The differential effects of vowel and onset consonant lengthening on speech segmentation: Evidence from Taiwanese Southern Min

Shu-chen Ou, and Zhe-chen Guo

Citation: The Journal of the Acoustical Society of America **149**, 1866 (2021); doi: 10.1121/10.0003751
View online: https://doi.org/10.1121/10.0003751
View Table of Contents: https://asa.scitation.org/toc/jas/149/3
Published by the Acoustical Society of America

ARTICLES YOU MAY BE INTERESTED IN

Effects of task and language nativeness on the Lombard effect and on its onset and offset timing
The Journal of the Acoustical Society of America **149**, 1855 (2021); https://doi.org/10.1121/10.0003772

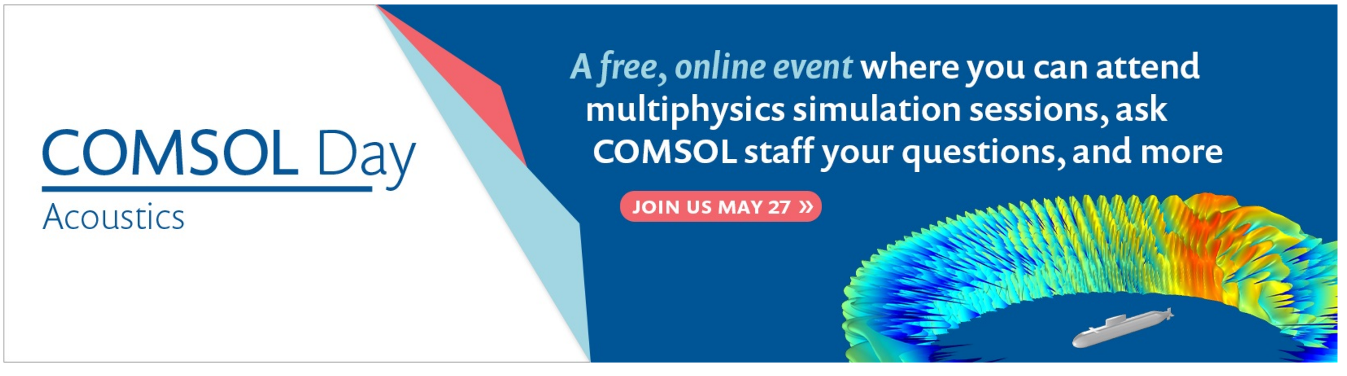
Prosodic focus marking in clefts and syntactically unmarked equivalents: Prosody–syntax trade-off or additive effects?
The Journal of the Acoustical Society of America **149**, 1390 (2021); https://doi.org/10.1121/10.0003594

Fundamental frequency range and other acoustic factors that might contribute to the clear-speech benefit
The Journal of the Acoustical Society of America **149**, 1685 (2021); https://doi.org/10.1121/10.0003564

Access to semantic cues does not lead to perceptual restoration of interrupted speech in cochlear-implant users
The Journal of the Acoustical Society of America **149**, 1488 (2021); https://doi.org/10.1121/10.0003573

An objective measure of quality for time-scale modification of audio
The Journal of the Acoustical Society of America **149**, 1843 (2021); https://doi.org/10.1121/10.0003753

Fluid-structure-acoustic interactions in an ex vivo porcine phonation model
The Journal of the Acoustical Society of America **149**, 1657 (2021); https://doi.org/10.1121/10.0003602



JASA ARTICLE

The differential effects of vowel and onset consonant lengthening on speech segmentation: Evidence from Taiwanese Southern Min

Shu-chen Ou[1,a)] and Zhe-chen Guo[2,b)]
[1]Department of Foreign Languages and Literature, National Sun Yat-sen University, Kaohsiung, 80424, Taiwan
[2]Department of Linguistics, University of Texas at Austin, Austin, Texas 78712, USA

ABSTRACT:
A review of previous speech segmentation research suggests the prediction that listeners of Taiwanese Southern Min (TSM), a lexical tone language, would exploit vowel lengthening and syllable-onset consonant lengthening to locate word ends and beginnings, respectively. Yet, correlations between segment duration and tone identity in tone languages along with some TSM-specific phonological phenomena may work against such use. Two artificial language learning experiments examined TSM listeners' use of the lengthening cues. The listeners heard the words of an artificial language (e.g., /ba.nu.me/) repeated continuously and identified them in a subsequent two-alternative forced-choice test. Experiment I revealed that their segmentation benefits from and only from word-initial onset lengthening or word-final vowel lengthening, supporting the prediction. Experiment II further demonstrated that these two cues in combination synergistically support segmentation at least when compared to word-initial onset lengthening alone, consistent with previous findings regarding complementary cues. These results furnish additional evidence that vowel and onset consonant lengthening affect segmentation in different ways, possibly reflecting a functional division between vowels and consonants that is supported by some prosody-computing mechanism. Additionally, vowel lengthening seems to affect segmentation to a greater extent than onset consonant lengthening. Possible explanations for this and further issues are discussed. © *2021 Acoustical Society of America*. https://doi.org/10.1121/10.0003751

(Received 10 November 2020; revised 18 February 2021; accepted 19 February 2021; published online 17 March 2021)
[Editor: Megha Sundara] Pages: 1866–1877

I. INTRODUCTION

Speech is continuous and segmenting it into separate words is indispensable for comprehension. Much work has shown that speech segmentation is guided by a wealth of prosodic cues, many of which are language-specific [e.g., Cutler *et al.* (1986), Jusczyk *et al.* (1993), Kim *et al.* (2012), McQueen *et al.* (2001), Tremblay *et al.* (2012), and Tyler and Cutler (2009)]. However, there is accumulating evidence that some of them are exploited cross-linguistically in a similar fashion (e.g., Endress and Hauser (2010) and Shukla *et al.* (2007)]. The current study aims to extend this line of research by examining two prosodic cues that are potentially useful for segmentation but in a functionally different way: vowel lengthening and syllable-onset consonant lengthening. To this end, we investigated listeners of Taiwanese Southern Min (TSM), a lexical tone language that has been understudied in the speech segmentation literature.

It is well-known that across languages, vowels in syllable nucleus or rimes tend to be lengthened word- or phrase-finally (Gussenhoven and Rietveld, 1992; Klatt, 1975; Lindblom, 1978; Vaissière, 1983). Over the past decades, studies have suggested that this cross-linguistic phenomenon inclines listeners to interpret longer vowels as cueing word-finality and exploit them to support segmentation. Saffran *et al.* (1996) presented English listeners with speech streams where CVCVCV nonsense words (e.g., *mupana*) were repeated continuously without pauses and found that the words were segmented better when their final vowels were lengthened than when no vowels were lengthened. Using the same experimental technique, Tyler and Cutler (2009) replicated this finding for English listeners as well as for Dutch and French listeners. Relatedly, Hay and Diehl (2007) showed that when hearing continuous speech sequences alternating in vowel duration, both English and French listeners preferred iambic grouping, parsing the sequences into units ending in longer vowels. These findings are reinforced by many demonstrations that for listeners of several languages (e.g., English, Basque, German, etc.), lengthening word-initial vowels does not facilitate segmentation and can even disrupt it (Ordin *et al.*, 2017; Saffran *et al.*, 1996; Tyler and Cutler, 2009). Longer vowels seem to generate a percept of word-finality cross-linguistically, making themselves a universal segmentation cue.

Nevertheless, this claim is undermined by recent findings showing that the use of vowel lengthening is constrained by language-specific phonology. In Ordin *et al.* (2017), Italian listeners' segmentation of continuous

a)Electronic mail: sherryou@mail.nsysu.edu.tw, ORCID: 0000-0002-9019-2389.
b)ORCID: 0000-0003-0870-8321.

 © 2021 Acoustical Society of America

nonsense words did not benefit from word-final lengthening. Instead, it was when the word-medial vowels were lengthened that their segmentation improved. Ordin *et al.* attribute this to lexical stress in Italian, which typically falls on the penultimate syllable (Krämer, 2009) and has longer vowel duration as its primary perceptual correlate (Bertinetto, 1980). Replicating the finding of Ordin *et al.*, White *et al.* (2020) demonstrated that Hungarian listeners did not benefit from final vowel lengthening either. They contend that this is because Hungarian has phonemic vowel length, inclining its listeners to interpret vowel duration differences as cues to phonemic identity rather than as segmentation cues. For listeners who exploit final vowel lengthening, vowel duration in their language does not have the same functions as in Italian or Hungarian. English, for example, does not use vowel duration phonemically. Besides, although English lexical stress is predominantly word-initial (Cutler and Carter, 1987), it is signaled by a combination of correlates including fundamental frequency (F0), vowel duration, amplitude, and vowel quality [e.g., Beckman (1986), Fry (1958), Gay (1978), Lieberman (1960), and Sluijter and van Heuven (1996)] and there is no consensus that vowel duration is the primary stress cue. Therefore, a weaker and more tenable version of the claim about vowel lengthening may be that this cue is generally perceived as signaling word-finality, but only if the listener's native language does not have phonemic vowel length or lexical stress that frequently occurs non-finally with vowel duration being the main stress correlate.

