Cognition 202 (2020) 104311

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Universals of listening: Equivalent prosodic entrainment in tone and non-tone languages

Martin Ho Kwan Ip[a,b,*], Anne Cutler[a,b]

[a] The MARCS Institute, Western Sydney University, Australia
[b] ARC Centre of Excellence for the Dynamics of Language, Australia

ARTICLE INFO

Keywords:
Prosody
Prosodic entrainment
Focus
Speech perception
Cross-linguistic comparisons

ABSTRACT

In English and Dutch, listeners entrain to prosodic contours to predict where focus will fall in an utterance. Here, we ask whether this strategy is universally available, even in languages with very different phonological systems (e.g., tone versus non-tone languages). In a phoneme detection experiment, we examined whether prosodic entrainment also occurs in Mandarin Chinese, a tone language, where the use of various suprasegmental cues to lexical identity may take precedence over their use in salience. Consistent with the results from Germanic languages, response times were facilitated when preceding intonation predicted high stress on the target-bearing word, and the lexical tone of the target word (i.e., rising versus falling) did not affect the Mandarin listeners' response. Further, the extent to which prosodic entrainment was used to detect the target phoneme was the same in both English and Mandarin listeners. Nevertheless, native Mandarin speakers did not adopt an entrainment strategy when the sentences were presented in English, consistent with the suggestion that L2 listening may be strained by additional functional load from prosodic processing. These findings have implications for how universal and language-specific mechanisms interact in the perception of focus structure in everyday discourse.

The speech stream is a continual cascade of information, from the physical properties of the speech sounds to the sequencing of words and the discourse context. To anticipate the likely continuation, listeners must constantly build up knowledge about the incoming signal by attending to cues from different parts of the language structure (Norris, McQueen, & Cutler, 2000). In the segmental domain, considerable research over the past decades has revealed both universal and language-specific mechanisms in speech perception. For example, across languages with differing phonological structures, there is evidence that listeners can use the same strategies to recognise words by tracking information based on their syllabic structure (e.g., Sonority Sequencing Principle: Gómez et al., 2014) or patterning of vowels and consonants (e.g., Possible Word Constraint: Brent & Cartwright, 1996; Cutler, Demuth, & McQueen, 2002; Norris, McQueen, Cutler, & Butterfield, 1997). At the same time, it is also well known that listeners are sensitive to language-specific features such as the transitional probabilities between syllables (Saffran, Aslin, & Newport, 1996), coarticulatory word-onset variations (Davis, Marslen-Wilson, & Gaskell, 2002), and phonotactic or allophonic regularities (Christiansen, Allen, & Seidenberg, 1998; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; McQueen, 1998; Vitevitch & Luce, 1999). Likewise, knowledge-based processing from higher-level domains (e.g., syntax, semantics) has also been shown to support perception of word boundaries (Gaskell & Marslen-Wilson, 1997; Mattys, Melhorn, & White, 2007), phoneme restoration (Samuel, 2001), and lexical selection and disambiguation (Altmann & Kamide, 1999; Seidenberg, Tanenhaus, Leiman, & Bienkowsky, 1982).

However, much less research has examined the role of prosodic prominence relations in sentence processing. Conversations between people can only occur if both speakers and listeners share a common understanding regarding some information about the world, and one way in which prosodic highlighting can facilitate communication is by conveying the speaker's state of mind through the focus structure, or the “information packaging” (Chafe, 1976), of the utterance. Speakers rarely assign equal acoustic weight to each word in the sentence; words with different discourse status (e.g., focus versus background) can be produced with different degrees of prosodic prominence to express the utterance semantic structure. In this way, even segmentally identical sentences can have different implications depending on how certain words are produced; as illustrated in (1), where “poodle” is prosodically highlighted to show that the new information being conveyed is about the Archduke's poodles, and not some other dog breed, compared to (2),

* Corresponding author at: Integrated Language Sciences and Technology (ILST), MindCORE, University of Pennsylvania, 3401-C Walnut Street, Suite 300, Philadelphia, PA 19104, United States of America.
E-mail address: mhkip@sas.upenn.edu (M.H.K. Ip).

https://doi.org/10.1016/j.cognition.2020.104311
Received 7 January 2019; Received in revised form 12 April 2020; Accepted 18 April 2020

where it is deaccented and the prosodic emphasis occurs later in the sentence. It is therefore important for listeners to identify both the location and features of different prosodic cues in order to understand the intended message.

(1) I was quite shocked to see the Archduke's POODLES eating

truffles for lunch.

(2) I was quite shocked to see the Archduke's poodles eating

TRUFFLES for lunch.

Prosodically highlighted words can speed up the sentence comprehension process, in part because the phonetic features of these words play an important role in perception. In English, for instance, where more than 60% of spoken words deviate from their citation form in at least one segment (Johnson, 2004), stressed syllables of focused words are realised with longer vowel duration, higher relative pitch, and greater peak amplitude and spectral clarity (e.g., de Jong, 2004; Lehiste, 1970; Sluijter & van Heuven, 1996). Conversely, unfocused words tend to have shorter duration, more centralised vowels, and lower pitch and intensity. These prosodic differences can be found across many languages where they serve a communicative function in allowing focused words to stand out from the background elements and making them clearer and easier to understand (e.g., Lieberman, 1963; Mattys & Samuel, 2000). Indeed, behavioural and ERP studies from various languages have shown that prosodic focus marking can provide many listening advantages. Prosodically highlighted words are recognised more rapidly and accurately (e.g., English: Cutler & Foss, 1977; McAllister, 1991; Japanese: Lee, Chiu, & Xu, 2017) and are processed more deeply in lexical activation (e.g., French: Brunellière, Auran, & Delrue, 2019; Mandarin Chinese: Li & Ren, 2012; English: Blutner & Sommer, 1988; Norris, Cutler, McQueen, & Butterfield, 2006). Accent can also speed up sentence comprehension, facilitate word learning, support processing of contextual alternatives, and help listeners identify different elements of the discourse structure (Dutch: Braun & Tagliapietra, 2010; English: Birch & Clifton, 1995; Dahan, Tanenhaus, & Chambers, 2002; Fowler & Housum, 1987; German: Braun, Asano, & Dehé, 2018; Gotzner, Spalek, & Wartenburger, 2013; Grassmann & Tomasello, 2007, 2010; Mandarin: Hsu, Evans, & Lee, 2015; Yan & Calhoun, 2019; Russian: Kushch, Igualada, & Prieto, 2018). In addition, cross-linguistic comparisons between typologically unrelated languages (e.g., English and Korean: Kember, Choi, Yu, & Cutler, 2019) have revealed better recognition memory for prosodically focused words (see also, Birch & Garnsey, 1995; Fraundorf, Watson, & Benjamin, 2010). All these findings indicate that prosodic focus may have similar processing effects across languages.

What is less clear, however, is whether there is also a common strategy that all listeners can use to forecast the location of a prosodically focused word, even before it is uttered. For Germanic languages (e.g., English and Dutch), Cutler and colleagues have discovered that listeners can anticipate an upcoming accented word by entraining to the ongoing utterance intonation contour (Akker & Cutler, 2003; Cutler, 1976; Cutler & Darwin, 1981; Cutler & Fodor, 1979). In a phoneme detection task, participants listened to a series of sentences in their native language and responded as fast as they could to words that began with a specified phoneme target (e.g., responded as soon as they heard the sound /d/ in "duck"). Listeners responded faster to the target phoneme in sentences where the preceding intonation contour predicted high stress on the target-bearing word, compared to sentences where the intonation predicted low stress. This response time advantage for sentences with predicted high stress contexts held even when the original target-bearing words in each context were replaced by an acoustically identical neutral version of the same word. Since the only difference was in the preceding intonation, it was concluded that listeners could attend to the preceding prosodic contour and entrain to it to predict the location of an upcoming focused word; their attention to the contour allowed them to *be transported along with it* to anticipate the prosodic form of an upcoming word.

