Applied Psycholinguistics (2021), **42**, 631–655
doi:10.1017/S0142716420000818

CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

The use of tonal coarticulation in segmentation of artificial language speech: A study with Mandarin listeners

Zhe-Chen Guo[1,*] and Shu-Chen Ou[2]

[1]University of Texas at Austin and [2]National Sun Yat-sen University
*Corresponding author. E-mail: zcadamguo@utexas.edu.

(Received 21 January 2020; revised 1 November 2020; accepted 13 November 2020; first published online 05 January 2021)

Abstract

Tonal carryover assimilation, whereby a tone is assimilated to the preceding one, is conditioned by prosodic boundaries in a way suggesting that its presence may signal continuity or lack of a boundary. Its possibility as a speech segmentation cue was investigated in two artificial language (AL) learning experiments. Mandarin-speaking listeners identified the "words" of a three-tone AL (e.g., [pé.tī.kù]) after listening to six long speech streams in which the words were repeated continuously without pauses. The first experiment revealed that segmentation was disrupted in an "incongruent-cues" condition where tonal carryover assimilation occurred across AL word boundaries and conflicted with statistical regularities in the speech streams. Segmentation was neither facilitated nor inhibited in a "congruent-cues" condition where tonal carryover assimilation occurred only within the AL words in 27% of the repetitions and never across word boundaries. A null effect was again found for the congruent-cues condition of the second experiment, where all AL word repetitions carried tonal carryover assimilation. These findings show that tonal carryover assimilation is exploited to resolve segmentation problems when cues conflict. Its null effect in the congruent-cues conditions might be linked to cue redundancy and suggest that it is weighted low in the segmentation cue hierarchy.

Keywords: artificial language learning; carryover assimilation; speech segmentation; tonal coarticulation

A crucial mental process during spoken language comprehension is the segmentation of continuous speech streams into discrete units such as words. Accumulated evidence has demonstrated that listeners exploit a wide variety of cues to facilitate this process (see, e.g., Cutler, 2012; Davis, Marslen-Wilson, & Gaskell, 2002, for reviews). These range from statistical regularities in the speech input to language-specific phonological patterns and acoustic-phonetic details. This study sets out to expand this line of investigation by experimentally testing whether tonal coarticulation, a low-level acoustic-phonetic phenomenon commonly observed in lexical tone languages, is used in speech segmentation. Before showing how tonal coarticulation could be useful, we present an overview of previous findings as awareness of what cues have been shown to support segmentation is helpful for designing an experiment for our purpose.

Despite its continuous nature, speech contains statistical regularities that provide useful boundary information. These regularities are usually expressed in terms of transitional probability (TP), which captures the likelihood that a pair of elements co-occur. In general, two consecutive syllables with a higher TP tend to be word-internal and are more likely to be perceived as being so, whereas two syllables with a lower TP tend to occur across words and are more likely to be perceived as straddling a boundary (Mirman, Magnuson, Estes, & Dixon, 2008; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Such a segmentation solution is not exclusive to adult listeners as young children and even infants are also able to compute TPs to extract possible word forms (Aslin, Saffran, & Newport, 1998; Estes, Evans, Alibali, & Saffran, 2007; Hay, Pelucchi, Estes, & Saffran, 2011; Saffran et al., 1997; Thiessen & Saffran, 2007). Tracking statistical regularities in speech is therefore thought to be an ontologically early segmentation strategy, permitting discovery of potentially meaningful units before the emergence of an adultlike lexicon.

With these statistical computations supporting segmentation from an early age, listeners further develop various language-specific segmentation solutions through increasing experience with native-language phonological patterns. For example, Dutch phonotactics prohibit word-internal [mr] sequences (e.g., *[mrɒk] is not a possible Dutch word) and Dutch listeners hearing these sequences would assume a word boundary between the two consonants (McQueen, 1998). Vowel harmony in Finnish dictates that word-internal vowels agree in frontness/backness (Karlsson, 1983), inclining Finnish listeners to segment speech in such a way that two syllables belong to a single word when their vowels agree in this feature but to different words when they do not (Suomi, McQueen, & Cutler, 1997; Vroomen, Tuomainen, & de Gelder, 1998). Phonological patterns that promote segmentation solutions may also arise from the distribution of a phonological entity, such as lexical stress. In English, the majority of the words begin with stressed syllables (Cutler & Carter, 1987); thus, English listeners treat stressed or prominent syllables as word onsets and segment speech accordingly (Cutler, 1990; Cutler & Butterfield, 1992; Cutler & Norris, 1998; Tyler & Cutler, 2009).

In addition to phonological patterns, fine-grained acoustic–phonetic aspects of speech sounds may also help resolve segmentation problems. English listeners use subtle durational differences in [l] to disambiguate *two lips* and *tulips* (Gow & Gordon, 1995). Besides, due to the prosodic structuring of speech, segments in the initial position of a larger prosodic constituent are produced with stronger articulatory strengthening than those in the initial position of a smaller prosodic constituent (Cho & Keating, 2001; Fougeron & Keating, 1997; Keating, Cho, Fougeron, & Hsu, 2004). The stronger strengthening effect associated with a larger constituent has been shown to facilitate the search of word onsets (Cho, McQueen, & Cox, 2007). Similarly, the degree of coarticulation between segments carries boundary information. Adjacent segments are more strongly coarticulated within words than between words (Byrd, 1996; Byrd & Saltzman, 1998), or across a smaller prosodic boundary than across a larger one (Cho, 2004; Fougeron & Keating, 1997). Listeners

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

can use such coarticulatory information to segment speech (Fernandes, Kolinsky, & Ventura, 2010; Fernandes, Ventura, & Kolinsky, 2007).

Fine-grained acoustic–phonetic details may modulate segmentation behavior to such an extent that they affect or even determine the use of a segmentation strategy that is motivated by a language-specific phonological pattern. Supporting evidence comes from recent studies with the use of tonal or fundamental frequency (F0) cues by Korean and Taiwanese Southern Min (TSM) listeners. Korean is thought to have a prosodic constituent called accentual phrase (AP), which frequently begins with a low (L) tone and ends in a high (H) tone (Jun, 1998) and Korean listeners tend to perceive H-L tone sequences as cueing an AP boundary (Kim, Broersma, & Cho, 2012; Kim & Cho, 2009). Moreover, Tremblay, Cho, Kim, and Shin (2019) suggest that the way in which the tone sequence is acoustic-phonetically realized affects how well it can be exploited for segmentation purposes. They found that Korean listeners' segmentation improved as the L tone in the tonal sequence became phonetically lower and more closely resembled the canonical realization of the AP-initial L tone in Korean. A related case was reported in Ou and Guo's (2019) study with TSM listeners. TSM is a tone language with an extensive tone sandhi process that restricts its only rising tone to the final position of the tone sandhi domain. It may thus be hypothesized that a cue as simple as a final rise in F0 suffices to signal finality for TSM listeners. Yet, given that the domain-final position is associated with phonetic final lengthening, it may be alternatively hypothesized that final lengthening is needed for a final F0 rise to be a sufficient finality cue. Ou and Guo's findings support the alternative hypothesis: final F0 rise alone did not improve TSM listeners' segmentation; instead, it was the combination of final F0 rise and final lengthening that did. Taken together, these two studies demonstrate that while some segmentation strategies are shaped by the distributions of phonological entities (e.g., the H-L tone sequence of the AP in Korean and the domain-final rising tone in TSM), they do not abstract away from the fine details of how those entities are acoustic-phonetically manifested.

