10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

The Opposite Effects of Vowel and Onset Consonant Lengthening on Speech Segmentation

Shu-chen Ou[1], *Zhe-chen Guo*[2]

[1]National Sun Yat-sen University, Taiwan
[2]The University of Texas at Austin, USA
sherryou@mail.nsysu.edu.tw, zcadamguo@utexas.edu

Abstract

This study examines the use of vowel and onset consonant lengthening in adult speech segmentation. It is well-documented that a longer vowel tends to be perceived as a finality cue, improving speech segmentation in the final position of a unit but leading to no facilitation in the initial position. Research on domain-initial strengthening suggests that a longer syllable-onset consonant may also be a segmentation cue, but one that signals initiality. We investigated this possibility with an artificial language (AL) learning experiment with Taiwanese Southern Min listeners. The listeners first learned an AL by listening for 10 to 12 minutes to speech streams in which the "words" of the AL (e.g., /ba.nu.me/) were concatenated without pauses and then identified the words. Higher identification accuracy indicated better segmentation during the learning. Results replicated previous findings on vowel lengthening and further demonstrated that the effects of onset consonant lengthening were opposite to those of vowel lengthening: a lengthened onset consonant improved segmentation in the initial position but resulted in no facilitation in the final one. It is assumed that lengthened vowels and onset consonants may be analyzed by some prosody-computing mechanism as signaling the end and beginning of a prosodic unit, respectively.

Index Terms: speech segmentation, initial onset consonant lengthening, artificial language learning

1. Introduction

A key process underlying spoken language comprehension is the segmentation of connected speech into discrete units such as words. This process is guided by various prosodic cues (e.g., [1–3]) and one well-established finding is the so-called "final vowel lengthening" effect. Listeners of English, French, Dutch, and Korean, for example, all segment continuous nonsense speech better when units in the speech have a lengthened final vowel ([3–5]). Such improvement is not found for initial vowel lengthening. For instance, in a study by [6] with Basque, German, Italian, and Spanish listeners, lengthening the initial vowels in the units of continuous speech leads to no facilitation or even hampers segmentation. These suggest a tendency toward perceiving longer vowels as signaling finality – a segmentation solution that presumably stems from the tendency for the final elements in a prosodic domain to be phonetically lengthened (e.g., [7, 8]). This study aims to replicate previous findings on vowel lengthening and, more importantly, examine another type of prosodic lengthening that has attracted less attention in segmentation research: onset consonant lengthening, for which effects opposite to those of vowel lengthening are expected.

The duration of a consonant in the syllable-onset position is potentially a segmentation cue as it correlates with prosodic boundary strength. One boundary-related phenomenon widely observed across languages is domain-initial strengthening as a result of prosodic organization of speech (e.g., [9–11]). Such strengthening affects articulation of segments in several ways, and one of its effects is that relative to those at the initial position of a lower-level prosodic unit, syllable-onset consonants at the initial position of a higher-level prosodic domain tends to be temporally expanded. For example, for nasal stops like /n/, the expansion is manifested by a longer nasal duration ([10, 12]); for voiced stops like /b/, it is manifested by a longer closure period and hence a greater negative voice onset time ([13]). It is then possible that listeners can exploit longer onset consonants to discover the beginnings of units in continuous speech.

The study that addresses questions most related to this possibility is [14], which investigates whether domain-initial strengthening facilitates segmentation of English words by English listeners. However, while their results provide some evidence for the facilitation, it is unclear whether and how onset consonant lengthening in the initial position per se contributes to segmentation since naturally produced speech was used in their study and their acoustic analysis revealed that there were other cues to domain-initial strengthening, such as increased peak amplitude during consonant articulation. A good initial approach to studying the effects of lengthening an initial onset consonant is to use the artificial language (AL) learning experiment, which is widely adopted by segmentation researchers and has provided much evidence for the final vowel lengthening effect ([3–5]). It requires subjects' segmentation of a nonsense language and has some methodological advantages over designs using real words (e.g., the priming experiments in [14]). For example, it helps minimize lexical confounds and prevent them from obscuring the effects of prosodic cues – an advantage that would be crucial as speech segmentation has been suggested to be driven primarily by lexical knowledge ([15, 16]).

