

SPEECH SEGMENTATION IS ADAPTIVE EVEN IN ADULTHOOD: ROLE OF THE LINGUISTIC ENVIRONMENT

Jui Namjoshi⁽¹⁾, Annie Tremblay⁽²⁾, Elsa Spinelli⁽³⁾, Mirjam Broersma⁽⁴⁾, Maria Teresa Martínez-García⁽²⁾,
Katrina Connell⁽²⁾, Taehong Cho⁽⁵⁾, & Sahyang Kim⁽⁶⁾

¹University of Illinois, ²University of Kansas, ³Université Pierre-Mendès France, ⁴Radboud University
Nijmegen, ⁵Hanyang University, ⁶Hongik University
atrembla@ku.edu

ABSTRACT

In this paper, we show that adult listeners who speak the same native language but live in different linguistic environments differ in their use of prosodic cues that signal word boundaries in the native language. Non-utterance-final word-final syllables have higher fundamental frequency in French. Adult native French listeners living in France or in the US completed an artificial-language segmentation task where fundamental frequency cued word-final boundaries (experimental). Other native French listeners living in France completed the corresponding task without prosodic cues (control). Results showed that France French listeners outperformed US French listeners and control French listeners, but US French listeners did not outperform control French listeners. The poorer performance of US French listeners is attributed to their regular exposure to (and thus interference from) English, a language where fundamental frequency signals word-initial boundaries. This suggests speech segmentation is adaptive, with listeners tuning in to the prosody of their linguistic environment.

Keywords: speech segmentation, artificial language, prosody, linguistic environment, French

1. INTRODUCTION

How adaptive is speech segmentation in adulthood? Existing research suggests that the native language (L1) plays an important role in determining how adult listeners segment an unfamiliar language into individual words. In artificial-language (AL) segmentation studies (e.g., [1,2,3]), listeners who speak different L1s are compared in their ability to use different segmentation cues. The typical finding is that listeners whose L1 patterns like the AL outperform listeners whose L1 patterns differ from the AL. In second-language (L2) segmentation studies (e.g., [4,5,6,7]), native and non-native listeners are compared in their ability to segment the native listeners' L1. These studies usually find that native listeners outperform non-native listeners in their ability to locate word boundaries in the target

language. From these segmentation studies, it is concluded that L1 exposure has an important impact on how listeners locate word boundaries in continuous speech.

Unclear, however, is the degree to which speech segmentation remains adaptive in adulthood. Finding that native listeners outperform non-native listeners in their use of segmentation cues in the target language does not necessarily imply that speech segmentation is not adaptive, in that native listeners differ from non-native listeners in their early (and most critical) experience with the target language. A better test of whether or not speech segmentation is adaptive in adulthood is one in which listeners who do not differ in their early experience with the native language but who differ in their subsequent exposure to other languages segment the same unfamiliar language.

The present study provides such a test by using an AL segmentation task to investigate native French listeners' use of fundamental frequency (F0) as a cue to word-final boundaries. In French, the last non-reduced syllable of the Accentual Phrase receives a pitch accent [8,9,10]. Syllables that receive a pitch accent are thus word-final in French. In natural French speech, non-utterance-final syllables with a pitch accent have higher F0 and are longer than the corresponding unaccented syllables [9,10]. Native French listeners have indeed been found to use both F0 and duration (together and on their own) as cues to word-final boundaries [2,6,11,12,13].

In this study, we examine whether native French listeners living in France and thus exposed to French on a regular basis segment speech differently from native French listeners who moved to the US in adulthood and thus are currently exposed to English on a regular basis. English differs from French in that accented syllables, which are lexically stressed in English, tend to be word-initial [14,15]. Whereas F0 provides a strong cue to word-initial boundaries in English, duration can signal both word-initial stress and phase-final lengthening [2,16]. Since French and English differ more markedly in how F0 signals word boundaries, the present study focuses on the use of F0 cues.

We hypothesize that speech segmentation remains adaptive in adulthood; as such, the speech processing system should abandon segmentation routines that are inefficient to segment speech in the linguistic environment to which listeners are primarily exposed, and possibly develop new segmentation routines that would prove more efficient for segmenting the new language. Thus, in an AL task where F0 signals word-final boundaries, native French listeners living in the US are predicted to perform more poorly than native French listeners living in France. Such a finding would suggest that French listeners' regular exposure to English interferes with their use of F0 cues to word-final boundaries in the AL.

