Use of Prosodic Cues in Speech Segmentation: The Effect of Recent Linguistic Exposure

Jui Namjoshi[1], Annie Tremblay[2], Mirjam Broersma[3,4], Sahyang Kim[5], Taehong Cho[6]

[1] Department of French, University of Illinois
[2] Department of Linguistics, University of Kansas
[3] Donders Centre for Cognition, Radboud University Nijmegen, The Netherlands
[4] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[5] Hongik University, Seoul, Korea
[6] Hanyang University, Seoul, Korea

namjosh2@illinois.edu, atrembla@ku.edu, mirjam.broersma@mpi.nl, sahyang@gmail.com, tcho@hanyang.ac.kr

Abstract

This research examines the relative influences of native language and recent linguistic exposure on listeners' use of prosodic cues in artificial-language speech segmentation. We tested listeners' use of fundamental-frequency (F0) rise as cue to word-final boundaries. Participants included French listeners who had spent various amounts of time in the US since their last stay in France, high-proficiency English learners of French who had spent various amounts of time in the US since their last French immersion, and English listeners without functional knowledge of French. The results suggest that recent exposure more accurately predicts listeners' segmentation accuracy than native language does.

Index Terms: speech segmentation, prosodic cues, artificial-language learning

1. Introduction

It is well established that speech segmentation is a language-specific skill: The cues to word boundaries that listeners use are those that are reliable in their native language (L1) and that make the native listening process a highly efficient one (for discussion, see [1]). Cross-linguistic studies have shown that speakers of a given language who perform a segmentation task in a language they do not know automatically transfer the use of L1 cues [2][3][4][5]. Studies using artificial-language (AL) learning paradigms have similarly found that listeners use L1 cues to locate word boundaries in the AL speech stream, and the use of these cues alongside transitional probabilities leads to higher segmentation accuracy than the use of transitional probabilities alone [6][7][8].

What is less clear is whether the linguistic input to which listeners have been recently exposed plays a significant role in their use of prosodic cues to speech segmentation. On the one hand, input in a second/foreign language (L2) could influence native listeners' use of segmentation cues in their L1, especially if they live in an environment where the L1 is not spoken. By the same token, native listeners might become increasingly more apt at using prosodic cues that are efficient for segmenting the L2 if they have gained substantial knowledge of that L2, and they might even transfer the use of these L2 cues to unfamiliar languages. Since previous research on speech segmentation has tended to conflate native language and recent linguistic exposure, the conclusions drawn from this research may have potentially overestimated the effect of L1 and underestimated the flexibility of the speech processing system viz. the linguistic input to which it is exposed.

The present study is an initial attempt at testing this hypothesis. It does so by employing an AL learning paradigm in which participants are exposed to an unfamiliar speech stream for approximately 20 minutes and subsequently tested on the words they heard in the speech stream. Such a paradigm is useful, because it can be administered to a large number of participants irrespective of their native language, and the characteristics of the speech stream can be manipulated independently.

Using an AL learning paradigm, [6] investigated Dutch and Korean listeners' use of fundamental-frequency (F0) rise (among other prosodic cues) as a cue to word-final boundaries. In Dutch, F0 rise can coincide with word-initial boundaries, whereas in Korean, F0 rise is a reliable cue to word-final boundaries. Their results show that after initial exposure to the AL speech stream, Korean listeners, but not Dutch listeners, were able to use F0 rise as a cue to word-final boundaries; yet, after a second exposure session (three days later), the Dutch listeners were also able to use F0 rise as a cue to word-final boundaries. One could interpret these findings as suggesting that recent linguistic exposure plays a critical role in listeners' ability to extract prosodic cues from the speech signal. However, in a similar study, [8] showed that Dutch listeners were able to use F0 rise as a cue to word-final boundaries after initial exposure to a different AL speech stream. In other words, it is unclear whether the Dutch listeners tested in both studies learned to locate word-final boundaries from minimal exposure to the speech stream, or whether this was something they were already able to do because of their native language.

We investigate the relative influences of native language and recent linguistic exposure by partially replicating the AL learning experiment used in [6] with native French listeners who have spent various amounts of time in the US since their last stay in France, high-proficiency English L2 learners of French who have spent various amounts of time in the US since their last French immersion, and native English listeners without functional knowledge of French (or Korean). By examining these participants' ability to segment an AL speech stream as a function of their recent exposure to English and as

a function of their knowledge of French, we may decouple the effects of native language from linguistic input on listeners' extraction of prosodic cues from the speech signal. The prosodic systems of French and English make this pairing of languages ideal for investigating the use of F0 rise as a cue to word-final boundaries.