To sum up, the literature has revealed at least three types of speech segmentation cues: (a) statistical regularities, (b) native-language phonological patterns, and (c) fine-grained acoustic–phonetic details. It is also found that acoustic–phonetic details may impact the use of phonological patterns, suggesting that they play a non-trivial role in shaping segmentation behavior. Research on acoustic–phonetic cues could thus contribute insight into how sensitive and resourceful listeners are in solving segmentation problems, a question that may inform theories and models seeking to reveal what cues are useful and how they are integrated (e.g., Mattys, White, & Melhorn, 2005). Yet, empirical work to date on this type of cues is mostly concerned with those at the segmental level (e.g., coarticulation of segments). Less attention has been paid to whether fine-grained tonal or F0 information is exploited. Perhaps the study that bears most on this question is that by Tremblay et al. (2019) discussed above. Nevertheless, while Tremblay et al. show that Korean listeners' use of the H-L tone sequence was affected by the phonetic manifestation of the L tone, the segmentation strategy that the listeners employed is phonologically motivated in nature. Subtle acoustic–phonetic changes to the scaling of the L tone do result in better use of the segmentation strategy if they create a more canonical realization of the tone, but they are not what lead Korean listeners to develop the strategy in the first place. However, there are detailed acoustic–phonetic tonal phenomena

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

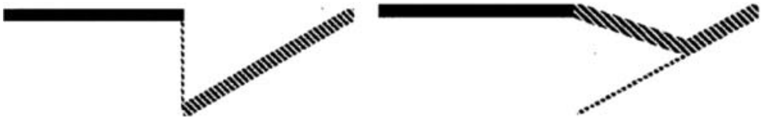

Figure 1. Possible transitions between a high-level tone and a rising tone (adapted from Xu, 1997, p. 63). The left figure represents a situation in which there is no tonal coarticulation. The right one illustrates tonal carryover assimilation, whereby the initial portion of the rising tone's F0 contour changes into a falling F0 transition due to assimilation to the preceding high-level tone.

in speech that do not seem to depend on a particular phonological entity and can potentially promote segmentation strategies on their own. One of them is tonal coarticulation. In this study, we focused on a specific type of tonal coarticulation and investigated its effect on speech segmentation. Below is an introduction to lexical tones and tonal coarticulation.

Tonal Coarticulation

Lexical tones are pitch patterns over a syllable that serve to differentiate word meanings. For example, in Mandarin, [ma] means "mother" when bearing a high-level tone (Tone 1) but "hemp" when bearing a rising tone (Tone 2). The primary acoustic correlate of lexical tones is F0, and the F0 realizations of one tone in connected speech may vary under the contextual influence of adjacent tones, resulting in the so-called tonal coarticulation. Production experiments have revealed much evidence for tonal coarticulation in Mandarin and other lexical tone languages (see, e.g., Chen, 2012; Xu, 2001, for a review). Recently, Hao, Zhang, Xie, and Zhang (2018) proposed a scheme for annotating tonal coarticulation and applied it to speech samples from a Mandarin corpus. About 51% of bitonal syllable sequences in their data were labeled as being tonally coarticulated, suggesting that tonal coarticulation is prevalent in connected speech, at least in Mandarin. The exact effect of one tone on another is commonly described in terms of (a) whether it is assimilatory or dissimilatory and (b) whether its direction is anticipatory or carryover (e.g., Brunelle, 2009; Chang & Hsieh, 2012; Cheng, 1968; Peng, 1997; Potisuk, Gandour, & Harper, 1997; Shen, 1990; Shih, 1988; Xu, 1994; Zhang & Liu, 2011). A logical corollary of this is that there are four theoretically possible types of tonal coarticulation: carryover assimilation, carryover dissimilation, anticipatory assimilation, and anticipatory dissimilation. As an example, the right panel of Figure 1 is a schematic illustration of carryover assimilation, whereby the F0 contour of a tone is partially assimilated to the preceding tone.

Among the possible types of tonal coarticulation, carryover assimilation is of particular interest to the current investigation for two reasons. First, cross-linguistically, carryover effects are generally assimilatory, as evidenced by the fact that carryover assimilation is attested in a broad range of lexical tone languages, including Mandarin (Shih, 1988; Xu, 1997), Taiwanese (Cheng, 1968; Wang, 2002), Tianjin Chinese (Zhang & Liu, 2011), Thai (Gandour, Potisuk, & Dechongkit, 1994), Cantonese (Li, Lee, & Qian, 2004), Vietnamese (Brunelle, 2009), and so on. In contrast, findings on whether anticipatory effects are assimilatory or dissimilatory are

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

relatively mixed as these effects are reported to vary across languages or even across different tones of the same language (Zhang & Liu, 2011). Second, and more importantly, it has been shown that as with segmental coarticulation, tonal carryover assimilation is conditioned by prosodic boundary strength. In Mandarin, it tends to be stronger when two adjacent tones span the boundary of a smaller prosodic unit than when they span that of a larger unit (Lai & Kuang, 2016). Such a tonal coarticulatory effect may even be completely eliminated when the neighboring syllables straddle a major prosodic break (Zhang & Kawanami, 1999). These suggest that tonal carryover assimilation may be useful for segmenting continuous speech into discrete units.

The Current Study

The goal of the present work is to experimentally test this possibility. As with many empirical studies, we attempt to draw conclusions about a single cue, which, in our case, is tonal carryover assimilation. Nevertheless, cues can rarely be isolated from each other even in well-controlled laboratory speech materials. For example, while Fernandes et al.'s (2007) listeners exploited segmental coarticulation in segmenting an artificial language (AL), TP information was always present in the stimuli. It is therefore instructive to consider two possible scenarios for segmentation in the presence of multiple cues as the considerations could inform the experimental design and result interpretation. One scenario is that the cues operate in cooperation. A possible outcome is that their effects are additive or even synergistic, enhancing segmentation to a greater extent than a single cue alone does. An example is Fernandes et al., in which segmentation was better when segmental coarticulation and TP information were present and congruent with each other than when TPs were the sole cue. Yet, cooperating cues may be redundant and may not produce extra facilitation. Bagou and Frauenfelder (2018) and Kim et al. (2012) showed that although French and Korean listeners benefited from final lengthening and final F0 rise in isolation, conjoining these two prosodic cues does not improve their performance further. The other scenario is when cues operate in conflict and there are again two possible outcomes. One is that the conflict leads to inhibition. For example, Ordin, Polyanskaya, Laka, and Nespor (2017) found that Italian listeners used vowel lengthening to locate word-medial positions; therefore, when it was the vowels in the word-initial positions that were lengthened, TP-based segmentation was disrupted, reducing performance to a level worse than that of a condition with TPs as the only cue. Alternatively, the cue conflict may neither facilitate nor inhibit segmentation.

With these two scenarios in mind, we investigated the use of tonal carryover assimilation with the AL learning technique. It is an experimental paradigm that has been widely adopted to explore how phonological patterns and acoustic–phonetic details guide segmentation (e.g., Bagou & Frauenfelder, 2018; Fernandes et al., 2007; Kim et al., 2012; Ordin & Nespor, 2016; Ordin et al., 2017; Toro, Pons, Bion, & Sebastián-Gallés, 2011; Tremblay et al., 2019; Tyler & Cutler, 2009). A typical AL learning experiment has a learning (or exposure) phase followed by a test phase. In the learning phase, participants learn an AL by listening

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

to long speech streams in which tokens of the “words” of the AL, which are meaningless syllable sequences, are concatenated without pauses in between. The basic cue for word segmentation is TP. For example, as mentioned, adjacent syllables with a lower TP are more likely to span a boundary. On top of TPs, additional cues may be introduced to examine how they impact segmentation. After listening to the speech streams, participants complete a two-alternative forced-choice test in which they hear a word of the AL and a sequence that is not part of the AL vocabulary and have to select the former. The proportion of correct selections is thought to reflect how well the AL speech streams were segmented during the learning.

Such an experiment has two important advantages for research concerned with phonological or acoustic–phonetic cues. First, as it has been suggested that segmentation is primarily lexically driven (Mattys & Bortfeld, 2017; Mattys et al., 2005), using nonsense speech prevents listeners from segmenting based on lexical knowledge (e.g., by lexical subtraction: White, Melhorn, & Mattys, 2010) and allows the researcher to obviate confounds such as word frequency. Second, with artificial speech, one can precisely control the acoustic–phonetic content of the additional non-TP cues. Two AL learning experiments were conducted in this study. Although their focus is on tonal carryover assimilation, as noted above, it is useful to also consider its possible effects in the presence of TP information, the basic segmentation cue in an AL learning task. We thus constructed conditions corresponding to the two scenarios discussed. Details about the design and the hypotheses to test are presented below.