Recent work also reveals that segmentation can as well be supported by lengthened consonants in syllable onset, which, contrary to lengthened vowels, signal where a word begins. Onset consonants in the initial position of a higher prosodic domain tend to undergo stronger articulatory strengthening and have a longer duration than those in the initial position of a lower prosodic domain (Cho, 2004; Cho and Keating, 2001; Fougeron and Keating, 1997; Hsu and Jun, 1998). These phenomena are attested in a number of languages, including English, Korean, French, and TSM (Keating *et al.*, 2004). They are exploited by listeners to facilitate the search for word beginnings, as shown in several experiments using naturally produced speech [e.g., Cho *et al.* (2007), Gow and Gordon (1995), Quené (1993), and Tagliapietra and McQueen (2010)]. Yet, such speech material contains multiple boundary cues and it cannot be unequivocally determined whether consonant duration by itself contributes to segmentation. The first direct evidence for the use of this cue comes from White *et al.* (2014, 2015). Similar to Saffran *et al.* (1996), they exposed English listeners to speech streams containing continuous repetitions of CVCVCV nonsense words. They found that lengthening the word-initial onset consonants boosted segmentation relative to a no-lengthening condition. Lengthening the second, medial onset consonants, however, led to no improvement. These suggest that a lengthened onset consonant *per se* is capable of supporting segmentation, but in a way different than that of vowel lengthening as it is perceived as word-initial.

Since lengthening due to domain-initial strengthening is observed cross-linguistically, it may be asked whether such use of onset consonants is universal. The question has been explored by White *et al.* (2020), who conducted similar experiments as in White *et al.* (2014, 2015) but included listeners of two languages in addition to English: Italian and Hungarian. They found that lengthening word-initial onset consonants enhanced segmentation performance for all listener groups. This is noteworthy considering that Italian has phonemic consonant length, which may militate against interpreting consonant duration as a segmentation cue. Therefore, unlike longer vowels, the use of which is subject to language-specific phonology, longer onset consonants seem to be universally exploited to locate word beginnings, regardless of the linguistic functions of consonant duration in the native language.

Taken together, the findings on the role of duration in speech segmentation suggest the following generalizations. For vowels, lengthening is interpreted as a word-finality cue by default except for (1) listeners of languages with phonemic vowel length (e.g., Hungarian) or (2) those in which lexical stress predominantly occurs in a non-final position with duration being the main stress correlate (e.g., Italian). In contrast, for onset consonants, lengthening is invariably interpreted as a word-beginning cue, overriding any language-specific phonological forces against such an interpretation. Thus, while vowel lengthening and onset consonant lengthening are both potentially useful segmentation cues, they are mirror images of each other in terms of their function and localization. Since TSM has no word stress and phonemic length distinctions, we expect from these generalizations that TSM listeners would use vowel lengthening and onset consonant lengthening to discover word ends and beginnings, respectively. Yet, it may be sensible to step back and rethink this prediction since, for reasons below, further exploration is required of the extent to which the functions of the two lengthening cues are universal or language-specifically constrained—an issue that also makes TSM an interesting focus of investigation.

First, TSM is a lexical tone language whereas the above generalizations are based on evidence from non-tonal languages. Segmentation studies that included tone-language listeners mostly examined how they use tonal cues [e.g., Caldwell-Harris *et al.* (2015) and Gómez *et al.* (2018)] and nearly none focused on durational cues. Differences in lexical prosody typology might lead tone-language listeners to exploit duration differently from listeners of well-studied languages like English. Unlike lexical stress, which may appear frequently or invariably in some position, lexical tones are distinctive pitch patterns over a syllable that can occur anywhere except under some phonological constraints. The unpredictability of a tone's occurrence may heighten the need to identify tones and encourage tone-language listeners to exploit segment duration for this purpose. This is not impossible since across tone languages, tones are related to vowel duration (Gandour, 1977; Howie, 1976; Kong, 1987; Yu, 2010) and to onset consonant duration (Maddieson, 1978; Zee, 1977). For example, vowel duration is inversely related to F0 and high-toned vowels tend to be shorter than low-toned ones. Such correlations might promote the use of variations in segment duration as tone identity cues to resolve lexical ambiguity, making them function like phonemic length contrasts. Nevertheless, this does not preclude that the variations would not be exploited for segmentation. Furthermore, it is well-established that F0 is the primary cue for tone perception [e.g., Fu and Zeng (2000), Kuo *et al.* (2008), Lin and Repp (1989), and Lin (1988)] and duration is perhaps a trivial one. All these possibilities highlight the need to investigate the use of vowel and onset consonant lengthening by tone-language listeners.

Second, with regard to onset consonant lengthening, it has to be noted that evidence for its universal interpretation as a word beginning cue is nascent and comes from a limited sample of languages. To date, only White *et al.* (2020) have examined onset consonant lengthening cross-linguistically and the languages they tested were English, Hungarian, and Italian, all of which have fixed or variable stress. It is unclear whether onset consonant lengthening functions as a word-beginning cue also for listeners of typologically distant languages such as tone languages, especially those with unique phonological properties that might work against this function. As we have shown, several past studies [e.g., Hay and Diehl (2007) and Tyler and Cutler (2009)] agreed that lengthened vowels are universally a cue to word-finality, but this view is challenged as emerging evidence points to an effect of language-specific phonology [e.g., Ordin *et al.* (2017) and White *et al.* (2020)]. The line of investigation on onset consonant lengthening might follow a similar path of development and uncover counterevidence against the cue's universality as researchers test more languages and consider more detailed aspects of their phonologies. Such a possibility calls for an investigation of an understudied language like TSM as it has phonological processes that appear to militate against our prediction. For example, when followed by an onsetless particle like –*a*, syllable-final consonants undergo gemination (and also voicing if they are voiceless, e.g., /ap a/ → [ab ba] "box"), resulting in word-medial long consonants that never appear word-initially (Ang, 1993; Chiang, 1992; Chung, 1996; Lin, 2004; Wang and Liu, 2010; Yang, 1991). Also, closed syllables ending in /ʔ/ in a non-final position change to open syllables with a phonetically longer vowel *via* /ʔ/-deletion (e.g., /baʔ kuã/ → [ba kuã] "jerky"), producing an effect resembling "non-final" vowel lengthening (Ang, 1993; Cheng, 1973; Chung, 1996; Xu, 1998). Currently, no research has shown that these phonological processes modulate TSM listeners' segmentation behavior. Yet, testing a language with such processes could add to the understanding of the extent to which the functionally different lengthening cues—especially onset consonant lengthening—are universal or language-specific.

Finally, the present work was also methodologically motivated, using an experimental design which may present an improvement over previous ones and provide more rigorous or conclusive results. As mentioned, White *et al.* (2014, 2015) found that English listeners benefited from lengthening of initial but not medial onset consonants. Yet, it could be argued that they showed this benefit not because longer onsets were interpreted as word-initial, but because lengthening any segment—whether it be a vowel or consonant—in a boundary-adjacent (i.e., word-initial or word-final) syllable directed listeners' attention to that syllable and hence to word edges. A similar explanation has been discussed by Toro *et al.* (2009), whose English and Spanish listeners showed worse segmentation performance with word-medial pitch rise (compared to word-initial and word-final pitch rise) possibly because the word-medial pitch cue drew their attention away from word boundaries. Thus, it would be necessary to consider vowel and onset consonant lengthening in both boundary-adjacent positions. With this in mind, we conducted experiment I to test our prediction for TSM listeners: namely, would they interpret vowel lengthening and onset consonant lengthening as cues to word ends and beginnings, respectively? If these cues are exploited as predicted, it would also be of interest to further ask whether their effects are synergistic—a question to which the answer has remained inconclusive due to methodological issues in previous research. White *et al.* (2014, 2015) demonstrated that lengthening both word-initial onset consonants and word-final vowels improved segmentation numerically but not significantly compared to a condition with only initial onset consonant lengthening. However, they did not include a condition with only word-final vowel lengthening for comparison. Experiment II therefore aimed to obtain more conclusive results by comparing both the two lengthening cues alone against their combination. We explored all these questions using the artificial language (AL) learning paradigm, described in detail below.