2. METHOD

2.1. Participants

Participants included: (i) 24 native French listeners living in France who heard an AL where F0 cued word-final boundaries (mean age: 21.6; 22 females); (ii) 20 native French listeners living in the US who heard exactly the same AL (mean age: 25.3; 13 females); and (iii) 23 native French listeners living in France who heard the corresponding AL without any prosodic cue (mean age: 19; 21 females). Groups (i)-(ii) serve as experimental groups and group (iii) serves as control group.

For all listeners, both parents spoke only French as L1. All listeners heard and spoke only French during the first 5 years of their life, and none of them reported traveling to an English-speaking environment before the age of 12. At the time of the study, the French listeners in (ii) had lived in the US for a consecutive period ranging from 3 months to 6 years; and their reported weekly language use was 37.5% (SD: 17.1%) for French and 60.3% (SD: 18.6%) for English.

2.2. Materials and Procedures

This study is a partial replication of [1] with French listeners. The experiment consisted of two phases: an exposure phase and a test phase.

In the exposure phase, the participants in the two experimental groups (i.e., (i)-(ii)) were exposed to an AL speech stream 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, ε/) 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. (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 participants in the control group (i.e., (iii)) heard the same AL but F0 was kept to 190 Hz throughout the speech stream.

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. 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.

2.3. Data Analysis

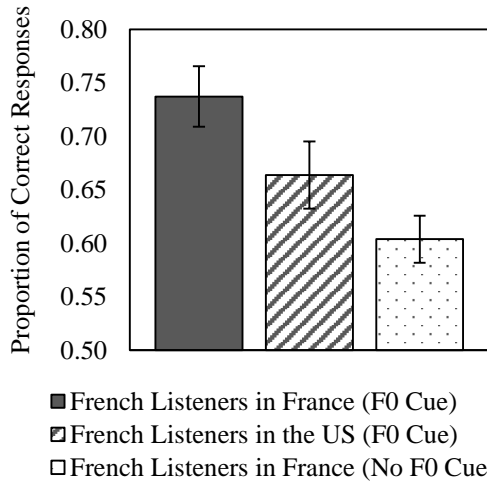
Two logit mixed-effects models were conducted on the participants' (binomial) accuracy, with participant and test item as crossed random variables (for details, see [17]). The first model was run on the data of the participants in the experimental groups (i.e., (i)-(ii)), with linguistic environment (France, US) as fixed variable. In this analysis, France served as baseline. The second model was run on the data of all three groups, with prosodic cue to word-final boundaries (F0, no F0) as fixed variable. In this analysis, the condition without F0 cue served as baseline.

3. RESULTS

Figure 1 shows the mean proportion of correct responses for each group of French listeners. As can

be seen from the results, all three groups performed above chance (chance = 0.5).

Figure 1: French listeners' proportion of correct responses (standard error bars)



The first logit mixed-effects model, reported in Table 1, revealed a marginal effect of linguistic environment, with French listeners living in France outperforming French listeners living in the US.

Table 1: Logit mixed-effects model testing for the effect of linguistic environment ($df= 1580$)

Variable	Estimate	z	p
(intercept)	1.19	6.52	.001
Linguistic environment	-0.41	-1.80	.073

The second logit mixed-effects model, reported in Table 2, revealed a significant effect of F0 cue only for French listeners living in France; French listeners living in the US, who heard the AL in which F0 cued word-final boundaries, did not perform significantly differently from French listeners who heard the AL without prosodic cues.

Table 2: Logit mixed-effects model testing for the effect of F0 cue ($df = 2407$)

Variable	Estimate	z	p
(intercept)	0.47	2.95	.003
F0 cue (France)	0.69	3.56	.001
F0 cue (US)	0.30	1.47	.141

Given the individual variability in the amount of time that the French listeners in (ii) spent in the US and in their percent use of French and English, we

performed correlations between these variables and the US French listeners' mean proportion of correct responses on the AL task (for time spent in the US, the correlation included only 17 listeners due to missing data). These correlations revealed a significant negative relationship between time spent in the US and performance on the AL task ($r = -.55$), but no significant relationship between percent use of French or English and performance on the task (respectively, $r = .11$ and $r = .003$). Thus, the longer French listeners have been in the US, the poorer their proportion of correct responses on the AL task.

We now turn to a discussion of these results and their implications for understanding the degree to which speech segmentation remains adaptive in adulthood.

