Second Language Perception of Word Segmentation

Proceedings of LSUGA's Second Interdisciplinary Conference in Linguistics

Aseel Alammar

George Mason University

Introduction.

Second language learners often find it difficult to parse a stream of speech into words. The speech signal is a continuous flow of sounds in which word boundaries may not be clearly defined. While native speakers are able to segment speech effortlessly, non-native speakers face a more difficult challenge. Finding where one word ends and another word begins is a demanding task for learners of a second language (Aljasser 2008; Hanulíková et al. 2011; Weber and Cutler 2006; Altenberg 2005; Ito and Strange 2009). This difficulty is possibly a result of having an established segmentation system that is specialized to their L1 and could interfere with L2 boundary detection. L1 transfer is found to play a crucial role in speech perception and segmentation (Flege and MacKay 2004; Brown 1998; Shoemaker and Rast 2013; Weber and Cutler 2005). For example, in a language with initial stress, L1 listeners are likely to use stress as a cue to signal a word boundary (Vroomen et al.1998). If a listener from that language tries to parse an L2 that does not have initial stress, their L1 segmentation strategy will not be of any use and could also hinder their segmentation ability. This leads to the question: Are L2 listeners flexible in learning a new segmentation strategy? Overall, the goal is to explore whether adult learners have learned to exploit their second language's segmentation cues to detect word boundaries.

Background.

Native speakers rely on various types of information to segment speech: syntactic, lexical, allophonic/ phonetic, or prosodic/ rhythmic information available in language. L1 speakers make use of all of these cues for accurate segmentation (Gow & Gordon 1995, Sanders & Neville 2000, McQueen 1998, Hanulíková et al. 2011, Cutler & Norris 1988, Nakatani & Dukes 1977, Mattys et al. 2005).

At the word level, each language has its own phonetic restrictions, which determine the positional occurrences and restrictions on segments. For example, English disallows the onset cluster /pw/. Therefore, one could predict that this cluster would enforce a word boundary between /p/ and /w/. Norris et al. (1997) proposed The Possible Word Model as a way in which listeners parse speech using a similar strategy. This model found that participants are able to spot a word such as *apple* easily after a nonsense word such as *vuffapple*, but not after *fapple.* They concluded that the reason was that *vuff* phonologically is a possible English word, compared to *f* which is not a possible English word and therefore would not enforce a boundary.

While native listeners use language specific phonetic cues for the purpose of speech segmentation, it remains unclear whether these cues are applied to the L2. One way native listeners make use of consonant sequences to identify word boundaries is by locating a boundary between illegal clusters. Cross-linguistic research has shown that certain phonotactic restrictions are not easily acquired by second language learners in word boundary detection (McQueen, Otake & Cutler 2001; Weber and Cutler 2006; Hanulíková et al. 2011).

Moreover, Nakatani and Dukes (1977) found a hierarchy of acoustic phonetic (allophonic) segmentation cues in English. They found glottal stops or laryngealization to be the strongest English segmentation cue, followed by aspiration on voiceless stops. They also found that the cues were prominent word initial; in onsets rather than codas. Altenberg (2005) conducted a cross-linguistic study to test the use of these different allophonic cues by Spanish learners of English using an adapted version of the study designed by Nakatani and Dukes (1977). Participants were presented with stimuli, such as *keeps talking* and *a nice man*, and were instructed to indicate whether they heard *keep **st**alking* or *keep**s** **t**alking*, and ***a ni**ce man* or ***an i**ce man*. The presence of aspiration /t^{h}/ in *keeps talking*, or the absence of it in *stalking*, would provide a strong segmentation cue. Similarly in *an ice man* the presence of a glottal stop or creaky voice preceding *ice* or the absence of it in *a nice*, provides a strong segmentation cue. Altenberg found that Spanish listeners' performance on the aspiration pairs was significantly less English-like compared to the glottal stop pairs as Spanish doesn't have aspiration on voiceless stops. Altenberg concluded that Spanish listeners were unable to accurately use these allophonic cues available to native English listeners. In addition to Spanish, Ito and Strange (2009)

performed a similar study with Japanese as the L1. Using the same stimuli, although their results were not identical to Altenberg's, they found that Japanese speakers were significantly less accurate in determining segmentation cues facilitated by aspiration compared to English speakers. These results leave an important question unanswered. In both studies listeners were familiar with the words in the stimuli, which was reported as a possible limitation to the experiment. It is not clear whether lexical familiarity, in addition to allophonic cues, played a role in boundary identification. In this present study, the use of natural speech is avoided to test whether allophonic cues (aspiration and glottal stop) can be a reliable signal for segmentation with no influence of meaning or content.