Experiment 1

Participants

Ninety-six adult native speakers of Mandarin (28 males and 68 females) with no self-reported history of hearing impairments were recruited from a university in Southern Taiwan. Their mean age in years was 20.3 (range: 18–23; standard deviation: 1.4). They had been learning English as a compulsory subject in school, and 9 of them had received musical training.[1] They were randomly assigned to one of the three experimental conditions (see the Design and stimuli section). The single-cue, congruent-cues, and incongruent-cues conditions had 31, 33, and 32 listeners, respectively.

Design and stimuli

In the learning phase, participants were exposed to a nonsense tonal language under three conditions, the design of which was modeled after Fernandes et al.’s (2007) study with segmental coarticulation. One was called the “single-cue” condition, in which participants could only rely on TPs to segment the AL speech streams. In addition to TPs, tonal carryover assimilation was introduced to the speech streams in the other two conditions. In the “congruent-cues” condition, tonal carryover assimilation occasionally occurred between the syllables within an AL word but never across word boundaries. Therefore, the tonal coarticulatory cue agreed with the TP cue. In the “incongruent-cues” condition, however, these cues were

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

Table 1. Words of the artificial language (AL) and partwords

Tone pattern	AL word	Partword	
High-mid-high	[kí.pī.tá]	[kí.tē.pá]	(from [tà.tū.kí] and [tē.pá.tī])
High-mid-low	[pé.tī.kù]	[tá.pū.kù]	(from [kí.pī.tá] and [pū.kù.pē])
Mid-high-mid	[tē.pá.tī]	[pē.kí.pī]	(from [pū.kù.pē] and [kí.pī.tá])
Mid-low-mid	[pū.kù.pē]	[tī.kù.tē]	(from [pé.tī.kù] and [tē.pá.tī])
Low-mid-high	[tà.tū.kí]	[pè.tē.pá]	(from [kì.kē.pè] and [tē.pá.tī])
Low-mid-low	[kì.kē.pè]	[kù.pē.tà]	(from [pū.kù.pē] and [tà.tū.kí])

Note: The acute (´), macron (¯), and grave (`) marks represent high-level, mid-level, and low-level tones, respectively. The dots indicate syllable boundaries.

pitted against each other by letting tonal carryover assimilation occur across AL word boundaries. The congruent-cues and incongruent-cues conditions corresponded to the situation in which the cues are in harmony and the situation in which they are in conflict, respectively.

As with several studies (e.g., Kim et al., 2012; Ordin et al., 2017; Saffran et al., 1996; Vroomen et al., 1998), we created an AL consisting of six trisyllabic words, which were meaningless sequences of consonant-vowel syllables, as listed in the second column of Table 1. The words were formed by four vowels ([a, i, u, e]), three consonants ([p, t, k]), and three level tones (high-, mid-, and low-level tones). These consonants and vowels are cross-linguistically common and occur in Mandarin at least at the phonetic level. Irrespective of the tones, all the used consonant–vowel syllables except for [ki] are phonotactically possible in Mandarin. One constraint imposed during the construction of the AL lexicon was that adjacent syllables in a word had to differ by one tone level. For example, a high-level tone could only be preceded and followed by a mid-level tone, not by itself or by the low-level tone. Thus, there could be only six tone patterns, as shown in the first column of Table 1. In these patterns, neighboring tones always had different tone heights, creating what is referred to in the literature as "conflicting tonal contexts" (e.g., Peng, 1997; Xu, 1994) and allowing us to implement tonal carryover assimilation for every two adjacent tones in an AL word (and also a partword, described below). The fact that adjacent tones differed by one tone level also enabled us to control for the magnitude of carryover assimilation, so that, for instance, there was no carryover assimilation between the high-level and low-level tones, which would span a wider F0 range than that between the middle-level and low-level tones.

The syllables making up the AL words were individually inserted into a carrier sentence (i.e., /wuo ʂuo _____ t͡ʂɤ kɤ t͡sɹ̩/ "I said the word _____.") and read by a male native speaker of Mandarin with phonetic training in a monotone into a Zoom H4n Handy Recorder. The recorded items were digitized at a sampling rate of 44.1 kHz and stored as a single WAV file. The syllables were excised from the carrier sentence and then underwent manipulations using Praat (Boersma & Weenink, 2018). Their root-mean-squared amplitudes were equalized and their durations were normalized to 335 ms, which was the mean duration of the original, unmanipulated syllables.

Next, their F0 contours were flattened, set to 126 Hz (the average F0 of the syllables prior to the manipulations), and resynthesized using the overlap-add method in Praat. The resulting flat F0 contour served as the mid-level tone. Following Caldwell-Harris, Lancaster, Ladd, Dediu, and Christiansen (2015), we created the high-level and low-level tones by shifting the F0 contour up and down, respectively, by 3.5 semitones. The manipulated syllables were concatenated to form the six AL words.

In addition, six "partwords," which are listed in the third column of Table 1, were created using the same set of manipulated syllables. They served as the distractor stimuli in the test phase and were trisyllabic sequences derived by combining the last syllable of an AL word with the first two syllables of another AL word, or by combining the last two syllables of an AL word with the first syllable of another AL word. They were constructed under the same constraint for the AL words and therefore had the same six tone patterns. This prevented participants from being able to easily reject the partwords by identifying novel tone patterns.

The learning-phase stimuli were six speech streams in which tokens of the AL six words were concatenated with no pauses in between. Each stream contained a total of 120 tokens, 20 for each AL word. These tokens appeared in a random order but under the restriction that the same word did not occur twice in a row. As in Tyler and Cutler (2009), the first and last 5 s of each stream were faded in and out to prevent participants from hearing the syllables at the beginning and the end of the stream and using them to discover word boundaries. Each stream was 2 min long and the total duration of the six streams (and hence the learning phase) was about 12 min. The TP for each pair of adjacent syllables *AB* in the streams was calculated using the formula proposed in Saffran et al. (1996); that is, it is equal to the frequency of *AB* divided by the frequency of *A*. For each AL word or partword, an average TP was computed by taking the average of the TP between the first and second syllables and that between the second and third syllables. The average TPs for the AL words ranged between 0.75 and 1.00 (mean: 0.88) and those for the partwords ranged between 0.32 and 0.62 (mean: 0.49). The six speech streams differed from each other in the order in which the AL words appeared but were the same across the three conditions except for the presence or absence of tonal carryover assimilation and the way in which it was introduced. In the single-cue condition, there was no carryover assimilatory effect from one tone on the next tone and listeners could only achieve segmentation by tracking TPs.

In the congruent-cues and incongruent-cues conditions, tonal carryover assimilation was introduced as an additional cue. In each stream, a fixed number of trisyllabic sequences was selected to receive tonal carryover assimilation. In the incongruent-cues condition, they corresponded to all instances (i.e., 100%) of the partwords in the speech stream. However, the partwords occurred only incidentally and made up just about 27% of the syllables in the stream. To ensure that the incongruent-cues and congruent-cues conditions differed only in the alignment of the tonal coarticulatory cue with word boundaries but not in the number of the syllables with the cue (as in Fernandes et al., 2007), only 27% of the tokens of the AL words in the congruent-cues condition were selected as the trisyllabic sequences that would receive tonal carryover assimilation. This tonal cue was implemented by performing the

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

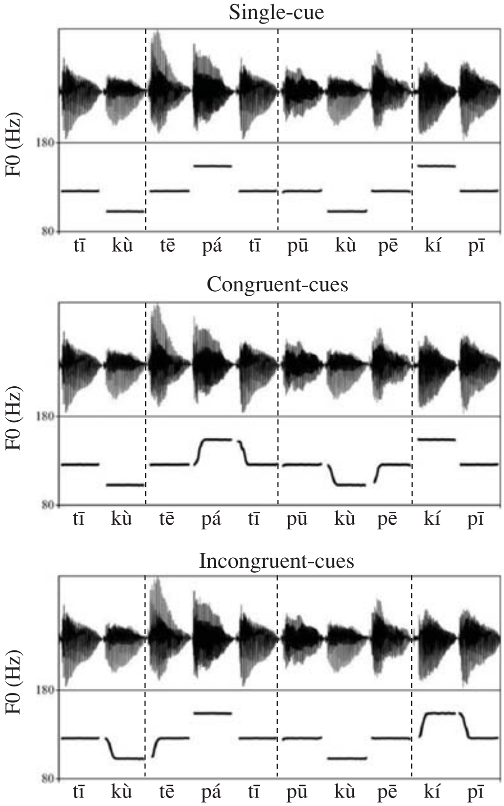

Figure 2. Samples of learning-phase speech streams under the single-cue, congruent-cues, and incongruent-cues conditions. The dashed lines indicate word boundaries.

