

Acoustic Cues Used by Learners of English

Danica Reid

Simon Fraser University

1 Introduction

Learners of a second language must acquire the ability to use acoustic and phonotactic information to segment continuous speech into meaningful words. In the process of learning to identify word boundaries in their second 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 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: “this table” and “this stable”. For a native English speaker, cues such as phoneme duration and voice onset time (VOT) can provide sufficient information to identify the word boundary in an example such as the one given above. A second language learner of English adapting 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 takes time and may also lead to erroneous segmentation.

Allophonic aspiration and phonetic consonant duration are the two main cues to the segmentation of English *s*+stop clusters. When native English speakers segment utterances containing *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 by these two cues: cross-boundary and word-initial clusters. In word-initial clusters the boundary precedes both segments of the cluster. They are cued by a longer /s/ and an absence of aspiration following the stop. In cross-boundary clusters the boundary falls between the two elements of the cluster where the /s/ is in the offset of one word and the stop is the onset of the following word. The environment for allophonic aspiration is met in these clusters and the boundary is marked by the presence of aspiration and a short /s/-duration.

The previous body of literature on the segmentation of *s*+stop clusters by second language speakers of English found that the acquisition of new word boundary cues is difficult compared to the adaptation of cues from a speaker’s first language into their second language (Altenberg, 2005; Ito & Strange, 2009; Shoemaker, 2014). The results of these studies suggested that the presence or absence of highly predictable cues, like aspiration, were weighted higher than other cues and provided the most relevant information to the listeners. Despite identifying the /s/-duration in *s*+stop clusters as a potential cue for word boundary identification, none of these studies discussed how the duration cue influenced boundary identification.

Listeners use cues like /s/-duration and VOT throughout the processing of a speech stream to identify the locations of word boundaries. 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 the continued input will identify non-viable candidates. During this process, the candidates compete for activation with more frequent or better matched words inhibiting activation of less favoured words until a single candidate can be identified (Alloppenna et al., 1998; Magnuson et al., 2007;

* This work was funded in part by a SSHRC Joseph-Armand Bombardier Canada Graduate Scholarship. I would like to thank Dr. Ashley Farris-Trimble and the members of the Simon Fraser University Phonological Processing Lab for their help on the project.

Tanenhaus et al., 2000). When activating potential candidates in phonetic inputs that differ in the type of *s*+stop cluster, the main difference in activation results from the relevant word boundary cues. The duration of the /s/ along with the presence or absence of aspiration cue listeners to the location of the word boundary which limits the candidate set.

Non-native speakers of English follow the same general processes of lexical activation upon hearing speech input, however, 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 et al., 2003; Weber & Cutler, 2004). This simultaneous activation of both languages causes non-native speakers of English to have slower overall processing as they are forced to suppress the lexical activation of words in their native language in favor of the non-native candidates. There have been several models proposed to detail bilingual word recognition (Dijkstra & van Heuven, 2002; Grosjean, 1997; Shook & Marian, 2013). The models of auditory word recognition suggest that upon hearing the first sound in a sequence, the listener activates words in both languages that have the same onset. As the input progresses, activation of words from both languages is influenced by a variety of different factors. For example, in the Bilingual Model of Lexical Access (BIMOLA; Grosjean, 1997), factors of bilingual activation relate to the language mode being used (monolingual or bilingual) and the phonetic similarity of sounds in the two languages.

The 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 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. The current study aims to fill this gap by investigating how speakers use the phonological properties of their first language to process speech 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 look at how these two cases are reflected in real-time processing in addition to accuracy of identification to further extend this field of research.

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 aspiration system implies that VOT is not used as a word boundary cue. These aspirated and unaspirated stops found in Mandarin fall into the same VOT categories as those found in the cross-boundary and word-initial *s*+stop clusters of English, long lag (60-100ms) and short lag (0-25ms) respectively. Unlike English, the phonotactic constraints of Mandarin do not allow for any word-initial or 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 *s*+stop cluster types found in English.

2.2 Participants A total of forty-one participants between the ages of nineteen and forty-five were recruited at Simon Fraser University and in the Metro-Vancouver area: twenty-one native English speakers and twenty native Mandarin speakers. All participants were required to have normal or corrected-to-normal vision as well as normal hearing to ensure that they could complete the eye-tracking portion of the study. Demographic information was collected from each participant regarding the age that the participant 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 to replace subjective measures of proficiency (Yeung & Lin, 2018). 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%; $t(39) = 5.6, p < .001$).

