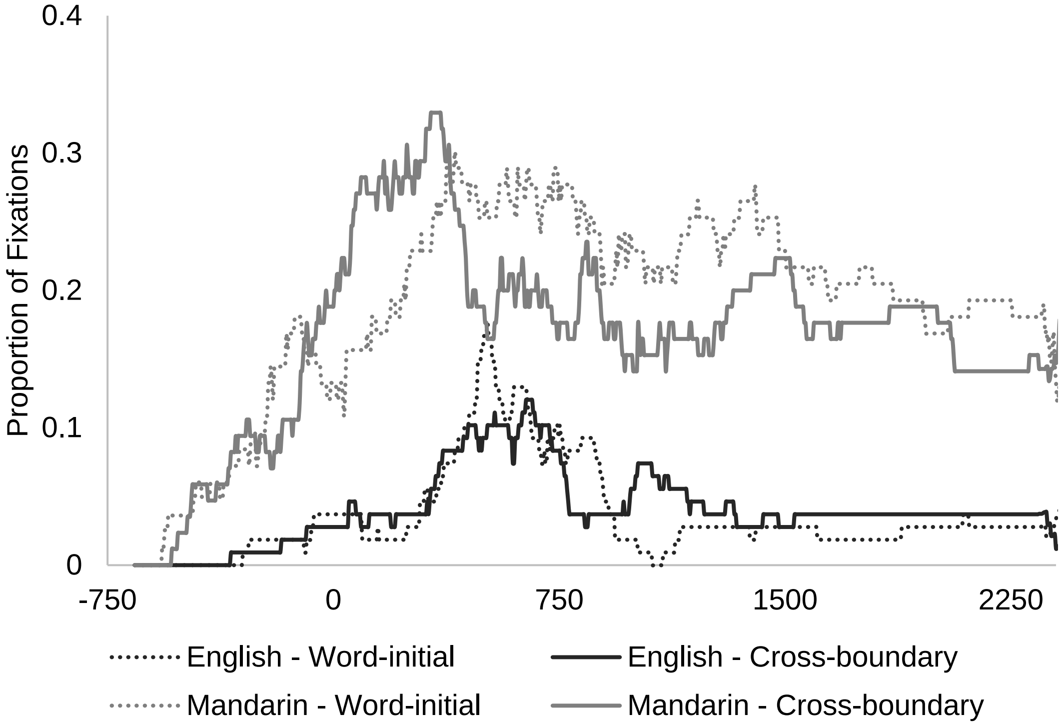

Figure 10. **Competitor fixations by language and cluster type**

In this experiment, target fixations were fit to a logistic curve given in Equation 1. There are four parameters of target fixation curves: the baseline (*b*); the peak (*p*); the crossover (*c*), which is the inflection of the curve where growth starts to slow; and the slope (*s*). The time course of competitor fixations were fit to a specialized gaussian curve given in Equation 2. These competitor curves have six parameters: onset baseline (b_1); onset slope (σ_1); midpoint (μ), which is the time of maximum fixations to the competitor; peak height (*p*); offset slope (σ_2); and offset baseline (b_2). The parameters of these equations for both target and fixation curves can be compared statistically (Farris-Trimble & McMurray, 2013). For the analyses of this experiment, seven parameters that reflect degree and timing of fixations are compared. For target fixations, degree of fixations are analysed using peak and timing of fixations are analysed using cross-over and slope. For competitor fixations, degree of fixations are analysed using peak and offset baseline and timing of fixations are analysed using onset slope and midpoint.

Equation 1. **Logistic function used to fit time course of target fixations**

$$P(target) = \frac{p-b}{1+\exp\left(4\cdot\frac{s}{p-b}\cdot(c-t)\right)}+b$$

Equation 2. **Gaussian function used to fit the time course of competitor fixations**

$$P(competitor) = \begin{cases} exp\left(\frac{(t-\mu)^2}{-2\sigma_1^2}\right)(p-b_1)+b_1 \;\; if\; t \leq \mu \\ exp\left(\frac{(t-\mu)^2}{-2\sigma_2^2}\right)(p-b_2)+b_2 \; if\; t > \mu \end{cases}$$

Unless otherwise stated, all of the following parameters of the fixation curves were analysed using repeated measures ANOVAs with language as the between-subjects factor. ANOVAs were performed on data averaged across subjects and across items. Only trials in which the participants selected the correct item were analysed. MinF' was calculated when an effect was significant in by-item analyses, by-subject analyses, or both and is reported where appropriate (Clark, 1973). In addition to the measures of fixation by-subject and by-item, I performed preliminary analyses on place of articulation and word frequency and determined that neither factor was a consistent predictor of fixations; as such place of articulation and word frequency will be excluded from the subsequent analyses except where significant.

Target Measures

Peak

The peak of the target fixations, which is the maximum proportion of fixations, is a measure of degree. Native English speakers had a higher proportion of fixations than native Mandarin speakers in both by-subject and by-item analyses ($F_1(1,39) = 14.4$, $p < .001$; $F_2(1,52) = 100.3$, $p < .001$; $minF'(1,50) = 12.6$, $p = .001$). There was no difference in maximum fixations by cluster type ($F_1(1,39) < 1$; $F_2(1,52) < 1$). Frequency of target items was marginally significant such that more frequent items received higher proportions of fixations than less frequent items ($F(1,95) = 3.5$, $p = .063$). Overall, this measure of degree showed that while native Mandarin speakers fixated the target less overall (likely due to uncertainty), their pattern of looking to cross-boundary and word-initial cluster targets mirrored that of the native English speakers.

Crossover

The crossover point of the target fixations is the inflection point of the curve where the growth starts to slow. Native English speakers had an earlier crossover point than the native Mandarin speakers in both by-subject and by-item analyses ($F_1(1,39) = 5.5$, $p = .024$; $F_2(1,52) = 31.6$, $p < .001$; $minF'(1,53) = 4.7$, $p = .035$). There was no difference in crossover point by cluster type ($F_1(1,39) = 2.3$, $p = .134$; $F_2(1,52) < 1$). Frequency of target items was marginally significant for crossover point such that when an item was more frequent, the participant had an earlier crossover point than when an item was less frequent ($F(1,95) = 3.2$, $p = .079$). This measure of timing showed that native Mandarin speakers were slower to recognize and fixate the target than the English speakers, but both groups treated word-initial and cross-boundary clusters the same way.

Slope

Target slope is another measure of timing of fixations that reflects the rate of increase in fixations to an item. Native English speakers had steer slopes than native Mandarin speakers for by-subject but not by-item analyses and this effect was not significant overall ($F_1(1,39) = 21.7$, $p < .001$; $F_2(1,52) = 1.4$, $p = .250$; $minF'(1,59) = 1.3$, $p = .256$). There was no difference in slope by cluster type ($F_1(1,39) < 1$; $F_2(1,52) = 1.4$, $p = .250$). These results suggested that native Mandarin and native English speakers did not differ in the rate of increase in fixations to the target items.

Competitor Measures

Peak

The peak of competitor fixations is the maximum proportion of fixations to competitor items. Native English speakers had a lower proportion of fixations to competitor items than native Mandarin speakers by-subject and by-item ($F_1(1,39) = 6.0$, $p = .019$; $F_2(1,52) = 78.6$, $p < .001$; $minF'(1,45) = 5.6$, $p = .023$). There was no difference in peak fixations by cluster type ($F_1(1,39) < 1$; $F_2(1,52) = 1.0$, $p = .315$)). This measure of degree indicates that while the native Mandarin speakers had a larger proportion of fixations to competitor items than the native English speakers, neither group differed in their peak fixations to cross-boundary or word-initial clusters.

Offset Baseline

The offset baseline is a measure of the final proportion of fixations to competitor items. Native Mandarin speakers had a higher final proportion of fixations to competitor items for both by-subject and by-item analyses ($F_1(1,39) = 12.9$, $p = .001$; $F_2(1,52) = 50.6$, $p < .001$; $minF'(1,59) = 10.3$, $p = .002$). There was no effect of cluster type on the offset baseline ($F_1(1,39) < 1$; $F_2(1,52) < 1$). That is, while native Mandarin speakers had a higher degree of fixations to competitor items at the end of a trial, neither group had a difference in final proportion of fixations based on the type of cluster they heard.

Onset Slope

The onset slope of competitor fixations is a measure that is representative of the rate increase in fixations to an item. Native Mandarin speakers had a marginally steeper slope than the native English speakers by-item, but not by-subject ($F_1(1,39) < 1$; $F_2(1,52) = 2.9$, $p = .096$). Cluster type had no effect on onset slope ($F_1(1,39) < 1$; $F_2(1,52) < 1$). Onset slope of competitor fixations was the only measured parameter where place of articulation was a predictor of fixations ($F(2,95) = 3.3$, $p = .041$) and where language and place of articulation showed a significant interaction ($F(2,95) = 3.2$, $p = .045$). Follow up univariate ANOVAs were run on each language independently with cluster type and place of articulation as fixed factors and word frequency as a covariate. These ANOVAs showed that the effect of place of articulation was present for the native Mandarin speakers ($F(2,47) = 5.5$, $p = .007$) but not the native English speakers ($F(2,47) < 1$). Independent sample t-tests showed that the native Mandarin speakers were slower to increase their fixations to /t/ stimuli than /p/ and /k/ stimuli (/p/ vs /t/: $t(34) = 2.1$, $p = .041$; /p/ vs /k/: $t(34) = 1.8$, $p = .089$; /t/ vs /k/: $t(34) = 3.1$, $p = .004$). This measure of timing showed that the native English and Mandarin speakers did not have an overall difference in the rate of increase to competitor items, however, place of articulation did predict the slope fixations for the native Mandarin participants.

Midpoint

This measure of timing specifies the time at which fixations to the competitor item peak. Native English speakers had earlier midpoint values than the native Mandarin speakers in a by-item, but not by-subject, analysis, and the effect was not significant overall ($F_1(1,39) < 1$; $F_2(1,52) = 6.0$, $p = .018$; $minF'(1,40) < 1$). There was no effect of cluster type ($F_1(1,39) < 1$; $F_2(1,52) < 1$). More frequent competitor items had earlier midpoints than less frequent competitor items (*F*(1,95) = 4.0, *p* = .048). This analysis of competitor fixations showed that while native Mandarin and native English participants did not differ in the time of their peak fixations to competitor items, both groups had earlier fixations to more frequent items than less frequent items.

Diversion Points

To further investigate the real-time processing of participants, I analysed the diversion points (the point at which one competing item seems to be chosen over another in the fixations) for each subject for both cross-boundary and word-initial clusters. The diversion point was calculated using the adjusted time from the fixation measures above (where 0 ms was the onset of the /s/ in “this”) to find the earliest point at least 200 ms after the onset of the /s/ at which fixations to the target and competitor items diverged by 10% and continued to be diverged by at least 10% for at least 52 ms. These criteria were chosen to ensure that any divergence between the target and competitor fixations would likely be the result of the two word boundary cues of interest in this experiment. Six participants (two native English and four native Mandarin) were removed from this analysis because their diversion points were two standard deviations away from the mean or were at 0 ms. Having a diversion point at 0 ms suggested that fixations to the target and competitor started to diverge before the participant heard the word boundary cues.