4. DISCUSSION AND CONCLUSION

This study investigated whether native French listeners living in France and thus exposed to French on a regular basis segment speech differently from native French listeners who moved to the US and thus are currently exposed to English on a regular basis. It did so by examining these French listeners' ability to use F0 as a cue to word-final boundaries in an AL, and by comparing their performance to that of French listeners who heard the corresponding AL without any prosodic cue.

The results revealed a marginally significant difference between French listeners living in France and French listeners living in the US in their ability to use F0 as cue to word-final boundaries in the AL. Furthermore, the results showed that French listeners living in the US, who heard the AL that contained F0 cues to word-final boundaries, did not differ significantly from French listeners who heard the corresponding AL without prosodic cues. Finally, the results showed that French listeners who spent more time in the US performed more poorly on the AL task where F0 signaled word-final boundaries than French listeners who spent less time in the US.

We attribute poorer performance of the French listeners in the US to their exposure to, and thus interference from, English speech. In English, F0 provides a reliable cue to word-initial boundaries. French listeners who are regularly exposed to English may have adapted their use of prosodic cues in speech segmentation if these cues prove to be inefficient for segmenting the language they hear on a regular basis. These adapted segmentation routines may then be transferred to the segmentation of the AL and thus result in poorer performance in the use of F0 cues to word-final boundaries.

One important question that arises from these results, however, is whether the poorer performance

of the French listeners with increasing time in the US is due to their inconsistent use of both French-like and English-like segmentation routines (where syllables with higher F0 are parsed as, respectively, word-final or word-initial) or to their abandoning a French-like segmentation routine without necessarily adopting an English-like one. Since this study did not examine these French listeners' segmentation of an AL in which F0 instead cued word-initial boundaries, this question cannot be answered from the present data. Further research should tease these two possibilities apart. What is clear from the present results, however, is that the French listeners in the US did not rely strictly on an English-like segmentation routine, in that such a routine would conflict with the transitional probabilities that signaled word boundaries in the AL and thus resulted in chance performance on the task.

Overall, the results of this study suggest that speech segmentation remains adaptive even in adulthood, at least to the extent that listeners do not necessarily adopt an L1 segmentation routine if this routine does not prove to be useful for segmenting the language to which these listeners are primarily exposed.

5. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grant no. BCS-1423905 awarded to Dr. Annie Tremblay. Support for this research also comes from a *Language Learning* small research grant awarded to Dr. Annie Tremblay.

6. REFERENCES

- [1] Kim, S., Broersma, M., Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar Language. *Studies in Second Language Acquisition*, 34, 415-444.
- [2] Tyler, M. D., Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America*, 126, 367-376.
- [3] Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The role of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, 38, 133-149.
- [4] Altenberg, E. P. (2005). The perception of word boundaries in a second language. *Second Language Research*, 21, 325-358.
- [5] Hanulíková, A., Mitterer, H., McQueen, J. M. (2011). Effects of first and second language on segmentation of non-native speech. *Bilingualism: Language and Cognition*, 14, 506-521
- [6] Tremblay, A., Coughlin, C. E., Bahler, C., Gaillard, S. (2012). Differential contributions of prosodic cues in the native and non-native segmentation of French speech. *Laboratory Phonology*, 3, 385-423.
- [7] Weber, A., Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, 119, 597-607.
- [8] Jun, S.-A., Fougeron, C. (2000). A phonological model of French intonation. In Antonis Botinis (Ed.), *Intonation: Analysis, modelling, and technology*. Dordrecht: Kluwer, 209-242.
- [9] Jun, S.-A., Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, 14, 147-172.
- [10] Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, 34, 343-371.
- [11] Bagou, C., Frauenfelder, U. H. (2006). Stratégie de segmentation prosodique: rôle des prééminences initiales et finales dans l'acquisition d'une langue artificielle. In *Actes des XXVIèmes Journées d'Étude sur la Parole*. Dinart, France, 571-574.
- [12] Christophe, A., Peperkamp, S., Pallier, C., Block, E., Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language*, 51, 523-547.
- [13] Michelas, A., D'Imperio, M. (2010). Accentual phrase boundaries and lexical access in French. In *Speech Prosody 2010* 100882: 1-4. Retrieved from <http://speechprosody2010.illinois.edu/papers/100882.pdf>.
- [14] Clopper, C. (2002). Frequency of stress patterns in English: A computational analysis. *IULC Working Papers Online*, 2. Available at <https://www.indiana.edu/~iulcwp/pdfs/02-clopper02.pdf>.
- [15] Cutler, A., Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133-142.
- [16] Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.
- [17] Baayen, H. D. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge University Press.