Similar prosodic entrainment strategies have also been observed in prediction of upcoming lexical forms. For example, Dilley and Pitt (2010) found that listeners can use contextual speech rate cues to predict the presence or absence of heavily coarticulated function words. Dilley and Pitt presented native English listeners with sentences containing a spectrally reduced function word, and manipulated the speech rate of the preceding prosody (e.g., *or* from *minor or* [maɪnɚː] in "Anyone must be a minor or child…"). Compared to sentences with normal speech rate, listeners were less likely to detect the function word when the preceding context was slowed, even though the target words were acoustically identical in both contexts. Conversely, speeding the speech rate caused listeners to hallucinate hearing a function word that was never spoken (e.g., *a* in "The company moved to (a) different…").

Subsequent experiments have further demonstrated that preceding speech rate can still facilitate listeners' anticipation of upcoming words even when the target words have been made clearer (e.g., by creating various degrees of amplitude dip at the word onset; Heffner, Dilley, McAuley, & Pitt, 2013). According to Dilley and colleagues, one way in which listeners can use such cues to anticipate upcoming word forms is by extracting the statistical (e.g., distributional) properties of the preceding prosody. For example, Baese-Berk and colleagues (Baese-Berk et al., 2014) examined the role of long-term exposure to varying speech rates and found that perceptual learning of contextual prosody can influence word perception. This indicates that human listeners are constantly updating their model of different prosodic cues to enable more accurate predictions about the upcoming signal. Consistent with this view, similar uses of speech rate have been replicated in other languages (e.g., Russian, Mandarin) in both native (L1) and non-native (L2) processing (Dilley, Morrill, & Banzina, 2013; Lai & Dilley, 2016), and other prosodic cues (e.g., rhythmic patterns) have also been found to support word recognition (Breen, Dilley, Devin McAuley, & Sanders, 2014; Brown, Salverda, Dilley, & Tanenhaus, 2011; Brown, Salverda, Dilley, & Tanenhaus, 2015; Dilley & McAuley, 2008; Dilley, Mattys, & Vinke, 2010; Kuijpers & van Donselaar, 1998; Morrill, Dilley, McAuley, & Pitt, 2014).

However, for focus perception it is still an empirical question whether preceding prosody can also facilitate such prediction across languages. For instance, the existing data on prosodic entrainment come from native speakers of English and Dutch, ruling out conclusions about universality and language-specificity given that the relation between prosody and focus is essentially the same in these two languages (Gussenhoven, 1983). More useful for examining such questions would be data from another language where listening is adapted to a different prosodic system; for instance, comparing English and Mandarin Chinese. Mandarin has features that are both similar to and different from English. Despite their typological distance, both languages express prosodic focus with much the same means (i.e., exaggerated pitch range/pitch accents, increased duration and intensity, and post-focal compression). However, recent work in our laboratory has revealed that the two languages can still differ in the degree to which different prosodic cues (e.g., pitch, intensity) are used to highlight focus (Ip & Cutler, 2016).

Further, other differences in phonological systems could prevent Mandarin speakers from showing the same entrainment effect. In English, sentences typically contain a focused constituent highlighted by a pitch accent. In Mandarin, however, both lexical tones and intonation share the same prosodic features, and to date, there is no consensus on how the two features co-exist. Xu (2005) argues that having a tonal system may not affect the use of pitch for other purposes because tones only require about one half of speakers' natural pitch range. Intonational effects in tone languages may also be phonetically layered on existing lexical tones and cause shifts in F_0 register or fluctuation of F_0 range (e.g., Mandarin: Xu, 1999; Yoloxóchitl Mixtec: DiCanio, Benn, & García, 2018). Similarly, some production studies suggest that prosody plays a dual role in the expression of information structure and lexical tones because features like F_0, intensity, and duration cues can be exaggerated to produce focus (e.g., Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2013). Contrasting with this view is the suggestion that much of the pitch contour would be exhausted in the phonetic expressions of contour tones, thereby resulting in a less elaborate intonational system (Hayes, 1995; Pierrehumbert, 1999) or no intonational system at all (Kratochvil, 1998). Research across various tone languages indeed shows that pitch accents can be minimal or absent (e.g., Mambila: Connell, 2017; Yoruba: Laniran & Clements, 2003), and in some cases not all tones may carry boundary tones (e.g., Akan: Kügler, 2017; Tswana: Zerbian, 2017). Particularly in the case of Mandarin, tones also co-specify lexical identity, and native speakers are sensitive to tonal differences in phonation, intrinsic duration, and amplitude (Blicher, Diehl, & Cohen, 1990; Fu, Zeng, Shannon, & Soli, 1998; Liu & Samuel, 2004; Whalen & Xu, 1992). Therefore, even if there is exaggeration of prosodic cues used for focus (e.g., Chen & Gussenhoven, 2008), it may be localised on only the focused word, with cues in the prefocus intonation contour preempted by tonal movements.

Indeed, some production research suggests that Mandarin speakers may not produce prefocus cues in the preceding intonation in a way that would support prosodic entrainment. Thus Xu (1999) found that the intonation contour before a Mandarin focused word tends to be acoustically similar to that of a neutrally produced sentence with no prosodic focus (see also, Liu & Xu, 2005; Yuan, 2004). There are also reports of other tone languages, such the Austronesian language Ma'ya (Remijsen, 2002), and some Otomanguean languages (Chávez-Peón, 2010; DiCanio & Hatcher, 2018), in which speakers only use duration to produce stress, due to the documented use of F_0 primarily for tonal contrasts. In addition, comparisons between tonal and non-tonal dialects of a single language (e.g., Kammu) show that intonation can be influenced by the tone combination in the sentence (Karlsson, House, Svantesson, & Tayanin, 2010). All these findings indicate that the richness of intonation cues can be constrained by the presence of tones.

Even if intonation cues are available, it is also possible that Mandarin listeners would be less likely to use these cues to predict the presence of an accented focused word. This view is supported by previous studies showing that competing F_0 contour adjustments by tones and intonation can hinder recognition of different intonational categories (e.g., statements versus questions; Liu & Xu, 2005; Yuan, 2011). Several experiments comparing tone and non-tone languages have also suggested that native speakers of tone languages are more likely to process pitch at a lexical level and are less sensitive to sentence intonation (e.g., Gandour et al., 2003; Gussenhoven & Chen, 2000). Finally, certain tones (e.g., Mandarin low-dipping tone) are more prone to F_0 restriction, and listeners are less likely to detect focus when focused syllables are produced with these tones (e.g., Lee, Wang, & Liberman, 2016). Therefore, even though suprasegmental features may indeed enjoy a dual function in the production of tone and focus (e.g., Ouyang & Kaiser, 2013), the presence of lexical tones may still place a limit on the degree to which speakers can produce, and listeners can perceive, preceding cues from which upcoming focus location may be predicted.

Thus the presence of lexical tones may impact both the production and perception of prefocus intonation. However, so far no studies have addressed predictive prosodic focus perception by Mandarin listeners. In the present study, we adopt the phoneme detection paradigm from Cutler and colleagues' experiments to compare English and Mandarin listeners' use of prosody in their anticipation of focus. Based on the phonological differences between English and Mandarin, Mandarin listeners may not have the ability to adopt an entrainment strategy. On the other hand, it is also possible that Mandarin listeners may still adopt the same entrainment strategy, but that the extent to which they can do so may be limited due to the presence of lexical tones, either because the intonation itself is less informative for focus detection, or because the listeners make less effective use of the intonational cues. A third possibility is that cues signalling prosodic focus may still assist Mandarin listeners to the same extent as the English listeners. This third view would suggest that prosodic entrainment may be a universal strategy that all listeners can adopt despite any differences in prosodic systems.

1. Experiment 1

1.1. Method

1.1.1. Participants

Two participant samples were tested: 23 native speakers of Australian English (M_{age} = 23.96 years, SD = 8.64 years; 16 females) and 23 native speakers of Mandarin Chinese (M_{age} = 25.02 years, SD = 3.78 years; 13 females). All of the English speakers reported that they were born and raised in Australia. The Mandarin speakers were born in Mainland China and had been living and studying in Australia for an average of one year and 5 months (SD = 25.44 months, range: 23 days–7.96 years). We tested three further participants but excluded their data for failing a follow-up recognition test (one Mandarin speaker), or due to technical issues (two English speakers). In addition, given the prosodic differences between the Mandarin spoken in Mainland China and other parts of the Sinophone world (e.g., Xu, Chen, & Wang, 2012), further data from two Mandarin speakers who grew up in communities outside of Mainland China (e.g., Taiwan) were not analysed. No participant reported any hearing or speech impairments.