For this analysis I ran a repeated measures ANOVAs with language as the between-subjects factor for data averaged across subjects and across items. Only trials in which the participant selected the correct item were analysed. MinF’ is reported where appropriate (Clark, 1973). Diversion points were found to be earlier in word-initial clusters than cross-boundary clusters by-subject, but marginally so by-item, and the effect was not significant overall (F_1(1, 33) = 6.9, *p* = .013; F_2(1, 52) = 3.7, *p* = .059; *minF’*(1, 84) = 2.4, *p* = .124). Native English participants had marginally earlier diversion points than native Mandarin speakers by-subject, but not by-item (F_1(1, 33) = 3.5, *p* = .069; F_2(1, 52) < 1). The results of this analysis showed that overall, the point at which target and cluster fixations diverged was not predicted by language background or cluster type. This suggests that the timing of the suppression of competitor fixations by both native English and native Mandarin speakers did not differ regardless of the differences in the fixation parameters found between the two groups.

2.9. Discussion

The identification measures analysed in this experiment showed that as expected, native English speakers had higher average accuracy than native Mandarin speakers. This difference was also reflected in the average reaction times of the two participant groups, with native English speakers responding faster than the native Mandarin speakers. In analyses of fixation measures, the measures of degree of fixations showed that the Mandarin participants had fewer fixations to target items and more fixations to competitor items. The measures of timing for target and competitor fixations illustrated the slower processing of target items for bilingual speakers which has been shown in previous studies. However, since there were no differences in any of the measures taken on these curves resulting from cluster type, it seems that the native Mandarin speakers were able to adapt their contrastive system of aspiration to use in the identification of word boundaries in English.

While the previous studies differed from this experiment in methodology, they tested participants using what is the same task fundamentally; participants heard a phrase and had to choose a response. Having the same fundamental task allows for a comparison of accuracy between the native Mandarin speakers tested in this experiment, and the native Japanese speakers tested previously (Ito & Strange, 2009). Japanese has a process of allophonic aspiration similar to that found in English; like the native Mandarin speakers tested in this experiment, the Japanese speakers tested in the previous study had a first language cue which could be adapted into English. Native Mandarin participants tested in this experiment showed a level of accuracy that was closer to the native English speaking baseline than it was to the Japanese participants tested previously. This difference may be due to the proficiency level of the non-native English speakers in the two studies. In the previous studies motivating this research, all participants were late learners or were in English language classes at the time of testing. However, in the current study, the native Mandarin participants were students at an English-speaking university and possessed a high level of English proficiency. Another possible explanation for the unexpectedly high accuracy of the native Mandarin participants is based on the type of contrast present in Mandarin compared to that found in Japanese. Having a phonemic contrast in Mandarin, where aspiration is used to distinguish minimal pairs, could cause native Mandarin speaking participants to be more sensitive to differences in aspiration than native Japanese speakers for whom aspiration is used only at word boundaries.

The Mandarin speakers tested in this study did not show any difference in their processing of cross-boundary and word-initial clusters. Because Mandarin uses an aspiration contrast with VOT categories that match those found in English, the participants of this experiment had experience using both short- and long-lag VOT in word recognition and were able to adapt this ability for use as a word boundary cue in English. This was expected based on native language experience, however, Mandarin speakers do not use aspiration as a word boundary cue, so their high accuracy in identification using this adapted cue is interesting.

One advantage to using the visual world paradigm is that measuring the target and competitor fixations can inform understanding of word activation in real-time speech processing. The models of activation for both monolingual (Gow & Gordon, 1995; Marslen-Wilson & Welsh, 1978; McQueen et al., 1995; Norris et al., 1995) and bilingual (Dijkstra & van Heuven, 2002; Grosjean, 1997; Shook & Marian, 2013) word recognition all make similar predictions with regards to how target and competitor activation progresses. The results of this study overall fit the patterns of activation predicted by these models. For both the native English and native Mandarin participant groups, fixations to the target and competitor items began to increase in parallel as the relevant linguistic information was processed. As the speech input continued, participants began to suppress their fixations to competitor items in favor of the target. However, when considered together, the results of the target and competitor fixations in this experiment showed that compared to the native English baseline, native Mandarin speakers were more unsure of their processing overall.

There were two main differences that highlight the uncertainty that native Mandarin speakers had during real-time processing: the initial activation of on-screen items and the overall activation of target and competitor items. Starting at the onset of trials, native English speakers had low fixations to target and competitor items until the phonetic input provided enough evidence to activate the on-screen candidates. For the native English speakers, this point in the phonetic input came after the frame “click on _____” suggesting that these speakers learned throughout the experiment that the information relevant to word recognition always followed this frame. For the native

Mandarin speakers on the other hand, fixations to the on-screen candidates began earlier in the input signal and increased gradually throughout the trials. This result suggests that even if the Mandarin speakers did learn throughout the experiment that the frame did not provide enough information to limit the candidate list, they began to fixate the on-screen items before the onset of the relevant information. In addition to the earlier fixations to the on-screen items, native Mandarin speakers also had a lower degree of fixations to the target items and a higher degree of fixations to the competitor items when compared to the native English speaking baseline. The differences in degree of fixations indicate that the native Mandarin speakers had more similar levels of activation to both items during processing following the presentation of the word boundary cues. These similar levels of activation reflect the uncertainty of their responses despite their high degree of accuracy in the task.

It was unclear from the results of this study whether the experimental methodology was not sensitive enough to discern any differences in the processing of cross-boundary and word-initial clusters or if the sample of native Mandarin speakers tested in this experiment were just too good at the task. Previous studies found a trend toward having higher accuracy in the identification of cross-boundary clusters as compared to word-initial clusters. In order to try to replicate this trend, a second experiment was run to test a previously tested language group using the same methodology used with the native Mandarin speakers in this experiment.

Chapter 3. Experiment 2

While native Mandarin speakers showed some differences from the native English baseline in the measures of real-time processing, the similarities found between the two groups raised the possibility that the chosen methodology was not sensitive to the differences that I expected to find. In order to test this I conducted a second experiment to test native speakers of a language with no aspiration contrast to see whether speakers who did not have an adaptable aspiration process and had to learn a new cue reflected this in their real-time processing.

This experiment was conducted separately from Experiment 1 due to difficulties in participant recruitment. The sample size of this experiment is not large enough to allow for any direct statistical comparison to the native English and native Mandarin participant groups in the previous experiment. As such, it is being treated here as a preliminary study and all comparisons to Experiment 1 will be observational.

3.1. Language of Interest

This experiment tested the accuracy and real-time processing of native French speakers. French speakers were tested in one of the previous studies motivating this research (Shoemaker, 2014), which allowed for a clearer comparison to the results of the previous studies. French has voiceless unaspirated stops which are distinguished from voiced stops by voicing and not by aspiration (Tranel, 1987; Walker, 2001). This means that in French, voiced stops fall in the lead VOT category and voiceless stops have short lag (Tranel, 1987). The phonotactics of French allow the word-initial *s*+stop clusters /sp, st, sk, sm, sn/ and allow the possibility of cross-boundary *s*+stop clusters because /s/ is a viable final consonant and stops are possible word-initially. Examples of word-initial and cross-boundary *s*+stop clusters are provided in Table 6.

Table 6. Word-initial and cross-boundary s+stop clusters in French

Word-initial *s*C clusters	**Cross-boundary *s*C clusters**
[spoʁtif] ‘athletic’	la piè[s] [p]rincipale ‘the main room’
[stilo] ‘pen’	la piè[s] [t]urquoise ‘the turquoise piece’
[skjœʁ] ‘skier’	la piè[s] [k]omptable ‘the accounting document’

While it is the case that native French speakers do not use aspiration as a word boundary cue in English, they do use duration. Unlike English, which associates greater loudness and, to a certain extent, longer duration with stressed syllables, stress in French is associated with a longer syllable duration (Tranel, 1987; Walker, 2001). Because the final syllable of words in French is typically stressed, this predictable pattern allows French speakers to use stress and its durational correlate to identify word boundaries (Cutler et al., 1986, 1992; Goetry & Kolinsky, 2000). A study on the segmentation of ambiguous sequences in French showed that when the first syllable of a word was stressed the participants were more likely to say they were hearing two monosyllabic words, but when the second syllable was stressed the participants were more likely to say they were hearing one bisyllabic word (Banel & Bacri, 1994). Although this finding mirrors the use of consonant duration as a word boundary cue in the segmentation of English *s*+stop clusters, native French speakers are expected to be able to adapt their knowledge of durational cues into English segmentation.

3.2. Hypotheses

Based on the results of the Shoemaker (2014) study that tested native French speakers on their ability to differentiate English phrases which differed in the presence or absence of aspiration, I expected that native French speakers would be less accurate than the native English and native Mandarin speakers tested in the previous experiment. Because French does not have any systematic aspiration, native French speakers who learn English have to learn to use allophonic aspiration as a word boundary cue. In addition to this, the properties of consonant duration found in French would bias English learners to look for a word boundary following a stressed syllable. With the trochaic stress pattern typical of English, this bias from durational cues could cause further difficulty for the native French participants.

Regarding the measures of real-time processing, native French participants were expected to have slower overall processing than the native English baseline and the native Mandarin participants because French does not have an aspiration contrast to rely on in segmentation in addition to the activation of within- and between-language competitors and potential uncertainty of interpretations (Blumenfeld & Marian, 2007, 2013; Farris-Trimble et al., 2014; Ju & Luce, 2004; Marian & Spivey, 2003a, 2003b; Marian et al., 2003; McMurray et al., 2010; Weber & Cutler, 2004). As with the previous experiment, this hypothesis would be supported in the data by slower fixations to target items and slower suppression of competitor items. The absence of an aspiration contrast in French was also expected to decrease the overall activation of target words compared to the language groups tested in Experiment 1. This hypothesis would be supported by a lower maximum proportion of fixation to target items.

3.3. Participants

A total of seven native French speakers were recruited from the Simon Fraser University campus and the Metro Vancouver area. As in Experiment 1, participants were between the ages of 19-45 and had normal or corrected-to-normal vision and normal hearing. Participants received $10 or Linguistics course credit for their participation. Demographic information collected for the native French participants compared to the native English and Mandarin participants from the previous experiment is provided in Table 7. The differences in proficiency between the three participant groups show that the native Mandarin and native French speakers performed similarly on the proficiency task when compared to the native English speakers.

Table 7. Experiment 2 demographic information

	Native English	Native Mandarin	Native French
Sample size	21	20	7
Mean proficiency	80% (45-95 %)	45% (10-95 %)	46% (0-80%)
Mean age began learning		7.15 (4-15)	9.43 (2-13)
Mean years of learning		14.6 (7-18)	12.33 (6-26)
Mean age of arrival		16.45 (4-21)	13.43 (0-40)
Mean years of residence		5.38 (1-15)	13.21 (0.25-35)

Demographic information presented as mean (range)

3.4. Stimuli

The stimuli used in this experiment are identical to those used in the first experiment.