following F0 manipulations on the selected trisyllabic sequences. First, the F0 onsets of the second and third syllables were raised or lowered to the F0 offsets of their immediately preceding syllables (i.e., the first and second syllables, respectively). Second, between the shifted F0 onset and the 25% time point into the F0 contour of the second or third syllable, a smooth F0 transition was interpolated quadratically using Praat. The erstwhile flat F0 contour of a level tone then had an F0 rise over the initial quarter of its F0 contour if its preceding tone was lower and an F0 fall if its preceding tone was higher. Note that in actual Mandarin tone production, the assimilatory effect exerted by the preceding tone can be far more extensive: for example, it may still be evident even at the 75% time point of the next tone (e.g., Xu, 1997). Manipulating only the initial 25% of the F0 contour of a syllable allowed us to evaluate the influence of tonal carryover assimilation conservatively. Shown in Figure 2 are samples of speech stream from each condition.

The test phase consisted of a two-alternative forced-choice test. In each trial, two stimuli (a word of the AL and a partword) were presented successively with 500 ms of silence in between. The stimuli did not have the tonal carryover assimilation cue and were the same for all conditions. Therefore, the conditions differed only in the

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

learning phase, specifically, in whether and how carryover assimilation was introduced to the speech streams. The orders in which the AL word and partword were presented in a trial were counterbalanced. There were 36 trials in total, yielded by pairing the six AL words exhaustively with the six partwords. E-prime 2.0 software (Psychology Software Tools, 2012) was used to control stimulus presentation and record responses.

Procedure

Participants were tested individually in front of a desktop computer in a sound-attenuated booth. They were told to learn an AL by listening to six prerecorded sound files of that language (i.e., the six learning-phase speech streams). They were not given any cues such as the length or number of the words in the AL. They were instructed to pay as much attention as possible to what they heard and made aware of an upcoming test that would assess their knowledge of the AL. They were allowed to take a short break after finishing listening to each sound file. After the learning phase, they immediately proceeded to the test, in which they heard two stimuli in a row in each trial. They were asked to select the one that they thought was a word of the AL by pressing the button on a response box that corresponded to the order of presentation of the word (i.e., button “1” or “2”). There was a 10-s response timeout after the second stimulus. Participants first completed three practice trials presenting nonsense syllable sequences not used in the AL to familiarize themselves with the procedure. They completed the practice by arbitrarily pressing any button on the response box but were reminded that they had to choose the AL words in the test proper.

Hypotheses and predictions

The single-cue condition served as the baseline for comparison with the other two conditions, based on which hypotheses regarding the effect of tonal carryover assimilation were tested. For the congruent-cues condition, the hypothesis was that tonal carryover assimilation in agreement with TPs contributes to segmentation above and beyond TPs. This predicted that listeners’ selection accuracy in the test would be significantly higher in the congruent-cues condition than in the single-cue one (as in Fernandes et al., 2007). As for the incongruent-cues condition, where the partwords received tonal carryover assimilation, the hypothesis of interest was that the conflict between the tonal cue and TPs would impede segmentation. Should this be the case, the listeners in the incongruent-cues condition would respond significantly less accurately compared with the single-cue one (much in the same way as the Italian listeners exposed to initial lengthening in Ordin et al., 2017). Findings lending support for the facilitation in the congruent-cues condition or the inhibition in the incongruent-cues condition may be interpreted as evidence for the use of tonal carryover assimilation.

Results and discussion of Experiment 1

The listeners’ responses in the test were analyzed. Timeouts (i.e., no responses within 10 s) accounted for about 0.43% of all observations and were excluded. In the remaining data, a response was coded as correct when the AL word in

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

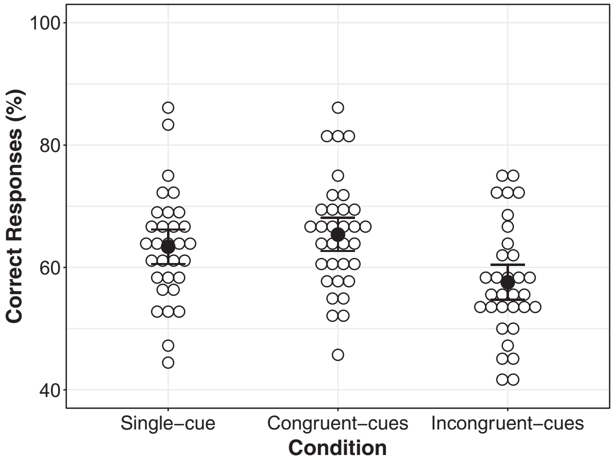

Figure 3. Percentages of correct responses of individual participants (empty circles) and the means of the single-cue, congruent-cues, and incongruent-cues conditions (filled circles). The bars represent 95% confidence intervals.

the trial was selected and as incorrect when the partword was selected. Displayed in Figure 3 are the mean percentage of correct responses of each condition along with those of individual participants. A linear mixed-effects logistic regression model was fitted to the data by using the glmer() function from the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2017) to examine the effects of the experimental conditions. The dependent variable was the response to each test trial, which was either correct or incorrect. The fixed effect of central interest was condition, with the single-cue condition being the baseline level. Two additional predictors were also entered as fixed effects to partial out their impact on responses. First, the order in which a given trial appeared in the test (trial) was included to control for possible fatigue or practice effects. Second, as recommended in Ou and Guo (2019), the log-transformed reaction time (LogRT) was included to capture any potential trade-offs between response accuracy and latency. Both trial and LogRT were centered and scaled. All the fixed-effect predictors were entered as main effects only. Following Barr, Levy, Scheepers, and Tily's (2013) recommendation, we used the maximal converging random-effects structure supported by the data. For Experiment 1, the random-effects structure consisted of a by-participant random intercept and a by-item random intercept for partwords.

Table 2 shows the results of the mixed-effects model. Trial was significant, with responses in later trials being less accurate than those in earlier trials, possibly due to fatigue. LogRT was significant as well, indicating an inverse correlation between response accuracy and latency (i.e., faster responses were more likely to be correct than slower ones). As suggested in Ou and Guo (2019), this correlation might be merely an artifact of response certainty: as the two stimuli were separated by 500 ms of silence, listeners might be ready to respond if they were sure that the first stimulus was an AL word or a partword. Most importantly, there was a significant main effect of condition, which indicated that response accuracy in the incongruent-

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

Table 2. Mixed-effects results of Experiment 1

Fixed effects		Estimate	*SE*	*z*	*p*
(Intercept)		0.563	0.166	3.395	<.001
Trial		–0.150	0.036	–4.129	<.001
LogRT		–0.231	0.039	–5.973	<.001
Condition (incongruent-cues)		–0.220	0.101	–2.178	.029
Condition (congruent-cues)		0.103	0.102	1.015	.310
Random effects		Variance	*SD*		
Participant	(Intercept)	0.037	0.192		
Partword	(Intercept)	0.134	0.366		
No. of observations:		3,441			

cues condition (mean: 57.58%) was generally lower than that in the single-cue one (mean: 63.38%). Therefore, pitting the tonal carryover assimilation cue against TP information (by letting the cue-bearing sequences span word boundaries) hinders segmentation, reducing the listeners' performance to a level below that of a condition where TPs are the sole cue. However, the other main-effect term of condition showed that response accuracy in the congruent-cues condition (mean: 65.39%), where tonal carryover assimilation occurred within word boundaries, was not significantly different from that of the single-cue one. There is thus no evidence that the listeners' segmentation is better or worse when the TP and tonal carryover assimilation cues occur in tandem and in a cooperative manner than when TP information is the only cue.