English and Arabic Segmentation Cues.

The difficulty of segmenting a second language stems from the fact that languages differ in the types of possible allophonic boundary cues. The difference in allophonic constraints between Arabic and English may hinder the ability of learners of English to accurately segment speech. Based on Nakatani and Dukes' (1977) conclusions regarding the use of acoustic phonetic variation as segmentation cues, the presence of glottal stops or (creaky voice) or of aspiration on voiceless stops in word-initial positions are strong segmentation cues for English speakers. Arabic also has both voiceless and voiced stops but differ VOT values. The VOT range for Arabic voiceless stops is closer to the range of English voiced stops (Kattab 2000). Long lag values (aspiration) are not attested in Arabic voiceless stops. This variation across the two languages will serve as a test as to whether Arabic L2 learners have acquired a segmentation cue that is not in their native language from their learning of English. The other strong English cue is the presence or absence of a glottal stop in the onset. Unlike aspiration, the glottal stop is a legal onset in both Arabic and English. Glottal stops are possible syllable onsets and are prominent word initially and weakened or dropped word-medial and word-final in non-standard Arabic (Watson 2007). Therefore, due to the nature of the glottal stop as weakened or dropped word medially in dialectal Arabic, I will assume that both Arabic and English speakers would be able to utilize the presence of a glottal stop as a strong segmentation cue.

Furthermore, both target conditions will be tested in the onset as onsets were found to be

the strongest segmentation position (Gow and Gorden 1994). The English group is predicted to perceive both target conditions as possible word boundaries. Whereas the Arabic group is predicted to pattern similarly with the English group in the glottal stop condition but not aspiration. Thus, if using segmentation cues is based on their L1 (i.e., L1 transfer), Arabic listeners in the present study would be expected to have a difficulty in perceiving aspiration as a word boundary. By examining the perception of Arabic listeners, the present study contributes new information on the use of allophonic cues for L2 word segmentation.

Method.

20 adult native speakers of American English living in the US served as the control group. An experimental group of 50 adult native speakers of Saudi Arabic living in Riyadh, Saudi Arabia who have learned English as a second language participated in this experiment.

Stimulus Materials.

To minimize the influence of lexical familiarity of real words, 48 nonsense two-syllable words (24 pairs) were chosen. In order to control for familiarity with the segments and sequence of segments in each syllable, the words were created from a list of highly frequent English nonsense syllables (Vitevitch et al. 1997, Vitevitch & Luce 2004). A female monolingual speaker of American English recorded the stimulus words. All stimuli were recorded in a single session in a sound-attenuated booth, in the George Mason University Linguistics program acoustics lab. The stimuli are divided into two groups: (a)- ten pairs of words containing aspiration cues (see Appendix 1- Table 1); (b)- ten pairs containing the glottal stop cue (see Appendix 1- Table 2).

Ten aspiration stimulus pairs were recorded so that the presence or absence of aspiration of a voiceless-stop at a syllable boundary would serve as a boundary signal. Each word consists of two syllables where the voiceless stop is an onset of the second syllable, 'e.g. [bis **kʰɜ**m]'. For each pair, only one word was recorded without aspiration. The word was then edited using Audacity 2.1.0-beta (Audacity Team 2015) to lengthen the VOT. Similarly, 10 glottal stop stimulus pairs were also recorded using the same method. The glottal stop was present in the

onset of the second syllable 'e.g. [lʌn ʔaɪp]'. The glottal stop was digitally added and characterized by a silent gap or period of creaky voice before the vowel (Ito and Strange 2009). These measures were taken to control for other prosodic segmentation cues, such as primary stress or final lengthening. Therefore, the words in every pair are identical except for VOT length or glottal stops.