1.1.2. Materials

The English and Mandarin sentences (see Appendices A and B) were each recorded by a female native speaker who did not know the purpose of the experiment. In both languages, 24 experimental sentences were recorded in three versions: predicted high stress, predicted low stress, and neutral. In the predicted high stress version, the target-bearing word received emphatic stress. In the predicted low stress version, emphatic stress was instead placed on a word that occurred later in the sentence than the target-bearing word, which, in consequence, received very reduced stress. In the neutral version, the target-bearing word and the sentence as a whole were produced in a way which resulted in no emphatic stress. In all of the experimental sentences, the phoneme target was a voiceless aspirated bilabial stop [p_h] occurring at the start of the target-bearing word's first syllable (e.g., "peanuts" [p^hi:nʌts]; "葡萄" *grapes* [p^hu2 t^hau0]). Further, the phoneme target in English always occurred on the word's lexically stressed syllable. Given the language differences in stop inventories, we only used one phoneme target for all sentence trials. For Mandarin, we also controlled the tone of the target-bearing words, such that half of the sentences had the phoneme target occurring on a high-rising second tone (e.g., "葡萄" *grapes* [p^hu2 t^hau0]) and half had the target on a falling fourth tone (e.g., "骗子" *swindler* [p^hjɛn4 ʂʐ̩0]).

Using Praat (Boersma & Weenink, 2018), the target-bearing words were extracted from all three versions of each experimental sentence, with the cuts being made at the nearest zero crossing to each end. The high- and low-stressed target-bearing words from the predicted high and low stress versions were then replaced by an acoustically identical token of the same target word from the neutral version. For both the English and Mandarin stimuli, two experimental conditions were constructed, each containing one version of each of the 24 spliced experimental sentences, plus an additional set of 24 filler sentences. The experimental and filler sentences were presented in a pseudo-random sequence and all participants heard them in the same order. Further, the English and Mandarin conditions had the same order of experimental and filler sentences. The experimental sentences with predicted high versus predicted low stress were counterbalanced across the two conditions (henceforth called "Version A" and "Version B").

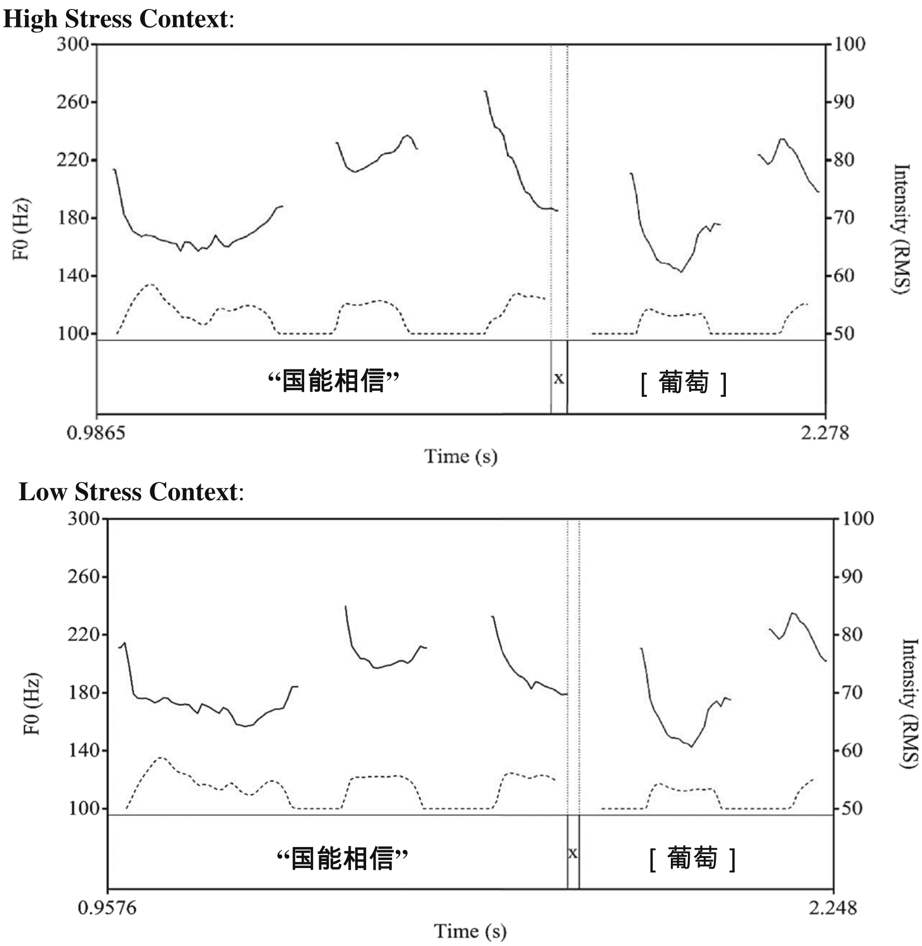


Fig. 1. Pitch and amplitude contours of an example experimental sentence in Mandarin predicted high and low stress contexts. Prosodic parameters (i.e., overall duration, mean and maximum F_0, F_0 range, mean and maximum intensity, and intensity range) three to four syllables preceding the target-bearing word (in this example, in “国能相信”) – were measured for our acoustic analyses. The “x” portion indicates the duration of the pretarget interval. In this example, the values were: 184 Hz (low) vs. 194 Hz (high) for mean F_0, 248 Hz (low) vs. 264 (high) for maximum F_0, 72 Hz (low) vs. 106 Hz (high) for F_0 range, 840 ms (low) vs. 800 (high) for overall duration, 20 ms (low) vs. 30 ms (high) for pretarget duration, 53.53 (low) vs. 53.42 (high) for mean intensity, 58.66 (low) vs. 58.40 (high) for maximum intensity, and 19.98 (low) vs. 14.71 (high) for intensity range.

The English and Mandarin experimental sentences were comparable in length, as measured in terms of the total number of syllables (English, $M = 17.92$, $SD = 3.92$; Mandarin, $M = 16.75$, $SD = 2.59$). Further, the number of syllables between the start of the sentence and the onset of the target-bearing word was comparable across the two languages (English, $M = 10.00$, $SD = 2.95$; Mandarin, $M = 9.04$, $SD = 2.35$), and was also similar to the set of English sentences used in the previous Cutler and Darwin (1981) experiments ($M = 10.30$, $SD = 3.16$). To avoid interference between the sentences, sentence beginnings were varied and semantic content that could be associated with another sentence in the set was avoided. We also varied the syntactic category of the word immediately preceding the target word, so that less than half of the target words were preceded by a determiner (and we used a variety of determiners). In addition, none of the sentences had any additional occurrence of voiced or voiceless bilabial stops beyond that in the target-bearing word. All of the sentences were produced at a natural fast-normal rate.

1.1.3. Procedures

All tests were conducted in the participant's native language in a sound-attenuated booth at the MARCS Institute, Western Sydney University. The phoneme-detection task was administered using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002) on a laptop computer, with attached to it a set of headphones and a Chronos® response device for button pressing.

Participants were informed that the experiment aimed to examine listeners' memory and language comprehension; they were further told that they would listen to a series of sentences and had two tasks: first, pay careful attention to the meaning of each sentence, and second, press a button as fast and as accurately as they could whenever they heard a word that began with the target sound $[p_h]$. Participants received two practice trials and feedback before starting the actual experiment. Instructions were written in the participants' native language (see Appendices C and D). The Chinese instructions were translated from the English version by a professional translator who was an instructor at the university's languages and translation department. The instructions contained no mention of sentence prosody.

At the end of the testing session, participants completed a follow-up recognition test in which they were asked to judge whether or not each of the 20 sentences in the list was from the experiment (see Appendices E and F). All participants scored 65% or above in the test (Mandarin speakers, $M = 84.13$, $SD = 10.51$, range: 65–100; English speakers, $M = 88.48$, $SD = 7.75$, range: 70–100).