3.5. Procedure

The procedures used in this experiment are identical to those used in the first experiment.

3.6. Results

Because the small sample size of this study does not lend itself to reliable statistical analysis, I will present primarily descriptive results in this section. However, statistical analyses will be presented for within-group comparisons of the two cluster types.

3.6.1. Production Measures

Recordings of participants underwent the same processing as the auditory stimuli and Experiment 1 recordings. Average /ɪ/-duration in “this” can be found in Figure 11, average /s/-duration across “this” and word-initial *s*+stop clusters can be seen in Figure 12, and the average VOT duration of the voiceless stops can be seen in Figure 13. To investigate the durational properties of French cross-boundary and word-initial *s*+stop clusters, paired t-tests were run to determine any significant differences. Like with the native English and Mandarin production analyses, there were excluded tokens for five French speakers (range of excluded tokens: 1 to 4) in cases where these participants produced a pause between the /s/ in “this” and the /s/ in word-initial *s*+stop clusters. As in first experiment, these participants were included in the analyses. The remaining two native French speakers produced all tokens in the production task. In terms of the vowel duration, /ɪ/ was shorter in cross-boundary clusters than in word-initial clusters ($t(6) = 3.0$, $p = .025$). This was also true of /s/-duration; /s/ was shorter in cross-boundary clusters than in word-initial clusters which was expected based on the properties of these clusters in English ($t(6) = 7.7$, $p < .001$). For VOT duration, VOT was longer in cross-boundary clusters than word-initial clusters which results from the environment of allophonic aspiration which is present for cross-boundary but not word-initial clusters ($t(6) = 3.3$, $p = .016$). Looking at the graph in Figure 13, the difference in VOT for the two cluster types for native French speakers was clearly smaller than that for native Mandarin and native English speakers. This difference is reflective of the use of aspiration contrasts across the three languages.

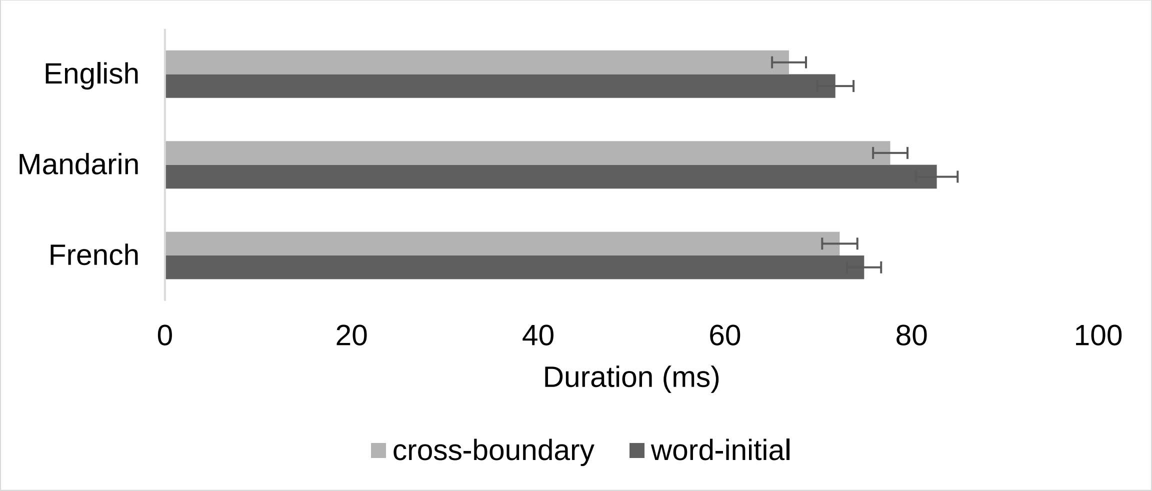

Figure 11. **Average /ɹ/ duration by group and by cluster type**

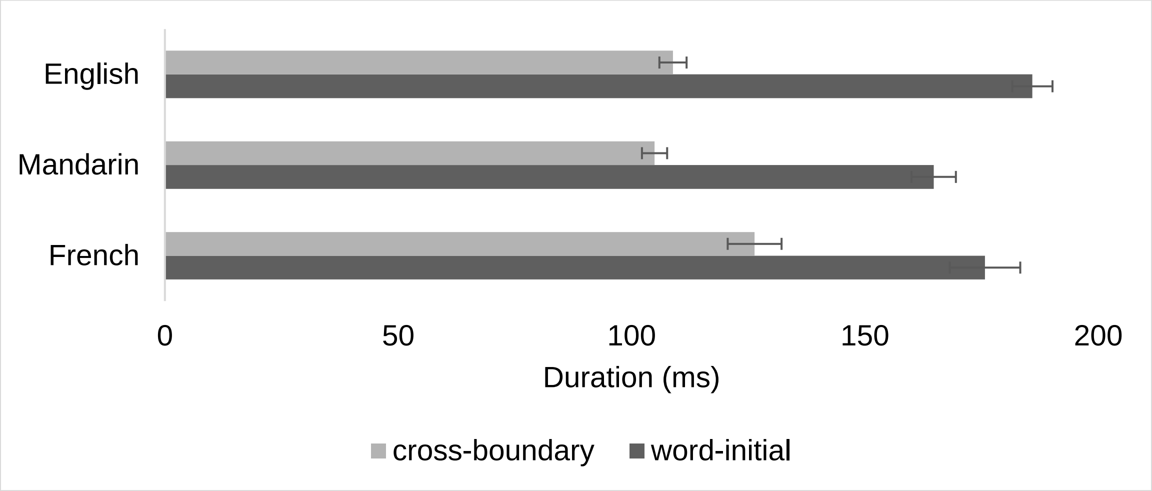

Figure 12. **Average /s/-duration by language and by cluster type**

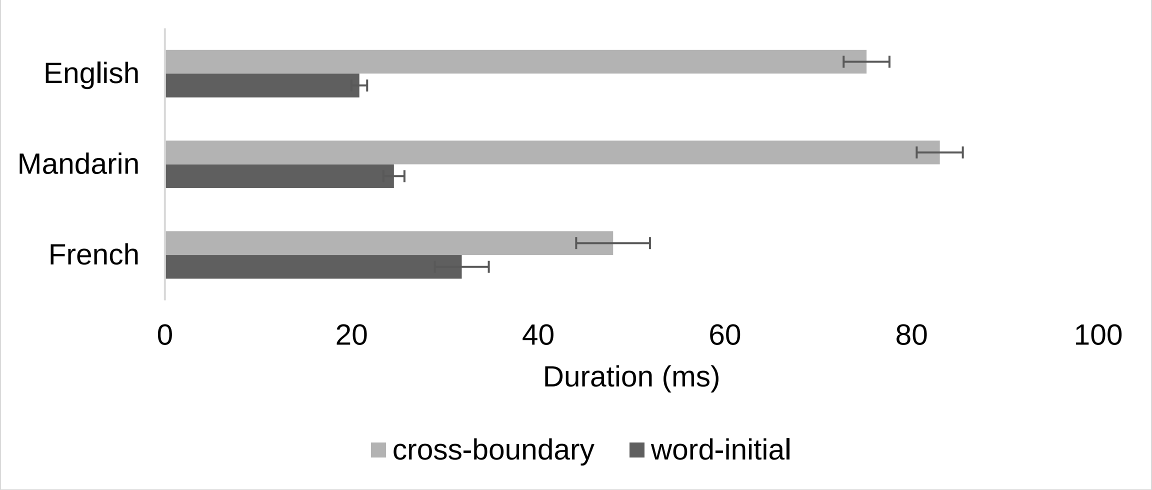

Figure 13. **Average VOT duration by language and by cluster type**

3.6.2. Identification Measures

A comparison of accuracy between cross-boundary and word-initial *s*+stop clusters showed no significant difference of cluster type on accuracy for the native French speakers (*t*(6) < |1|). This was the same result found for both native Mandarin and Native English speakers. Overall accuracy of the French speakers was quite close to that of the native Mandarin speakers (93%, compared to 92% in the Mandarin group) and the range of identification accuracy was also similar (81-98% for French speakers compared to 82-99% for Mandarin speakers). Figure 14 shows accuracy by cluster type for each of the three tested language groups. A comparison of the two cluster types for reaction time for the native French speakers once again showed no significant difference (*t*(6) < |1|). The overall reaction time of native French speakers was closer to the average reaction time for native Mandarin than native English participants (2068 ms, compared to 2072 ms for the Mandarin group and 1707 ms for the Engslish group). Again the range of reaction times for native French compared to native Mandarin speakers was similar (1707-2640 ms for French speakers, 1674-2589 ms for Mandarin speakers). Figure 15 shows reaction time by cluster type for each of the three tested language groups.

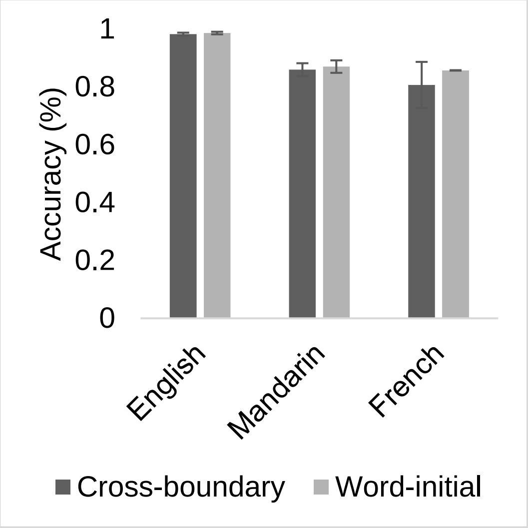

Figure 14. Accuracy by cluster type comparison to French

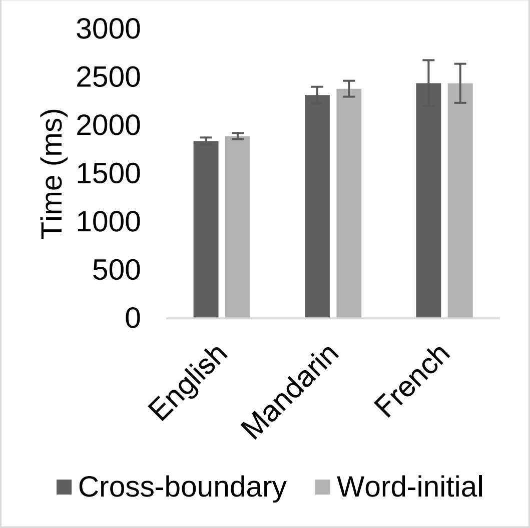

Figure 15. Reaction times by cluster type comparison to French

3.6.3. Fixation Measures

In this discussion of the real-time processing of native French speakers I will begin with a visual comparison of French fixations to the fixations of the two language groups tested in Experiment 1. Figure 16 shows the overall target fixations for all three languages. Based on this graph it is clear that the degree of fixations shown by the native French participants was somewhere between the native English and native Mandarin speakers. However, native French speakers appear to have a slope that was more closely matched to the native Mandarin speakers tested in the previous experiment. From the target fixations alone it seems like the native French speakers had similar timing of fixations to the native Mandarin speakers, but that they were more confident in their final responses as reflected in their maximum proportion of fixations.