2.3 Stimuli The word-list for this experiment includes ten pairs of nouns for each word-initial voiceless stop in English. These pairs differed in onset with one member beginning with an *s*+stop cluster and the other beginning with a voiceless aspirated stop. In addition to these thirty pairs, sixty phonetically and semantically

unrelated filler items were selected – two per pair – to create four-item experimental sets. All words used in the experiment were a maximum of two syllables in length with trochaic stress. Comparisons of word frequencies for experimental pairs showed that there was no significant difference for each place of articulation ($t(9) < |1|$ in all comparisons). Additionally, word frequency of filler items overall compared to the frequency of all experimental items was not significant ($t(118) < |1|$). Examples of experimental sets include: park, spark, ladder, mug; tack, stack, bell, hand; and kale, scale, moose, rose.

2.3.1 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. The filler items were recorded in the frame “click on a ____.” Using two separate frames gave participants information to quickly rule out filler items making it possible to determine more precisely how the relevant word boundary cues were being used.

2.3.2 Visual Stimuli The majority of the visual stimuli used in this experiment were taken from a database of images developed by the Mechanisms of Audio-Visual Categorization Lab at the University of Iowa and maintained by the Simon Fraser University Phonological Processing Lab. Images found in this database are held to a standard of clarity, typicality, and salience associated with eye-tracking in the visual world paradigm; images not found in the database were held to the same standard. In each experimental set, the pictures used were all clearly different from one another and had roughly equivalent visual salience.

2.4 Procedure

2.4.1 Production Task The production task of this experiment served two 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 participants were producing the relevant word boundary cues. In this task, participants were shown screens that contained a picture with the associated word and a short sentence underneath. Participants read the word followed by the sentence for each item. All 120 items were pseudo-randomized such that word pairs were never consecutive. Each participant was presented with these items in the same order. Productions were digitally recorded in Audacity using a Blue Yeti microphone at 44,100 Hz.

2.4.2 Visual World Paradigm Task Participants were seated in front of a screen displaying one image in each corner. Items were presented four times, each time with a different auditory token for a total of 480 trials. The experiment was presented in 15 blocks of 32 trials with a break between each block. The trials were randomized individually for each participant to ensure that order of presentation was not a factor in the results. In addition, the location of the items on the screen was randomized for each trial. Eye-movements were recorded using an EyeLink1000 eye tracker which sampled participant fixations every 4ms from the start of the trial until the response.

2.5 Results Based on a preliminary analysis of average group accuracy, three phrase pairs, and their associated filler items, were removed from the following 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.

2.5.1 Production Measures In analyses of the production task, three participants (one native English and two native Mandarin speakers) were excluded because they paused between the /s/ in “this” and the /s/ in all word-initial cluster items so the full /s/-duration could not be measured. For nine participants (two native English and seven native Mandarin speakers), individual tokens were excluded from analyses where the same pauses were found (range of excluded tokens: 1 to 10). The remaining eighteen native English and eleven native Mandarin speakers produced all tokens in the production task.

Recordings of participants were annotated to measure the vowel /i/ in “this”, the full /s/-duration, and the VOT of the voiceless stops. Graphs of these measures by language and cluster type can be found in Figure 1. Repeated measures ANOVAs were run on all three duration measures with language as the between-subjects factor. For the vowel duration, the vowel was shorter in cross-boundary than 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 ($F(1,36) = 308.2, p < .001$) and there was no difference in duration between the two language groups ($F(1,36) = 2.6, p = .118$). 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 /s/-duration and VOT for the two cluster types. The effect of cluster type on the /t/-duration was not expected, but the direction of the difference could be related to the /s/-duration found in word-initial and cross-boundary s+stop clusters.