Similarly to Korean, accented syllables in French occur at the right edge of the Accentual Phrase (AP), and thus are word-final; in non-utterance-final positions, these accented syllables have higher F0 and longer duration than the corresponding unaccented syllables [9][10]. Native French listeners use these prosodic cues to parse accented syllables as word-final syllables [11][12][13]. It should be mentioned, however, that F0 rise (without the concurrent presence of increased duration) can also mark the beginning of the AP in French [9][10], and native French listeners also use this prosodic information to locate word-initial boundaries [14][15][16]. Conversely, in English, accented syllables tend to be word-initial [17][18], with F0 rise, increased amplitude and, to some extent, increased duration signaling this word-initial prominence, but with increased duration also being a reliable cue to word-final boundaries [8][19][20]. Native English listeners have been found to use this prosodic information to identify word-initial boundaries in continuous speech [21][22][23].

Our participants heard the condition of the AL experiment in [6] where F0 rise marked word-final boundaries. Given the distinct prosodic systems of French and English, if the L1 plays an important role in listeners' use of prosodic cues in speech segmentation, native French listeners should outperform both high-proficiency English L2 learners of French and native English listeners without functional knowledge of French (or Korean), even if the majority of the native French listeners were tested in the US and had spent various amounts of time in the US since their last stay in France. By contrast, if it is knowledge of French that matters, then native French listeners and high-proficiency English L2 learners of French should outperform native English listeners without functional knowledge of French (or Korean), and the L2 learners' performance should improve as their proficiency in French increases. A third possibility is that the listeners' ability to use F0 rise as a cue to word-final boundaries may decrease as the amount of time they have spent in the US since their last stay in France (for the native French listeners) or since their last substantial (i.e., 3+ months) French immersion (for the English L2 learners of French) increases.

2. Method

2.1. Participants

Ten native French speakers from France (L1 French group; mean age: 26.7, standard deviation (SD): 5.5), 10 native speakers of American English at high proficiencies in French (L2 French group; mean age: 27.3, SD: 2.3), and 10 native speakers of American English without functional knowledge of French or Korean (L1 English group; mean age: 24.3, SD: 4.3) participated in this study. Eight of the 10 native French speakers lived in the US at the time of the testing and had not returned to France for 31.2 months (SD: 38.4). The English L2 learners of French had first been exposed to French at the mean age of 13 (SD: 5.5); they had lived in a French environment for 14.3 months (SD: 10.7); they used French 21.6% of the time every week (SD: 16.1); and they had been in the US for 49 months (SD: 31.4) since their last substantial (3+ months) French immersion. The L2 learners' proficiency in French was established on the basis of a cloze (i.e., fill-in-the-blank) test [24]. The L2 learners received a mean cloze test score of 33 (SD: 4.3), which is in the range of students who pursue graduate studies in French in the US.

2.2. Materials and Procedures

The experiment consisted of two phases: an exposure phase and a test phase.

In the exposure phase, participants were exposed to the AL speech stream used in [6] where F0 rise marked word-final boundaries. The AL consisted of six trisyllabic words. Four consonants (/p, t, k, m/) and four vowels (/a, i, u, ɛ/) were used to create 16 syllables, which were then combined into six trisyllabic words: [tikepu], [petami], [mupaki], [kapime], [kutepa], [pimatu]. Ten repetitions of the syllables were recorded individually by a female native speaker of Korean, who recorded the stimuli for [6]. (Since the selected consonants and vowels are similar in French and Korean, we do not anticipate that this should adversely affect the results.) The syllables selected for the AL had their duration, intensity, and F0 normalized to the average value of all syllables. All syllables were 252 ms long and had a baseline F0 of 190 Hz.

The syllables were then combined to create the six trisyllabic words, and the last syllable of each word had its F0 raised to 250 Hz. The words were randomly concatenated such that each word would be heard a total of 126 times throughout the AL. No word occurred twice in a row, and there was no pause between any of the words. Syllable-to-syllable transitional probability ranged from 0.5 to 1 within words and from 0.03 to 0.44 between words. The total duration of the AL was approximately 10 minutes, and the participants listened to it twice (total = 20 mins.). There were 20-ms fade in and fade out periods at the beginning and end of the speech stream so that listeners could not use the onset of the initial word and the offset of the final word to locate word boundaries.

In the test phase, the participants heard 36 pairs of trisyllabic sequences, and for each pair they identified which word they thought they heard in the AL. These 36 pairs were created by comparing the six AL words to three part-words and three non-words. Part-words had an additional syllable added to the last two syllables of a legal word (the transitional probability between the second and third syllables was 0.39-0.44). Non-words had syllables in a completely unfamiliar order, the transitional probability within them being zero. All syllables in the test phase had a baseline F0 of 190 Hz.