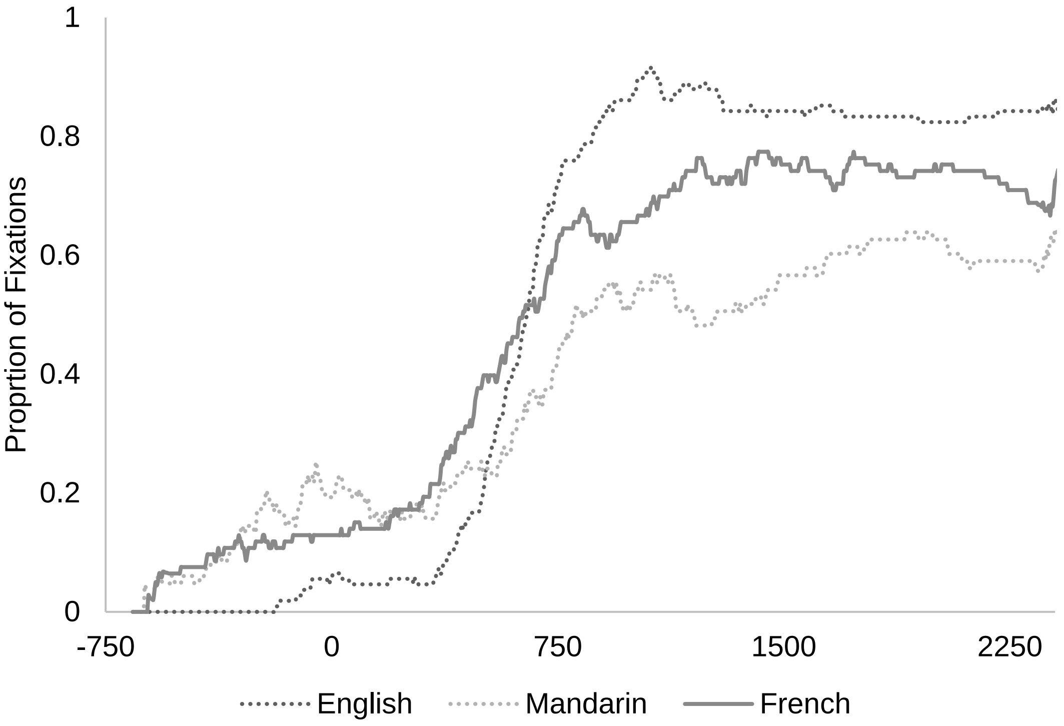

Figure 16. French, English, and Mandarin target fixations

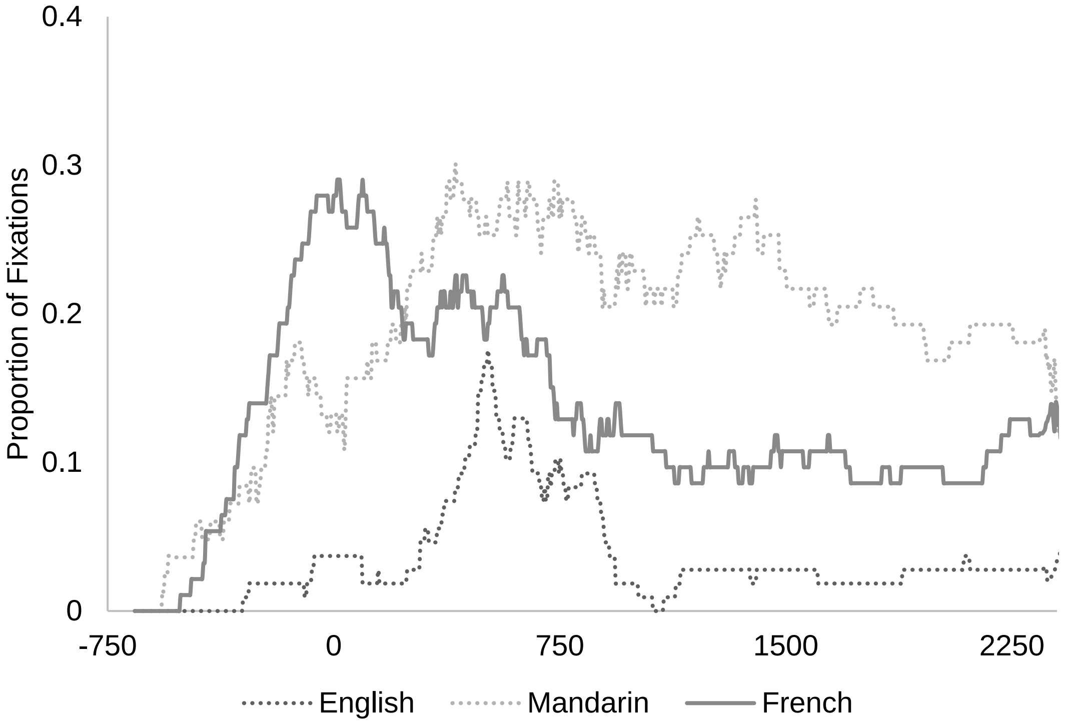

Figure 17. **French, English, and Mandarin competitor fixations**

When looking at the competitor fixations in Figure 17, the native French speakers had a similar degree of fixations to the native Mandarin speakers, however the maximum proportion of fixations appears to have been earlier in the processing for the French speakers. The timing of fixations for native French speakers did not clearly match either of the previously tested groups as the French speakers had a much earlier peak in their fixations to the competitors which were gradually suppressed for the remainder of processing and resulted in a lower final baseline of fixations when compared to the native Mandarin speakers. Based on these visual comparisons, the native French speakers shared similar timings of fixations to the native Mandarin speakers. Where the two groups differed, as with the degree of fixations to target and competitor items, may have been a by-product of the demographic characteristics of the two groups. Although the native Mandarin speakers started learning English at an earlier age, and spent more time actively learning English, approximately half of the native French speakers were French-Canadian and the other half had spent more time living in North America than the native Mandarin speakers. Having spent more time in North America would have given the native French speakers more passive English exposure than the native Mandarin speakers.

To compare the effects of the two cluster types on target and competitor fixations for the native French participants, paired t-tests were run on the degree and timing measures for both target and competitor fixations. For target fixations (see Figure 18), cluster type showed no effect on peak of target fixations (*t*(6) < |1|), crossover point (*t*(6) = 1.5, *p* = .194), or slope (*t*(6) < |1|). The result was the same for competitor fixations (see Figure 19), with no effect of cluster type on peak of competitor fixations (*t*(6) < |1|), final degree of competitor fixations (*t*(6) < |1|), time of peak fixations (*t*(6) = 1.6, *p* = .161), or onset slope (*t*(6) = 1.3, *p* .237). This lack of difference showed that the native French speakers were able to use the word boundary cues used in cross-boundary and word-initial clusters to process the incoming speech signal.

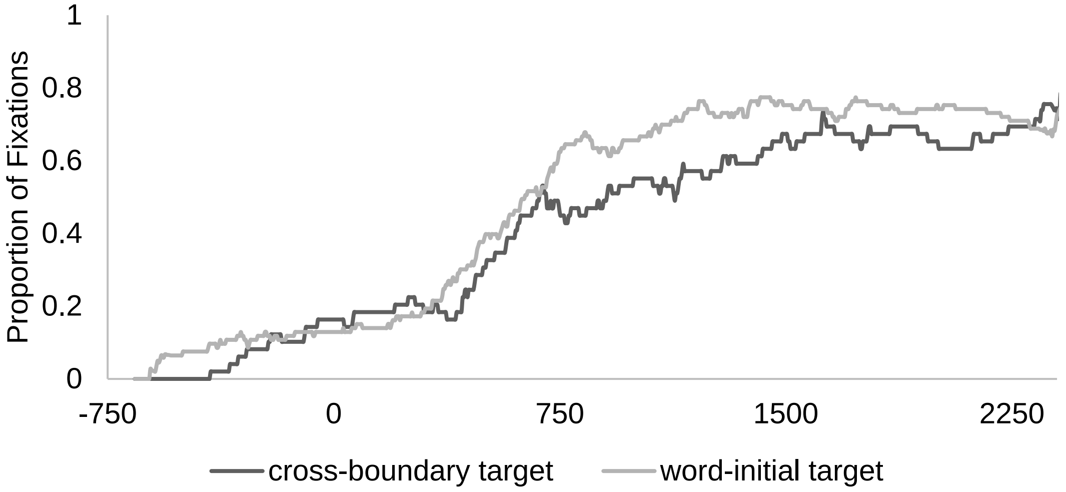

Figure 18. **French target fixations by cluster type**

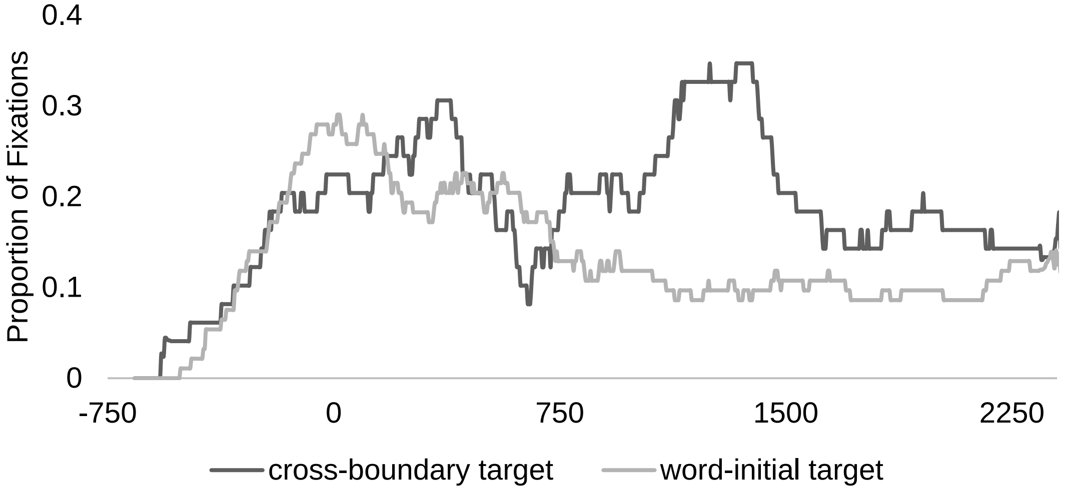

Figure 19. **French competitor fixations by cluster type**

3.7. Discussion

The small sample size of this experiment makes it difficult to make direct statistical comparisons to the results of Experiment 1. However, an observational analysis suggested that the native French participants did not pattern in the expected way. Based on the fact that French does not have any systematic process of aspiration, I expected the native French participants to have lower accuracy and slower overall processing than both participant groups tested in Experiment 1. However, the native French speakers showed a level of accuracy and processing which patterned similarly to the native Mandarin participant group. One particularly interesting observation was that the proportion of fixations to both target and competitor items by native French speakers appeared to fall between the proportion of fixations for the other two language groups. These differences suggest that the native French speakers had less sustained competition between the on-screen candidates as compared to the native Mandarin participants. Overall, the findings of this experiment suggest that the native French speakers were able to use one or more of the English word boundary cues relevant to the segmentation of cross-boundary and word-initial *s*+stop clusters.