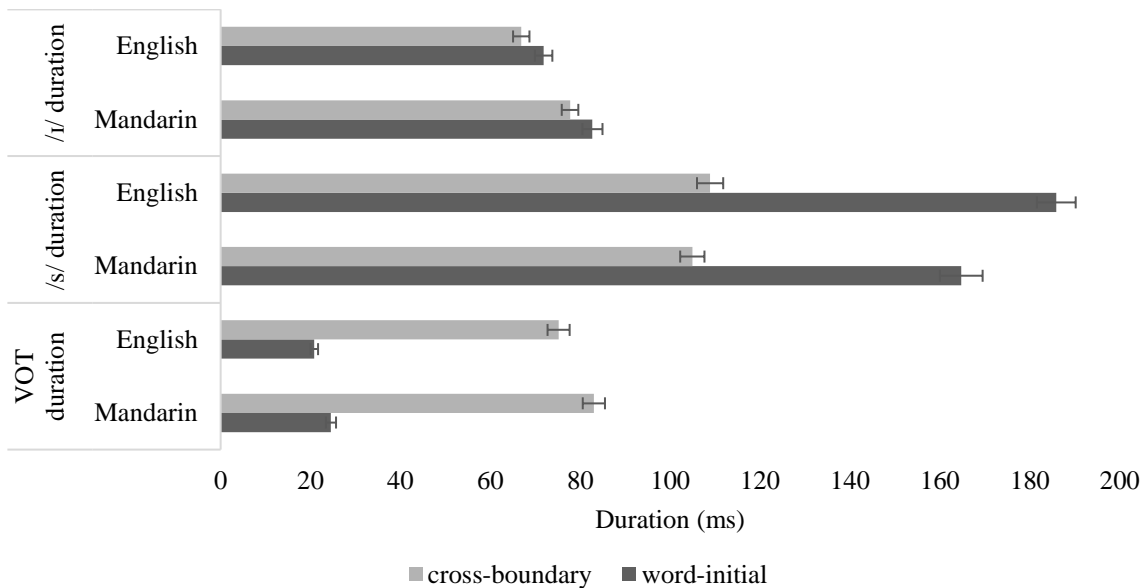


Figure 1. Average durational values by language and by cluster type

2.5.2 Identification Measures Overall accuracy and reaction times were calculated for each participant. Native English speakers had a mean accuracy of 99% ($SD = 0.91\%$) and Native Mandarin speakers were significantly less accurate, with a mean accuracy of 92% ($SD = 5.23\%$; $t(39) = 6.1, p < .001$). Native English speakers also had faster reaction times (1707ms; $SD = 113$ ms) than native Mandarin speakers (2072ms; $SD = 252$ ms; $t(39) = 6.1, p < .001$). Accuracy did not differ by cluster type (word-initial or cross-boundary) for either language group (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 ($t(20) = 2.4, p = .027$); differences in reaction time were not found in the native Mandarin speakers ($t(19) = 1.5, p = .147$). This difference suggested that the native English speakers may have been more reliant on the presence of aspiration in the cross-boundary clusters than the native Mandarin speakers.

2.5.3 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 an indication of how they are processing the input. However, when data are averaged across trials for target and competitor separately, smooth curves emerge. Despite following different curves, target and competitor fixations typically increase together approximately 200ms after the onset of acoustic information because it takes roughly 200ms to plan and launch an eye-movement (Viviani,

1990). Figure 2 shows a plot of target fixations by language group and 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 0ms represents the onset of the /s/ in “this” because this is the onset of the first word boundary cue; any fixation 200ms or more after the adjusted 0ms 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 accounted for 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 200ms, levelling off around 1000ms. Between the fixations for native English and native Mandarin participants, the English participants have a steeper slope and higher proportion of fixations to target items.