II. EXPERIMENT I

A. Method

1. Participants

One hundred and fifty sequential bilingual speakers of TSM and Taiwan Mandarin (90 females and 60 males) participated in the study, all recruited from a university in Southern Taiwan. Their mean age in years was 20.6 (standard deviation: 1.93; range: 18–33). They had learned English as a foreign language and compulsory subject in school for at least six years. None reported history of hearing disorders. All signed a written consent form before the experiment and received a small honorarium after they finished it. They were randomly allocated to the five conditions (see Sec. II A 2) with an equal number of participants ($N = 30$) in each condition.

2. Experimental design

As a typical AL learning experiment, experiment I consisted of a learning phase followed by a test phase. First, listeners learned the "words" of an AL by listening to long utterances comprised of continuous repetitions of the words.

Next, they recognized the words in a two-alternative forced-choice test. Recognition accuracy was used as a measure of segmentation performance during the learning.

In experiment I, listeners learned the AL under either of the following five conditions. In the "no lengthening (NL)" condition, which served as the baseline for comparison with the other conditions, no vowels or onset consonants in the AL words were lengthened and segmentation could only be achieved by tracking transitional probabilities between syllables. In the "final vowel lengthening (FVL)" and "initial onset consonant lengthening (IOL)" conditions, the vowels in word-final syllables and the syllable-onset consonants in word-initial syllables, respectively, were lengthened. If TSM listeners exploited lengthening cues as the generalizations predicted, their segmentation performance in these two conditions would be better than in the NL one. Finally, in the "initial vowel lengthening (IVL)" and "final onset consonant lengthening (FOL)" conditions, the vowels in word-initial syllables and onset consonants in word-final syllables, respectively, were lengthened. These represented situations in which the lengthening cues were misplaced in view of the generalizations and their effects on segmentation should be neutral, or even harmful, relative to the NL condition. Including these two conditions allowed us to assess whether there could be an alternative explanation for improvement in the FVL and IOL conditions (if any): that is, as discussed, lengthening any segment in boundary-adjacent syllables enhances listeners' segmentation as it directs their attention to word edges and it does not matter whether that segment is a vowel or a consonant. To establish that vowel and onset consonant lengthening are indeed interpreted as cueing word-finality and word-initiality, respectively, one would need to demonstrate that TSM listeners benefit only from the FVL and IOL conditions.

3. Stimuli

As in several studies (e.g., Caldwell-Harris *et al.*, 2015; Kim *et al.*, 2012; Saffran *et al.*, 1996), our AL comprised of six trisyllabic CVCVCV words: /ba.nu.me/, /bi.mɔ.na/, /ge.ni.gɔ/, /mi.ma.bu/, /ne.bɔ.gi/, and /nɔ.ga.mu/ (the dots denote syllable boundaries). They were constructed from five vowels (/a, i, e, u, ɔ/) and four consonants (/b, g, m, n/). Note that as domain-initial articulatory strengthening has been shown to affect stops and fricatives to different degrees (Fougeron, 2001), we used only (oral and nasal) stops to minimize the effect of segment type. For the same purpose, we did not include approximants like /l/, which could also make it difficult to pinpoint consonant-vowel boundaries. Voiceless stops such as /p/ or /pʰ/ were not used since the aspiration period after the release of a voiceless stop may be perceived as part of the following vowel (Gussenhoven and Zhou, 2013; Van Wijk, 1999). A male native speaker of TSM and Taiwan Mandarin with phonetic training provided the stimuli in a sound-attenuated booth in front of a desktop computer. He was instructed to insert each of the component CV syllables of the AL words into a TSM carrier sentence (/gua kɔŋ ____ tsit e dʑi/ "I said the word ____.") and read it aloud. He was also told to produce a flat monotone at about the same pitch height as the mid-level tone in TSM for the target syllables. The recordings were made with a Zoom H4n Handy Recorder and digitized at a sampling rate of 44.1 kHz in WAV format using Audacity.

The CV syllables were isolated from the carrier sentence and manipulated with PRAAT (Boersma and Weenink, 2018). They were equated for root-mean-squared amplitude and their F0 contours were flattened at 119 Hz, the mean F0 of the syllables before manipulation. The durations of all segments were normalized to 150 ms. Thus, each CV syllable had a duration of 300 ms, which was within the range of the lengths of (non-lengthened) syllables used in previous research [e.g., 252 ms in Kim *et al.* (2012) and 450 ms in Gómez *et al.* (2018)]. These syllables[1] were concatenated to form the base version of the six AL words, which were used in the NL condition. The words for the other four conditions were created from the base version by lengthening the critical segments according to their conditions by a factor of 1.5 (i.e., from 150 to 225 ms). This magnitude of lengthening was comparable to those used in studies on vowel lengthening (e.g., 1.52 times in Tyler and Cutler, 2009) and onset consonant lengthening [e.g., 1.54 times in White *et al.* (2020), (2015)]. Figure 1 gives a comparison of the conditions.

For each condition, six speech streams were generated for use in the learning phase by concatenating the six words without inserting pauses. In each steam, the words were repeated 20 times each in a pseudorandom order such that the same word did not appear in a row. The words appeared in different orders across the six speech streams, giving rise to six different orders. Yet, all the orders were used in each experimental condition. Thus, across conditions, the speech streams differed only in whether a segment in a word was lengthened and which one was lengthened. The total length of the streams was about 11–12 min. The first and last five seconds of each stream were faded in and out, respectively, with logarithmic ramps to prevent listeners from hearing the very first and last syllables and using them to easily locate word beginnings and ends.

Listeners completed a two-alternative forced-choice task in the test phase and heard two stimuli in each trial: one was one of the six words of the AL and the other a "partword." A partword was a trisyllabic CVCVCV sequence formed by concatenating substrings of two AL words (e.g., /ma.bu.ge/ = /mi.ma.bu/ + /ge.ni.gɔ/). There were six different partwords: /ma.bu.ge/, /bɔ.gi.nɔ/, /ni.gɔ.mi/, /gi.ba.nu/, /na.ne.bɔ/, and /me.bi.mɔ/. Computed using the formula of Saffran *et al.* (1996), the average transitional probability between two adjacent syllables ranged between 0.59 and 0.61 for partwords whereas it was always one for AL words. No segments in the test-phase words and partwords were lengthened; as a result, the test was the same for all conditions. There were totally 36 trials, comprised of all possible pairings of the six words with the six partwords and presented randomly. There was an equal

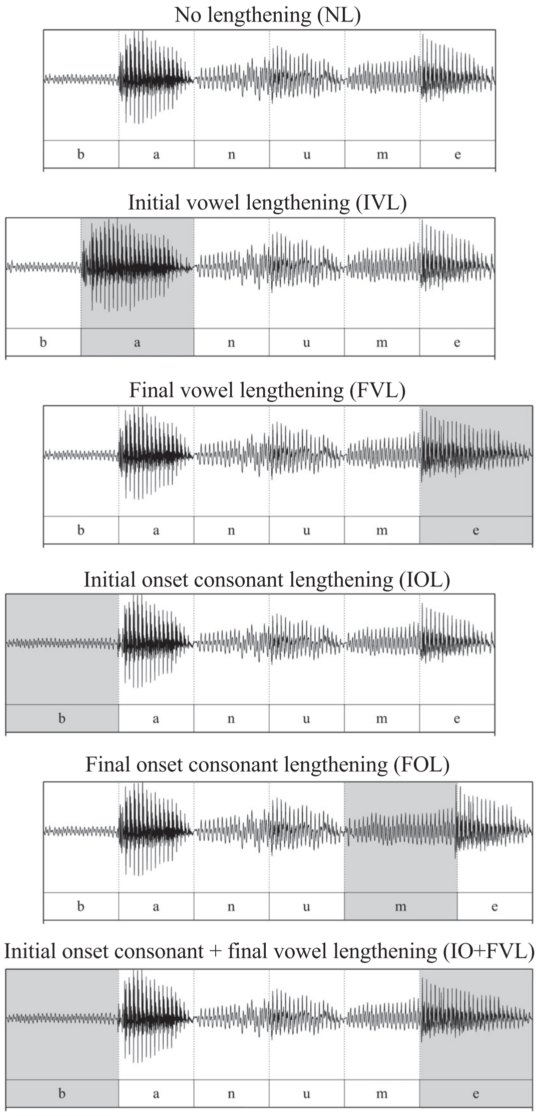


FIG. 1. Waveforms of an example word /ba.nu.me/ under the NL, IVL, FVL, IOL, and FOL conditions (experiment I) and the IO+FVL condition (experiment II). Lengthened segments are shaded.

number of trials for the two orders in which the two stimuli in a trial could appear (i.e., word–partword or partword–word). E-PRIME 2.0 (Psychology Software Tools, 2012) was used to present stimuli and record responses.