The participants were told that they would be listening to an AL. They were told that the AL was not French or like French, and that they should not be looking for French words in the speech stream. However, because many of the participants were recruited from French classrooms, and given the affiliation of the first author (who administered the experiment), the participants knew that French was likely to be involved one way or another.

2.3. Data Analysis

For the purpose of statistical analysis, since we had only 10 participants per group, we replaced missing and outlying data as follows.

Since one native French listener's number of months spent in the US since her last stay in France was higher than two SDs away from the mean of all the participants who had functional knowledge of French (her time spent in the US: 120; mean: 37, SD: 35.3), this outlying value was replaced by

the value of two SDs above the mean (107.5 months) so that it would not drive any effect we may find.

We defined substantial immersion in a French-speaking environment as greater than 3 months. Two L2 learners had never had substantial French immersion. Thus, for them, time spent in the US would have corresponded to their age multiplied by 12 months. These values would have been much too high as compared to the others, potentially driving any effect that we may find. These two L2 learners were therefore assigned a number of months in the US that corresponded to two SDs above the mean for the group of L2 learners (assignment: 90 months; mean; 25.8, SD: 25.8).

Two L2 learners did not complete the cloze test because they knew the answers to the test. We therefore assigned them the mean score that the other eight participants received on the test (i.e., 33). These L2 learners were graduate students in French, so we believe this score assignment is a close reflection of their proficiency in French.

First, we performed one-sample *t*-tests on the participants' mean accuracy to determine if each group had obtained accuracy rates that were significantly above chance (50%). Then, we ran logit mixed-effects models on the participants' accuracy in selecting the word that they thought had been heard in the AL (for discussion, see [25]). All models had participant and test item as crossed random variables, but they differed in their dependent and fixed variables. The first two models were performed on all the responses (1=correct, 0=incorrect), with either L1 (French, English) or functional knowledge of French (yes, no) as fixed variable. L1 and functional knowledge of French were not included in the same model because they were partially correlated: For native French listeners, L1 always coincided with knowledge of French, whereas for the native English listeners who had no functional knowledge of French, it never did. The third model was performed on the responses of the participants who had functional knowledge of French, with both L1 and months spent in the US since their last stay in France or since their last substantial French immersion as fixed variables. One should note here that the native French listeners as a group, including the two French listeners who were tested in France, had been in the US for a smaller number of months than the L2 learners (respectively, 23.7 vs. 49 months). Finally, a fourth model was performed on the L2 learners' responses, with proficiency (cloze test scores) and months of French immersion as fixed variables. The two fixed variables in this model were not correlated (r=.05).

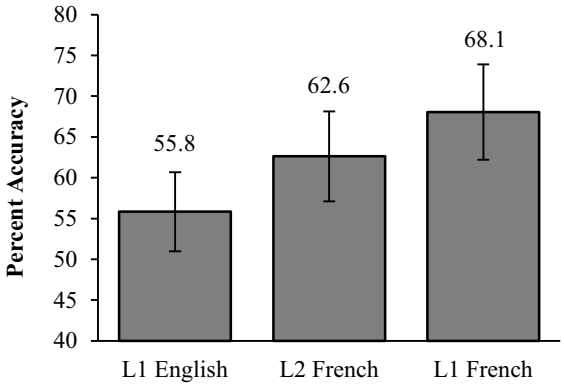
3. Results

Figure 1 shows the percentage of correct responses and the standard error that each group obtained on the word identification task. The percentages and standard errors were computed from the participant means.

One-sample *t*-tests revealed that the L1 French group obtained accuracy rates that were significantly above chance, $t(9)=3.4$, $p<.008$; however, the L2 French group performed only marginally significantly above chance, $t(9)=1.9$, $p<.089$, and the L1 English group's performance was only at chance, $t(9)=1.2$, $p<.217$.

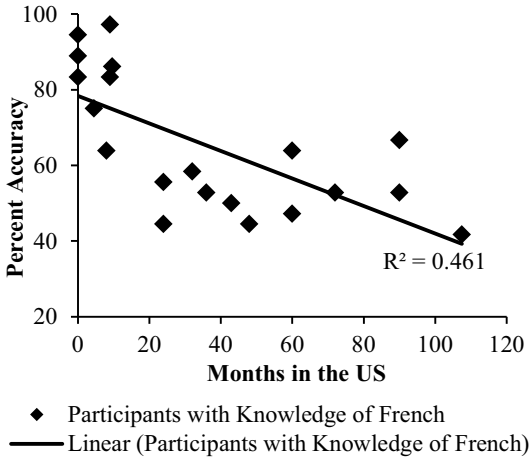
The first logit mixed-effects model (on all listeners' accuracy) revealed a marginally significant effect of L1, $z(1079)= 1.8$, $p<.1$. The second logit mixed-effects model (on all listeners' accuracy) revealed no significant effect of knowledge of French, $z(1079)= 1.5$, $p>.1$.