An interesting observation resulting from the visual comparison of competitor fixations (Figure 19) for the native French speakers was the apparent double peak in fixations to competitors in trials where the target was a cross-boundary cluster. Because these trials occurred over a very short period of time (a maximum of three seconds), it was not expected that participants would have two separate increases in fixations to any item on the screen. To further investigate the apparent double peak in competitor fixations, I plotted individual participant fixations to competitor items for cross-boundary target trials (Figure 20). This figure showed that the later peak in competitor fixations in these trials was likely driven by the competitor fixations of two of the seven native French speakers. Additionally, in looking at the curvefit values for the timing of peak fixations to competitor items in these trials, the same two native French speakers had later peak fixations than the other speakers.

Figure 21**Error! Reference source not found.** shows the relevant competitor fixations for the two participants who did not match the pattern of the majority of the French speakers. In this fixation plot it does seem as though these two participants did each have two peaks in fixations to the competitor items. It is unlikely that these two peaks represent two increases in fixations to the same competitor in the same trial. It is more likely that the separate peaks reflect a bimodal distribution of fixations across trials; in some trials these participants were able to fixate earlier, and in others, they delayed. Some potential reasons for a bimodal distribution of fixations could be a difference in high- and low- frequency or familiarity words. In looking at the demographic information collected from these two speakers, they were the participants who had the oldest ages when arriving in North America, they had lived in North America for the least amount of time, and had the lowest scores on the proficiency test. Based on the differences in the demographic characteristics of these two French speakers compared to the other five French speakers, it seems likely that exposure to English in an immersion environment may have been a factor in real-time processing. Having less exposure to English in daily life may have caused these speakers to increase fixations to the word-initial cluster competitor items after hearing the aspiration in the cross-boundary target because the presence of aspiration in the input was unexpected based on language experience and created uncertainty.

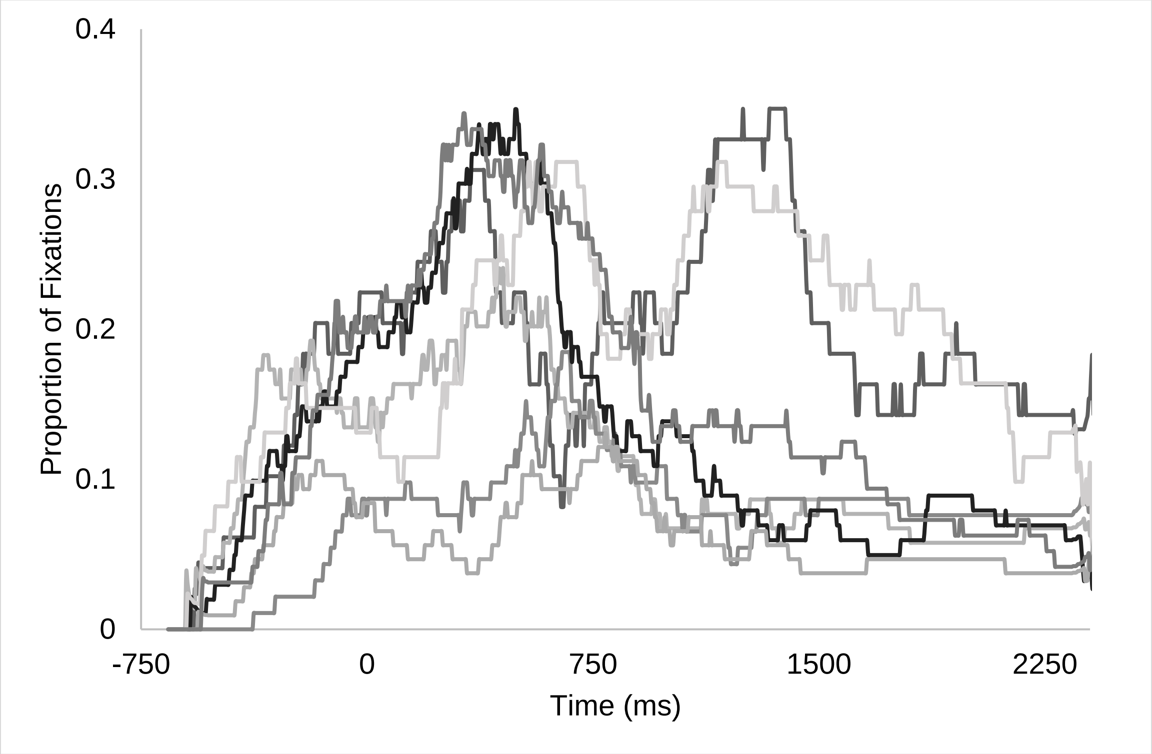

Figure 20. **French individual competitor fixations in cross-boundary target trials**

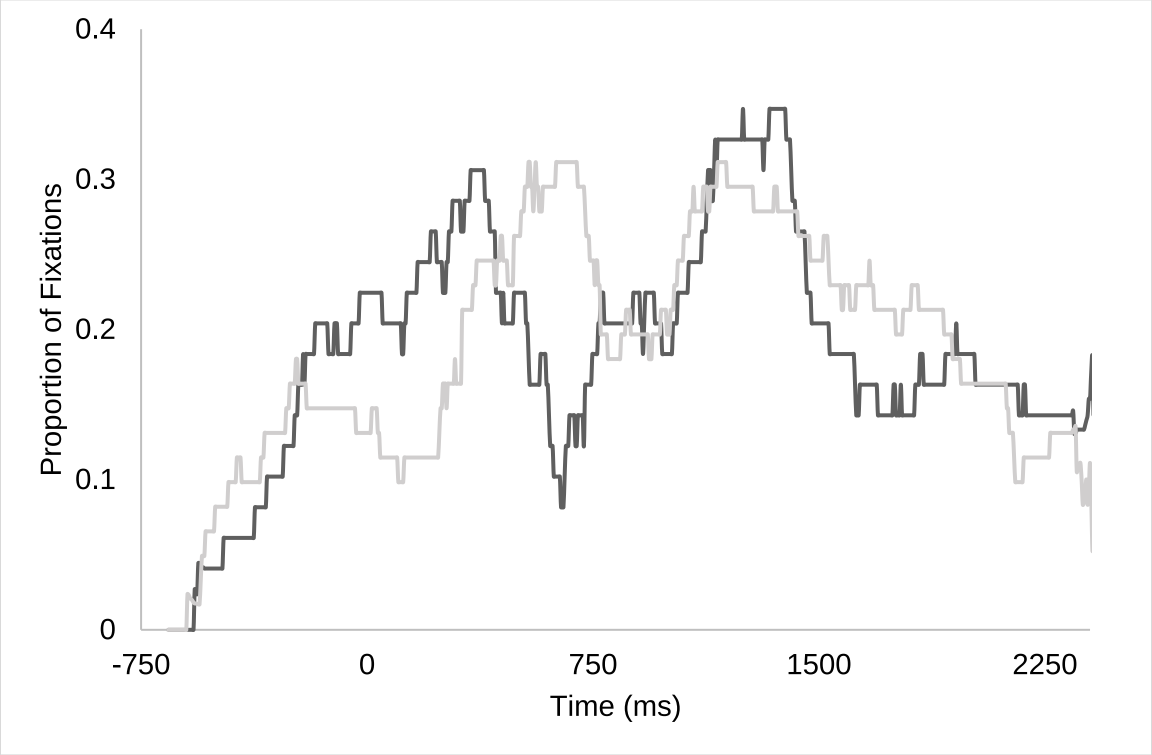

Figure 21. **French individual competitor fixations in cross-boundary target trials for two outlier participants**

Although the results of this experiment are representative of a small sample size, a comparison to the accuracy of native French speakers tested in a previous study (Shoemaker, 2014) can tentatively be made. The native French participants tested in this experiment show a level of accuracy that is more similar to the native English baseline of Experiment 1 than to the accuracy of native French participants tested by Shoemaker (2014). As was the case in Experiment 1, this difference in accuracy found between the native French speakers in the two studies likely resulted from differences in proficiency levels of the two groups. Shoemaker tested French speakers living in France studying in an English language program, while the participants in this study were all residents of the Metro Vancouver area where the majority language is English.

When connecting the fixation results of this experiment to the process of word activation, the native French speakers followed the structure described by models of bilingual word activation (Dijkstra & van Heuven, 2002; Grosjean, 1997; Shook & Marian, 2013). Activation of the on-screen candidates increased in parallel as the input signal was processed. As the signal continued, participants began to suppress their activation of the competitor item leading to a higher activation of the target item. Similar to the processing of the native Mandarin speakers, the native French speakers showed more uncertainty in their fixations to target and competitor items. Both language groups had earlier fixations to on-screen items, a lower degree of fixations to target items, and a higher degree of fixations to competitor items compared to the native English speakers. However, the native French speakers did seem to suppress their fixations to the competitor items earlier and to a greater extent than the native Mandarin speakers which suggested that the French speakers were able to limit their candidate sets more as they received the auditory signal.

Chapter 4. General Discussion

Overall, the results of these two experiments do not provide clear support for the idea that adapting a cue gives a processing advantage compared to learning a new cue. Native French and Mandarin participants showed similar results for accuracy, reaction time, and real-time processing, which was unexpected based on the results of previous studies (Altenberg, 2005; Ito & Strange, 2009; Shoemaker, 2014). Both of these tested language groups were also equally good at the segmentation of word-initial and cross-boundary *s*+stop clusters, suggesting that they were able to use at least one of the two relevant English word boundary cues.

Despite not following the predicted patterns based on language background, the results of these two experiments did provide support for the models of monolingual (Gow & Gordon, 1995; Marslen-Wilson & Welsh, 1978; McQueen et al., 1995; Norris et al., 1995) and bilingual word activation (Dijkstra & van Heuven, 2002; Grosjean, 1997; Shook & Marian, 2013). When hearing the onset of the auditory input, the two competing on-screen items were both activated in parallel. As more of the signal was received, the candidate set was updated and fixations to the competitor item were suppressed in favor of the target item. This difference in activation was found to be the case for all three language groups tested. The main differences found between monolingual and bilingual processing were the decreased fixation of target items and increased fixation of competitor items in bilingual word recognition.