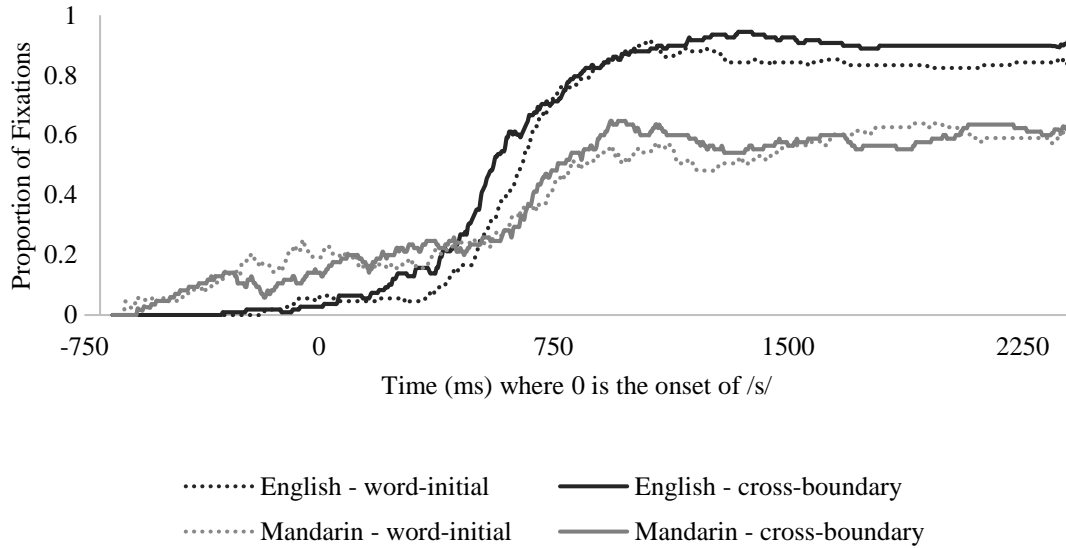


Figure 2. Target fixations by language and cluster type

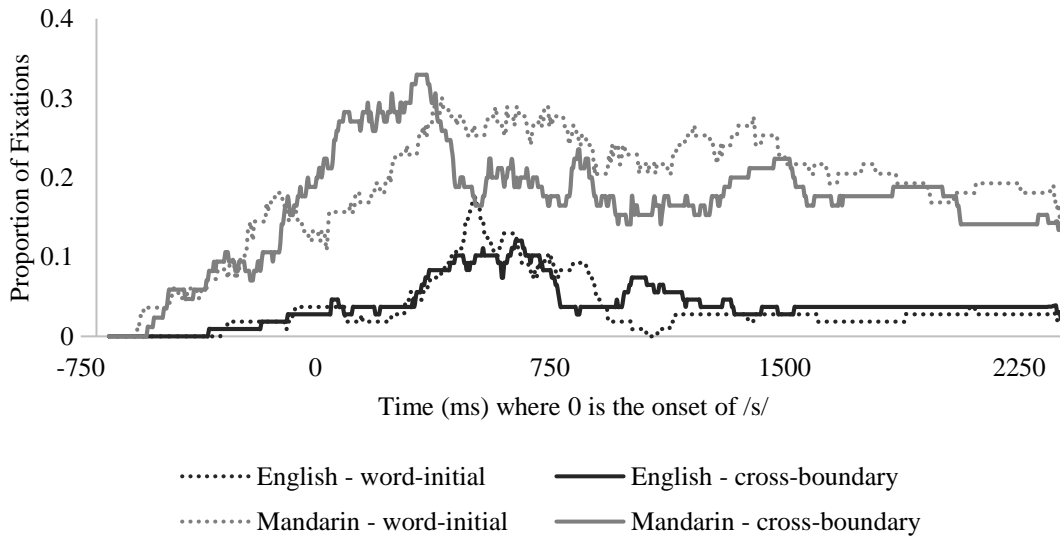


Figure 3. Competitor fixations by language and cluster type

Figure 3 shows a plot of competitor fixations by language group and by cluster type. 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 200ms to their peak at around 500ms and then decrease. The native English speakers 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 maintained plateau in fixations.

In this experiment, target fixations were fit to a logistic curve which allowed four parameters to be measured: the baseline; the peak; the crossover, which is the inflection of the curve where growth starts to slow; and the slope. Competitor fixations were fit to a specialized Gaussian curve which allowed six parameters to be measured: the onset baseline; the onset slope; the midpoint, which is the time of maximum fixations to the competitor; the peak height; the offset slope; and the offset baseline. The parameters for both target and competitor curves can be compared statistically (Farris-Trimble & McMurray, 2013). For the analyses of this experiment, seven parameters that reflect degree and timing of fixations were compared. For target fixations, degree of fixations were analysed using peak and timing of fixations were analysed using cross-over and slope. For competitor fixations, degree of fixations were analysed using peak and offset baseline and timing of fixations were analysed using onset slope and midpoint.

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). Only measures with significant effects will be reported here; ANOVAs for all measures can be found in Reid (2019). No significant differences were found for either language group between the two cluster types.

Significant effects for measures of target fixations reflected both degree and timing of fixations. The peak of 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 ($F_1(1,39) = 14.4, p < .001$; $F_2(1,52) = 100.3, p < .001$; $\text{minF}'(1,50) = 12.6, p = .001$). The crossover point of 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 ($F_1(1,39) = 5.5, p = .024$; $F_2(1,52) = 31.6, p < .001$; $\text{minF}'(1,53) = 4.7, p = .035$). These measures of degree and timing showed that native Mandarin speakers were slower to recognize and fixate the target and were less confident in their responses than the native English speakers.

Significant effects for measures of competitor fixations only reflected the degree of fixations. 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 ($F_1(1,39) = 6.0, p = .019$; $F_2(1,52) = 78.6, p < .001$; $\text{minF}'(1,45) = 5.6, p = .023$). The offset baseline is a measure of the final proportion of fixations to competitor items. Native Mandarin speakers had a higher degree of fixations to competitor items for both by-subject and by-item analyses ($F_1(1,39) = 12.9, p = .001$; $F_2(1, 52) = 50.6, p < .001$; $\text{minF}'(1,59) = 10.3, p = .002$). These measures of degree indicated that the native Mandarin speakers fixated the competitor items more than the native English speakers.