Experiment 1 examined the use of tonal carryover assimilation by Mandarin listeners in speech segmentation with an AL learning task. One proposed hypothesis predicted that relative to that of the single-cue condition, the response accuracy of the incongruent-cues condition would be significantly lower and this prediction was borne out. The finding is consistent with Fernandes et al.'s (2007) study with segmental coarticulation, in which segmentation in the incongruent-cues condition was worse than in the single-cue one. Analogous results have also been reported by AL learning research demonstrating that compared with a TP-only condition, segmentation is disrupted when an additional prosodic cue appears in a position that is unexpected in view of phonological patterns in the listeners' native language (e.g., Ordin et al., 2017). The Mandarin listeners' disrupted segmentation performance in the incongruent-cues condition of the present study may be interpreted as reflecting their attempts to use the tonal assimilation cue, even though such use is in conflict with TP regularities. Such disruption of segmentation should not be possible if the cue had not been exploited at all.

With regard to the congruent-cues condition, we hypothesized that the congruent tonal carryover assimilation cue would facilitate segmentation above and beyond the effects of TPs. The results did not support the hypothesis as there was no significant difference in response accuracy between the single-cue and congruent-cues conditions. This is not consistent with Fernandes et al. (2007), who did find that with segmental coarticulation as the additional non-TP cue, segmentation under the congruent-cues condition was clearly better than under the single-cue one. Rather, it seems compatible with an alternative view: adding congruent tonal carryover assimilation would not facilitate segmentation because word boundaries are redundantly cued by the tonal and statistical information. Higher TPs between adjacent syllables within an AL word already signal that a boundary between them is unlikely and there might be no need for conveying similar information via tonal coarticulation. Such a cue redundancy hypothesis may lead one to expect no significant difference in the listeners' accuracy between the single-cue and congruent-cues conditions (as in the case with the final lengthening and F0 rise in Bagou & Frauenfelder, 2018; Kim et al., 2012).

However, the lack of a significant effect of cue congruence could potentially be attributed to a confounding factor: cue reliability. Recall that only 27% of the AL word tokens in the congruent-cues condition carried tonal carryover assimilation. This percentage was rather low considering the percentage of bitonal sequences that showed a tonal coarticulatory effect (i.e., about 51%) as reported by Hao et al. (2018) for Mandarin corpus speech. Therefore, an alternative view was that our listeners were unable to benefit from tonal carryover assimilation under the congruent-cues condition simply because it was not reliably present, not because it was redundant with TP information. Specifically, as cue reliability is associated with cue weight (e.g., Mattys et al., 2005; Seidl, 2007; Tremblay, Spinelli, Coughlin, & Namjoshi, 2018), they might have allocated a relatively low weight to the unreliably present tonal carryover assimilation cue. This is not entirely impossible given that, however prevalent tonal coarticulation is in Mandarin, the listeners were instructed to learn a completely novel AL and they might develop a cue hierarchy for that AL, one in which tonal carryover assimilation was so lowly weighted that it failed to produce any extra gain in segmentation performance.

The main goal of Experiment 2 is then to test this cue reliability hypothesis, that is, to examine whether enhancing the reliability of the tonal carryover assimilation cue in the congruent-cues condition would facilitate segmentation and provide an alternative explanation for the null effect in Experiment 1. Currently, it is unclear as to how many cue-bearing tokens would count as sufficient for obtaining the kind of facilitation effects reported by Fernandes et al. (2007), who did not provide the exact percentage of segmentally coarticulated tokens in their congruent-cues condition. Yet, it would be insightful to test an experimental condition in which the reliability of the tonal carryover assimilation cue is maximized, namely, one in which all the tokens of the AL words receive the cue. Such a condition was included in Experiment 2. In addition, Experiment 2 included the same single-cue and incongruent-cues conditions from Experiment 1. This was done to ensure that participants had been randomly assigned to different conditions while the conditions were being compared and to examine whether the findings of Experiment 1 could be replicated.

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

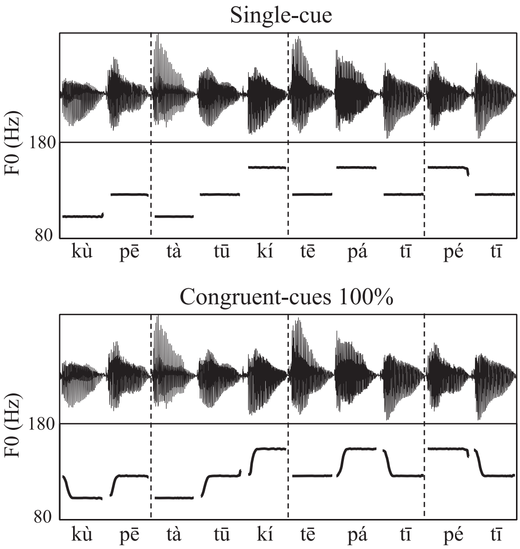

Figure 4. A sample of learning-phase speech streams under the congruent-cues 100% condition (lower panel). The corresponding portion in the single-cue condition is included for comparison. The dashed lines indicate word boundaries.

Experiment 2

Participants

Ninety new adult native listeners of Mandarin (35 males and 55 females) from the same population as those in Experiment 1 were recruited. Their mean age in years was 20.7 (range: 18–25; standard deviation: 2.0) and 15 of them had received musical training. They were randomly and equally assigned to the single-cue, congruent-cues 100%, and incongruent-cues conditions. None participated in Experiment 1.

Design and stimuli

The overall design of Experiment 2 was as in Experiment 1. The difference was that for the congruent-cues condition, all instances of the AL words (not just 27%) in the learning-phase speech streams received the tonal carryover assimilation cue, as shown in the lower panel of Figure 4. We refer to this condition as “congruent-cues 100%.” The test-phase stimuli were the same as in Experiment 1.

Procedure

The procedure was identical to that of Experiment 1.

Hypothesis and prediction

If it was cue (un)reliability that underpinned the lack of a significant effect of cue congruence in the previous experiment, it was hypothesized that the tonal carryover

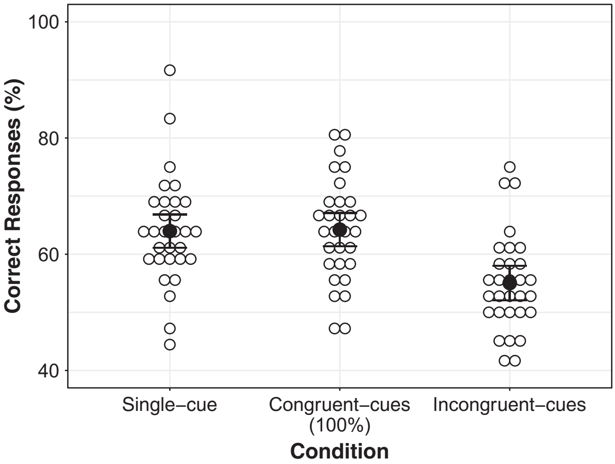

Figure 5. Percentages of correct responses of individual participants (empty circles) and the means of the single-cue, congruent-cues 100%, and incongruent-cues conditions (filled circles). The bars represent 95% confidence intervals.

assimilation cue would be effectively exploited if it was always present in the AL words. This predicted that listeners' response accuracy in the test would be significantly higher in the congruent-cues 100% condition than in the single-cue one.

Results and discussion of Experiment 2

Again, responses in the test phase were analyzed. Timeouts (less than 0.13% of the data) were discarded, and a response was correct when the AL word was selected and incorrect when the partword was selected. Shown in Figure 5 are the percentage of correct selections of each participant and the mean of each condition. A linear mixed-effects logistic regression model with the same fixed-effects was fitted; that is, it contained the main effects of condition (baseline: single-cue), trial, and LogRT. As in Experiment 1, the maximal converging random-effects structure was used. This time it included random intercepts for participants, AL words, and partwords.

The results of the analysis are in Table 3. The effects of trial and LogRT were significant and in the same direction as in Experiment 1. Response accuracy was lower in later trials, possibly because of fatigue; faster responses tended to be correct and this might be an artifact of response certainty. Most crucial were the two condition main effects. First, the response accuracy in the incongruent-cues condition (mean: 55.05%) was significantly lower than that in the single-cue one (mean: 63.97%), as in Experiment 1. Second, and more importantly, the response accuracy in the congruent-cues 100% condition (mean: 64.23%) was neither significantly better nor worse than that in the single-cue one, again suggesting no evidence that the congruence between statistical and tonal coarticulatory cues would enhance segmentation performance.