The present research uses the AL learning paradigm to investigate how lengthening of segments – particularly onset consonants – is exploited in segmentation. Two results are expected. The first one is a replication of previous findings on vowel lengthening: that is, the cue should facilitate segmentation in the final position ([3, 4]) but hinder it or have no effect in the initial position ([3, 6]). Second, since a longer onset consonant is one acoustic correlate of domain-initial strengthening, it may be associated with initiality and show the opposite pattern: it would facilitate segmentation in the initial position but hinder it or have no effect in the final position. In other words, vowel lengthening and onset consonant lengthening are mirror images of each other in

10.21437/SpeechProsody.2020-16

terms of their expected effects. In this study, we focus on listeners of Taiwanese Southern Min (TSM), which is among the several languages for which final lengthening has been reported ([17]) and, crucially, onset consonants are found to be lengthened due to domain-initial strengthening ([11]).

2. Experiment

2.1. Experimental design

The AL learning experiment conducted to test the expectations consisted of a learning phase and a subsequent test phase. In the learning phase, subjects learned the words of an AL, which were meaningless syllable sequences, by listening to long, continuous speech streams in which tokens of the words were concatenated without pauses. Next, they received a two-alternative forced-choice test in which they identified the AL words. Higher identification accuracy indicated more successful segmentation during the learning phase.

The experiment was conducted under five conditions. One was a baseline condition – called "TP-only – in which no segments in the learning-phase speech streams were lengthened and the only segmentation cue was the transitional probability (TP) between adjacent syllables. In addition to the TP cue, segmental lengthening was introduced in the other four conditions. In the "initial vowel lengthening (IVL)" and "final vowel lengthening (FVL)" ones, the lengthened segments were the vowels in the initial and final syllables of the AL words, respectively. In the "initial onset lengthening (IOL)" and "final onset lengthening (FOL)" conditions, the lengthened segments were the onset consonants in the initial and final syllables of the AL words, respectively.

2.2. Hypotheses and Predictions

A well-established hypothesis is that a longer vowel is interpreted as a finality cue. This predicts that segmentation would be facilitated by a final lengthened vowel but hampered, or at least not facilitated, by an initial one. The prediction is supported if, compared with that of the baseline TP-only condition, the response accuracy in the test phase is significantly higher in the FVL condition but significantly lower or not different in the IVL one. We also hypothesize that a lengthened onset consonant is perceived as signaling initiality and expect the opposite pattern: segmentation would be facilitated by an initial lengthened onset consonant but inhibited or not facilitated by a final one. Support for this would be that compared with that of the TP-only condition, the test response accuracy is significantly higher in the IOL condition but significantly lower or not different in the FOL one.

2.3. Stimuli

Five vowels (/a, i, e, u, o/) and four consonants (/b, g, m, n/) were used to create the AL lexicon, which comprised of six trisyllabic words with a CVCVCV structure: /ba.nu.me/, /bi.mo.na/, /ge.ni.go/, /mi.ma.bu/, /ne.bo.gi/, and /no.ga.mu/ (the dots indicate syllabic boundaries). All the consonants and vowels occur in TSM and the component syllables of each AL word were read individually by a male sequential bilingual speaker of TSM and Taiwan Mandarin in a carrier sentence (/gua kɔŋ ___/ 'I said ___.'). Recoded items were digitized at a sampling rate of 44.1k Hz. The syllables were then excised from the sentence and subjected to manipulations using Praat ([18]). First, their amplitude was equalized and their F0 contours were flattened at 119 Hz, the average F0 of the original unmanipulated syllables. Next, the consonant and vowels were normalized to have the same duration (as in [3, 19]) and the duration was set to 150 milliseconds (ms). The resulting syllables were concatenated to form the AL words.

In all conditions except TP-only, the consonants or vowels in the AL words further underwent lengthening according to their conditions. In the IOL one, for example, the onset consonants in the first syllables of the trisyllabic AL words were lengthened. Following some AL studies on vowel lengthening (e.g., [3]), we lengthened all the critical segments by 1.5 times. This kept the lengthening amount comparable across the consonants and vowels. A 1.5-times lengthening for onset consonants was within a normal range of lengthening due to domain-initial strengthening ([11, 13]). Figure 1 shows how the same AL word differed across conditions.