4. Procedure

Participants were tested individually in front of a desktop computer in a sound-proof booth and told that their task was to learn an AL by just listening to six sound files (i.e., the six speech streams) of that language. They were warned that there were no pauses between the words in the sound files; they were not provided any information about the AL, such as the number of its words. All they were instructed to do was to try to keep their attention on what they heard. They could take a short break every time they finished listening to a sound file. After the learning phase, they immediately moved on to the test. They were told that they would hear two stimuli separated by a half-second silence in each trial and have to press button "1" on a response device if they thought the first stimulus sounded more like a word of the AL and button "2" if they thought the second stimulus sounded more like a word of the AL. A response had to be made within five seconds after the second stimulus was presented, and they might guess if they were not sure. Before the test proper, three trials using syllable sequences irrelevant to the AL were presented to help familiarize participants with the test procedure.

B. Results and discussion

Listeners' responses in the test phase were analyzed. Time-outs, or trials in which no responses were made within the allotted five seconds, accounted for 0.5% of all data of experiment I and were excluded. In the remaining data, a response was coded as correct when the word of AL was chosen and incorrect when the partword was chosen. Shown in Fig. 2 are individual listeners' accuracy rates and boxplots for all the conditions.

We analyzed listeners' responses, which were either correct or incorrect, with a Bayesian mixed-effects logistic regression model implemented by using the brms package (Bürkner, 2017) of the R programming language (R Core Team, 2017). Bayesian analysis was used here as it has a number of advantages over traditional frequentist approaches relying on null-hypothesis significance testing [e.g., Kruschke *et al.* (2012), Vasishth *et al.* (2018), and Wagenmakers *et al.* (2018)]. An important one is that it estimates the probabilities of parameters given the data and model, which can be conceptually more meaningful and interpretable than frequentist *p*-values. In most scientific investigations, researchers hypothesize about some quantities at the population level (i.e., the parameters) and collect a sample of data to draw inferences about them. Usually, they are most interested in the parameters rather than the data themselves. The Bayesian methods meet these interests as their outputs are probability distributions of the parameters given the data and model, allowing researchers to infer the probability at which a parameter would fall within a certain range of values. In contrast, the *p*-value is a single point

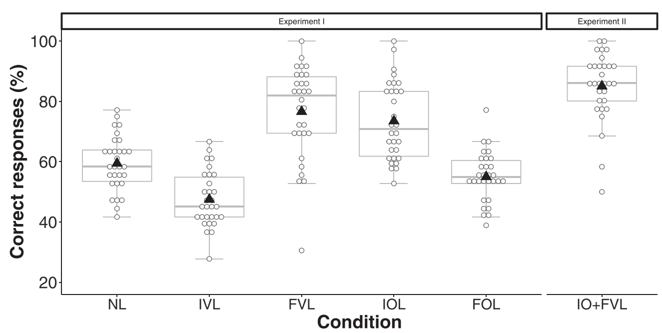


FIG. 2. Test-phase response accuracy rates of individual participants (circles), along with boxplots and means (triangles) for the NL, IVL, FVL, IOL, and FOL conditions (experiment I) and the IO+FVL condition (experiment II). The horizontal line in each box represents the median.

https://doi.org/10.1121/10.0003751

estimate that quantifies the probability of the *data* given that the null hypothesis (e.g., the parameter value is zero) is true. Bayesian inference also allows one to integrate prior knowledge and evidence in estimating parameter probabilities. See, among others, Kruschke (2014) and McElreath (2018) for more about the Bayesian approach.

The fixed-effect predictor of primary interest was condition (dummy-coded; baseline: NL condition). Following Ou and Guo (2019), we also included the ordinal rank of each trail (Trial) and log-transformed response time (LogRT) as fixed effects to partial out potential training or fatigue effects and to capture the relationship between response time and accuracy, respectively. These two variables were centered and scaled and all fixed-effect predictors were entered as main effects without interactions. The random effects included random intercepts for participants, words, and partwords as well as a by-participant random slope for words. Since no similar previous research has been conducted on TSM listeners, weakly informative regularizing priors (i.e., normal distribution with $\mu = 0$ and $\sigma = 10$) were applied to all fixed-effect parameters. The same priors were used on the standard deviation parameters of the random intercepts except they were left-truncated at zero (as standard deviations cannot be negative). Following Vasishth *et al.* (2018), we also assigned a weakly informative LKJ(2) prior (Lewandowski *et al.*, 2009) to the correlation parameter between the random intercept and random slope. Four Monte Carlo Markov Chains (MCMCs), each with 2000 iterations and the initial 1000 iterations as warm-ups, were run to sample the joint posterior distribution. The R-hat value (Gelman and Hill, 2006), or the ratio of between-chain variance over within-chain variance, reached one for all parameters, indicating that the chains had converged.

Figure 3 shows the marginal posterior distributions of the fixed-effect parameters. $\beta_{\text{Intercept}}$ represented the posterior probability distribution of log-odds ratio for the baseline condition, which was NL here, at the mean values of LogRT and Trial. A common decision criterion for determining whether a predictor has an effect (i.e., whether its parameter value is non-zero) in Bayesian inference is to check whether the 95% highest density interval (HDI), which includes 95% of the posterior distribution, excludes zero or not. The entire posterior distribution of β_{LogRT} (mean: –0.24, 95% HDI: [–0.31, –0.17]) fell in the negative region and its 95% HDI excluded zero, indicating that slower responses were more likely to be incorrect. As Ou and Guo (2019) suggest, this inverse relationship between response time and accuracy might reflect response confidence: as the two stimuli in a trial were separated by a half-second silence, listeners could already be ready to respond if they were sure whether the first stimulus was a word or a partword. The posterior on β_{Trial} (mean: –0.01, 95% HDI: [–0.07, 0.05]) provided no evidence for training or fatigue effects. Most crucial were the condition parameters. As can be seen in the posteriors on $\beta_{\text{condition (FVL)}}$ (mean: 0.90, 95% HDI: [0.46, 1.42]) and $\beta_{\text{condition (IOL)}}$ (mean: 0.68, 95% HDI: [0.30, 1.06]), almost all values were positive and zero was implausible, suggesting

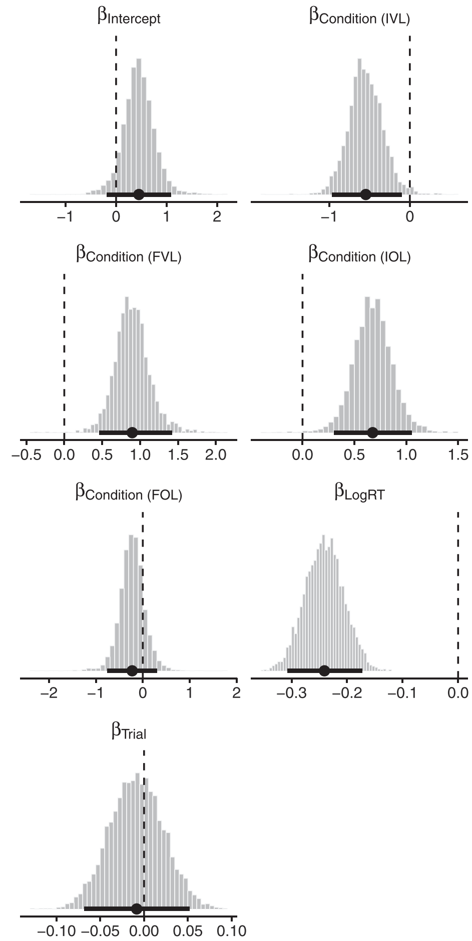


FIG. 3. Marginal posterior distributions of the fixed-effect parameters in experiment I. The bars below the distributions represent the 95% HDIs, with the dots indicating the mean values. Dashed vertical lines mark zero.

that segmentation performance in the FVL and IOL conditions was superior to that in the NL condition. In contrast, only a small fraction of the values in the posterior distributions of $\beta_{\text{condition (IVL)}}$ (mean: –0.55, 95% HDI: [–0.97, –0.10]) and $\beta_{\text{condition (FOL)}}$ (mean: –0.23, 95% HD: [–0.75, 0.31]) were positive, providing no evidence that segmentation was better in the IVL and FOL conditions than in the NL one. In fact, segmentation even worsened in the IVL condition.