2.6 Discussion The identification measures analysed in this experiment showed that, as expected, native English speakers had a 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. The results of the perception task fit the overall patterns of lexical activation predicted by models of monolingual (Gow & Gordon, 1995; Marslen-Wilson & Welsh, 1978; McQueen et al., 1995; Norris et al., 1995) and bilingual word recognition (Dijkstra & van Heuven, 2002; Grosjean, 1997; Shook & Marian, 2013). For both the English and Mandarin speaker groups, fixations to 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 favour of the target. However, suppression of competitor fixations was stronger for native English than native Mandarin participants. 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.

3 Experiment 2

While the native Mandarin speakers showed some differences from the native English baseline in the measures of real-time processing, the similarities between the two groups raised the possibility that the methodology was not sensitive to differences in cluster type. In order to test this, a second experiment was run with 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.

3.1 *Language of Interest* This experiment tested the accuracy and real-time processing of native French speakers. French has voiceless unaspirated stops which are distinguished from voiced stops by voicing and not by aspiration (Tranel, 1987; Walker, 2001). As such, in French, voiced stops fall in the lead VOT category and voiceless stops have short lag (Tranel, 1987). The phonotactics of French allow both word-initial and cross-boundary s+stop clusters.

While it is the case that native French speakers do not natively use aspiration as a word boundary cue, 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 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). Based on this, it is possible that native French speakers could adapt their knowledge of durational cues into English segmentation.

3.2 *Participants* A total of seven native French speakers were recruited at Simon Fraser University and in the Metro-Vancouver area. As in the first experiment, participants were between the ages of nineteen and forty-five and had normal or corrected-to-normal vision and normal hearing. The same demographic information was collected from the native French speakers regarding the age that the participants began learning English, years of English learning, age of arrival in North America, and length of residence in North America. The same proficiency test was also used (Yeung & Lin, 2018). The differences in proficiency between the three participant groups showed that the native Mandarin and native French speakers performed similarly on the proficiency task when compared to the native English speakers (French mean accuracy: 46%; Mandarin mean accuracy: 45%; English mean accuracy: 80%).

3.3 *Stimuli* The stimuli used in this experiment were identical to those used in the first experiment.

3.4 *Procedure* The procedures used in this experiment were identical to those used in the first experiment.

3.5 *Results* The sample size for this experiment is not large enough to allow for direct statistical comparison to the native English and native Mandarin participant groups in the previous experiment. As such, all comparisons to the first experiment are observational. However, statistical analyses are presented for within-group comparisons of the two cluster types.

3.5.1 *Production Measures* Recordings of participants underwent the same processing as the recordings from the first experiment. Average values for /i/-duration, /s/-duration, and VOT duration can be found in Figure 4. As with the native English and Mandarin production analyses, individual tokens were excluded 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. The remaining two native French speakers produced all tokens in the production task. To investigate the durational properties of the cross-boundary and word-initial s+stop clusters, paired t-tests were run. For the vowel duration, /i/ 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 than 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 resulting from the environment of allophonic aspiration being present for cross-boundary but not word-initial clusters ($t(6) = 3.3, p = .016$). Looking at the graph in Figure 4, 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.