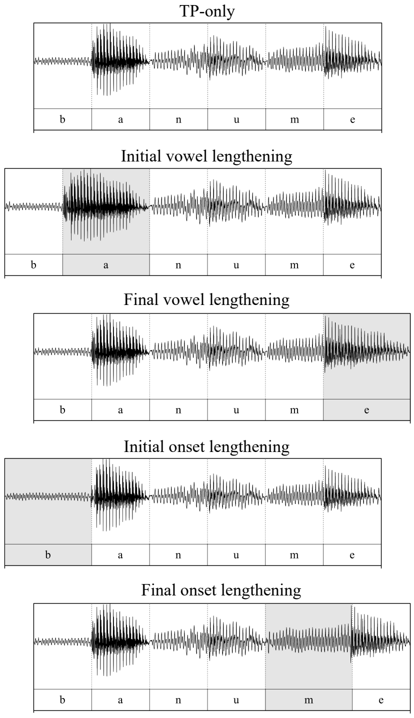

Figure 1: *Waveforms of a sample AL word (/ba.nu.me/) under the five conditions. Lengthened segments are shaded.*

The learning-phase stimuli were six speech streams that each contained 20 repetitions of each of the six AL words. The tokens of the words were concatenated without pauses and in a pseudo-random order such that the same word did not follow itself. As in [3], each stream was faded in and out over a five-second interval to make the very first and last syllables inaudible. Across conditions, the speech streams were identical except they differed in whether and how segmental lengthening was introduced. Their total duration was about 10 to 12 minutes. For any two adjacent syllables in the streams, the TP was the conditional probability of the second syllable given the first one ([5]). An average TP was calculated for each AL word by taking the mean of the TPs of adjacent syllables. All AL words had an average TP of one.

Each trial in the two-alternative forced-choice test presented two stimuli with a 500-ms interstimulus interval: one was an AL word and the other a partword. The partword was a trisyllabic sequence that straddled two AL words (e.g., /bo.gi.no/, which combined the last two syllables of /ne.bo.gi/ and the first one of /no.ga.mu/). There were six partwords (/bo.gi.no/, /gi.ba.nu/, /ma.bu.ge/, /me.bi.mo/, /na.ne.bo/, and /ni.go.mi/) and they had a mean average TP of 0.58 (range: 0.57–0.60). The lengthening cues were removed to prevent subjects from responding based on acoustic matching between the learning and test stimuli (as in [20]). Thus, none of the AL words and partwords presented in the test contained a lengthened a segment and the test was identical for all conditions. The test consisted of 36 trials, generated from all possible pairings of the AL words and partwords. The orders in which the AL words appeared were counterbalanced.

2.4. Procedure

Tested individually in a sound-proof booth, subjects were first asked to learn the words of the AL just by listening to the six speech streams over headphones. They were not given any information regarding the length or number of the words or any cues to word boundaries. All they had to do was to pay as much attention as possible to what they listened to. They were made aware of an upcoming test that would assess if they learned any words of the AL. Optional short breaks were interposed between the speech streams. They preceded to the test immediately after the learning phase. In each trial, they heard the two stimuli and pressed a button labeled “1” on a response box if they thought the first stimulus was an AL word and a button labeled “2” if they thought second stimulus was an AL word. They had five seconds to respond after the presentation of the second stimulus. E-prime 2.0 was used to control stimulus presentation and record responses.

2.5. Participants

One hundred and fifty adult sequential bilingual speakers of TSM and Taiwan Mandarin (60 males and 90 females) were recruited from a university in Southern Taiwan and allocated randomly and equally to the five conditions. They had learned English as a compulsory subject in school and their ages ranged between 18 and 22. None reported history of hearing impairments.