Experiment I examined TSM listeners' use of segment lengthening in different positions in an AL learning task. The results were consistent with what we predicted based on previous findings: TSM listeners would interpret vowel lengthening as word-final and onset consonant lengthening as word-initial. This is not only directly reflected in the improvement in segmentation they showed in the IOL and FVL conditions, but also indirectly supported by the lack of evidence for such improvement in the FOL and IVL conditions. Consequently, the patterns of results cannot be explained as merely the result of using any lengthened segments in boundary-adjacent syllables to facilitate segmentation. They are better accounted for if it is assumed that segment lengthening is interpreted in the way predicted.

https://doi.org/10.1121/10.0003751

To better understand (word-final) vowel lengthening and (word-initial) onset lengthening, both of which are proved useful for TSM listeners, we turn to the second question of this study. That is, would these two cues in combination have a synergistic effect, boosting the listeners' segmentation to a greater extent than either of them in isolation? As discussed above, previous studies [e.g., White *et al.* (2015)] do not seem to provide a firm conclusion to this question as not all the relevant conditions were included for comparison. Nevertheless, there are findings supporting the idea that the combination of two such lengthening cues could synergistically support segmentation. In an AL learning experiment with French listeners, Bagou and Frauenfelder (2018) observed that two cues in combination facilitated segmentation compared to either cue alone when they were "complementary" (i.e., in different positions, such as word-initial phonotactic cue and word-final accent). In contrast, no such facilitation was found when the cues were "redundant" (i.e., in the same position, such as word-final lengthening and word-final pitch rise). Word-final vowel lengthening and word-initial onset lengthening were complementary and it could be predicted based on Bagou and Frauenfelder's findings that combining these cues would have a synergistic effect. We conducted another AL learning experiment to test this prediction in experiment II. Supporting evidence for it would be better segmentation performance in a condition supplying the two lengthening cues simultaneously than in both FVL and IOL conditions or at least one of them.

III. EXPERIMENT II

A. Method

Thirty new sequential bilingual speakers of TSM and Taiwan Mandarin (14 females and 16 males) who were from the same population as in experiment I but did not participate in it before were recruited. Their mean age in years was 20.2 (standard deviation: 1.73; range: 18–24). The stimuli, procedure, and experimental design were as in experiment I except that the 30 new participants learned the AL under only one condition: "initial onset consonant plus final vowel lengthening (IO+FVL)." The initial onset consonants and final vowels in the AL words of this condition were simultaneously lengthened (by 1.5 times), as shown in Fig. 1.

B. Results and discussion

Again, responses in the test phase were analyzed and time-outs (0.6% of all data of experiment II) were discarded. In order to determine whether the combined cue had a synergistic effect, the data from the IO+FVL condition of experiment II and those from the IOL and FVL conditions of experiment I were analyzed together using a Bayesian mixed-effects logistic regression model (see Fig. 2 for the accuracy rates of individual participants and the means of the three conditions). The fixed- and random-effects and the priors on their corresponding parameters were as in experiment I, except the condition predictor had only three levels and the IO+FVL condition served as the baseline level. The R-hat statistic reached one for all parameters, indicating chain convergence.

Figure 4 gives the marginal posterior distributions of the fixed-effect parameters. The posterior on $\beta_{\text{Intercept}}$ again represented the probability distribution of log-odds ratio for the baseline condition, which was IO+FVL in experiment II (when LogRT and Trial were at their mean values). As in the analysis of experiment I, the posterior distribution of β_{LogRT} (mean: –0.54, 95% HDI: [–0.67, –0.43]) contained only negative values as slower responses were more likely to be incorrect. This might be due to response confidence, as explained. This time the posterior on β_{Trial} (mean: –0.15, 95% HDI: [−0.25, –0.06]) contained almost only negative values with its 95% HDI excluding zero, indicating that trials later in the test were less likely to be correct, possibly due to fatigue. The most important parameters were again those related to condition. Note that because the IO+FVL condition was designated as the baseline of this factor, a negative value for the condition parameters meant that the likelihood of correct responses increased in the IO+FVL condition. The posterior on $\beta_{\text{condition (IOL)}}$ (mean: –0.71, 95% HDI: [–1.33, –0.10]) revealed that response accuracy was indeed higher in the IO+FVL condition than in the IOL one. Yet, based on the 95% HDI criterion, the posterior distribution of $\beta_{\text{condition (FVL)}}$ (mean: –0.49, 95% HDI: [–1.14, 0.18]) provided no evidence that the IO+FVL condition had higher accuracy compared to the FVL one. However, given our model and data, there was still a rather high posterior probability (i.e., 94%) that the value of $\beta_{\text{condition (FVL)}}$ was

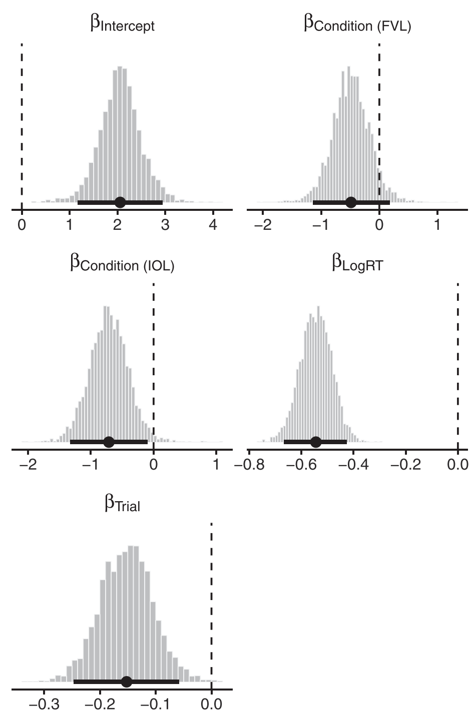


FIG. 4. Marginal posterior distributions of the fixed-effect parameters in experiment II. The bars below the distributions represent the 95% HDIs, with the dots indicating the mean values. Dashed vertical lines mark zero.

negative. These results suggested that the combination of word-initial onset consonant lengthening and word-final vowel lengthening synergistically supports segmentation at least when compared with the former cue by itself, providing partial but clear support for the prediction based on cue complementarity. In Sec. IV, we present a general discussion of the findings from the two experiments.

IV. GENERAL DISCUSSION

The current study examines the use of segment lengthening in speech segmentation by TSM listeners with two AL learning experiments. A review of previous work points to the conclusion that lengthened vowels are by default interpreted as word-final except when the listener's native language uses vowel duration phonemically or has predominantly non-final word stress cued primarily by longer duration. In contrast, lengthened onset consonants seem to be universally interpreted as word-initial, potentially overriding any language-specific phonological factors that could militate against this interpretation. These generalizations led us to predict that TSM listeners would exploit vowel lengthening to locate word ends and onset consonant lengthening to locate word beginnings. The prediction was supported by experiment I: TSM listeners showed improved segmentation in the FVL and IOL conditions relative to the NL one. On the other hand, their segmentation did not improve or even worsened in the IVL and FOL conditions, ruling out the possibility that they were simply using any lengthened segments in boundary-adjacent syllables to discover word edges. Experiment II further investigated whether word-initial onset lengthening and word-final vowel lengthening, which were complementary cues, would synergistically support TSM listeners' segmentation when they were simultaneously present. There was evidence for this synergistic effect: the IO+FVL condition enhanced segmentation when compared to the IOL one. Besides, given our model and data, there was a high probability that IO+FVL condition produced a benefit relative to the FVL one.

Through experimentation with listeners of TSM, a less studied tone language, the present research furnishes additional support for the generalizations based on previous findings from non-tonal languages and potentially also for a functional division between consonants and vowels. Although there are correlations between tones and segment durations (e.g., Gandour, 1977; Maddieson, 1978), they do not seem to prevent TSM listeners from using onset and vowel lengthening for segmentation. Moreover, as predicted, the listeners interpret longer vowels as word-final and longer onset consonants as word-initial despite the possibility that some productive TSM-specific phonological processes (e.g., word-medial gemination and non-final vowel lengthening due to dropping of glottal coda) could militate against such interpretations. Our findings reinforce the idea that longer onset consonants and longer vowels both potentially serve a demarcative function, but in a different way. Such a view has been expressed by White *et al.* (2015) in their study with English listeners, who behaved just like our TSM listeners. It also finds support from numerous demonstrations that consonants and vowels play distinct roles at various levels and stages of speech processing [the CV hypothesis: Nespor *et al.* (2003)]. For example, at the level of phoneme recognition, consonants are perceived more categorically than vowels (Pisoni, 1973; Stevens *et al.*, 1969). During word recognition, information carried by consonants constrains lexical access more than that carried by vowels (Cutler *et al.*, 2000). Together with those from previous studies, our findings from TSM listeners may be interpreted as reflecting a similar division of labor between consonants and vowels during the speech segmentation process. Yet, it has to be noted that the demarcative functions of segment lengthening—particularly that of vowel lengthening—can possibly be suppressed due to pressure from other functions required by language-specific phonology, such as the need to keep phonemic vowel length contrasts salient.