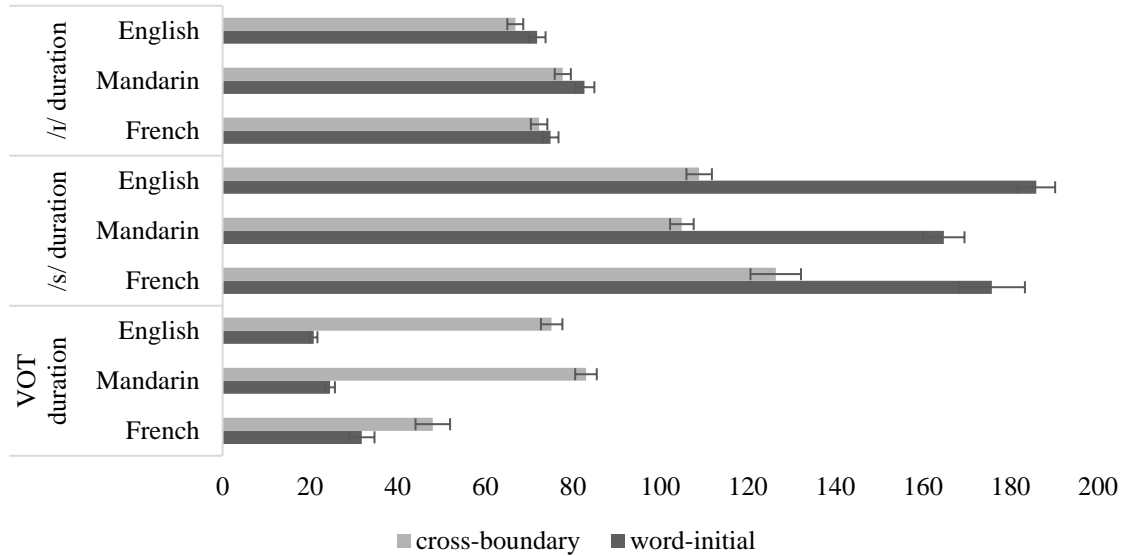


Figure 4. Average durational values by language and by cluster type

3.5.2 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 the native Mandarin and native English speakers. Overall accuracy of the French speakers (92%, $SD = 6.5\%$) was close to that of the native Mandarin speakers (92%, $SD = 5.2\%$); see left graph of Figure 5. A comparison of the two cluster types for reaction time for the native French speakers showed no significant difference ($t(6) < |1|$). The overall reaction time of native French speakers (2068ms, $SD = 313ms$) was also close to the average reaction time for native Mandarin participants (2072ms, $SD = 252ms$); see right graph of Figure 5.

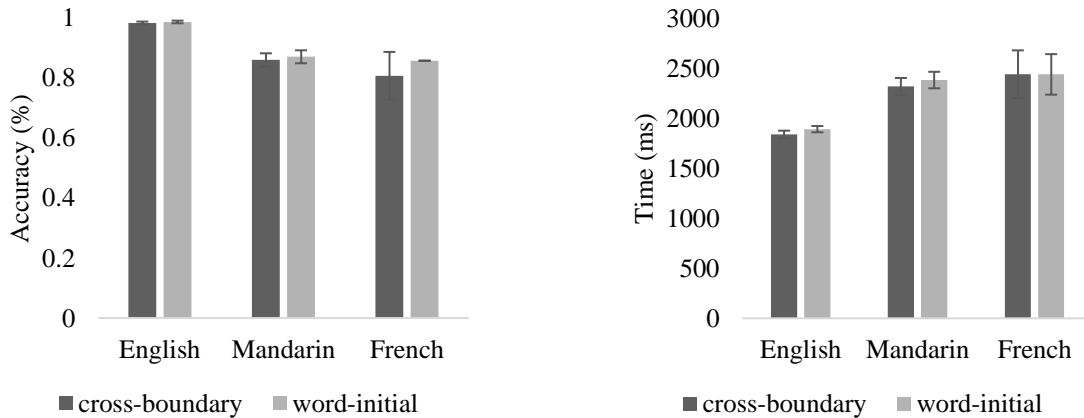


Figure 5. Average accuracy (left) and reaction time (right) by language and cluster type

3.5.3 Fixation Measures Figure 6 shows the overall target fixations for all three languages. Based on this plot, the degree of fixations shown by the native French participants was somewhere between the native English and native Mandarin speakers. However, the 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.