3. Results

Subjects’ responses in the test were correct when they selected the AL word in a trial and incorrect when they selected the partword. Omissions (i.e., failures to respond within the allotted time) accounted for 0.5% of all data and were discarded. Figure 2 shows the mean percentages of correct responses of the five conditions. One-sample *t*-tests indicated that subjects performed significantly better than chance (50%) in all the conditions except for the IVL one (TP-only: $p < 0.001$; IVL: $p = 0.141$; FVL: $p < 0.001$; IOL: $p < 0.001$; FOL: $p < 0.003$), suggesting that they were able to learn some words of the AL under most of the conditions.

To examine if the predictions were supported, a generalized linear mixed-effects logistic regression analysis was performed by using the glmer function from the lme4 package ([21]) of R ([22]). Of main interest was the fixed effect Condition, a dummy-coded factor with the TP-only condition as the baseline level. Following the recommendations of [23], we included log-transformed reaction time (LogRT) and the ordinal number of a trial (Trial) to partial out the influence of potential speed-accuracy tradeoff and practice or fatigue during the test. The random effects included random intercepts for subject, AL word, and partword.

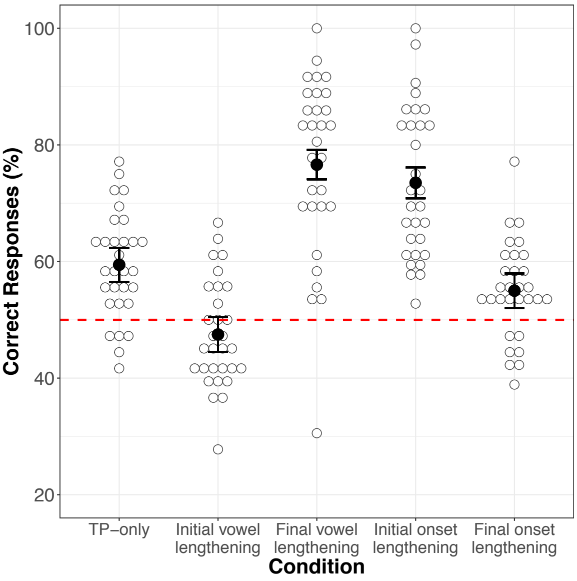

Figure 2: *Individual subjects’ percentages of correct responses (empty circles) and the means of the five conditions (solid circles). The error bars represent 95% confidence intervals and the red horizontal dashed line indicates chance level (50%).*

The fixed-effect results of the analysis are in Table 1. There was a significant effect of LogRT, suggesting that responses with a longer latency were less likely to be correct and subjects did not trade response speed for accuracy (as was found in [23]). Trial was not significant, providing no evidence for training or practice effects. Importantly, the Condition terms indicated that when compared with that of the TP-only condition (mean: 59%), the accuracy of the IVL condition (mean: 47%) was significantly lower whereas that of the FVL (mean: 77%) was significantly higher. Furthermore, relative to the TP-only one, the IOL condition also showed a significantly higher accuracy (mean: 73%). Yet, there was no significant difference between the TP-only and FOL conditions, although the latter had a slightly lower accuracy (mean: 55%). In general, the predictions based on the hypothesized effects of vowel and onset consonant lengthening were supported.

Table 1: *Fixed-effect results of mixed-effects logistic regression analysis of subjects’ responses in the test.*

	Estimate	**SE**	***z***	***p***
(Intercept)	2.131	0.335	6.360	< 0.001
LogRT	-0.258	0.038	-6.865	< 0.001
Trial	-0.001	0.003	-0.241	0.809
Condition (IVL vs. TP-only)	-0.554	0.141	-3.941	< 0.001
Condition (FVL vs. TP-only)	0.857	0.147	5.829	< 0.001
Condition (IOL vs. TP-only)	0.674	0.145	4.644	< 0.001
Condition (FOL vs. TP-only)	-0.222	0.141	-1.582	0.114

4. Discussion and Conclusion

The present study investigates the effects of vowel and onset consonant lengthening on speech segmentation. An AL learning experiment with TSM listeners furnishes additional support for the well-documented finding on vowel lengthening: that is, increasing the duration of the vowel in the final position is favorable to segmentation ([3–5]). Besides, as in [3, 6], lengthening an initial vowel does not result in facilitation; in fact, it even hampers segmentation for our TSM listeners. The novel aspect of this study is that it also examines lengthening of onset consonants. It is hypothesized that this cue would show result patterns opposite to those of vowel lengthening and this indeed is supported: lengthening onset consonants improves TSM listeners' segmentation in the initial position whereas it has no significant effect in the final one. This provides evidence for a segmentation strategy whereby longer onset consonants are perceived as signaling initiality and exploited to locate the beginning of units in continuous speech.