The factors of activation presented in BIMOLA (Grosjean, 1997) seem to best account for the unexpected similarity of identification and processing for native and non-native English speakers. One of the characteristics of this model is that the language mode influences the degree of activation for the *'base'* (native) and *'guest'* (non-native) languages. The language mode exists on a scale where monolingual in one language exists at one end, bilingual language mode exists in the middle, and monolingual in the other language exists at the other end. The environment in which an interaction is taking place dictates the level of activation on this scale. That is, if a bilingual speaker goes into an experiment where they are addressed in English, are reading English instructions, and are hearing English words, they will likely be in a monolingual English mode and as a result, English will be more active than their other language. For both language groups tested in these experiments, the environment of the study encouraged use of the English monolingual language mode as opposed to the Mandarin or French monolingual mode. As a result, these participants would have had stronger activation of English than of their native language which may have added to the similarities to the native English speaking baseline.

Another characteristic of BIMOLA is that it assumes that the level of activation for between-language competitors depends on the degree of similarity between the sounds of the two languages. For example, the difference between English word-initial /p/ and French word-initial /p/ would be strong enough to support the activation of English words without activating French words as strongly. This characteristic can help to explain the differences in observed processing between the native Mandarin and French participants. For the native Mandarin participants, voiceless aspirated and unaspirated stops in Mandarin and English occupy the same categories along the VOT continuum. This similarity would lead to the activation of both English and Mandarin candidates during word recognition. On the other hand, as was explained in the example above, native French speakers would not be expected to activate French words upon hearing a word-initial voiceless stop in English. As a result, levels of activation for within- and between-language competitors would differ between the native Mandarin and native French participants such that the native French participants would have fewer between-language competitors. These characteristics of BIMOLA are able to explain the visually observed differences between activation in Mandarin and French speakers while still accounting for the high levels of accuracy found in both groups.

Native Mandarin and French speakers showed more overall uncertainty in their fixations, which was not reflective of their overall accuracy of cluster identification. This uncertainty is clear despite the analysed fixations being representative of only trials where participants selected the target as the item that they heard. The main parameters that reflected this uncertainty were the measures of degree of target and competitor fixations. The native Mandarin and French speakers had a lower degree of target fixations and a higher degree of competitor fixations across trials when compared to the native English speakers. This difference in fixations suggested that targets were less active and competitors were more active than is typical in native speakers. Maintaining a lower degree of target activation and a higher degree of competitor activation is a strategy for dealing with uncertainty that has also been found in the process of word recognition by participant groups with atypical language (Farris-Trimble et al., 2014; McMurray et al., 2010). In these clinical populations, the source of uncertainty is general signal degradation due to language or cognitive impairment or the use of cochlear implants. These previous studies have shown that one possible strategy for dealing with this uncertainty was to decrease target activation and increase competitor activation. By adopting this strategy, populations dealing with signal degradation have a clear repair strategy for incorrect processing; maintaining a higher level of activation to competitor items would make it easier to adjust their interpretation of the input. For the non-native English speakers tested in the present experiment, there are a variety of possible sources of uncertainty including having an incomplete vocabulary, limited listening experience, or gaps in the surrounding context (Field, 2003, 2008). Despite the difference in the source of uncertainty compared to the clinical populations, the native Mandarin and French participants of this study applied the same strategy of maintaining competitor activation that would allow for easier repairs to the interpretation. This comparison is not to say that bilingualism is a clinical problem, however, this similarity in the strategies for dealing with uncertainty in processing suggests that humans in general have a way of adapting to uncertainty regardless of its source.

4.1. Limitations and Future Directions

The main limitation of this study was the small sample size for the native French participant group. Having such a small sample made statistical comparisons between all three participant groups impossible due to a lack of statistical power. A continuation of the second experiment is needed to allow for these comparisons to be made in order to provide a clearer picture of how language background influences second language speech segmentation.

To further connect the results of real-time processing to the previous studies on second language segmentation of English *s*+stop clusters, further research would benefit from recruitment of second language learners of a lower proficiency than those tested in this research. Although the demographic characteristics of participants reported in previous studies and the current study were not all the same, the participants in previous studies were either late learners of English or were enrolled in English language studies at the time of testing (Altenberg, 2005; Ito & Strange, 2009; Shoemaker, 2014). The participants in the current study were all residing in the Metro

Vancouver area and were enrolled in university classes taught in English or were speaking English at their places of employment. A population of speakers which may be better suited for comparison to the previous studies could be recruited from Fraser International College and other advanced English language classes in the Metro Vancouver area.

The use of naturally produced stimuli in this experiment allowed participants to use all available sublexical cues in the process of speech segmentation. While this presented a more naturalistic analysis of real-time processing, using natural speech made it difficult to identify which word boundary cues were more or less important to the participants from the three tested language groups. In order to determine the weight of these cues for participants from different language backgrounds it would be necessary to run a study where /s/-duration and VOT duration of the stimuli are manipulated. This could be done by removing the effect of one of the two cues or by manipulating the relevant durations to maintain a specific value across the individual tokens of each cluster type. By making one of these manipulations to the stimuli and using the same methodology it would be possible to identify which cue(s) were more important to a participant group as the timing of fixations could be more clearly matched to that of the relevant cues.

In the analyses of this study I chose not to look at individual differences in participants and items because the main questions that I was asking related to the speech processing of groups of speakers. However, the collected data will allow for these analyses to be done in the future. Some analyses of interest are to investigate how word frequency of individual items influenced the processing of the pairs of experimental items and to look into individual differences in processing for the native Mandarin and French speaker groups based on the collected demographic information. It would also be interesting to look at whether an individual's productions are related to their use of a specific cue in the segmentation of cross-boundary and word-initial clusters.

In addition to the experimental directions that can be taken based on the results of this study, there are further factors that can be investigated which were not controlled for in the current experiments. During the development of the stimuli used in these experiments, between-language competitors were not controlled for. As a result, it is possible that competitor activation for native Mandarin and French participants could have been influenced by the items present on the screen during each trial. By controlling for between-language competitors and the influence of cognates, future research would be better able to follow the process of activation resulting from the on-screen competitors.

The results of this study motivate further research into the adaptation and acquisition of the word boundary cues used by English speakers in the segmentation of cross-boundary and word-initial *s*+stop clusters. For example, future research might look at how English second language speakers of Mandarin learn to use the acoustic properties of their allophonic aspiration process in a contrastive process, or how English learners of French learn to rely more on stress and duration as the cues to word boundary location. Additional research is also needed to look at how other cues are used in speech segmentation and how those cues affect processing. The previous studies which motivated the current research (Altenberg, 2005; Ito & Strange, 2009; Shoemaker, 2014) all made comparisons between identification of boundaries marked by aspiration and boundaries marked by glottalization before word-initial vowels. The results of all three experiments showed greater accuracy in identification of phrases differentiated by glottalization which led to a hypothesis that glottalization is a “universal” phonetic effect that can be used regardless of language background. However, as highlighted in the motivations for the current research, these studies were focused on a small number of languages with specific acoustic cues. Future research would benefit from an investigation of accuracy and real-time processing of word boundary identification using cues like glottalization by speakers from language backgrounds where different boundary cues are more or less prevalent.

The main goal of this study was to determine how different language backgrounds influence the real-time processing of English as a second language. The results of the two conducted experiments did not lend support to the idea that having a contrast that could be adapted as a word boundary cue (as in Mandarin) as opposed to learning a new word boundary cue (as in French) would lead to more efficient second language processing. Instead, this study provided further support for models of language activation that have been proposed for monolingual and bilingual speech processing.

Reference

Allopenna, P., Magnuson, J., & Tanenhaus, M. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/http://dx.doi.org.proxy.lib.sfu.ca/10.1006/jmla.1997.2558

Altenberg, E. (2005). The perception of word boundaries in a second language. *Second Language Research*, *21*(4), 325–358. https://doi.org/10.1191/0267658305sr250oa

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8 month old infants. *Psychological Science*, *9*(4), 321–324. https://doi.org/http://dx.doi.org/10.1111/1467-9280.00063

Banel, M.-H., & Bacri, N. (1994). On metrical patterns and lexical parsing in French. *Speech Communication*, *15*, 115–126. https://doi.org/http://dx.doi.org/10.1016/0167-6393(94)90046-9

Barry, W. (1981). Internal juncture and speech communication. In W. J. Barry & K. J. Kohler (Eds.), *Beitrage zur experimenteelen und angewandten phonetetik* (pp. 229–289). Kiel, Germany: AIPUK.

Berko Gleason, J., & Bernstein Ratner, N. (2013). *The Development of Language* (8th ed.). Upper Saddle River, NJ: Pearson.

Best, C., & Tyler, M. (2007). Nonnative and Second Language Speech Perception. Commonalities and complementarities. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (Vol. 10389, pp. 13–34). https://doi.org/10.1075/lllt.17.07bes

Bion, R., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, *54*(1), 123–140. https://doi.org/10.1177/0023830910388018

Blumenfeld, H., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633–660. https://doi.org/10.1080/01690960601000746

Blumenfeld, H., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology*, *25*(5), 547–567. https://doi.org/10.1080/20445911.2013.812093

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer. Retrieved from http://www.praat.org/

Brown, C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, *14*(2), 136–193. https://doi.org/10.1191/026765898669508401

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Carroll, S. (2004). Segmentation: Learning how to “hear words” in the L2 speech stream. *Transactions of the Philological Society*, *102*(2), 227–254. https://doi.org/10.1111/j.0079-1636.2004.00136.x

Chao, K., & Chen, L. (2008). A Cross-Linguistic Study of Voice Onset Time in Stop Consonant Productions. *Computational Linguistics and Chinese Language Processing*, *13*(2), 215–232.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229. https://doi.org/http://dx.doi.org.proxy.lib.sfu.ca/10.1006/jpho.1999.0094

Christie, W. M. (1974). Some cues for syllable juncture perception in English. *The Journal of the Acoustical Society of America*, *55*(4), 819–821. https://doi.org/10.1121/1.1914606

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, *31*(3), 585–598. https://doi.org/10.1016/S0095-4470(03)00040-8

Church, K. (1987). Phonological parsing and lexical retrieval. *Cognition*, *25*(1), 53–69. https://doi.org/10.1016/0010-0277(87)90004-7

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/http://dx.doi.org/10.1016/S0022-5371(73)80014-3

Cutler, A. (2002). Listening to a second language through the ears of a first. *Interpreting*, *5*(1), 1–23. https://doi.org/10.1075/intp.5.1.02cut

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236. https://doi.org/10.1016/0749-596X(92)90012-M

Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2*(3–4), 133–142. https://doi.org/10.1016/0885-2308(87)90004-0

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's different role in segmentation of French and English. *Journal of Memory and Language*, *25*, 385–400. https://doi.org/http://dx.doi.org/10.1016/0749-596X(86)90033-1

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*(3), 381–410. https://doi.org/10.1016/0010-0285(92)90012-Q

Cutler, A., & Norris, D. (1988). The Role of Strong Syllables in Segmentation for Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113–121. https://doi.org/10.1037/0096-1523.14.1.113

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Duanmu, S. (2000). *The Phonology of Standard Chinese*. New York, USA: Oxford University Press.