1.2. Results and Discussion

1.2.1. General overview

Response times (RTs) were measured as the duration between the release of the target stop consonant and participants' button presses. We compared participants' RT to the target phoneme in predicted high stress sentences with their RT in predicted low stress sentences. No participants had RT shorter than 100 milliseconds (i.e., false alarms); RT datapoints longer than 2500 milliseconds (possibly indicating a reprocessing of the sentence; Ratcliff, 1993) were excluded from final analyses. Both the predicted high stress and low stress contexts had two datapoints over 2500 milliseconds in Mandarin and there was one such datapoint in a predicted high stress context sentence in English. No participant had more than two instances of RT longer than 2500 milliseconds. All of the raw data can be accessed from the following link: osf.io/zyfah/quickfiles.

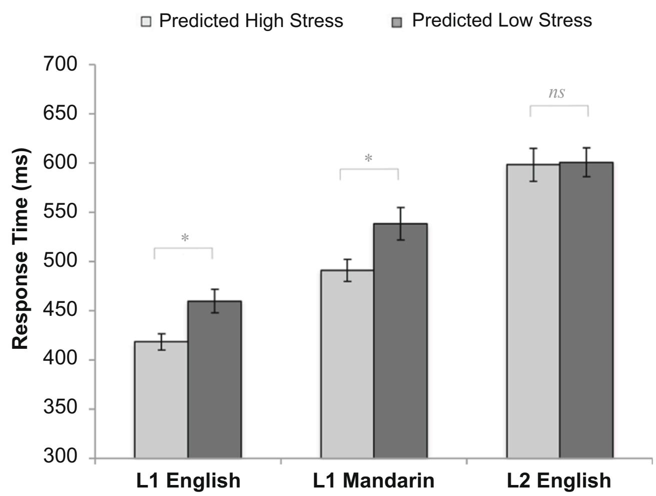


Fig. 2. Response time (ms) as a function of intonationally predicted high versus low stress in Experiments 1 (L1 English, L1 Mandarin) and 2 (L2 English). Error bars represent standard error of the mean. $*p \leq .05$.

The primary aim of our statistical analyses was to examine whether RT differed across the predicted high and low stress prosodic contexts. Another aim was to test for language-specific differences in listeners' RTs across the prosodic contexts and the experimental trials. We also conducted acoustic analyses of the prefocus cues in the preceding prosody of the stimuli sentences, by (a) examining the duration, F_0, and intensity cues in the prefocus region of each stimulus sentence (i.e., two to four syllables before the onset of the target phoneme; see Fig. 1 for an example in Mandarin), and (b) measuring the pre-target interval, i.e., the duration of the silence between offset of the preceding word and the release of the target stop consonant. Previous studies have shown that listeners can still predict upcoming stress even when certain preceding cues (e.g., stop closure duration, F_0) have been made uninformative (Cutler & Darwin, 1981). However, it is still uncertain whether there is a relationship between listeners' prediction of upcoming stress and any of the preceding prosodic cues, and whether languages may differ in the type of prosodic cues provided by the preceding intonation.

1.2.2. Response time

Using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015, version 1.1–7), Linear Mixed-Effects (LME) regression models were constructed to obtain the best fitting model predicting participants' response time (RT). The raw RT data formed a skewed distribution and were transformed to their inverse RTs using the Box-Cox procedure (Box & Cox, 1964). This transformation approach has been argued to be best suited for psycholinguistic data (Lo & Andrews, 2015) and is suitable to our analyses since it provides a better approximation to normal-distribution and homoscedasticity assumption for linear models compared to simple logarithmic transformation (see Balota, Aschenbrenner, & Yap, 2013). Analyses were therefore performed with the Box-Cox-transformed RT data as dependent variable, but for the reader's convenience, all the RT means, standard deviations, fixed effects estimates (β), and standard errors reported in the main text and figures and tables will be in their raw values (in milliseconds).

A baseline model was used as a starting point, including by-subject and by-item random intercepts as well as by-subject and by-item random slopes for the effect of prosodic context. Predictors were then added in a step-wise fashion to determine model fit, conducted using chi-squared tests of model log-likelihoods. Predictors that did not yield significant improvement in the model comparisons were dropped from the model before additional predictors were added. This was determined based on the *p*-values of the chi-squared tests and/or differences in Akaike Information Criterion (AIC), with the latter being more useful in cases where the complexity of the model cannot be justified by the additional variance explained (Shaw et al., 2018). Leave-one-out comparisons were used to ensure that each predictor yielded a significant gain in log likelihood with all other predictors in the model.

Fixed effects for prosodic context, language, and all of the acoustic variables were coded with mean-centered contrast codes. Participant gender was included in the analyses as a categorical (factor) predictor. Trial sequence order was included in the model as a continuous predictor, where each level was labelled according to its trial order across the experimental trials (i.e., from 1 to 24). Due to its large eigenvalue, we rescaled this variable by centering the trial order levels into numeric values from 0 to 1.

Before testing for the effect of prosodic context, we first examined language and subject gender (male vs. female) as control variables. The best fitting control model contained a significant effect of language. The average RT for English listeners was 438.89 ms, versus 514.54 ms for Mandarin listeners (χ^2 (1) = 5.95, p = .015 in leave-one-out comparisons; β = 118.42, t = 2.50). However, there was no significant effect of gender (χ^2 (1) = 1.18, p = .277). The best fitting control model therefore consisted of language as the only fixed factor, and subject and sentence item as random factors.

Once the best fitting control model was obtained, we examined the effect of prosodic context (predicted high vs. low stress context). The addition of prosodic context to the best fitting control model revealed a significant gain in model log-likelihood (χ^2 (1) = 16.07, $p < .001$). As shown in Fig. 2 (see also Table 1), there was a significant main effect of prosodic context (β = 67.37, t = 4.62): RTs to the target phoneme in both English and Mandarin were faster for sentences with predicted high stress contexts (English: M = 418.46 ms, SD = 139.04 ms; Mandarin: M = 491.01 ms, SD = 181.71 ms) compared to those with predicted low stress (English: M = 459.63 ms, SD = 196.73 ms; Mandarin: M = 538.23 ms, SD = 273.65 ms). However, there was no significant interaction between prosodic context and language (χ^2 (1) = 0.72, p = .398; β = 16.24, t = −1.01). The results of the model comparisons are summarised in Table 2.

1.2.3. Response time across sentence trials

We also examined whether there were any language differences in the pattern of listeners' RT across the 24 experimental sentence trials (see Supplementary Data). The effect of trial order was tested against the updated best fitting model with the prosodic context variable added as fixed factor. Our analyses show that adding the main effect of trial order did not significantly improve model fit (χ^2 (1) = 0.16, p = .686; β = 6.23, t = 0.42). There was a significant 2-way interaction between trial and language (χ^2 (1) = 9.08, p = .003; β = −32.62, t = −3.35:

Table 1
Response time (ms) to the target phoneme [p^h] in Experiments 1 and 2.

Experiment	Sample	Mean Response Time (SD)	
		Predicted high stress	Predicted low stress
Experiment 1: L1 phoneme detection	Native English speakers (n = 23)	418.46* (139.04)	459.63 (196.73)
	Native Mandarin speakers (n = 23)	491.01* (181.71)	538.23 (273.65)
Experiment 2: L2 phoneme detection	Native Mandarin speakers (n = 24)	598.18 (274.93)	600.71 (245.12)

* $p \leq .05$.

Table 2

Experiment 1: Results of the linear mixed-effects model analyses for RTs. See Appendix G for random effects and Appendix H for Box-Cox converted beta and standard error values. The χ^2 values and corresponding p-values are based on leave-one-out model comparisons. Analyses were based on 1088 datapoints from 46 participants and 24 items. *Note*: *p < .05, **$p < .01$.