The mechanism that underlies TSM listeners' segmentation behavior may be one that is similar to the Prosody Analyzer, which is proposed by [14] to account for English listeners' use of domain-initial strengthening. The Prosody Analyzer calculates the prosodic structure of perceived speech based on available information and yields segmentation hypotheses. In the current study, lengthened vowels and lengthened onset consonants may be analyzed by a similar mechanism as signaling the end and beginning of a prosodic constituent, respectively. This significantly improves segmentation when the vowels are indeed in the final position (of an AL word) and the consonants in the initial position. The lack of facilitation in the IVL and FOL conditions is then the result of a conflict between prosodic analysis and statistical regularities. In the FOL condition, the prosody-computing mechanism treats the longer onset consonant in the final syllable of an AL word as signaling initiality, thus placing a boundary between the first two syllables and the last one of the word. Such an analysis is at odds with the relatively higher TPs for adjacent syllables in the AL words, which suggest that the syllables within a word should constitute a unit and contain no boundaries. Due to these competing segmentation hypotheses, no facilitation is observed. A similar case can be made for the IVL condition.

The lack of improvement in the IVL and FOL conditions is itself another noteworthy finding as it highlights a slight asymmetry between vowel and onset consonant lengthening in the extent to which they affect TSM listeners' segmentation. While both the IVL and FOL conditions produce no facilitation, only the former significantly reduces segmentation when compared with TP-only. Also, IVL is the only condition in which subjects' overall accuracy is not significantly better than chance. On the other hand, while the two favorable lengthening conditions – FVL and IOL – both significantly enhance segmentation performance (compared with the TP-only one), the former does so slightly better than the latter (see Figure 2). Therefore, vowel lengthening appears to have a more drastic impact than onset consonant lengthening does. A potential explanation for this asymmetry might be cue weighting. That is, vowels might have been assigned a greater cue weight than onset consonants; thus, they are more beneficial when occurring in a position favorable to segmentation and more deleterious when occurring in an unfavorable one. Such an explanation, however, would need to be tested with further experiments, preferably with ones in which cues are systematically pitted against each other (e.g., [16]).

To conclude, this study replicates the effects of vowel lengthening as reported in previous research and further demonstrates that onset consonants with a longer duration are treated as a cue to the beginnings of units in connected speech. Some issues may be explored further to examine how this conclusion would fit into a fuller picture of the role of prosodic lengthening in segmentation. First, given that both lengthening vowels and onset consonants facilitate segmentation, it is of interest to see whether they would produce a synergistic effect when they co-exist. Naturalistic speech seldom contains only one cue to a particular boundary; more often than not, multiple cues may redundantly signal the same boundary. An example scenario is when a lengthened final vowel is followed by a lengthened initial onset consonant and thus the boundary between them is doubly signaled by the lengthening cues. One possible outcome is that the redundancy would improve segmentation to a greater extent than either cue alone. Alternatively, it is may be the case that multiple prosodic cues would have no synergistic effect, as has been observed in some AL learning studies (e.g. [1, 4]). Another further issue concerns whether using longer onset consonants to locate initial positions is a cross-linguistic segmentation solution. Although the current study focuses only on TSM listeners, it is possible that similar results can be replicated with listeners of other languages such as English, French, and Korean as onset consonant lengthening induced by domain-initial strengthening is also found in these languages ([11]). This can be tested to gain improved insight into the universal and language-specific aspects of speech segmentation.

5. Acknowledgments

We would like to thank Kai-ting Chung with subject recruitment and three anonymous reviewers for their comments on an earlier version of this paper. We would also like to thank the Ministry of Science and Technology, Taiwan, for funding the research.

6. References

[1] O. Bagou and U. H. Frauenfelder, "Lexical segmentation in artificial word learning: The effects of converging sublexical cues," *Language and Speech*, vol. 61, no. 4, pp. 3–30, 2018.