Figure 1. *Percent accuracy rates*



The third logit mixed-effects model (on the accuracy of only the listeners who had functional knowledge of French) revealed a significant effect of months spent in the US after last stay in France or last substantial immersion in a French-speaking environment, $z(717) = –3.4$, $p<.001$, but no effect of L1 and no interaction between months spent in the US and L1, z's<|1|. Figure 2 illustrates the relationship between the participants' number of months spent in the US and their accuracy on the segmentation task. As can be seen from this figure, the accuracy means for each participant show a sharper decline during the first 20 months of stay in the US than thereafter, suggesting that this relationship may not be linear.

Figure 2. *Months in the US as predictor of accuracy*



Finally, the fourth logit mixed-effects model (on only the L2 learners' accuracy) revealed no significant effect of proficiency or months spent in a French speaking environment, and no interaction between the two, z's<|1|.

These results indicate that time spent away from a French-speaking environment is the best predictor of the participants' ability to use F0 rise as cue to word-final boundaries in an AL. The effect of L1 did have a marginal effect on the analysis of all the accuracy, but it was also the case that the group of native French listeners had spent less time in the US after their last stay in France than the English L2 learners of French after their last substantial immersion. In other words, it is unclear whether the observed effect of L1 is indeed a mild one or whether it can be attributed to the native French listeners' more recent exposure to French as compared to that of L2 learners.

4. Discussion and Conclusion

This study investigated the relative effects of L1 and recent linguistic exposure on listeners' use of prosodic cues in an AL. Our results shed new light on this relative influence: They indicate that recent linguistic exposure is a strong predictor of listeners' ability to extract prosodic cues from an unfamiliar speech stream, perhaps even more so than L1, and the relationship between listeners' use of prosodic cues and linguistic exposure appears stronger near the onset of new linguistic input. This suggests that the speech processing system is extremely flexible and can readily adapt to new incoming linguistic input.

These findings raise important questions for research on non-native speech segmentation: If listeners' processing system is so flexible, why is it that L2 learners infrequently reach native-like levels in their use of prosodic cues that are efficient for segmenting the L2 but not the L1? On the one hand, our participants did not reach very high accuracy rates, suggesting that their use of prosodic cues to word boundaries is not always successful. On the other hand, the present results may be in part due to the AL learning paradigm used in this and many other studies [6][7][8]. AL speech streams are manipulated such that they contain a very limited number of specific (e.g., prosodic and transitional) cues to word boundaries. No other information is available in the speech stream that listeners can use to make inferences about the identity of the actual words. Hence, listeners cannot use lexical, syntactic, or semantic information to infer where word boundaries are located. This makes it easier for them to extract the relevant segmentation cues, including prosodic ones.

Research on L2 speech segmentation shows that L2 learners are in fact very successful at using segmentation cues derived from lexical words (e.g., phonotactic and segmental cues) for locating word boundaries [26][27][28][29], and they show little transfer of L1 prosodic cues when segmenting an L2 that they have learned or are in the process of learning [26][29]. This seems to suggest that lexical information plays a much stronger role than prosodic information in L2 speech segmentation, something which has also been suggested for native listeners [30]. Hence, research that focuses on the role of segmentation cues in the learning of an actual language (e.g., [13]) is also crucial to establish how non-native listeners use these different types of cues in online speech processing.

5. Acknowledgements

This research was supported by funding from the Department of French at the University of Illinois, by a Korea Research Foundation grant from the Government of Korea to the fifth author (KRF-2009-32A-A00060), and by a Veni grant from the Netherlands Organization for Scientific Research (NWO) to the third author. We would like to thank all our participants at the University of Illinois and in France, as well as three anonymous reviewers for their valuable comments.

6. References

[1] Cutler, A. "Listening to a second language through the ears of a first", Interpreting, 5:1-23, 2000/2001.

[2] Cutler, A. "The syllable's role in the segmentation of stress languages", Lg. Cog. Proc., 12:839-845, 1997.

[3] Cutler, A., Mahler, J., Norris, D. G., and Segui, J. "The syllable's differing role in the segmentation of French and English", J. Mem. Lg., 25:385-400, 1986.