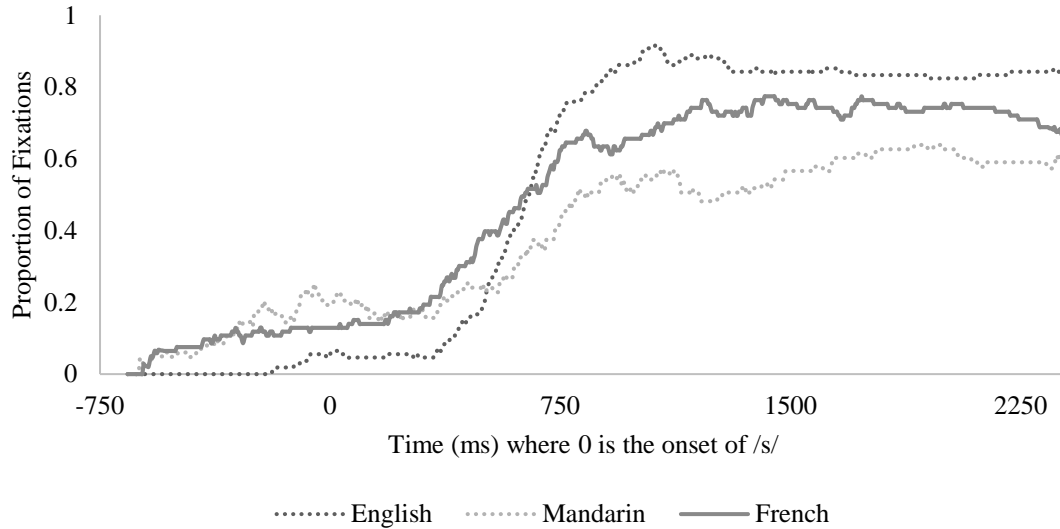


Figure 6. French, English, and Mandarin target fixations

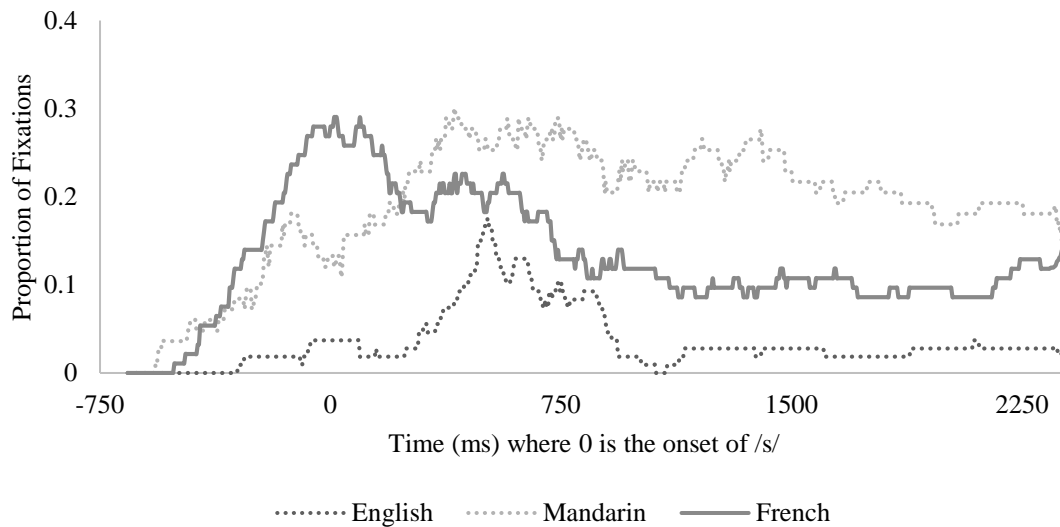


Figure 7. French, English, and Mandarin competitor fixations

When looking at the competitor fixations in Figure 7, 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 the native French speakers did not clearly match either of the previously tested language groups. The French speakers had a much earlier peak in their fixations to competitor items 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.

3.5.4 Discussion The small sample size of this experiment made it difficult to make direct statistical comparisons to Experiment 1. However, observational analyses suggested that the native French speakers had a level of accuracy and processing which patterned similarly to the native Mandarin participant group. This was unexpected based on the different uses of VOT by the two language groups. 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.

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: /s/-duration and VOT.

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. 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) 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 each language. The language mode exists on a scale with monolingual mode for each language at opposite ends and bilingual mode in the middle. The environment in which an interaction is taking place dictates the level of activation on this scale. If a bilingual speaker is addressed in English, is reading English instructions, and is 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 the language groups tested in these experiments, the environment of the study encouraged the use of the English monolingual language mode as opposed to the Mandarin or French monolingual mode which may have influenced the similarities found between the three language groups. Another characteristic of BIMOLA is the assumption that the level of activation for between-language competitors depends on the degree of similarity between the sounds of the two languages. This can help explain the observed differences in processing between the native Mandarin and French participants. For the native Mandarin participants, voiceless stops in Mandarin and English fall into the same VOT categories leading to the activation of candidates from both languages. On the other hand, native French speakers would not be expected to activate French words upon hearing a word-initial voiceless stop in English because there are clear differences in VOT between the two languages. As a result, native French participants would have fewer between-language competitors than the native Mandarin participants. 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 clear despite the analysed fixations being representative of trials with correct responses. The uncertainty is most obvious in 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 compared to the native English speakers. This difference 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 been found in the process of word recognition by speakers with atypical language (Farris-Trimble et al., 2014; McMurray et al., 2010). Adopting this strategy gives populations dealing with signal degradation a repair strategy for incorrect processing; maintaining a higher level of activation to competitor items would make it easier to adjust interpretations of the input. Despite differences in the sources of uncertainty between clinical and non-native populations, the native Mandarin and French participants of this study applied the same repair strategy. 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.

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 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.