This functional division view may prompt some further questions: what could possibly be the underlying mechanism that supports the different functions of vowels and onset consonants? How can it explain the performance patterns across the conditions of the current experiments? Here, it is suggested that the mechanism may be the same as or akin to the Prosody Analyzer proposed by Cho *et al.* (2007) to account for English listeners' use of domain-initial strengthening effects in segmenting meaningful speech. The Prosody Analyzer utilizes the suprasegmental information in the signal and computes the prosodic structure of the current utterance. Given our findings, it is possible that this prosody-computing mechanism analyzes a longer onset consonant as signaling the left edge of some prosodic constituent and a longer vowel as signaling the right edge. TSM listeners' improvement in the IOL and FVL conditions of experiment I could then be attributed to an agreement between prosodic analysis and statistical regularities. Recall that the basic segmentation cue in an AL learning task was transitional probability. Specifically, sequences of word-internal syllables tend to have a higher transitional probability than across-word sequences. In the IOL and FVL conditions, the constituent edges in the prosodic representations coincided with "dips" in transitional probability, leading to a stronger percept of boundary. The lack of improvement in the FOL and IVL conditions arises from a conflict between the prosodic and statistical sources of information. In the FOL condition, for example, lengthening the onsets of the third syllables of the AL words misled the Prosody Analyzer to place a left-edge boundary before the onsets, interfering with computations of statistical regularities. This resulted in no facilitation or even inhibition (as in the IVL condition).

Moreover, the findings from experiment II seem to suggest that two cues in different positions might be analyzed by the Prosody Analyzer separately even if they signal exactly the same boundary. In the learning-phase speech streams of the IO+FVL condition, the lengthened vowel of an AL word was always followed by the lengthened onset consonant of the next word and hence every word boundary in the streams was always doubly cued by the lengthened segments. However, this does not seem to constitute a case of cue redundancy: consistent with the scenario of cue complementarity in Bagou and Frauenfelder (2018), word-initial onset lengthening and word-final vowel lengthening in combination synergistically enhance segmentation compared to the former cue by itself and there is a high probability that this is also the case when compared to the latter cue by itself. This is possibly because the two lengthening cues correspond to distinct entities in the representations generated by the Prosody Analyzer. Specifically, when detecting a lengthened onset, the analyzer assumes a left edge independently from the right edge signaled by the preceding lengthened vowel. The IO+FVL condition then yields the strongest percept of boundary as three different sources of information converge to cue boundaries: dip in transitional probability, prosodic left edge, and prosodic right edge. However, further work may be needed to more conclusively establish whether the combined lengthening cues do have a synergistic effect. As mentioned, there seems to be no statistically robust evidence for an improvement in performance in the IO+FVL condition relative to the FVL one. Yet, given that the mean accuracy in our IO+FVL condition was 85%, which is rather high considering the range of mean values (50%–80%) typically reported in AL learning studies [e.g., Caldwell-Harris *et al.* (2015), Ordin *et al.* (2017), Tyler and Cutler (2009), and White *et al.* (2015)], this might partly be due to close-to-ceiling performance.

In addition to the differential effects of vowel and onset lengthening, there is a less apparent finding which merits discussion: vowel lengthening affects TSM listeners' segmentation to a greater extent than onset lengthening. While word-final vowel lengthening and word-initial onset lengthening are both useful, as experiment I revealed, the former benefits segmentation slightly more than the latter does. This can be seen not only by visually inspecting the mean accuracy of the FVL (77%) and IOL (73%) conditions in Fig. 2, but also by computing the posterior distribution of $\beta_{\text{condition (FVL)}} - \beta_{\text{condition (IOL)}}$, which represents the difference between the two conditions in the amount of improvement they produced relative to NL, from the output of experiment I's analysis. The posterior probability of the difference being positive was 82%, providing some evidence that vowel lengthening produces a greater gain than onset lengthening in positions favorable to segmentation. Consistently, we had statistically sufficient evidence for a synergistic effect of the combined lengthening cues only when the IO+FVL condition (mean: 85%) was compared against the IOL condition, but not when it was compared against the FVL one. Furthermore, in the conditions where lengthening was "misplaced" in view of our prediction (i.e., IVL and FOL), only misplaced vowel lengthening impeded segmentation. These all seem to suggest that vowel lengthening is more influential than onset consonant lengthening, improving segmentation more when appearing in the expected position but inhibiting it more when misplaced. If this view is accepted, a question then arises to what leads the two types of lengthening to contribute to segmentation differently, at least for TSM listeners. Below we explore two possible accounts.

The first one relates to the concept of "perceptual center" (P-center: Morton *et al.*, 1976), which has been invoked by White *et al.* (2015) in an alternative explanation for their findings. The P-center is the moment at which a syllable is perceived to occur. It is situated around the left edge of the vocalic rime and its location may be perturbed by the preceding onset consonants. In general, lengthening the onset pulls the P-center slightly leftward away from the end of the syllable, but this shift is relatively small compared to the increase in onset duration such that the more obvious effect of onset lengthening is that it delays the P-center with respect to the whole syllable's beginning (and the P-center of the preceding syllable). These are captured in a formula proposed by Marcus (1981) for predicting P-center locations and schematically illustrated in Figs. 5(a) and 5(b). As a concrete example using the formula, lengthening the consonant in a CV syllable from 100 to 200 ms moves the P-center leftward from syllable offset by only 35 ms, but causing it to occur 65 ms later relative to the syllable's beginning. Yet, as White *et al.* point out, such a delay in P-center timing occurs not only when a syllable's onset is longer, but also when the vowel or rime of its preceding syllable is longer. Thus, instead of assuming a sharp functional division between vowels and consonants, an alternative account discussed by White *et al.* for why word-final vowel lengthening and word-initial onset lengthening are exploited in segmentation is that both cues lead to a lag in the P-center of the upcoming syllable, which listeners may interpret as signaling a prosodic boundary.

However, a point not discussed by White *et al.* (2015) is that vowel lengthening and onset lengthening may have different consequences for the perceived weight of a syllable, which might explain their different contributions to segmentation. This idea derives from a theory proposed by Ryan (2014) to account for the positive correlation between

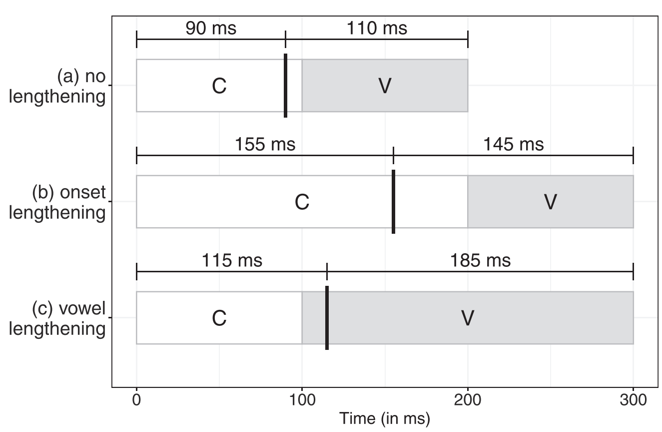


FIG. 5. Predicted P-center locations (black vertical lines in the bars) when (a) no segment in the CV syllables is lengthened; (b) when the onset consonant is lengthened by 100 ms; (c) when the vowel is lengthened by 100 ms.

https://doi.org/10.1121/10.0003751

onset size and a syllable's propensity for stress in languages such as English and Russian (e.g., CCV syllables are more likely to be stressed than CV or V syllables). Ryan proposes that the domain over which the perceived weight of a syllable is computed begins at its P-center. Since, as mentioned, lengthening the onset (via addition of consonants or increase in acoustic duration) moves the P-center slightly leftward, it increases the domain of weight calculation, making syllables with longer onsets preferred docking sites for stress. Yet, Ryan also notes that the onset exerts a weaker effect on perceived weight than the vowel. As the above example shows, doubling the duration of the 100-ms consonant results in only a 35-ms leftward shift of the P-center and hence a 35-ms increase in the weight calculation domain. In contrast, the vowel in the rime contributes almost in its entirety: doubling the duration of a 100-ms vowel adds 75 ms to weight calculation [see Fig. 5(c)]. As a result, given equal magnitude of lengthening, syllables with vowel lengthening are perceptually heavier than those with onset lengthening. It is perhaps this extra perceptual weight conferred on vowel-lengthened syllables that makes vowel lengthening a more salient and influential segmentation cue. Clearly, Ryan's theory does not set out to explain segmentation behavior. However, it is possible that the perceptual principle he proposes for syllable weight calculation in stress systems also plays a role in the segmentation process for all listeners, including those with a tone-language background.