Farris-Trimble, A., & McMurray, B. (2013). Test–Retest Reliability of Eye Tracking in the Visual World Paradigm for the Study of Real-Time Spoken Word Recognition. *Journal of Speech Language and Hearing Research*, *56*(4), 1328–1345. https://doi.org/10.1044/1092-4388(2012/12-0145)

Farris-Trimble, A., McMurray, B., Cigrand, N., & Tomblin, J. B. (2014). The process of spoken word recognition in the face of signal degradation. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 308–327. https://doi.org/10.1037/a0034353

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, *57*(4), 325–334. https://doi.org/10.1093/elt/57.4.325

Field, J. (2008). Revising segmentation hypotheses in first and second language listening. *System*, *36*(1), 35–51. https://doi.org/10.1016/j.system.2007.10.003

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). Timonium, MD: York Press.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*(2), 126–152. https://doi.org/10.1177/002383095800100207

Goetry, V., & Kolinsky, R. (2000). The role of rhythmic cues for speech segmentation in monolingual and bilingual listeners. *Psychologica Belgica*, *40*(3), 115–152. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2001-16248-001&site=ehost-live&scope=site

Gow, D. W., & Gordon, P. C. (1995). Lexical and Prelexical Influences on Word Segmentation: Evidence From Priming. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(2), 344–359. https://doi.org/10.1037/0096-1523.21.2.344

Grosjean, F. (1997). Processing mixed language: issues findings and models. In A. M. B. Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: psycholinguistic perspectives* (pp. 225–254). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Ito, K., & Strange, W. (2009). Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *The Journal of the Acoustical Society of America*, *125*(4), 2348–2360. https://doi.org/10.1121/1.3082103

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, *15*(5), 314–318. https://doi.org/10.1111/j.0956-7976.2004.00675.x

Klatt, D. (1975). Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters. *Journal of Speech, Language, and Hearing Research*, *18*, 686–706. https://doi.org/http://dx.doi.org/10.1044/jshr.1804.686

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208–1221. https://doi.org/10.1121/1.380986

Lehiste, I. (1960). An acoustic – phonetic study of internal open juncture. *Phonetica*, *5*, 5–54. https://doi.org/10.1159/000258062

Lin, H. (2001). *A grammar of Mandarn Chinese*. Munich, Germany: Lincom Europa.

Lin, Y.-H. (2007). *The Sounds of Chinese*. New York, USA: Cambridge University Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133–156. https://doi.org/10.1080/03640210709336987

Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, *24*(2), 173–193. https://doi.org/10.1017/S0142716403000092

Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97–115. https://doi.org/10.1017/S1366728903001068

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, *86*(1), 70–82. https://doi.org/10.1016/S0093-934X(02)00535-7

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63. https://doi.org/http://dx.doi.org/10.1016/0010-0285(78)90018-X

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception and Psychophysics*, *62*(2), 253–265. https://doi.org/10.3758/BF03205547

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91–121. https://doi.org/10.1016/S0010-0277(00)00109-8

Mattys, S. L., & Melhorn, J. F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *The Journal of the Acoustical Society of America*, *122*(1), 554–567. https://doi.org/10.1121/1.2735105

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. https://doi.org/10.1037/0096-3445.134.4.477

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

McQueen, J. M. (1998). Segmentation of Continuous Speech Using Phonotactics. *Journal of Memory and Language*, *39*(1), 21–46. https://doi.org/http://dx.doi.org.proxy.lib.sfu.ca/10.1006/jmla.1998.2568

McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of Continuous Speech Recognition and the Contents of the Vocabulary. *Language and Cognitive Processes*, *10*(3–4), 309–331. https://doi.org/10.1080/01690969508407098

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in Spoken Word Recognition: Spotting Words in Other Words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 621–638. https://doi.org/10.1037/0278-7393.20.3.621

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108*(1), 271–280. https://doi.org/10.1016/j.cognition.2008.02.003

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of the USA*, *62*(3), 714–719. https://doi.org/10.1121/1.381583

Nakatani, L. H., & Schaffer, J. A. (1978). Hearing “words” without words: prosodic cues for word perception. *The Journal of the Acoustical Society of America*, *63*(1), 234–245. https://doi.org/10.1121/1.381719

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and Segmentation in Spoken-Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1209–1228. https://doi.org/10.1037/0278-7393.21.5.1209

Redford, M. A., & Randall, P. (2005). The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics*, *33*(1), 27–46. https://doi.org/10.1016/j.wocn.2004.05.003

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/http://dx.doi.org/10.1126/science.274.5294.1926

Saffran, J., Newport, E., & Aslin, R. (1996). Word Segmentation : The Role of Distributional Cues. *Journal of Memory and Language*, *621*(35), 606–621.

Saffran, J., Newport, E., Aslin, R., Tunick, R., & Barrueco, S. (1997). Incidental language learning:. Llistening (and learning) out of the comer of your ear. *Psychological Science*, *8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Sanders, L. D., & Neville, H. J. (2000). Lexical, Syntactic, and Stress-Pattern Cues for Speech Segmentation. *Journal of Speech Language and Hearing Research*, *43*(6), 1301–1321. https://doi.org/10.1044/jslhr.4306.1301

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research*, *15*(3), 228–240. https://doi.org/10.1016/S0926-6410(02)00195-7

Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002). Speech Segmentation by Native and Non-Native Speakers: The Use of Lexical, Syntactic, and Stress-Pattern Cues. *2Journal of Speech, Language, and Hearing Research*, *45*(3), 519–530. https://doi.org/10.1044/1092-4388(2002/041)

Shea, C. E., & Curtin, S. (2010). Discovering the relationship between context and allophones in a second language: Evidence for distribution-based learning. *Studies in Second Language Acquisition*, *32*(4), 581–606. https://doi.org/10.1017/S0272263110000276

Shea, C. E., & Curtin, S. (2011). Experience, representations and the production of second language allophones. *Second Language Research*, *27*(2), 229–250. https://doi.org/10.1177/0267658310375753

Shoemaker, E. (2014). The Exploitation of Subphonemic Acoustic Detail in L2 Speech Segmentation. *Studies in Second Language Acquisition*, *36*(4), 709–731. https://doi.org/10.1017/S027226311400014X

Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism*, *16*(2), 304–324. https://doi.org/10.1017/S1366728912000466

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557–580. https://doi.org/10.1023/A:1026464108329

Tranel, B. (1987). *The sounds of French*. Cambridge: Cambridge University Press.

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*(4), 397–440. https://doi.org/10.1006/jpho.2000.0123

Viviani, P. (1990). Eye movements in visual search: cognitive, perceptual and motor control aspects. *Reviews of Oculomotor Research*, *4*, 353–393.

Walker, D. (2001). *French Sound Structure*. Calgary, Alberta: University of Calgary Press.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*, *119*(1), 597–607. https://doi.org/10.1121/1.2141003

Wells, J. C. (1990). Syllabification and Allophony. *Studies in the Pronunciation of English, a Commemorative Volume in Honour of AC Gimson*, 76–86.

Yeung, H., & Lin, C. (2018). Mandarin-English Proficiency Test. Unpublished Manuscript. Burnaby, BC.

Appendix A. List of Stimuli and Logarithmic Word Frequencies

	Singletons		***s*+stop Clusters**		**Filler**		**Filler**	
Labial	park	3.5657	spark	2.5065	ladder	2.6749	mug	2.5441
	peach	2.5119	speech	3.2880	duck	3.1017	gift	3.5173
	pie	3.1664	spy	3.0103	harp	2.1303	doll	3.1017
	pill	2.7810	spill	2.6365	wreath	1.7404	church	3.5507
	pine	2.5011	spine	2.4683	watch	4.2261	lobster	2.5740
	pit	2.8293	spit	2.9948	donut	1.8633	knife	3.3780
	pool	3.3797	spool	1.4314	dog	3.9928	window	3.6422
	port	2.8704	sport	3.0069	cherry	2.8414	net	2.8998
	pot	3.0607	spot	3.4971	nest	2.7536	gear	2.9122
	pout	1.7482	spout	1.7243	garden	3.1319	hammock	1.8573
Alveolar	table	3.7314	stable	2.8287	mitten	1.3802	bike	3.1209
	tack	2.0374	stack	2.4942	bell	3.3025	hand	4.1542
	tag	2.8506	stag	1.9138	rocket	2.7818	ghost	3.2711
	tar	2.2068	star	3.6180	knot	2.2765	cheetah	2.0719
	team	3.8767	steam	2.8370	gate	3.2135	wheel	3.1402
	tear	3.1392	stair	1.8451	match	3.4017	whistle	2.8971
	tear	3.1392	steer	2.4857	hose	2.8848	robot	2.7938
	tick	2.5694	stick	3.6950	lion	2.8943	cello	1.9823
	tool	2.7396	stool	2.2553	whale	2.7597	bone	3.1239
	top	3.8329	stop	4.5572	gong	2.0969	monkey	3.2330
Velar	cab	3.2617	scab	1.5185	ring	3.6750	dart	1.9956
	car	4.3916	scar	2.6365	bee	2.7235	raft	2.3820
	collar	2.7300	scholar	2.2718	chimney	2.3304	waffle	2.0374
	cone	2.1761	scone	1.4150	leaf	2.4249	hanger	1.8451
	coop	2.7235	scoop	2.4624	knight	3.1355	bat	3.0224
	core	2.7007	score	3.1912	bottle	3.4131	lock	3.4603
	cot	2.0128	scot	1.2304	hammer	2.8041	gnome	1.5185
	crew	3.3847	screw	3.2817	chair	3.4000	map	3.2106
	kale	1.4624	scale	2.6866	moose	2.4518	rose	3.4322
	key	3.6465	ski	2.6170	lemon	2.7882	door	4.1731