[2] M. Shukla, M. Nespor, and J. Mehler, "An interaction between prosody and statistics in the segmentation of fluent speech," *Cognitive Psychology*, vol. 54, no. 1, pp. 1–32, 2007.

[3] M. D. Tyler and A. Cutler, "Cross-language differences in cue use for speech segmentation," *The Journal of the Acoustical Society of America*, vol. 126, no. 1, pp. 367–376, 2009.

[4] S. Kim, M. Broersma, and T. Cho, "The use of prosodic cues in learning new words in an unfamiliar language," *Studies in Second Language Acquisition*, vol. 34, no. 3, pp. 415–444, 2012.

[5] J. R. Saffran, E. L. Newport, and R. N, Aslin, "Word segmentation: The role of distributional cues," *Journal of Memory and Language*, vol. 35, no. 4, pp. 606–621, 1996.

[6] M. Ordin, L. Polyanskaya, I. Laka, and M. Nespor, "Cross-linguistic differences in the use of durational cues for the segmentation of a novel language," *Memory & Cognition*, vol 45, no. 5, pp. 863–876, 2017.

[7] B. Lindblom, "Final lengthening in speech and music," in E. Gårding, G. Bruce, and R. Bannert (Eds.), *Nordic Prosody*, pp. 85–100, 1978.

[8] J. Vaissière, "Language-independent prosodic features," in A. Cutler and D. R, Ladd (Eds.), *Prosody: Models and Measurements*, pp. 53–66, 1983.

[9] T. Cho, "Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English," *Journal of Phonetics*, vol. 32, no. 2, pp. 141–176, 2004.

[10] C. Fougeron and P. A. Keating, "Articulatory strengthening at edges of prosodic domains," *Journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3728–3740, 1997.

[11] P. Keating, T. Cho, C. Fougeron, and C. Hsu, "Domain-initial strengthening in four languages," in *Papers in Laboratory Phonology 6: Phonetic Interpretations*, pp. 145–163, 2003.

[12] T. Cho and P. A. Keating, "Articulatory and acoustic studies of domain-initial strengthening in Korean," *Journal of Phonetics*, vol. 29, no. 2, pp. 155–190, 2001.

[13] C.-S. Hsu and S.-A. Jun, "Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic?" *UCLA Working Papers in Phonetics*, vol. 96, pp. 69–89, 1998.

[14] T. Cho, J. M. McQueen, and E. A. Cox, "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English," *Journal of Phonetics*, vol. 35, no. 2, pp. 210–243.

[15] S. L. Mattys and H. Bortfeld, H. "Speech segmentation," in Gaskell, M. G. and Mirković, J. (Eds.), *Speech Perception and Spoken Word Recognition*, pp. 55–75, 2017.

[16] S. L. Mattys, L. White, and J. F. Melhorn, "Integration of multiple speech segmentation cues: a hierarchical framework," *Journal of Experimental Psychology: General*, vol. 134, no. 4, 477–500, 2005.

[17] J. Tsay, J. Charles-Luce, and Y.-S., Guo. "The syntax-phonology interface in Taiwanese: acoustic evidence," *In ICPhS 1999 – 14th International Congress of Phonetic Sciences, San Francisco, August 1-7, USA, Proceedings*, 1999, pp. 2407–2410.

[18] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," Version 6.0.37, retrieved from http://www.praat.org/, 3 February, 2018.

[19] M. Ordin and M. Nespor, "Transition probabilities and different levels of prominence in segmentation," *Language Learning*, vol. 63, no. 4, pp. 800–834, 2013.

[20] T. Fernandes, P. Ventura, and R. Kolinsky, "Statistical information and coarticulation as cues to word boundaries: A matter of signal quality," *Perception & Psychophysics*, vol. 69, no. 6, pp. 856–864, 2007.

[21] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[22] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, retrieved from https://www.R-project.org/, December 6, 2017.

[23] S.-C. Ou and Z.-C. Guo, "The language-specific use of fundamental frequency rise in segmentation of an artificial language: Evidence from listeners of Taiwanese Southern Min," *Language and Speech*, 2019.