The primary goal of Experiment 2 was to test the cue reliability hypothesis proposed above. In particular, it investigated whether Mandarin listeners' segmentation

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

Table 3. Mixed-effects results of Experiment 2

Fixed effects		Estimate	*SE*	*z*	*p*
(Intercept)		0.558	0.113	4.944	<.001
Trial		–0.103	0.037	–2.793	.005
LogRT		–0.220	0.041	–5.432	<.001
Condition (incongruent-cues)		–0.293	0.101	–2.893	.004
Condition (congruent-cues 100%)		0.021	0.102	0.207	.8361
Random effects		Variance	*SD*		
Participant	(Intercept)	0.030	0.174		
Word	(Intercept)	0.014	0.119		
Partword	(Intercept)	0.031	0.177		
No. of observations:		3,235			

would be facilitated under a condition in which all occurrences of the words in the AL speech streams carried the tonal carryover assimilation cue, as compared with the single-cue condition. The results provided no support for the hypothesis. Furthermore, the experiment replicated the inhibitory effect of cue incongruence. In the General Discussion section below, we will summarize the results from the two experiments, consider some possible explanations, and offer a unified account for the findings.

General Discussion

This study investigates the role of tonal coarticulation in the segmentation of continuous speech. Previous work has demonstrated that subtle acoustic–phonetic details are exploited in solving segmentation problems (e.g., Cho et al., 2007; Gow & Gordon, 1995) and can even modulate the use of segmentation strategies motivated by the distribution of a phonological entity (e.g., Ou & Guo, 2019; Tremblay et al., 2019). To extend this line of work, we examined the effect of tonal carryover assimilation, a cross-linguistically attested type of tonal coarticulation, on Mandarin listeners' segmentation with two AL learning experiments. The findings are summarized in Table 4. Experiment 1 revealed that response accuracy in the test was significantly lower in the incongruent-cues condition, where tonal carryover assimilation was pitted against TPs, than in the single-cue condition, where TP information was the only segmentation cue. This reflects the listeners' attempts to exploit the tonal cue despite its conflict with TPs. Yet, in the congruent-cues condition, where the tonal cue and TPs agreed with each other, response accuracy did not differ significantly from that of the single-cue condition. This seems consistent with the view that congruent tonal carryover assimilation is unable to improve segmentation further because it is redundant in the presence of TPs. Experiment 2

Table 4. Summary of the findings

Conditions with tonal carryover assimilation cue		Effect on segmentation relative to single-cue condition
Experiment 1	Congruent-cues	No facilitation or inhibition
	Incongruent-cues	Inhibition
Experiment 2	Congruent-cues 100%	No facilitation or inhibition
	Incongruent-cues	Inhibition

examined Mandarin listeners' segmentation in the congruent-cues 100% condition, where tonal carryover assimilation in agreement with TPs was introduced to all instances of the AL words. The results again showed no significant accuracy difference between this condition and the single-cue one, discounting the possibility that the null effect of cue congruence can be attributed to cue reliability. Experiment 2 also replicated the inhibitory effect of cue incongruence. Below we explore a few possible explanations for the findings, discuss their implications for tonal coarticulation as a fine-grained acoustic–phonetic cue in speech segmentation, and point out some further issues.

One noteworthy finding from Experiment 1 was the null effect of cue congruence. It was found again in Experiment 2, ruling out the cue reliability hypothesis. Yet, one might wonder whether the null effect is simply a methodological artifact stemming from the stimuli used in the test phase. Recall that the test stimuli did not contain the carryover assimilation cue and were identical across the conditions. It could be argued that the listeners did not show significantly higher accuracy in the two congruent-cues conditions due to mismatch between the learned representations and the test stimuli. That is, the representations of the AL words that they built up during the learning phase contained the tonal coarticulatory cue and somewhat deviated from the AL words actually presented in the test. This possibility gains support from episodic or exemplar-based theories of phonology and speech perception (Bybee, 2000; Goldinger, 1996; Goldinger & Azuma, 2004; Johnson, 1997; Pierrehumbert, 2001). These theories hold that representations of categories are constructed from remembered instances, or exemplars, of those categories. Detailed acoustic–phonetic traces associated with these exemplars may be incidentally encoded in memory and affect speech processing, even though they are not crucial for category distinctions. For example, words and sentences presented before are recognized slower or less accurately when they are presented again in a novel voice than in a familiar voice, suggesting that speaker voice information is retained in memory (Craik & Kirsner, 1974; Geiselman & Bellezza, 1977; Palmeri, Goldinger, & Pisoni, 1993). A similar case could possibly be made for the listeners in our study: they did not benefit from cue congruence because all or at least some exemplars in their AL word representations contained tonal coarticulatory distortions, preventing them from effectively recognizing the stimuli in the test.

Such an exemplar-based explanation, however, is untenable in view of the present results for two reasons. First, if the listeners' responses were driven by exemplars, one would expect their test accuracy to be significantly higher in the single-cue

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

condition than in any other condition as the tokens of the AL words presented in the learning phase of the single-cue condition were acoustically identical to those in the test. This expectation was not borne out. Second, several AL learning studies (e.g., Fernandes et al., 2007; Kim et al., 2012; Ordin & Nespor, 2016; Tremblay et al., 2019) also presented "uncued" stimuli (i.e., stimuli that did not contain the cues of interest) in the test and thus the test AL words were always the same as those in the learning phase of their single-cue or TP-only conditions. None of them reported that listeners performed the best in these conditions.

What seems to be a more plausible explanation for the null effect in Experiments 1 and 2 is one of cue redundancy. In the two congruent-cues conditions, TPs between syllables already signal the presence or absence of boundaries and similar information is redundantly provided by tonal carryover assimilation. The consequence is that the congruence between the TP and tonal cues does not yield an extra gain in segmentation performance for Mandarin listeners. Comparable findings have been reported in previous studies with F0 and lengthening cues (e.g., Bagou & Frauenfelder, 2018; Kim et al., 2012). Nevertheless, the redundancy hypothesis needs to be reconciled with the findings of Fernandes et al. (2007), who show that, as mentioned, segmental coarticulation in agreement with TPs leads to extra facilitation when compared with TP information alone. In their study, cue redundancy does not seem to detract from the efficacy of segmental coarticulation as a useful segmentation cue.

Before suggesting how this discrepancy may be accommodated, it should be noted that tonal and segmental coarticulation are possibly rather different in nature, and this needs to be borne in mind in directly comparing the two phenomena. Although there are no prior studies analyzing how they would differentially affect segmentation, production experiments have revealed some differences between the two. For example, while tonal coarticulation is restricted to two contiguous tones (as in our AL), segmental coarticulatory effects can extend up to four segments (Shen, 1990). The scope of segmental coarticulation was not controlled in Fernandes et al. (2007). As a result, direct comparisons between tonal and segmental coarticulation would be the most appropriate only when factors like this are kept equal or systematically manipulated. Yet, an explanation for the discrepant findings on the tonal and segmental coarticulation may offer insight for rethinking previous models of speech segmentation cues.

We assume that redundancy affects tonal coarticulation but not segmental coarticulation in the presence of TPs possibly because tonal information is a less powerful cue than segmental information in speech segmentation, at least for Mandarin listeners. It has been shown that for listeners of Mandarin or other lexical tone languages, tonal cues are relatively disadvantaged compared with segmental cues as the former become available at later stages of auditory processing (Cutler & Chen, 1997; Sereno & Lee, 2015; Taft & Chen, 1992; Ye & Connine, 1999). Support for this comes from, for example, Sereno and Lee (2015), whose auditory priming and lexical decision experiments found that the prime facilitated recognition of the target when the two had overlapping segments but mismatching tone. By contrast, no facilitation was found when the prime and targets had overlapping tone but mismatching segments. Tong, Francis, and Gandour (2008) present converging evidence from Mandarin listeners and offer an explanation from an informationtheoretic perspective. They argue that the tone disadvantage may be attributed to the fact that the tonal inventory is smaller than the segmental inventory. As a result, tones exert fewer constraints on lexical access and are less informative. Due to their experience with the weaker role of tonal information in speech processing in general, listeners may have allocated a relatively low weight to tonal segmentation cues such as tonal carryover assimilation. Such cues may be ignored when the boundary information they provide is redundant.