Appendix B. Duration Values for Auditory Stimuli

pair	token	/ɪ/	/s/	VOT	token	/ɪ/	/s/	VOT
park **spark**	park1	0.058418	0.088967	0.068974	spark1	0.066538	0.12059	0.010033
	park2	0.064063	0.084833	0.051936	spark2	0.059253	0.135781	0.013682
	park3	0.072804	0.10008	0.073201	spark3	0.060866	0.113727	0.008428
	park4	0.062918	0.114296	0.060878	spark4	0.057726	0.132701	0.009689
peach **speech**	peach1	0.07714	0.109506	0.06259	speech1	0.062243	0.149137	0.006952
	peach2	0.082023	0.072749	0.085791	speech2	0.073787	0.185786	0.014852
	peach3	0.077713	0.088086	0.088949	speech3	0.074678	0.176305	0.016896
	peach4	0.079147	0.104968	0.096348	speech4	0.070368	0.180275	0.019138
pie **spy**	pie1	0.088298	0.091801	0.076458	spy1	0.04943	0.140829	0.008506
	pie2	0.086501	0.087151	0.078771	spy2	0.060235	0.151541	0.008039
	pie3	0.078096	0.086454	0.075457	spy3	0.055334	0.125713	0.007907
	pie4	0.085382	0.102321	0.062923	spy4	0.07214	0.136707	0.011686
pill **spill**	pill1	0.070293	0.074797	0.055061	spill1	0.069344	0.165242	0.012621
	pill2	0.068028	0.090521	0.066502	spill2	0.066608	0.172504	0.010906
	pill3	0.086313	0.077839	0.058919	spill3	0.071789	0.160635	0.017593
	pill4	0.078903	0.100911	0.101944	spill4	0.070754	0.148319	0.014561
pine **spine**	pine1	0.073869	0.118917	0.093366	spine1	0.073407	0.178438	0.017543
	pine2	0.067478	0.086505	0.112881	spine2	0.07677	0.1901	0.011197
	pine3	0.076534	0.085417	0.078591	spine3	0.078199	0.170144	0.012373
	pine4	0.077956	0.120074	0.091551	spine4	0.065456	0.134302	0.014199
pit **spit**	pit1	0.081364	0.11199	0.048717	spit1	0.057767	0.166088	0.016731
	pit2	0.076768	0.091931	0.056282	spit2	0.073915	0.160064	0.01501
	pit3	0.0667	0.090372	0.053425	spit3	0.074349	0.164069	0.01582
	pit4	0.078438	0.08151	0.049055	spit4	0.074343	0.156756	0.024008
pool **spool**	pool1	0.060886	0.0702	0.055976	spool1	0.070523	0.143936	0.016093
	pool2	0.054333	0.078894	0.070526	spool2	0.067188	0.143794	0.016056
	pool3	0.0678	0.082251	0.060484	spool3	0.076008	0.196045	0.021703
	pool4	0.054948	0.081462	0.07025	spool4	0.073957	0.143367	0.024843
port **sport**	port1	0.060767	0.080422	0.084774	sport1	0.067667	0.155611	0.021277
	port2	0.060282	0.088054	0.067934	sport2	0.059161	0.176255	0.01392
	port3	0.059324	0.068674	0.071966	sport3	0.046716	0.175614	0.027394
	port4	0.06556	0.082396	0.065339	sport4	0.072243	0.191783	0.016007
pot **spot**	pot1	0.070738	0.103244	0.103682	spot1	0.056695	0.155257	0.008551
	pot2	0.07094	0.093825	0.129647	spot2	0.06796	0.140642	0.012384
	pot3	0.078332	0.076917	0.075533	spot3	0.078727	0.185005	0.014466
	pot4	0.078943	0.066331	0.058798	spot4	0.068172	0.167972	0.007547

pair	token	/ɪ/	/s/	VOT	token	/ɪ/	/s/	VOT
pout spout	pout1	0.057045	0.06584	0.032391	spout1	0.057492	0.139806	0.016647
	pout2	0.062172	0.0851	0.088987	spout2	0.063537	0.154634	0.010042
	pout3	0.060566	0.104966	0.086513	spout3	0.050243	0.169081	0.011386
	pout4	0.055383	0.089931	0.061791	spout4	0.070265	0.130892	0.008129
table stable	table1	0.080617	0.093056	0.069274	stable1	0.075665	0.162471	0.015596
	table2	0.067327	0.111456	0.065065	stable2	0.071489	0.151516	0.016109
	table3	0.083644	0.107381	0.060099	stable3	0.068343	0.146887	0.017818
	table4	0.070057	0.126823	0.062546	stable4	0.067932	0.143111	0.018386
tack stack	tack1	0.073657	0.101054	0.050502	stack1	0.070749	0.16771	0.022951
	tack2	0.068584	0.095902	0.074647	stack2	0.071968	0.167588	0.026387
	tack3	0.061131	0.09457	0.0627	stack3	0.088408	0.184348	0.019072
	tack4	0.076996	0.098551	0.05939	stack4	0.079931	0.175504	0.027095
tag stag	tag1	0.098318	0.110459	0.073806	stag1	0.083352	0.147204	0.020995
	tag2	0.095816	0.129412	0.112596	stag2	0.070544	0.153162	0.015776
	tag3	0.071406	0.112807	0.074836	stag3	0.067256	0.149898	0.016728
	tag4	0.07865	0.101581	0.082508	stag4	0.061085	0.158732	0.019197
tar star	tar1	0.067201	0.087609	0.086383	star1	0.067122	0.166979	0.022337
	tar2	0.08065	0.089644	0.075852	star2	0.063426	0.170972	0.020998
	tar3	0.068099	0.121034	0.051456	star3	0.070752	0.19659	0.015834
	tar4	0.082948	0.115944	0.100442	star4	0.077166	0.214981	0.017937
team steam	team1	0.07486	0.099891	0.070728	steam1	0.078651	0.167125	0.023322
	team2	0.072418	0.102491	0.063125	steam2	0.078114	0.17627	0.019195
	team3	0.084333	0.090915	0.083941	steam3	0.070567	0.196548	0.02113
	team4	0.079888	0.11656	0.075077	steam4	0.067812	0.217798	0.019216
tear stair	tear1	0.062968	0.138118	0.073183	stair1	0.064139	0.182442	0.017719
	tear2	0.076458	0.130148	0.105036	stair2	0.065822	0.180341	0.019504
	tear3	0.077129	0.140953	0.097021	stair3	0.083681	0.185818	0.01411
	tear4	0.087511	0.117112	0.079599	stair4	0.077066	0.171121	0.019649
tear steer	tear1	0.070135	0.107925	0.071301	steer1	0.064023	0.20349	0.023838
	tear2	0.061386	0.132575	0.082859	steer2	0.061353	0.197278	0.01909
	tear3	0.081894	0.105683	0.072123	steer3	0.081986	0.169718	0.022218
	tear4	0.06721	0.120541	0.093946	steer4	0.078866	0.183058	0.026919
tick stick	tick1	0.071961	0.110618	0.067223	stick1	0.073167	0.178313	0.031182
	tick2	0.073905	0.10283	0.041726	stick2	0.07065	0.205095	0.022281
	tick3	0.072888	0.091692	0.062077	stick3	0.084314	0.19825	0.021858
	tick4	0.061687	0.10289	0.07508	stick4	0.076157	0.161735	0.023302
tool stool	tool1	0.096133	0.155563	0.114637	stool1	0.072245	0.17191	0.028838
	tool2	0.069469	0.127388	0.085862	stool2	0.087332	0.184756	0.023296
	tool3	0.081851	0.124791	0.077859	stool3	0.071035	0.156109	0.024806
	tool4	0.088693	0.107912	0.093404	stool4	0.098685	0.197045	0.029357

pair	token	/ɪ/	/s/	VOT	token	/ɪ/	/s/	VOT
top **stop**	top1	0.071994	0.107634	0.061806	stop1	0.085459	0.17032	0.014732
	top2	0.087405	0.113829	0.049757	stop2	0.073372	0.200075	0.021119
	top3	0.072859	0.099039	0.044672	stop3	0.078308	0.155899	0.017046
	top4	0.071598	0.095226	0.0671	stop4	0.072751	0.15823	0.019546
cab **scab**	cab1	0.062285	0.095086	0.087843	scab1	0.073166	0.202299	0.022597
	cab2	0.073803	0.108333	0.075314	scab2	0.086495	0.197597	0.032716
	cab3	0.073023	0.090496	0.071043	scab3	0.088344	0.177836	0.025276
	cab4	0.065342	0.10193	0.076364	scab4	0.080713	0.169517	0.024923
car **scar**	car1	0.065825	0.093432	0.074737	scar1	0.068077	0.143263	0.023708
	car2	0.073142	0.088415	0.073376	scar2	0.073364	0.144876	0.019268
	car3	0.061779	0.110608	0.093624	scar3	0.072308	0.169258	0.021058
	car4	0.061967	0.100757	0.06503	scar4	0.08296	0.144473	0.019763
collar **scholar**	collar1	0.069757	0.117467	0.060403	scholar1	0.092997	0.166402	0.019241
	collar2	0.067033	0.103166	0.10183	scholar2	0.098665	0.173225	0.021121
	collar3	0.066519	0.111031	0.075684	scholar3	0.089522	0.174685	0.018904
	collar4	0.072875	0.125723	0.089892	scholar4	0.070138	0.167721	0.020038
cone **scone**	cone1	0.053337	0.083431	0.065995	scone1	0.073158	0.160764	0.021098
	cone2	0.063554	0.096381	0.086348	scone2	0.070863	0.171791	0.019795
	cone3	0.067069	0.076045	0.089645	scone3	0.073522	0.171852	0.014463
	cone4	0.063377	0.111649	0.090866	scone4	0.066937	0.175553	0.018474
coop **scoop**	coop1	0.081327	0.114861	0.103827	scoop1	0.070853	0.16046	0.027731
	coop2	0.070644	0.102153	0.113984	scoop2	0.068021	0.189595	0.039063
	coop3	0.078245	0.122424	0.116165	scoop3	0.087791	0.186161	0.03008
	coop4	0.060233	0.100275	0.087018	scoop4	0.070568	0.180398	0.026626
core **score**	core1	0.064487	0.091155	0.08408	score1	0.069739	0.163896	0.021588
	core2	0.065665	0.106389	0.104599	score2	0.078865	0.157812	0.019306
	core3	0.071206	0.123711	0.102566	score3	0.07481	0.138082	0.023387
	core4	0.062354	0.123876	0.097901	score4	0.082616	0.159738	0.032226
cot **scot**	cot1	0.084126	0.12376	0.073721	scot1	0.060675	0.181871	0.024308
	cot2	0.07619	0.099492	0.093002	scot2	0.067042	0.177802	0.030289
	cot3	0.071423	0.108048	0.093382	scot3	0.074064	0.18617	0.023521
	cot4	0.064104	0.099832	0.098258	scot4	0.078138	0.174872	0.026035
crew **screw**	crew1	0.063583	0.112306	0.120467	screw1	0.074541	0.144813	0.033473
	crew2	0.070765	0.133799	0.11546	screw2	0.065189	0.141511	0.042297
	crew3	0.077274	0.111523	0.110444	screw3	0.07081	0.151746	0.031108
	crew4	0.059782	0.129673	0.10263	screw4	0.079754	0.167989	0.036616
kale **scale**	kale1	0.073021	0.112268	0.08394	scale1	0.074084	0.160891	0.023442
	kale2	0.063436	0.105987	0.063208	scale2	0.072756	0.147835	0.025363
	kale3	0.068519	0.096656	0.06708	scale3	0.062128	0.14755	0.018841
	kale4	0.07115	0.082612	0.070915	scale4	0.08117	0.170288	0.024588

pair	token	/ɪ/	/s/	VOT	token	/ɪ/	/s/	VOT
key **ski**	key1	0.061701	0.095511	0.087069	ski1	0.06794	0.186793	0.039064
	key2	0.075759	0.107995	0.070961	ski2	0.07345	0.185705	0.025559
	key3	0.070192	0.091801	0.076072	ski3	0.07345	0.196974	0.031801
	key4	0.078151	0.097825	0.078676	ski4	0.081064	0.181821	0.063614