Fixed effects	β	SE	χ^2 (1)	p-Value
(Intercept)	454.45	14.87		
Language	118.42	41.17	5.95	0.015*
Gender	38.70	28.80	1.18	0.277
Prosodic context	67.37	21.96	16.07	6.117^{e-05}*
Language × Prosodic context	16.24	57.95	0.72	0.398
Trial	6.23	18.03	0.16	0.686
Trial × Prosodic context	−9.79	21.73	0.004	0.948
Trial × Language	−32.62	17.30	9.08	0.003*
English listeners (542 datapoints)				
(Intercept)	422.96	17.30		
Prosodic context	62.57	23.28	10.63	0.001**
Preceding duration × Prosodic context	114.86	84.41	2.80	0.246
Pretarget interval duration × Prosodic context	−103.28	133.35	4.65	0.098
Mean F0 × Prosodic context	77.63	273.91	1.03	0.599
Maximum F0 × Prosodic context	120.42	188.43	1.67	0.435
F0 Range × Prosodic context	49.89	62.37	2.16	0.339
Mean intensity × Prosodic context	152.68	586.84	0.79	0.675
Maximum intensity × Prosodic context	223.33	733.27	1.31	0.519
Intensity range × Prosodic context	−2.81	98.82	2.38	0.304
Mandarin listeners (546 datapoints)				
(Interval)	488.14	22.00		
Prosodic context	74.43	40.41	6.55	0.011*
Preceding duration × Prosodic context	−121.64	215.16	0.42	0.811
Pretarget interval duration × Prosodic context	27.16	132.27	3.96	0.138
Mean F0 × Prosodic context	507.53	408.46	2.60	0.272
Maximum F0 × Prosodic context	314.36	574.59	1.02	0.601
f0 range × Prosodic context	33.48	115.38	0.19	0.911
Mean intensity × Prosodic context	663.27	503.98	5.71	0.058
Maximum intensity × Prosodic context	860.79	529.73	7.36	0.025*
Intensity range × Prosodic context	68.12	128.85	0.51	0.774

see Fig. 6 in Supplementary Data). However, there was no significant 2-way interaction between trial order and prosodic context (χ^2 (1) = 0.004, p = .948; β = −9.79, t = −0.57).

1.2.4. Acoustic analyses

Acoustic analyses of the stimuli recordings were conducted using Praat () based on inspection of both the waveform and the spectrogram well as the pitch tracks and amplitude envelopes. The preceding prosodic features of each stimuli sentence were examined by looking at parts of the sentence that were two to four syllables before the onset of the target-bearing word. For each sentence's preceding prosody, we measured duration, mean F_0, maximum F_0, F_0 range, root-mean-square (RMS) mean intensity, maximum intensity, and intensity range (see Fig. 1). We also measured the pre-target interval, i.e., the duration of the silence between offset of the preceding word and the release of the target stop consonant.

The acoustic results for the preceding duration, F_0, and intensity are summarised in Tables 3 and 4. Using two-tailed pairwise t-tests, evaluation of the acoustic data for the Mandarin stimuli found a significant difference in F_0 range between the predicted high and low stress contexts, such that syllables before target-bearing words had greater F_0 range in predicted high stress sentences than in predicted low stress contexts, $t(23) = 3.78$, $p = .001$. Maximum F_0 was also greater in predicted high stress sentences in Mandarin, $t(23) = 2.65$, $p = .014$. There was also a longer pre-target interval for high stress context sentences, $t(23) = 4.99$, $p < .001$. No significant differences were observed for mean F_0, overall duration, or any of the intensity cues. In contrast, in English, the preceding prosody of predicted high stress sentences was produced with higher values on all measures except for intensity range. Compared to predicted low stress contexts, the preceding prosody of English high stress context sentences had higher mean F_0, $t(23) = 2.23$, $p = .036$, higher maximum F_0, $t(23) = 3.78$, $p = .001$, greater F_0 range, $t(23) = 4.61$, $p < .001$, longer overall duration, $t(23) = 2.23$, $p = .036$, longer pause duration, $t(23) = 4.46$, $p < .001$, greater mean intensity, $t(23) = 4.88$, $p < .001$, and greater maximum intensity, $t(23) = 5.30$, $p < .001$.

We also conducted additional 2 (language: English versus Mandarin) X 2 (prosodic context: high versus low stress) mixed-model ANOVAs for maximum F_0, F_0 range, and pre-target interval duration. This was to examine whether the magnitude of these prosodic differences between high and low stress contexts was different across the English and Mandarin sentences, despite these parameters having shown significant differences in both languages. However, none of the analyses showed a significant interaction between language and prosodic context. Therefore, there were no crosslanguage differences in the degree to which the English and Mandarin speaker used these acoustic parameters to differentiate the high and low stress contexts.

1.2.5. Relation between preceding prosodic cues and response time

Further analyses were conducted to examine whether the faster RT found in the predicted high stress contexts could be explained by any of the acoustic features in the preceding prosody. Given that there were language differences in the acoustic features of the preceding prosody, separate LME regression models were conducted for the English and Mandarin RT datasets. The model comparisons and specifications for the English and Mandarin datasets are summarised in Table 2. In English (see Fig. 3), there was no significant interaction between prosodic context and any of the preceding cues. In Mandarin (see Fig. 4), however, there was a marginal significant interaction between prosodic context and preceding mean intensity (vii) (χ^2 (2) = 5.71, p = .058; β = 663.27, t = 2.30) and a significant interaction between prosodic context and preceding maximum intensity (viii) (χ^2 (2) = 7.36, p = .025; β = 860.79, t = 2.58).

To complement the LME regression analyses, we also conducted a series of Pearson's two-tailed correlation analyses to examine whether there was any link between the strength of the different prosodic cues in each sentence and the degree to which listeners showed a RT difference between high and low stress contexts. For each sentence, we calculated each prosodic parameter's proportional difference (i.e., percentage change) between high and low stress contexts. For each sentence, we also calculated the proportional difference in RT averaged across the participants. Consistent with our LME model comparisons, there were no significant correlations between RT difference and any of the parameters in English, but in Mandarin, there were significant negative correlations between proportional differences in RT and mean intensity ($r = -0.57$, $p = .004$) and maximum intensity ($r = -0.58$, $p = .003$).

1.2.6. Discussion

Overall, both English and Mandarin listeners responded faster to the target phoneme in sentences where the preceding prosody predicted high stress on the target-bearing word. Further, there was no significant interaction between prosodic context and language. This indicates that there was no language-specific difference in the degree to which high stress contexts facilitated RT, despite the acoustic data showing more cues being available in the English stimuli. Thus, this listening strategy appears to be used to an equivalent extent in each language. Also, in the acoustic analyses of the preceding prosodic measures (maximum F_0 and F_0 range and pretarget duration) that were significant in the stimuli of both languages, there were no cross-linguistic differences in the degree to which they differentiated the prosodic high and low stress contexts.

However, all of the Mandarin-speaking participants were proficient in English and had been living and studying in an English-speaking country. Exposure to English as an L2 might have helped the Mandarin speakers develop a non-native listening strategy that they could apply when listening to their native language. To test this competing

Table 3
Preceding prosody F_0 (mean, maximum, and range in Hz) and duration (in milliseconds) three or four syllables before target onset in predicted high versus low stress contexts.

Stimuli	Mean prosodic variables (SD) [Range]									
	Mean F_0		Maximum F_0		F_0 Range		Overall Duration		Pre-target Interval Duration	
	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress
English (24 sentence pairs)	180.84* (15.43) [161–223]	176.11 (14.60) [154–201]	213.97** (22.57) [175–286]	203.25 (25.99) [165–255]	58.38** (20.08) [19–100]	44.67 (20.02) [17–90]	585.04* (159.22) [385–1000]	553.58 (142.91) [317–940]	74.35** (10.91) [55–95]	61.71 (13.91) [33–89]
Mandarin (24 sentence pairs)	200.97 (22.85) [140–251]	197.36 (19.29) [152–252]	252.62* (22.25) [195–291]	242.42 (17.10) [200–293]	106.43** (42.04) [23–204]	85.41 (35.61) [37–176]	745.67 (130.83) [500–1101]	755.04 (140.76) [510–1070]	66.67** (26.09) [14–120]	49.04 (19.10) [4–71]

* $p \leq .05$ significant differences from predicted low stress contexts (two-tailed).
** $p \leq .001$ significant differences from predicted low stress contexts (two-tailed).

explanation, we conducted Experiment 2 to examine whether Mandarin speakers would also respond faster to phoneme targets due to high stress contexts in the English sentences. The same pattern of response in English by Mandarin speakers may indicate that they have acquired this prediction strategy from their L2 experience with English, but it could also mean that prosodic entrainment is a general strategy that all listeners can use in any language that has prosodic cues to upcoming focus.