The account that appeals to the P-center and perceived syllable weight likely predicts that vowel lengthening is universally more salient than onset consonant lengthening regardless of the listener's native language. Nevertheless, one may also take a language-specific view on the finding: the different contributions of vowel and onset lengthening observed in our TSM listeners may be ascribed to their experience with lexical tones. In a series of AL learning experiments, Gómez *et al.* (2018) found that listeners of Cantonese, a lexical tone language, were able to use transitional probability information carried by vowels for segmentation, but not that carried by consonants. This is in contrast to the findings obtained from non-tonal languages such as French, whose listeners exploit transitional probabilities between consonants but not those between vowels (Bonatti *et al.*, 2005). It is thus possible that tone experience might promote sensitivity to vocalic segments, at least for a task that requires segmentation of continuous speech based on its statistical regularities. As a consequence, any prosodic cues on vowels incidentally become more salient and are exploited better. This account may also explain some differences between previous findings and ours. Recall that, unlike our TSM listeners, the English listeners in White *et al.* (2015, 2014) showed no synergistic effect for word-initial onset lengthening and word-final vowel lengthening in combination when compared to the former cue in isolation. This could mean that onset lengthening is already a powerful cue, which may be because English listeners, like French ones, rely on transitional probabilities between consonants rather than between vowels. Yet, it has to be stressed that this interpretation is only speculative as White *et al.* did not actually test how word-final vowel lengthening alone fares against the cue combination.

The above discussion points to a few issues for future research. First, although the two accounts just considered do not seem contradictory and might in fact complement each other, more work could be done to investigate which one is more tenable. But before this investigation, it needs to be established in the first place whether vowel and onset lengthening indeed contribute to segmentation differently. Our hypothesis regarding the different contributions of the two lengthening cues is at best speculative as we did not include conditions for assessing their relative powers, which should preferably be ones in which the cues are directly pitted against each other (e.g., word-initial onset lengthening plus word-initial vowel lengthening). With such conditions, one can then examine which cue is more dominant. Second, while the available evidence seems to support our generalizations regarding segment lengthening, the current understanding of the role of onset consonant lengthening might still be limited as it is not thoroughly and widely investigated as vowel lengthening. Further studies could test whether longer onset consonants are also interpreted as word-initial by listeners of languages of other word-prosodic typological categories, particularly by those of languages that, like TSM, have language-specific phonological processes that may work against the interpretation of consonant duration as a structural cue.

V. CONCLUSION

Based on previous findings, it was predicted that TSM listeners would exploit vowel lengthening and onset consonant lengthening as cues to word-finality and word-initiality, respectively. The results of experiment I are consistent with the prediction as both word-final vowel lengthening and word-initial onset consonant lengthening improve the listeners' segmentation. Experiment II further demonstrates that these cues in combination synergistically enhance segmentation, at least when compared to word-initial onset consonant lengthening. Taken together, these findings may be interpreted as reflecting a functional division between vowels and consonants, with longer vowels analyzed by the Prosody Analyzer as signaling a prosodic right edge and longer onsets as signaling a prosodic left edge. We also found that vowel lengthening seems to exert a greater influence on TSM listeners' segmentation than onset lengthening. Two accounts have proposed: one invokes the concepts of P-center and perceived syllable weight while the other attributes the finding to language-specific experience with lexical tones. Possible directions for future research include establishing whether the two types of lengthening indeed contribute to segmentation to different degrees and investigating the putatively universal interpretation of onset lengthening as a word beginning cue in a broader range of languages.

[1]Inspection of the syllables revealed that there was some degree of nasality in the vowels after nasal onsets. The extent to which the vowels were nasalized could potentially affect TSM listeners' segmentation strategies as syllable-initial voiced stops are in complementary distribution with syllable-initial nasals in TSM: the former precede only oral stops and the latter only nasal vowels (e.g., [ba] and [nẽ] can occur but not *[bẽ] and *[mi]) (Wang, 1995, 1999). It is unclear how the patterns of vowel nasality in the stimuli would compare to those typically found in spontaneous TSM speech and whether TSM listeners would perceive the AL vowels as all being oral (thus perceiving the AL as not respecting the complementary distribution). Whichever the case, though, it is unlikely that vowel nasality would yield a net effect on our results when the different lengthening conditions were compared since the stimuli in all the conditions were created from the same stimulus set. Differences between the conditions are more likely due to the lengthening manipulation.

ACKNOWLEDGMENTS

We would like to thank Kai-ting Chuang for assistance with participant recruitment. This work was supported by a grant from the Ministry of Science and Technology, Taiwan, to the first author (Grant number: MOST 109-2410-H-110-069-MY3).

Ang, U.-J. (**1993**). "Táiwān hélǎoyǎ shēngdiào yánjiū" ["A study of tones in Taiwanese Holo"], The Independence Evening Post.

Bagou, O., and Frauenfelder, U. H. (**2018**). "Lexical segmentation in artificial word learning: The effects of converging sublexical cues," Lang. Speech **61**(1), 3–30.

Beckman, M. E. (**1986**). *Stress and Non-Stress Accent* (De Gruyter Mouton, Berlin).

Bertinetto, P. M. (**1980**). "The perception of stress by Italian speakers," J. Phon **8**(4), 385–395.

Boersma, P., and Weenink, D. (**2018**). "Praat: Doing phonetics by computer" [computer program], http://www.praat.org/ (Last viewed 2/3/2018).

Bonatti, L. L., Peña, M., Nespor, M., and Mehler, J. (**2005**). "Linguistic constraints on statistical computations," Psychol. Sci **16**(6), 451–459.

Bürkner, P. C. (**2017**). "brms: An R package for Bayesian multilevel models using Stan," J. Stat. Softw. **80**(1), 1–28.

Caldwell-Harris, C. L., Lancaster, A., Ladd, D. R., Dediu, D., and Christiansen, M. H. (**2015**). "Factors influencing sensitivity to lexical tone in an artificial language: Implications for second language learning," Stud. Second Lang. Acquis **37**(2), 335–357.

Cheng, R. L. (**1973**). "Somoe notes on tone sandhi in Taiwanese," Linguistics **11**(100), 5–25.

Chiang, W.-Y. (**1992**). *The Prosodic Morphology and Phonology of Affixation in Taiwanese and Other Chinese Languages* (University of Delaware, Newark).

Cho, T. (**2004**). "Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English," J. Phon **32**(2), 141–176.

Cho, T., and Keating, P. A. (**2001**). "Articulatory and acoustic studies on domain-initial strengthening in Korean," J. Phon. **29**(2), 155–190.

Cho, T., McQueen, J. M., and Cox, E. A. (**2007**). "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English," J. Phon **35**(2), 210–243.

Chung, R.-F. (**1996**). *The Segmental Phonology of Southern Min in Taiwan* (Crane Publishing, Taipei City).

Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang **2**(3–4), 133–142.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (**1986**). "The syllable's differing role in the segmentation of French and English," J. Mem. Lang. **25**(4), 385–400.

Cutler, A., Sebastian-Galles, N., Soler-Vilageliu, O., and Van Ooijen, B. (**2000**). "Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons," Mem. Cogn. **28**(5), 746–755.

Endress, A. D., and Hauser, M. D. (**2010**). "Word segmentation with universal prosodic cues," Cogn. Psychol. **61**(2), 177–199.

Fougeron, C. (**2001**). "Articulatory properties of initial segments in several prosodic constituents in French," J. Phon. **29**(2), 109–135.

Fougeron, C., and Keating, P. A. (**1997**). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. **101**(6), 3728–3740.

Fry, D. B. (**1958**). "Experiments in the perception of stress," Lang. Speech **1**(2), 126–152.

Fu, Q.-J., and Zeng, F.-G. (**2000**). "Identification of temporal envelope cues in Chinese tone recognition," Asia Pacific J. Speech Lang. Hear. **5**(1), 45–57.

Gandour, J. (**1977**). "On the interaction between tone and vowel length: Evidence from Thai dialects," Phonetica **34**(1), 54–65.

Gay, T. (**1978**). "Physiological and acoustic correlates of perceived stress," Lang. Speech **21**(4), 347–353.

Gelman, A., and Hill, J. (**2006**). *Data Analysis Using Regression and Multilevel/Hierarchical models* (Cambridge University Press, Cambridge).

Gómez, D. M., Mok, P., Ordin, M., Mehler, J., and Nespor, M. (**2018**). "Statistical speech segmentation in tone llanguages: The role of lexical tones," Lang. Speech **61**(1), 84–96.

Gow, D. W., and Gordon, P. C. (**1995**). "Lexical and prelexical influences on word segmentation: Evidence from priming," J. Exp. Psychol. Hum. Percept. Perform **21**(2), 344–359.

Gussenhoven, C., and Rietveld, A. C. M. (**1992**). "Intonation contours, prosodic structure and preboundary lengthening," J. Phon. **20**(3), 283–303.

Gussenhoven, C., and Zhou, W.-C. (**2013**). "Revisiting pitch slope and height effects on perceived duration," in *Interspeech 2013, 14th Annual Conference of the International Speech Communication Association*, pp. 1365–1369.

Hay, J. S. F., and Diehl, R. L. (**2007**). "Perception of rhythmic grouping: Testing the iambic/trochaic law," Percept. Psychophys. **69**(1), 113–122.