Procedure.

Survey Gizmo was used to present the stimuli to the participants. The survey starts with a background questionnaire. After that, the participants were asked to listen to an audio clip with the stimuli. For example they heard a clip with positive aspiration [bis **kʰɝ**m], and another clip with negative aspiration [bis **kɝ**m], then answer for each if they heard one word or two. Although English does have words where the target condition occurs in the second syllable, such as [pəthɛɾo], research has shown that aspiration and glottal stops are perceived more often as an onset (Gow and Gorden 1994). Therefore, the aim of this task is to demonstrate whether there is a greater likelihood to perceive a strong segmentation cue as a word boundary without meaning or context. The audio clips were randomized in their order.

Results.

As acknowledged previously, there is no correct answer, rather we will be tabulating how often a respondent reports hearing one word or two words when cues are present or absent. The mean score of responses by American English and Saudi Arabic participants for each stimulus type is presented in Table 3.

Table 3. Mean score for each stimulus type (positive and negative).

Stimulus type	English subjects (*n*=20)			Arabic subjects (*n*=50)		
+aspiration (*n=10)*	Mean	SD	SE	Mean	SD	SE
	7.5	1.7	0.38	5.2	2.1	0.3
-aspiration (*n=10)*	7.4	1.5	0.34	5.7	2.1	0.3

+glottalstop (*n*=10)	8.3	1.1	0.26	7.8	1.8	0.26
-glottalstop (*n*=10)	8.7	0.9	0.2	7.5	2	0.3

The native English group performance overall was higher (80%) in both conditions compared to the Arabic group (66%). This suggests that the native English control group was able to detect a boundary in the positive conditions by reporting hearing two words, and one word in the negative conditions.

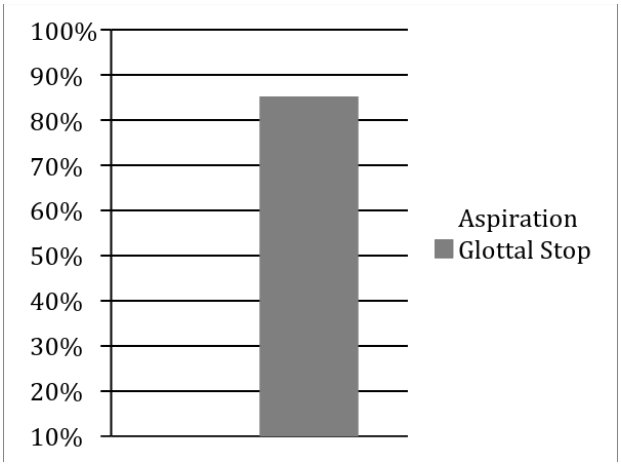

Figure 1. English group 's scores on glottal stop cues (85%) and aspiration cues (74%)

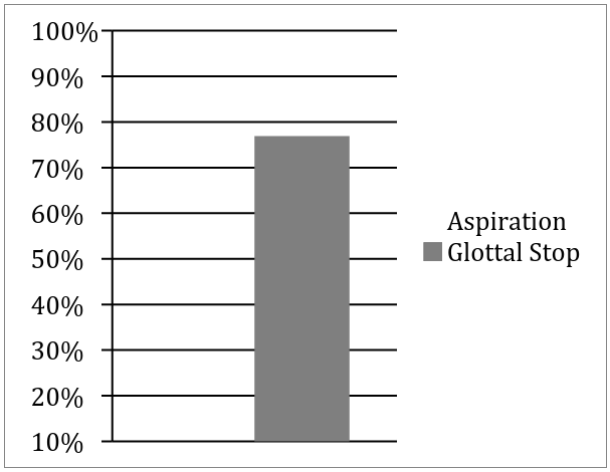

Figure 2. Arabic group's scores on glottal stop cues (77%) and aspiration cues (55%)

Signal detection measures were used to take into account participants' sensitivity towards answering one word or two words for both conditions (MacMillan and Creelman 1991). The English group's *d′* values were higher in both conditions, aspiration (*M* =1.3, *SD* = 1.1) and glottal stop (*M* =0.6, *SD* = 0.9), compared to the Arabic group (*M* =0.2, *SD* = 1.3) and glottal stop (*M* =0.5, *SD* = 0.8). An ANOVA analysis on *d′* values showed no effect of cue type (*F* (1, 136) = 3.65, *p* <.056).This indicates that there was minimal bias towards reporting hearing one word or two.