2. Experiment 2

2.1. Method

2.1.1. Participants

Participants in Experiment 2 were 24 native Mandarin speakers who were born and raised in Mainland China (M_{age} = 25.13, *SD* = 4.09; 14 females), of whom 14 had also taken part in Experiment 1. We aimed to capture a wider range of Mandarin speakers with different amounts of exposure to English. To account for participants' degree of exposure to English, we calculated how long each participants have been living in Australia (i.e., date of testing minus date of arrival in Australia), since length of stay abroad is a reliable indicator of L2 proficiency (e.g., Dwyer, 2004; Félix-Brasdefer, 2004; Ife, Vives, & Meara, 2000). All participants spoke English as their second language and had been living and studying in Australia for an average of 2.45 years (*SD* = 2.63 years; range: 3 months to over 10 years).

2.1.2. Materials and procedures

The procedures were identical to those in Experiment 1, except in that the English sentences and recognition test as used for the native English speakers in Experiment 1 were now presented to the native Mandarin speakers. As in the L1 English group from Experiment 1, all participants scored at 70% or above on the follow-up recognition test (*M* = 78.33, *SD* = 9.40, range: 70–100). To optimise comparability with the L1 English speakers from Experiment 1, we excluded additional data from participants who scored below 70% and three participants whose average RT scores were over 1000 milliseconds.

2.2. Results and Discussion

From the predicted high stress data set, we removed two RT response longer than 2500 milliseconds and three false alarm responses (i.e., RT shorter than 100 milliseconds). Similarly, we also excluded six false alarm responses and one response longer than 2500 milliseconds from the predicted low stress set. As in Experiment 1, we used a baseline control model with subject, item, and experimental version as random factors, with predictors added in a stepwise fashion to determine model fit; predictors that did not yield significant improvement were dropped before additional predictors were added. Based on our LME regressionanalyses (see Table 5), the RT for the 14 participants who had previously participated in the Mandarin condition of Experiment 1 did not significantly differ from that for the 10 new participants without experience of similar experiments: adding experience from Experiment 1 as a fixed predictor into the model did not significantly improve model fit (χ^2 (1) = 2.47, *p* = .116). Data from all participants were therefore included in the main analyses.

In striking contrast to Experiment 1, the RTs of Experiment 2 revealed no effect of predicted high (*M* = 598.18, *SD* = 274.93) versus low stress (*M* = 600.71, *SD* = 245.12) (χ^2 (1) = 0.97, *p* = .323; see Fig. 2 and Table 5). Thus native Mandarin speakers' phoneme detection in English did not display the entrainment that they had demonstrated in their native language. We also tested for effects of intensity on RT. Given the interaction of intensity and prosodic context for L1 Mandarin in Experiment 1, it is worth asking whether this interaction also holds

Table 4
Preceding prosody intensity (mean, maximum, and range in RMS) three or four syllables before target onset in predicted high versus low stress contexts.

Stimuli	Mean Prosodic Variables (SD) [Range]					
	Mean Intensity		Maximum Intensity		Intensity Range	
	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress
English (24 sentence pairs)	53.63*** (2.09) [50-58]	52.46 (1.99) [48–56]	59.03*** (1.88) [56-62]	57.32 (1.97) [53–62]	26.94 (7.17) [19–41]	25.63 (6.03) [14–40]
Mandarin (24 sentence pairs)	54.44 (3.60) [51–64]	55.43 (4.36) [51–63]	59.06 (3.85) [56–69]	59.75 (4.37) [55–68]	26.47 (8.37) [15–42]	27.23 (8.61) [14–44]

*** $p \leq .001$ significant differences from predicted low stress contexts (two-tailed).

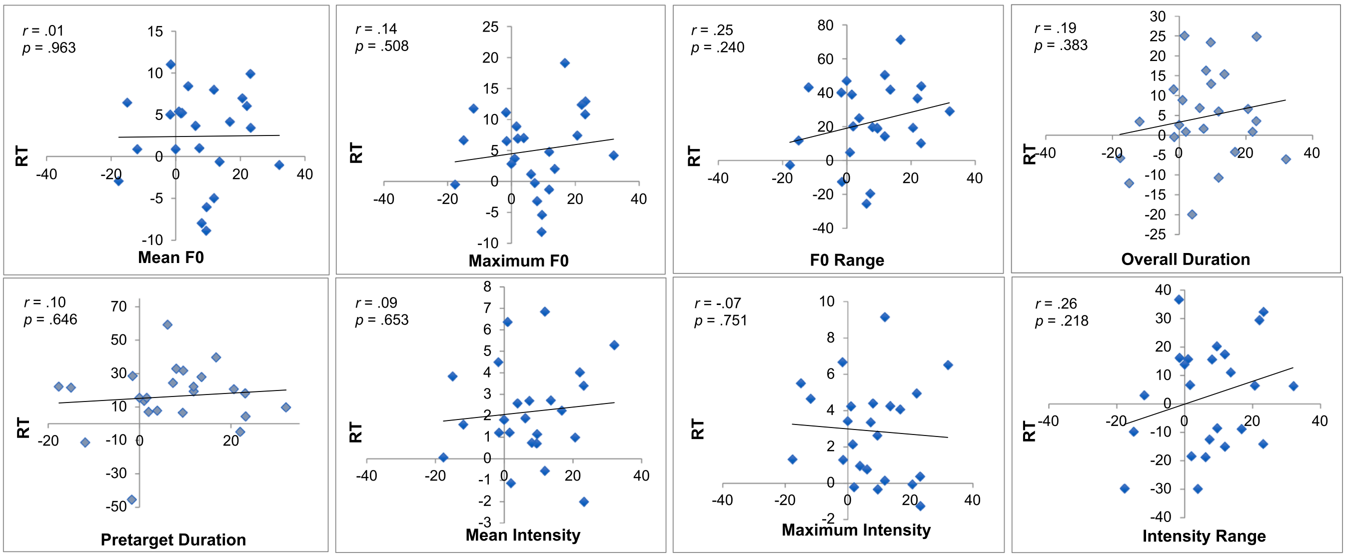


Fig. 3. Scatterplots between RT and each acoustic variable in English. All values are displayed as proportion differences (in %) between predicted high and low stress contexts.

for L2 English. A positive interaction here would indicate that Mandarin speakers learn the relevant acoustic cues to focus from their L1 and can generalise these cues to their L2, even though they might still have trouble learning the new relevant cues from their L2. However, there were no such significant interactions. Further, we investigated the role of L2 exposure and proficiency by adding length of stay in Australia and recognition scores as fixed factors; no significant improvements in model fit appeared (length of stay: χ^2 (1) = 0.37, p = .543; recognition test scores: χ^2 (1) = 2.84, p = .092). To complement these results, Pearson's correlation analyses (see Fig. 5) revealed no significant association between the proportion of RT difference between high and low stress contexts and either participants' length of stay in Australia (i.e., date of testing minus date of arrival; r = 0.054, n = 24, p = .801) or their scores on the recognition test (r = 0.285, n = 24, p = .178).

For the sample of the Mandarin speakers who participated in the Mandarin experiment in Experiment 1, there was also no significant correlation between their length of stay in Australia and the proportion of RT difference between the high and low stress context conditions (r = −0.266, n = 23, p = .219). Again, this Pearson's r correlation result was consistent with our LME model analysis, which did not yield a significant improvement after length of stay in Australia was added as a fixed factor (χ^2 (1) = 0.37, p = .543). With both the correlation and LME regression models results taken into account, the RT differences we observed between the high and low stress contexts in Mandarin seemed very unlikely to be due to the Mandarin listeners' amount of L2 exposure to English.