The idea that tonal coarticulation is weighted low or at least lower than segmental coarticulation suggests a few refinements to the current understanding of segmentation cue weight. Based on a series of experiments, Mattys et al. (2005) propose a three-tier segmentation cue hierarchy in which lexical cues are top-ranked, followed by segmental cues such as segmental coarticulation and then by prosodic cues such as word stress. It is not clear how tonal coarticulation can fit into this framework. On the one hand, tonal coarticulation is neither a segmental phenomenon nor a phonological or metrical feature like word stress. On the other hand, as with segmental coarticulation, it is a fine-grained acoustic–phonetic detail, and it involves variations in F0, which is also an acoustic correlate of stress (e.g., Beckman, 1986; Fry, 1958; Gay, 1978; Lieberman, 1960). Tonal coarticulation, which seems to have a lower weight than segmental coarticulation, may be accommodated by expanding the hierarchy with an additional cue category between the segmental and prosody tiers. Alternatively, it can be subsumed under the prosody category, but it has to be noted that the prosodic cue in Mattys et al.'s model currently refers to lexical stress and evidence for its relative importance comes only from experimentation with stress in English. Further investigation can be conducted to pinpoint where tonal coarticulation stands in the cue hierarchy. Yet, as far as the present findings are concerned, some modifications to current models of cue weight or ranking may be necessary, especially if they are to be adapted to account for the segmentation behavior of tone-language listeners.

Although the tonal carryover assimilation cue, unlike the segmental coarticulatory one in Fernandes et al. (2007), did not significantly improve segmentation when congruent with TPs, it did have an appreciable effect in the incongruent-cues conditions of the two experiments. In these conditions, tonal carryover assimilation favored the segmentation of the AL speech streams into what were defined as part-words, which contained a dip in TP (as they spanned an AL word boundary). This resulted in a conflict between prosodic and statistical information and hence hampered the listeners' segmentation. It is argued that the mechanism that drives the listeners to use tonal carryover assimilation might be one akin to the prosody analyzer proposed by Cho et al. (2007). The prosody analyzer computes the prosodic structure of an utterance using available suprasegmental information and generates possible segmentation hypotheses. In the case of the present study, the tonal carryover assimilatory effect of one syllable on the next syllable may be analyzed by a similar mechanism as signaling continuity or the lack of a prosodic boundary between the two syllables, therefore leading the listeners to perceive the two syllables as belonging to a unit. Segmentation is disrupted when the boundaries of prosodically and statistically defined units do not align. Similar cases of disruption have been reported in previous studies examining the effects of prosodic grouping and statistical regularities. For example, Shukla, Nespor, and Mehler (2007) exposed

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

listeners to recurrent nonsense words presented in continuous speech with recurrent F0 frames that spanned several syllables. In a subsequent test, it was found that these nonsense words were recognized more poorly if they previously straddled the boundary of two F0 frames than if they did not.

It is concluded that tonal carryover assimilation can be exploited as a speech segmentation cue at least by Mandarin listeners. The use may be the result of a segmentation mechanism similar to the prosody analyzer, one that interprets the cue as signaling the absence of a prosodic boundary. An issue for further exploration concerns the extent to which such a mechanism is cross-linguistic. Tonal carryover assimilation was chosen for investigation because carryover effects tend to be assimilatory across languages (Zhang & Liu, 2011). Yet, there are exceptions. One of them is Malaysian Hokkien. Chang and Hsieh (2012) elicited productions of disyllabic words by speakers of this language and found that the carryover effect of the first tone on the second one was not assimilatory and even slightly dissimilatory. They attribute this to the final prominence in Malaysian Hokkien tone sandhi system, which requires the tone in the final position to be faithfully realized and militates against coarticulatory distortions by the preceding tone. Analogous findings have been reported by Chen, Wiltshire, and Li (2018) for Nanjin Chinese. Thus, it is possible that tonal carryover assimilation would be not be used by Malaysian Hokkien and Nanjin Chinese listeners, whether it is congruent with TPs or not. Support for such a possibility would suggest that the use of some putatively cross-linguistic acoustic–phonetic cues can be overridden by language-specific phonology. There has been some evidence for this. For example, while final lengthening has been thought to be universal and cross-linguistically useful for segmentation (e.g., Hay & Diehl, 2007; Klatt, 1975; Lindblom, 1978; Oller, 1973; Tyler & Cutler, 2009), recent evidence shows that Italian listeners exploit medial but not final lengthening presumably because stress in Italian falls predominantly on the penultimate syllable of a word (Ordin et al., 2017). Given that the use of a segmentation strategy motivated by phonological patterns is affected by subtle acoustic–phonetic details (e.g., Ou & Guo, 2019; Tremblay et al., 2019), it may be of interest to further examine the influence in the opposite direction, that is, how segmentation solutions motivated by fine-grained acoustic–phonetic information are constrained by language-specific phonology.

Still, it has to be recognized that as far as the current findings are concerned, the contribution of tonal carryover assimilation to segmentation is somewhat limited. It is not observed in the case of cue congruence and, as discussed, this might be linked to cue redundancy and the relatively minor role of tonal information in speech processing. Further issues also arise as to whether the limited contribution of tonal carryover assimilation can also be attributed to methodological factors, such as the fact that only the initial 25% of a tone's F0 contour was manipulated. Such manipulation allowed us to evaluate the impact of tonal carryover assimilation conservatively but might also lead us to underestimate it as the carryover assimilatory effect in naturalistic speech can be more extensive (e.g., Xu, 1997). In addition, despite being commonly adopted, AL learning is not the only experimental technique for studying speech segmentation. The AL learning task assesses segmentation hypotheses about a nonsense language using an offline forced-choice test. It would be insightful to further examine whether tonal carryover assimilation would have a more salient

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

effect in the online segmentation of real meaningful speech, which can be captured by using time-sensitive measures such as eye-tracking fixations (e.g., Tremblay et al., 2018) or lexical decision latencies (e.g., Gow & Gordon, 1995; Mattys et al., 2005).

Conclusion

The present work aims to add to the understanding of the role of fine-grained acoustic–phonetic information in speech segmentation by examining how tonal carryover assimilation is used by Mandarin listeners in segmenting continuous speech streams of an AL. Experiments 1 and 2 found that their segmentation performance was hampered in the incongruent-cues conditions, suggesting that the tonal carryover assimilation cue was exploited despite its conflict with TPs. This might reflect a segmentation mechanism that analyzes the assimilatory effect of one tone on the next tone as cueing the absence of a boundary. However, the contribution of such an effect may be limited as Experiment 1 revealed that tonal carryover assimilation did not facilitate segmentation when agreeing with TPs. This finding cannot be attributed to cue reliability, as Experiment 2 suggests. Nor can it be accounted for by an exemplar-based view of phonological representations. It is assumed that tonal carryover assimilation is redundant in the presence of congruent statistical cues, and the discrepancy with previous studies with segmental coarticulation may be linked to the relatively lower weight of tonal information in speech processing. Further work can be done to investigate whether the current findings would be replicated cross-linguistically and to experiment with different acoustic implementations of tonal carryover assimilation and paradigms other than AL learning.

Acknowledgments. We would like to thank the associate editor Miquel Simonet and four anonymous reviewers for many thoughtful comments and suggestions that have helped us significantly improve this paper. Any errors that remain are entirely our own. We would also like to thank Kai-ting Chuang, for her assistance with participant recruitment. This work was partially supported by Grant Number 109-2410-H-110-069-MY3 from the Ministry of Science and Technology, Taiwan, to the second author.