To test the effect of the two segmentation cues; aspiration and glottal stop, a mixed-effects regression model was used. In this full model, the type of cue (aspiration and glottal stop) and native language was included as fixed effects. Subjects and stimuli item were included as random effects. To test the significance of the cue type (aspiration and glottal stop), the cue type was removed and compared to the full model. This comparison revealed a significance of cue type ($\chi 2 = 23.95$, $p < .001$). Similarly, another model was conducted without native language, which revealed a significance of native language ($\chi 2 = 59.72$, $p < .001$) when compared to the full model. Finally, a significant interaction between the two fixed effects; type of cue and native language was found ($\chi 2 = 6.64$, $p < 0.01$).

Discussion.

A central finding of the study is that second language learners do not segment speech in

the target language the way that native speakers do. The native English group was more likely to report two syllables with a strong segmentation cue between them as two words in both conditions. This was not the case for second language learners. Listeners in that group did not use the available cues as effectively as native listeners did. The Arabic group was able to perceive and treat glottal stops as cues, which are available in their L1 more than aspiration (long lag VOT), which is absent from their L1. Thus, it is possible to conclude that second language learners are unable to use the full set of allophonic cues available to native speakers.

Transfer from L1 could be a key factor responsible for the Arabic speakers' differential scores on the aspiration and glottal stop stimuli; since a glottal stop occurs in syllable onsets in Arabic while aspiration does not. This would corroborate previous findings regarding the effect of language specific phonetic and phonotactic cues. However, there is one important note on the glottal stop condition, which was detected more than aspiration by both groups. It seems like certain conditions facilitate segmentation more than the other even if both are available in the L1. One reason could be that vowel initial utterances in English are often preceded by a glottal stop to create an onset. The process of glottal stop insertion preceding vowel initial words is a common phenomenon found in different languages (Borden et al. 2003). Thus, it might be the case that the unmarked process of glottal insertion is learned more easily as a segmentation cue by both native and non-native speakers. In terms of the results present here, transfer, markedness, or a combination of both might be contributing to both groups' higher scores in detecting the glottal stop stimuli more than the aspiration stimuli. Finally, even with absence of content or meaning in the stimuli, native and non-native speakers utilized allophonic cues to a certain extent. However, the English group's score did not exceed 85%. This could possibly suggest that lexical knowledge cannot be excluded as a crucial influence in speech segmentation. Although, the results in this study present an initial investigation in the use of two segmentation cues by Arabic learners of English. Further research is required to address the limitations of the study and draw solid conclusions.

Limitation and Further Analysis.

Future investigation in L2 acquisition of phonetic cues will require supplemental tasks

that could better represent the subjects' perception of speech boundaries. The present study examined the use of only two of the many allophonic cues used in English word segmentation and only in the onset position. It is necessary to include not only different cues, but also consider their position in the syllable. Furthermore, L2 word segmentation studies have not addressed issues such as coarticulation effects across word boundaries, such as assimilation and deletion, which may cause significant changes in the perception of phonetic segmentation. Additionally, phonetic segmentation of Arabic dialects has not been previously studied. Before comparing L2 acquisition of non-native segmentation cues, it is important to know the native segmentation patterns to better understand and compare their application of L2 segmentation strategies.

References.

Al-jasser, Faisal. 2008. The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System* 36*,* 94–106.

Altenberg, Evelyn P. 2005. The perception of word boundaries in a second language. *Second Language Research*, 21(4), 325-358.

Audacity Team. 2015. Audacity, version 2.1.0-beta. Online: http://audacity.

sourceforge.net/ (accessed 31 March 2015)

Borden, G. J., Harris, K. S; Raphael, L. J. 2003. *Speech Science Primer*, 4th ed. Lippincott Williams and Wilkins, Philadelphia, PA

Brown, Cynthia. A. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14, 136–93.