3. General Discussion

The reported results offer insight into how both language-universal and language-specific mechanisms influence the sentence comprehension process. Consistent with previous findings from English and Dutch (e.g., Akker & Cutler, 2003; Cutler, 1976), native Mandarin listeners too can entrain to the intonation contour to forecast an upcoming accent, although in their language much of the same prosodic information in speech also conveys lexical tones. As in the preceding studies, the entrainment was established by the fact that the original target-bearing words had been replaced by neutrally produced words, so that in both

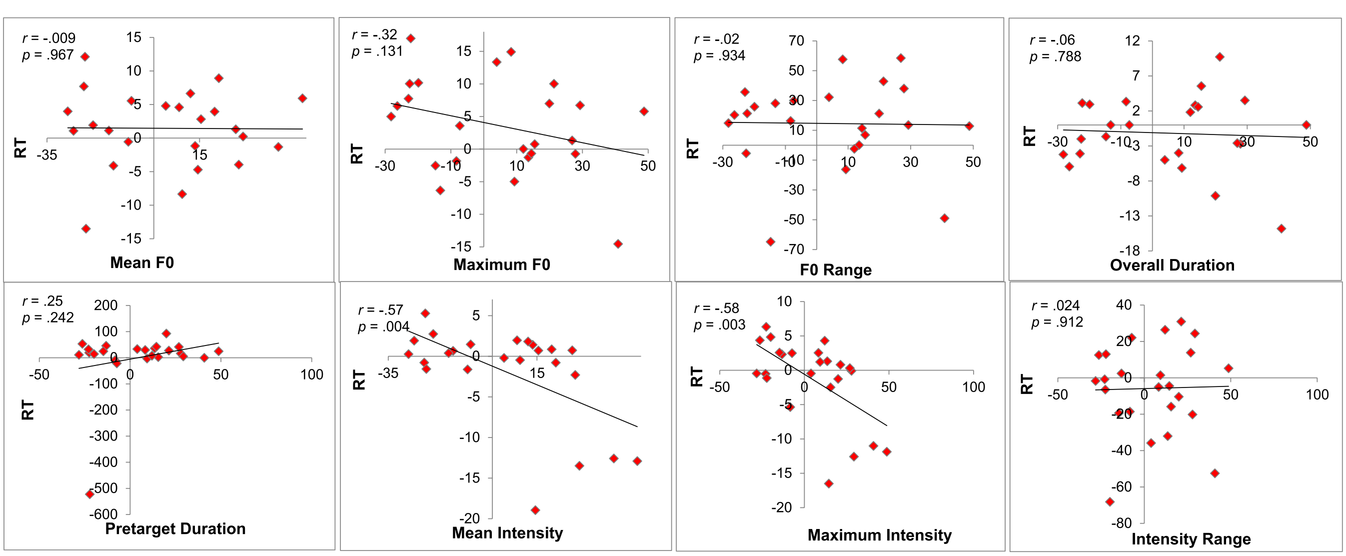


Fig. 4. Scatterplots between RT and each acoustic variable in Mandarin. All values are displayed as proportion differences (in %) between predicted high and low stress context.

Table 5
Experiment 2: Results of the linear mixed-effects model of RTs. See Appendices J and K for random effects and Appendix L for Box-Cox converted beta and standard error values. The χ^2 values and corresponding p-values come from leave-one-out model comparisons. Analyses were based on 548 datapoints from 24 participants and 24 items.

Fixed effects	β	SE	χ^2 (1)	*p*-Value
(Intercept)	604.01	32.61		
Prosodic context	1.14	39.86	0.97	0.323
Participation in Experiment 1	−84.46	57.38	2.47	0.116
Length of stay	−16.18	40.77	0.37	0.543
Post-test recognition scores	383.29	242.04	2.84	0.092
Preceding duration × Prosodic context	−158.15	114.61	2.27	0.519
Pretarget interval duration × Prosodic context	−327.37	209.13	2.95	0.400
Mean F0 × Prosodic context	114.36	376.78	1.32	0.726
Maximum F0 × Prosodic context	87.05	289.40	2.01	0.570
F0 Range × Prosodic context	−41.79	88.29	1.99	0.574
Mean Intensity × Prosodic context	485.44	815.48	5.55	0.136
Maximum Intensity × Prosodic context	449.00	1078.50	4.80	0.187
Intensity range × Prosodic context	−19.46	129.02	4.23	0.238

sentence contexts the targets being reacted to were acoustically identical. This finding for Mandarin listeners in their native language could not be ascribed to these listeners' ability to speak another language which uses the entrainment strategy, since no such strategy was adopted when Mandarin speakers were listening to sentences in English. In light of these results, our findings support the view that a common strategy may exist in listeners' prefocus entrainment to prosody despite differences in phonological systems.

Languages with lexical tone, such as Mandarin, might be thought to have less scope for a complex intonational system, given that tonal processing for distinguishing words preempts much of the intonational contour (see e.g., Hayes, 1995; Nolan, 2006; Pierrehumbert, 1999, on this issue). Indeed, Mandarin listeners do fail to distinguish between intonational categories when the relevant features conflict with cues to tone (Liu & Xu, 2005). This might suggest a processing priority for lexical tones over sentence-level intonation in situations of conflict. However, such conflicts are not the norm; prosodic cues to focus and lexical cues to tones do co-exist in Mandarin speech, but often in a way in which focus production does not at all interfere with tonal identity (e.g., by exaggeration of pitch register while maintaining pitch contour shape; see Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2013; Xu, 1999). In the present study we have provided perceptual confirmation of this peaceful co-existence: preceding prosodic cues to focus are available in Mandarin just as in English, and although Mandarin listeners certainly process the lexical tones in all words of an utterance, they also, like English listeners, make use of the sentence-level intonation contour to predict the location of upcoming accents and thereby direct attention to a focused word. Distal intonation is useful for focus location even when the pitch contour is simultaneously conveying lexical tones.

In our study, prosodic context facilitated prediction of upcoming accents in both English and Mandarin to the same extent, further supporting the argument that the entrainment strategy is universal. This was despite the fact that the languages differed in where prosodic support was available. Acoustic analyses of the English sentences with predicted high stress revealed reliable support in all preceding duration, pitch and intensity cues, but the same analyses for Mandarin found only some pitch and duration cues (longer pretarget intervals, greater F_0 range expansion, heightened F_0 peaks). Note that no difference of prosodic realization led to any detectable difference in listener reliance on the cues; neither in the linear mixed-effects models nor the correlation analyses was there evidence of faster RT as a function of the specific degree to which individual items provided such cues.

In English, RT was not directly related to the cue strength of any of the measures (see Table 2 and Fig. 3). RT in Mandarin was related only to the preceding mean and maximum intensity. Each of these intensity values was lower preceding the predicted high stress than preceding the predicted low stress items in our Mandarin recordings, such that the intensity contour would function to ensure greater relative highlighting of the constant target-bearing word in the predicted high-stress case. In agreement with this, items with lower preceding mean and maximum intensity tended to have faster RT compared to sentences with higher preceding intensity (see Fig. 4).

These findings suggest that the Mandarin listeners made effective use of intensity cues, but across the board, there was no other particular