[4] Otake, T., Hatano, G., Cutler, A., and Mehler, J. "Mora or syllable? Speech segmentation in Japanese", J. Mem. Lg., 32:258-278, 1993.

[5] Otake, T., Halano. G., and Yoneyallla, K. "Speech segmentation by Japanese listeners", in T. Otake and A. Cutler [Eds.] Phonological Structure and Language Processing: Cross-Linguistic Studies, 183-201, Mouton de Gruyter, 1996.

[6] Kim, S, Broersma, M., & Cho, T. "The use of prosodic cues in processing an unfamiliar language", St. Sec. Lg. Acq., to appear.

[7] Saffran, J. R., Newport, E. L., and Aslin, R. N. "Word segmentation: The role of distributional cues", J. Mem. Lg., 35:606-621, 1996.

[8] Tyler, M. D., and Cutler, A. "Cross-language differences in cue use for speech segmentation", J. Acoust. Soc. Am., 126:367-376, 2009.

[9] Jun, S.-A., & Fougeron, C. "Realizations of accentual phrase in French intonation", Probus, 14:147-172, 2002.

[10] Welby, P. "French intonational structure: Evidence from tonal alignment", J. of Phonetics, 34:343-371, 2006.

[11] Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. "Phonological phrase boundaries constrain lexical access: I. Adult data", J. Mem. Lg., 51:523-547, 2004.

[12] Michelas, A., and D'Imperio, M. "Accentual phrase boundaries and lexical access in French", in Proceedings of the Speech Prosody 2010 Conference, Chicago. Online: http://speech prosody2010.illinois.edu/papers/100882.pdf, 2010.

[13] Tremblay, A., Coughlin, C. E., Bahler, C., & Gaillard, S. "Differential contributions of in the native and non-native segmentation of French speech", Lab. Phon., to appear.

[14] Spinelli, E., Grimault, N., Meunier, F., and Welby, P. "An intonational cue to word segmentation in phonemically identical sequences", Att. Perc. Psychophys, 72:775-787, 2010.

[15] Spinelli, E., Welby, P., and Schaegis, A.-L. "Fine-grained access to targets and competitors in phonemically identical spoken sequences: The case of French elision", Lg. Cog. Proc. 22:828-859, 2007.

[16] Welby, P. "The role of early fundamental frequency rises and elbows in French word segmentation", Sp. Com. 49:28-48, 2007.

[17] Clopper, C. "Frequency of stress patterns in English: A computational analysis", in IULC Working Papers. Online: https://www.indiana.edu/~iulcwp/ pdfs/02-clopper02.pdf, 2002.

[18] Cutler, A., and Carter, D. M. "The predominance of strong initial syllables in the English vocabulary", Comp. Sp. Lg. 2:133-142, 1987.

[19] Beckman, M. E. Stress and Non-Stress Accent, Foris, 1986.

[20] Lieberman, P. "Some acoustic correlates of word stress in American English", J. Acoust. Soc. Am. 32: 451-454, 1961.

[21] Cutler, A., and Norris, D. "The role of strong syllables in segmentation for lexical access", J. of Exp. Psych.: Human Perc. & Perf., 14:113-121, 1988.

[22] Mattys, S. L. "Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation", J. Exp. Psych: Human Perc. Perf., 30: 397-408, 2004.

[23] McQueen, J. M., Norris, D., and Cutler, A. "Competition in spoken word recognition: Spotting words in other words", J. Exp. Psych.: Learn. Mem. Cog., 20:621-638, 2004.

[24] Tremblay, A. "Proficiency assessment standard in second language acquisition research: "Clozing" the gap", St. Sec. Lg. Acq., 33:339-372, 2011.

[25] Baayen, R. H. Analyzing Linguistic Data: A Practical Introduction to Statistics. Cambridge University Press, 2008.

[26] Hanulíková, A., Mitterer, H., and McQueen, J. M. "Effects of first and second language on segmentation of non-native speech", Biling. Lg. Cog., 14:506-521,2011.

[27] Tremblay, A. "Learning to parse liaison-initial words: An eye-tracking study", Biling. Lg. Cog., 14:257-279, 2011.

[28] Weber, A., and Cutler, A. "First-language phonotactics in second-language listening", J. Acoust. Soc. Am., 119:597-607, 2006.

[29] White, L., Melhorn, J. F., & Mattys, S. L. "Segmentation by lexical subtraction in Hungarian speakers of second language English", Quart. J. Exp. Psych., 63:544-554, 2010.

[30] Mattys, S. L., White, L., and Melhorn, J. F. "Integration of multiple segmentation cues: A hierarchical framework", J. of Exp. Psycho.: Gen., 134:477-500, 2005.