Cutler, Anne; Dennis Norris; Norris, D. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human perception and performance* 14(1), 113.

Flege, James Emil; Ian RA MacKay. 2004. Perceiving vowels in a second language. *Studies in second language acquisition* 26(01), 1-34.

Gow, Jr, 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, 344–359.

Hanulikova, Adriana; Holger Mitterer; James M. McQueen. 2011. Effects of first and second language on segmentation of non-native speech. *Bilingualism: Language and Cognition* 14(04), 506-521.

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.

Khattab, Ghada. 2000. VOT production in English and Arabic bilingual and monolingual children. *Amsterdam studies in the theory and history of linguistic science series* 4, 1-38.

MacMillan, Neil A; C. Douglas Creelman. 1991. Detection theory: A user's guide. New York: Cambridge University Press.

Mattys, Sven L., ;Laurence White; James F. Melhorn. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*. 134, 477–500.

McQueen James M; Takashi Otake; Anne Cutler. 2001. Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language*, 45, 103–132.

McQueen, James. M. 1998. Segmentation of continuous speech using phonotactics. *Journal of memory and language*. 39, 21–46.

Nakatani, Lloyd H. ;Kathleen D. Dukes. 1977. Locus of segmental cues for word juncture. *The Journal of the Acoustical Society of America*. 62, 714–719.

Norris, Dennis. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189– 234.

Norris, Dennis, James M. McQueen, Anne Cutler, and Sally Butterfield. 1997. The possible word constraint in the segmentation of continuous speech. *Cognitive Psychology*, 34(3), 191-243.

Sanders, Lisa D ; Helen J. Neville. 2000. Lexical, syntactic, and stress-pattern cues for speech segmentation. *Journal of Speech, Language, and Hearing Research*. 43, 1302–1321.

Shoemaker, Ellenor; Rebekah Rast. 2013. Extracting words from the speech stream at first exposure. *Second Language Research*, 29(2), 165-183.

Vitevitch, Michael S ;Paul A. Luce. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. Behavior Research Methods, Instruments, and Computers, 36, 481-487. Online: http://129.237.66.221/VLbrmic.pdf

Vitevitch, Michael S. ; Paul A. Luce; Jan Charles-Luce; David Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and speech*, 40(1), 47-62.

Vroomen, Jean; Jyrki Tuomainen; Beatrice de Gelder. 1998. The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*. 38, 133–149.

Watson, Janet CE. 2007. *The phonology and morphology of Arabic*. Oxford university Press.

Weber, Andrea; Anne Cutler. 2006. First-language phonotactics in second-language listening.

Journal of the Acoustical Society of America, *119*(1), 597-607. Table

APPENDIX 1.

Table 1. Aspiration stimuli:

+aspiration	**- aspiration**
1. [bis**k**^{h}ɝm]	1. [bis**k**ɝm]
2. [dʌs**p**^{h}aɪm]	2. [dʌs**p**aɪm]
3. [fɝs**k**^{h}ed]	3. [fɝs**k**ed]
4. [hes**k**^{h}is]	4. [hes**k**is]
5. [fis**p**^{h}eb]	5. [fis**p**eb]
6. [ris**t**^{h}aɪs]	6. [ris**t**aɪs]
7. [nes**p**^{h}im]	7. [nes**p**im]
8. [hʌs**t**^{h}en]	8. [hʌs**t**en]
9. [dis**p**^{h}ʌm]	9. [dis**p**ʌm]
10.[ses**t**^{h}ʌl]	10.[ses**t**ʌl]

Table 2. Glottal stop stimuli:

+glottal stop	**- glottal stop**
1. [hinʔɝt]	1. [hinɝt]
2. [taɪsʔʌl]	2. [taɪsʌl]
3. [rinʔaɪk]	3. [rinaɪk]
4. [sʌvʔin]	4. [sʌvin]
5. [vinʔɝm]	5. [vinɝm]
6. [lʌnʔaɪp]	6. [lʌnaɪp]
7. [bilʔʌb]	7. [bilʌb]
8. [baɪnʔim]	8. [baɪnim]
9. [θinʔʌf]	9. [θinʌf]
10.[remʔaɪd]	10.[remaɪd]