Howie, J. M. (**1976**). *Acoustical Studies of Mandarin Vowels and Tones* (Cambridge University Press, Cambridge).

Hsu, C.-S. K., and Jun, S.-A. (**1998**). "Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic?," UCLA Work. Pap. Phon. **96**, 69–89.

Jusczyk, P. W., Cutler, A., and Redanz, N. J. (**1993**). " 'Infants' preference for the predominant stress patterns of English words," Child Dev. **64**(3), 675–687.

Keating, P., Cho, T., Fougeron, C., and Hsu, C.-S. (**2004**). "Domain-initial articulatory strengthening in four languages," in *Phonetic Interpretation Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 145–163.

Kim, S., Broersma, M., and Cho, T. (**2012**). "The use of prosodic cues in learning new words in an unfamiliar language," Stud. Second Lang. Acquis. **34**(3), 415–444.

Klatt, D. H. (**1975**). "Vowel lengthening is syntactically determined in a connected discourse," J. Phon. **3**(3), 129–140.

Kong, Q.-M. (**1987**). "Influence of tones upon vowel duration in cantonese," Lang. Speech **30**(4), 387–399.

Krämer, M. (**2009**). *The Phonology of Italian* (Oxford University Press, Oxford).

Kruschke, J. K. (**2014**). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan* (Academic Press, New York).

Kruschke, J. K., Aguinis, H., and Joo, H. (**2012**). "The time has come: Bayesian methods for data analysis in the organizational sciences," Organ. Res. Methods **15**(4), 722–752.

Kuo, Y.-C., Rosen, S., and Faulkner, A. (**2008**). "Acoustic cues to tonal contrasts in Mandarin: Implications for cochlear implants," J. Acoust. Soc. Am. **123**(5), 2815–2824.

Lewandowski, D., Kurowicka, D., and Joe, H. (**2009**). "Generating random correlation matrices based on vines and extended onion method," J. Multivar. Anal. **100**(9), 1989–2001.

Lieberman, P. (**1960**). "Some acoustic correlates of word stress in American English," J. Acoust. Soc. Am. **32**(4), 451–454.

Lin, H.-B., and Repp, B. H. (**1989**). "Cues to the perception of Taiwanese tones," Lang. Speech **32**(1), 25–44.

Lin, M.-C. (**1988**). "Putonghua shengdiao de shengxue texing he zhijue zhengzhao" ["The acoustic characteristics and perceptual cues of tones in Standard Chinese"], Zhongguo Yuwen [Chin. Linguist.] **204**, 182–193.

Lin, Y.-H. (**2004**). "Chinese affixal phonology: Some analytical and theoretical issues," Lang. Linguist. **5**(4), 1019–1046.

Lindblom, B. (**1978**). "Final lengthening in speech and music," in *Nordic Prosody*, edited by E. Gårding, G. Bruce, and R. Bannert (Department of Linguistics, Lund University), pp. 85–101.
Maddieson, I. (**1978**). "Tone effects on consonants," J. Phon. **6**(4), 327–343.
Marcus, S. M. (**1981**). "Acoustic determinants of perceptual center (P-center) location," Percept. Psychophys. **30**(3), 247–256.
McElreath, R. (**2018**). *Statistical Rethinking: A Bayesian Course with examples in R and Stan* (Chapman and Hall/CRC, New York).
McQueen, J. M., Otake, T., and Cutler, A. (**2001**). "Rhythmic cues and possible-word constraints in Japanese speech segmentation," J. Mem. Lang. **45**(1), 103–132.
Morton, J., Marcus, S., and Frankish, C. (**1976**). "Perceptual centers (P-centers)," Psychol. Rev. **83**(5), 405–408.
Nespor, M., Peña, M., and Mehler, J. (**2003**). "On the different roles of vowels and consonants in speech processing and language acquisition," Ling. Ling. **2**(2), 203–229.
Ordin, M., Polyanskaya, L., Laka, I., and Nespor, M. (**2017**). "Cross-linguistic differences in the use of durational cues for the segmentation of a novel language," Mem. Cogn. **45**(5), 863–876.
Ou, S.-C., and Guo, Z.-C. (**2019**). "The language-specific use of fundamental frequency rise in segmentation of an artificial language: Evidence from listeners of Taiwanese Southern Min," Lang. Speech (published online).
Pisoni, D. B. (**1973**). "Auditory and phonetic memory codes in the discrimination of consonants and vowels," Percept. Psychophys. **13**(2), 253–260.
Psychology Software Tools. (**2012**). "E-Prime 2.0."
Quené, H. (**1993**). "Segment durations and accent as cues to word segmentation in Dutch," J. Acoust. Soc. Am. **94**(4), 2027–2035.
R Core Team (**2017**). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, https://www.r-project.org/ (Last viewed 12/6/2017).
Ryan, K. M. (**2014**). "Onsets contribute to syllable weight: Statistical evidence from stress and meter," Language **90**(2), 309–341.
Saffran, J. R., Newport, E. L., and Aslin, R. N. (**1996**). "Word segmentation: The role of distributional cues," J. Mem. Lang. **35**(4), 606–621.
Shukla, M., Nespor, M., and Mehler, J. (**2007**). "An interaction between prosody and statistics in the segmentation of fluent speech," Cogn. Psychol. **54**(1), 1–32.
Sluijter, A. M. C., and van Heuven, V. J. (**1996**). "Acoustic correlates of linguistic stress and accent in Dutch and American English," in *Proceedings of the Fourth International Conference on Spoken Language Processing*, pp. 630–633.
Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., and Öhman, S. E. G. (**1969**). "Cross language study of vowel perception," Lang. Speech **12**(1), 1–23.
Tagliapietra, L., and McQueen, J. M. (**2010**). "What and where in speech recognition: Geminates and singletons in spoken Italian," J. Mem. Lang. **63**(3), 306–323.
Toro, J. M., Sebastián-Gallés, N., and Mattys, S. L. (**2009**). "The role of perceptual salience during the segmentation of connected speech," Eur. J. Cogn. Psychol. **21**(5), 786–800.
Tremblay, A., Coughlin, C. E., Bahler, C., and Gaillard, S. (**2012**). "Differential contribution of prosodic cues in the native and non-native segmentation of French speech," Lab. Phonol. **3**(2), 385–423.
Tyler, M. D., and Cutler, A. (**2009**). "Cross-language differences in cue use for speech segmentation," J. Acoust. Soc. Am. **126**(1), 367–376.
Vaissière, J. (**1983**). "Language-independent prosodic features," in *Prosody: Models and Measurements*, edited by A. Cutler and R. D. Ladd (Springer, Berlin), pp. 53–66.
Van Wijk, M. (**1999**). "The perception of aspiration: A cross-linguistic study," in *Proceedings of the 14th International Congress of Phonetic Sciences*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey (University of California, Berkeley), pp. 853–856.
Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., and Kong, E. J. (**2018**). "Bayesian data analysis in the phonetic sciences: A tutorial introduction," J. Phon. **71**, 147–161.
Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., and Morey, R. D. (**2018**). "Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications," Psychon. Bull. Rev **25**(1), 35–57.
Wang, S. H. (**1995**). *Experimental Studies in Taiwanese Phonology* (Crane Publishing, Taipei City).
Wang, S. H. (**1999**). "A concept formation experiment for Taiwan Min voiced stop consonants," Symp. Ser. Inst. Linguist. Acad. Sin **2**, 233–249.
Wang, S. H., and Liu, J. H. C. (**2010**). "The morphologization of liaison consonants in Taiwan Min and Taiwan Hakka," Lang. Linguist. **11**(1), 1–20.
White, L., Benavides-Varela, S., and Mády, K. (**2020**). "Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues?," J. Phon. **81**, 100982.
White, L., Mattys, S. L., Stefansdottir, L., and Jones, V. (**2015**). "Beating the bounds: Localized timing cues to word segmentation," J. Acoust. Soc. Am. **138**(2), 1214–1220.
White, L., Mattys, S., Stefansdottir, L., and Jones, V. (**2014**). "Lengthened consonants are interpreted as word-initial," in *Proceedings of the 7th International Conference on Speech Prosody 2014*, pp. 477–481.
Xu, J.-D. (**1998**). *Táiwāny ǔ Gàilùn* [An Introduction to Taiwanese] (Avanguard Book, Taipei City).
Yang, H.-F. (**1991**). *Táiwān Mǐnnány ǔ Yǔfǎ Gǎo* [A Grammar of Taiwan Minnan] (Da-an Press, Taipei City).
Yu, A. C. L. (**2010**). "Tonal effects on perceived vowel duration," in *Laboratory Phonology 10*, edited by C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (De Gruyter Mouton, Berlin), pp. 151–168.
Zee, E. (**1977**). "The effect of F0 on the duration of [s]," UCLA Work. Pap. Phon. **36**, 122–127.