Note

[1] Musical background has been widely reported to confer an advantage in lexical tone identification and pitch discrimination (e.g., Delogu, Lampis, & Belardinelli, 2010; Lee & Hung, 2008; Xie & Myers, 2015). It is thus possible that the musically trained participants might exploit the tonal coarticulatory cue more effectively (or at least differently). To test this, we conducted mixed-effects analyses with musical training and its interactions with condition also included as fixed effects. Results indicated that for both experiments, none of these newly added fixed effects was significant and the patterns of statistical significances of the other fixed effects remained the same. There was no evidence that the participants with musical training behaved differently.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, **9**, 321–324.

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**, 255–278.

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48.

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer (Version 6.0.37) [Computer program]. Retrieved from http://www.praat.org/

Brunelle, M. (2009). Northern and Southern Vietnamese tone coarticulation: A comparative case study. *Journal of Southeast Asian Linguistics*, **1**, 49–62.

Bybee, J. L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-Based models of language* (pp. 65–85). Stanford, CA: CSLI Publications.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, **24**, 209–244.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, **26**, 173–199.

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

Chang, Y.-C., & Hsieh, F.-F. (2012). Tonal coarticulation in Malaysian Hokkien: A typological anomaly? *Linguistic Review*, **29**, 37–73.

Chen, S., Wiltshire, C., & Li, B. (2018). An updated typology of tonal coarticulation properties. *Taiwan Journal of Linguistics*, **16**, 79–114.

Chen, Y. (2012). Tonal variation. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 103–114). Oxford: Oxford University Press.

Cheng, R. L. (1968). Tone sandhi in Taiwanese. *Linguistics*, **6**, 19–42.

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, **32**, 141–176.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, **29**, 155–190.

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

Craik, F. I., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, **26**, 274–284.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 105–121). Cambridge, MA: MIT Press.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge: MIT Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, **31**, 218–236.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, **2**, 133–142.

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, **59**, 165–179.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 113–121.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **28**, 218–244.

Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, **22**, 46–61.

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, **18**, 254–260.

Fernandes, T., Kolinsky, R., & Ventura, P. (2010). The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, & Psychophysics*, **72**, 1522–1532.

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception and Psychophysics*, **69**, 856–864.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, **101**, 3728–3740.

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

Gandour, J., Potisuk, S., & Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics*, **22**, 474–492.

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

Geiselman, R. E., & Bellezza, F. S. (1977). Incidental retention of speaker's voice. *Memory & Cognition*, **5**, 658–665.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22**, 1166–1183.

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review*, **11**, 716–722.

Gow, D. W., Jr., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 344–359.

Hao, L., Zhang, W., Xie, Y., & Zhang, J. (2018). A preliminary study on tonal coarticulation in continuous speech. In *Proceedings of INTERSPEECH* (pp. 3017–3021), Hyderabad, India, September 2–6, 2018.

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

Hay, J. S. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, **63**, 93–106.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.

Jun, S. A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, **15**, 189–226.

Karlsson, F. (1983). *Suomen kielen äänne—ja muotorakenne [Finnish phonological and morphological structure]*. Helsinki, WSOY.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. S. (2004). Domain-Initial articulatory strengthening in four languages. *Phonetic Interpretation: Papers in Laboratory Phonology*, **6**, 143–161.

Kim, S., Broersma, M., & Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language. *Studies in Second Language Acquisition*, **34**, 415–444.

Kim, S., & Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean. *Journal of the Acoustical Society of America*, **125**, 3373–3386.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, 129–140.

Lai, W., & Kuang, J. (2016). Prosodic grouping in Chinese trisyllabic structures by multiple cues–tone coarticulation, tone sandhi and consonant lenition. *In Proceedings of Tonal Aspects of Languages* **2016** (pp. 157–161). Buffalo, NY, May 24–27, 2016.

Lee, C.-Y., & Hung, T.-H. (2008). Identification of Mandarin tones by English-Speaking musicians and nonmusicians. *Journal of the Acoustical Society of America*, **124**, 3235–3248.

Li, Y, Lee, T, & Qian, Y. (2004). Analysis and modeling of F0 contours for Cantonese text-to-speech. *ACM Transactions on Asian Language Information Processing*, **3**, 169–180.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, **32**, 451–454.

Lindblom, B. (1978). Final lengthening in speech and music. In E. Gårding, G. Bruce, & R. Bannert (Eds.), *Final lengthening in speech and music* (pp. 85–100). Lund, Sweden: Lund University.

Mattys, S. L., & Bortfeld, H. (2017). Speech segmentation. In M. G. Gaskell & J. Mirković (Eds.), *Speech perception and spoken word recognition* (pp. 55–75). London: Routledge.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, **134**, 477–500.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, **39**, 21–46.

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, **108**, 271–280.

https://doi.org/10.1017/S0142716420000818 Published online by Cambridge University Press

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, **54**, 1235–1247.

Ordin, M., & Nespor, M. (2016). Native language influence in the segmentation of a novel language. *Language Learning and Development*, **12**, 461–481.

Ordin, M., Polyanskaya, L., Laka, I., & Nespor, M. (2017). Cross-Linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory & Cognition*, **45**, 863–876.

Ou, S.-C., & Guo, Z.-C. (2019). The language-specific use of F0 rise in segmentation of an artificial language: Evidence from listeners of Taiwanese Southern Min. *Language and Speech*. Advance online publication. doi: 10.1177/0023830919886604

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 309–328.

Peng, S.-H. (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics*, **25**, 371–400.

Pierrehumbert, J. (2001) Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam: Benjamins.

Potisuk, S., Gandour, J., & Harper, M. P. (1997). Contextual variations in trisyllabic sequences of Thai tones. *Phonetica*, **54**, 22–42.

Psychology Software Tools. (2012). *E-Prime 2.0*. Pittsburgh, PA: Author.

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, **35**, 606–621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, **8**, 101–105.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, **57**, 24–48.

Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, **58**, 131–151.

Shen, X.-S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics*, **18**, 281–295.

Shih, C.-L. (1988). Tone and intonation in Mandarin. Working Papers of the Cornell Phonetics *Laboratory*, **3**, 83–109.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, **54**, 1–32.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, **36**, 422–444.

Taft, M., & Chen, H.-C. (1992). Judging homophony in Chinese: The influence of tones. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp.151–172). Amsterdam: Elsevier.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, **3**, 73–100.

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, **23**, 689–708.

Toro, J. M., Pons, F., Bion, R. A., & Sebastián-Gallés, N. (2011). The contribution of language-specific knowledge in the selection of statistically-coherent word candidates. *Journal of Memory and Language*, **64**, 171–180.

Tremblay, A., Cho, T., Kim, S., & Shin, S. (2019). Phonetic and phonological effects of tonal information in the segmentation of Korean speech: An artificial-language segmentation study. *Applied Psycholinguistics*, **40**, 1221–1240.

Tremblay, A., Spinelli, E., Coughlin, C. E., & Namjoshi, J. (2018). Syntactic cues take precedence over distributional cues in native and non-native speech segmentation. *Language and Speech*, **61**, 615–631.

Tyler, M. D., & Cutler, A. (2009). Cross-Language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America*, **126**, 367–376.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, **38**, 133–149.

Wang, H. S. (2002). The prosodic effects on Taiwan Min tones. *Language and Linguistics*, **3**, 839–852.

White, L., Melhorn, J. F., & Mattys, S. L. (2010). Segmentation by lexical subtraction in Hungarian speakers of second-language English. *Quarterly Journal of Experimental Psychology*, **63**, 544–554.

Xie, X., & Myers, E. (2015). The impact of musical training and tone language experience on talker identification. *Journal of the Acoustical Society of America*, **137**, 419–432.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, **95**, 2240–2253.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, **25**, 61–83.

Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics*, **17**, 1–31.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, **14**, 609–630.

Zhang, J., & Liu, J. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica*, **68**, 161–191.

Zhang, J.-S., & Kawanami, H. (1999). Modeling carryover and anticipation effects for Chinese tone recognition. In Proceedings of the Sixth European Conference on Speech Communication and Technology (pp. 747–750), Budapest, Hungary, September 5–9, 1999.

Cite this article: Guo, Z.-C. and Ou, S.-C. (2021). The use of tonal coarticulation in segmentation of artificial language speech: A study with Mandarin listeners. *Applied Psycholinguistics* **42**, 631–655. https://doi.org/10.1017/S0142716420000818