References

- Allopenna, Paul, James Magnuson & Michael Tanenhaus (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.
- Altenberg, Evelyn (2005). The perception of word boundaries in a second language. *Second Language Research*, 21(4), 325–358.
- Barry, William (1981). Internal juncture and speech communication. In William J. Barry & Klaus J Kohler (Eds.), *Beitrage zur experimentellen und angewandten phonetic* (pp. 229–289). Kiel, Germany: AIPUK.
- Blumenfeld, Henrike & Viorica Marian (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.
- Blumenfeld, Henrike & Viorica Marian (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology*, 25(5), 547–567.
- Christie, William M. (1974). Some cues for syllable juncture perception in English. *The Journal of the Acoustical Society of America*, 55(4), 819–821.
- Clark, Herbert, 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.
- Cutler, Anne, Jacques Mehler, Dennis Norris & Juan Segui (1986). The syllable's different role in segmentation of French and English. *Journal of Memory and Language*, 25, 385–400.
- Cutler, Anne, Jacques Mehler, Dennis Norris & Juan Segui (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, 24(3), 381–410.
- Dijkstra, Ton & Walter J. B. van Heuven (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175–197.
- Duanmu, San (2000). *The Phonology of Standard Chinese*. New York, USA: Oxford University Press.
- Farris-Trimble, Ashley & Bob McMurray (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.
- Farris-Trimble, Ashley, Bob McMurray, Nicole Cigrand & J. Bruce Tomblin (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.
- Goetry, Vincent & Régine Kolinsky (2000). The role of rhythmic cues for speech segmentation in monolingual and bilingual listeners. *Psychologica Belgica*, 40(3), 115–152.
- Gow, David W. & Peter C. Gordon (1995). Lexical and Prelexical Influences on Word Segmentation: Evidence From Priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21(2), 344–259.
- Grosjean, Francois (1997). Processing mixed language: issues, findings, and models. In Annette M. B. Groot & Judith F. Kroll (Eds.), *Tutorials in bilingualism: psycholinguistic perspectives* (pp. 225–254). Mahwah, New Jersey:

- Laurence Erlbaum Associates, Inc., Publishers.
- Ito, Kikuyo & Winifred Strange (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.
- Ju, Min & Paul A. Luce (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, 15(5), 314–318.
- Klatt, Dennis (1975). Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters. *Journal of Speech, Language, and Hearing Research*, 18, 686–706.
- Lehiste, Ilse (1960). An acoustic – phonetic study of internal open juncture. *Phonetica*, 5, 5–54.
- Lin, Hua (2001). *A grammar of Mandarin Chinese*. Munich, Germany: Lincom Europa.
- Lin, Yen-Hwei (2007). *The Sounds of Chinese*. New York, USA: Cambridge University Press.
- Magnuson, James S., James A. Dixon, Michael K. Tanenhaus & Richard N. Aslin (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31(1), 133–156.
- Marian, Viorica & Michael Spivey (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, 24(2), 173–193.
- Marian, Viorica & Michael Spivey (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, 6(2), 97–115.
- Marian, Viorica, Michael Spivey & Joy Hirsch (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, 86(1), 70–82.
- Marslen-Wilson, William D. & Alan Welsh (1978). Processing interactions during word recognition in continuous speech. *Cognitive Psychology*, 10, 29–63.
- McMurray, Bob, Vicki M. Samelson, Sung Hee Lee & J. Bruce Tomblin (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, 60(1), 1–39.
- McQueen, James M., Anne Cutler, Ted Briscoe & Dennis Norris (1995). Models of Continuous Speech Recognition and the Contents of the Vocabulary. *Language and Cognitive Processes*, 10(3–4), 309–331.
- Nakatani, Lloyd H. & Kathleen D. Dukes (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of the USA*, 62(3), 714–719.
- Norris, Dennis, James M. McQueen & Anne Cutler (1995). Competition and Segmentation in Spoken-Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5), 1209–1228.
- Reid, Danica (2019). *Acoustic Cues Used by Learners of English*. (Unpublished Master's thesis). Simon Fraser University, Burnaby, Canada.
- Shoemaker, Ellenor (2014). The Exploitation of Subphonemic Acoustic Detail in L2 Speech Segmentation. *Studies in Second Language Acquisition*, 36(4), 709–731.
- Shook, Anthony & Viorica Marian (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism*, 16(2), 304–324.
- Tanenhaus, Michael K., James S. Magnuson, Delphine Dahan & Craig Chambers (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.
- Tranel, Bernard (1987). *The sounds of French*. Cambridge, Cambridgeshire: Cambridge University Press.
- Viviani, Paolo (1990). Eye movements in visual search: cognitive, perceptual and motor control aspects. *Reviews of Oculomotor Research*, 4, 353–393.
- Walker, Douglas (2001). *French Sound Structure*. Calgary, Alberta: University of Calgary Press.
- Weber, Andrea & Anne Cutler (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1–25.
- Yeung, Henny & Cathy Lin (2018). Mandarin-English Proficiency Test. Unpublished Manuscript. Burnaby, British Columbia.