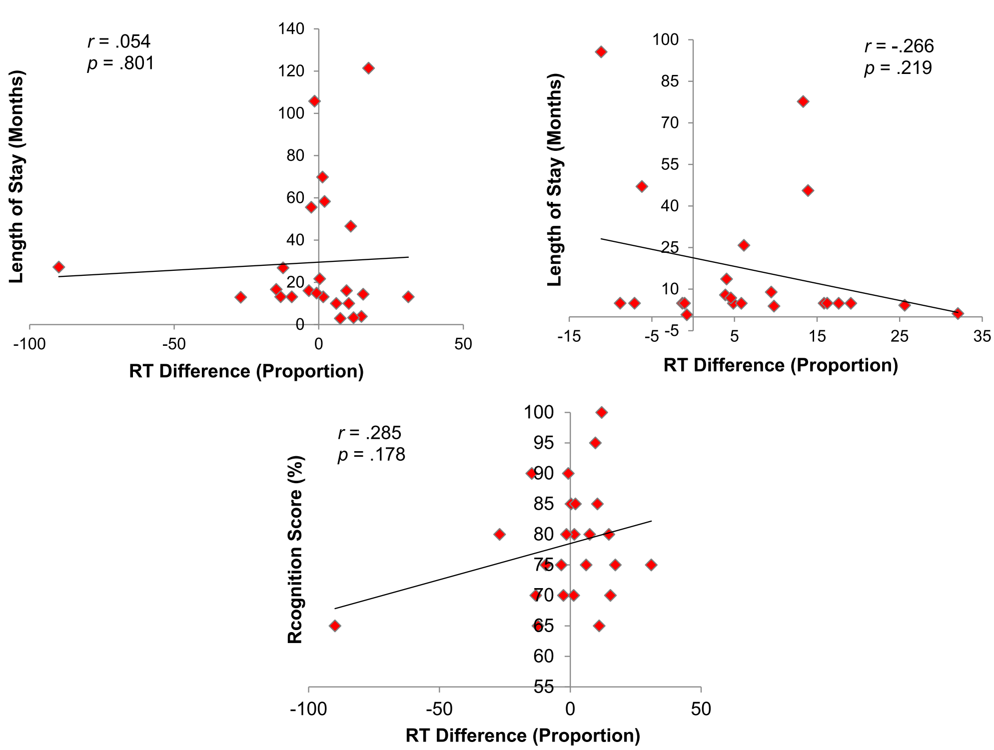


Fig. 5. Non-significant correlations between Mandarin-speaking participants' response time difference between high and low stress prosodic contexts (in proportions) and length of stay in Australia (in months) in Experiment 2 (top left) and Experiment 1 (top right), and their post-test recognition scores (between 70% to 100%) in Experiment 2 (bottom centred).

cue of which either listener group made systematic and equivalent use. This is not surprising, since our acoustic analyses, in agreement with previous research, showed that several cues were at work to signal focus and their deployment varied across the items. Other studies in our laboratory have also shown that talkers vary in which cues they tend to deploy (Ip & Cutler, 2018). Listeners have experience with such variation across talkers and utterances, and have been shown to deal with it. Thus previous research on English (Cutler & Darwin, 1981) has indicated listener flexibility in the present type of entrainment; listeners in that study did not depend upon any single prosodic dimension, and continued to respond faster in predicted high stress sentences when some strong cues for the high stress placement were rendered uninformative. The use of prosodic cues for detecting upcoming focus could thus involve evaluation of an overall prosodic pattern, whereby the proportional contribution of its component features can differ (as long as these do not contradict one another, presumably, so that a final pattern is explicit). If need be, listeners might then accomplish efficient processing of upcoming focus on the basis of just one informative cue in the preceding prosody, whatever that cue in a given utterance by a given talker might be.

In this flexibility, the processing of prosody is of course not particularly different from any other part of speech processing. It has long been known that categorical distinctions in speech are frequently signaled by multiple cues, with the cues engaging in trading relations such that they are evaluated not independently but relative to one another (Pisoni & Luce, 1987). Such cue trading is also found in the processing of prosodic cues to syntactic structure or word identity (e.g., Beach, 1991; Howell, 1993). We did not impose any such deliberate acoustic tradeoffs in our present stimuli in either language, but our results strongly suggest that listeners would be capable of dealing with whatever cue structure was presented.

This flexibility in native prosodic processing renders the result of our second experiment all the more paradoxical. Mandarin users of L2 English did not make use of preceding prosodic cues to direct their attention to the location of focus in the English input. These were the same English materials in which English listeners had found the prosodic cues and exploited them efficiently, and the cues in these materials included the cues that Mandarin L1 listeners had found, and made effective use of, in the materials in their native language.

As L2 speakers, the participants in Experiment 2 naturally had lower levels of overall English proficiency than the English-speaking participants in Experiment 1. As expected, the average RTs were slower across both high and low stress contexts, and the scores on the recognition test lower in Experiment 2 than for either participant group in Experiment 1. However, lower English proficiency levels cannot be the explanation of the lack of nonnative transference, as our mixed-model analyses for Experiment 2 did not detect any significant relationship between listeners' RT in this study and either their recognition scores or their amount of exposure to English (measured as length of time in Australia). The explanation for the asymmetry between the Experiment 1 and Experiment 2 patterns must lie outside the test situation.

Research on speech processing in L2 has comprehensively analysed the extent to which its success (or failure) depends on L1-L2 similarities and differences. For phonemic processing, the pattern has been mapped for situations where listeners know only one of those languages (Best, 1995) and for situations where they know both, as L1 versus L2 (Flege, 1995). There is a hierarchy of difficulty, with contrasts which are effectively the same in two languages being easily discriminated by listeners of each, while L2 contrasts which do not map exactly to any contrast of the L1 are harder, with contrast between two L2 sounds which would map to a single L1 category being the hardest of all. Importantly, this latter ranking of the difficulty of contrast types holds both for listening to unfamiliar languages and for listening to late-acquired L2, as in the present case (Best & Tyler, 2007). If the processing of sentence-level intonation is analogous to the processing of phonemic structure, then use of the same range of prosodic cues to focus in two languages might be thought to predict that these cues could be used in utterances in either language by listeners of each language. Since this was not the case in our Experiment 2, we propose that the processing of prosody in this manner is not analogous to the processing of phonemes.

Other current approaches to accounting for L2 listening difficulty compare not so much the categories (at any level) of the L1 versus the L2, but the relative usefulness of different speech information for processing L1 versus L2. Thus cue weighting theory (e.g., Tremblay, Broersma, & Coughlin, 2017) proposes that mastery of the use of a given cue in processing the L1 may enable and indeed enhance listeners' use of the same cue in L2, even in cases where the cue in question serves a different purpose in the two languages. The theory allows for both segmental and suprasegmental cues to be repurposed in this manner; in Tremblay et al.'s study, native listeners of Dutch were found to transfer their L1 use of F_0 cues for lexical stress to the perception of word-final boundaries in French. In other studies, better encoding of English lexical stress by Mandarin listeners than by Korean listeners was ascribed to Mandarin listeners' enhanced use of the same suprasegmental cue to process L1 lexical tones (Connell et al., 2018; Lin, Wang, Idsardi, & Xu, 2014). Again, however, it is difficult to apply a cue-weighting perspective to the situation in our study; the same cues that were used in L1 and were present in L2 proved ineffective in the latter case.

Instead, we would interpret our findings in a larger perspective than the recognition of speech structure, either phonemic or prosodic. We suggest that the source of the L2 users' failure to deploy skills from their L1 is to be found in the *conjunction* of prosodic processing with the processing of the further structure of speech. Prosodic structure is rarely taught in the classroom, either at school for the L1 or at any age for the L2, resulting in a widespread lack of awareness of prosody in general, and many failures to exploit prosodic information in practice (Reed & Michaud, 2014). Failures to make use of prosody have in fact been demonstrated in a number of L2 listening studies. Vanlancker-Sidtis (2003) found that L2 listeners perform less well than L1 in discriminating (prosodically cued) idiomatic from literal readings of word sequences. Pennington and Ellis (2000) presented native Cantonese speakers of L2 English with a set of English sentences, and then tested their recognition of what they had heard. In the recognition test, the prosody alone might be altered; even highly proficient L2 users were poor at distinguishing between prosodically differing versions when they were not made aware of the different intonation patterns. Using a similar task to that in the present study, Akker and Cutler (2003) found that in the L1, the use of distal intonation to direct attention to a focused word is largely suspended if focus is separately cued by preceding discourse information; that is, the prosodic and the discourse effects interacted. This finding held true in both English and Dutch L1 experiments. In an L2 experiment (Dutch listeners and English materials), however, Akker and Cutler found that the interaction failed to appear; instead, both effects were observed, suggesting that although these (proficient) L2 listeners were able to process the prosodic contour as well as the discourse information, they were unable to integrate these two components of the sentence processing task in an native-like manner.

In short: the processing of L2 speech is difficult, and prosodic processing may be abandoned in the interest of correctly ascertaining the sequence of segments, even when attention to prosodic information could in fact significantly assist in the task of understanding the utterance in its larger discourse context. This functional load account places our otherwise puzzling Experiment 2 finding in the context of similar findings of failure to exploit prosody in L2 despite its successful use in L1.

The processing of L2 prosody needs more research attention, as does the processing of prosody in relation to other speech structure in general. The developmental trajectory shows a number of interesting prosodic effects which deserve further investigation. Thus, while language learners are generally sensitive to statistical structures in the language input (e.g., Kleinschmidt & Jaeger, 2011; Saffran et al., 1996; Vallabha, McClelland, Pons, Werker, & Amano, 2007), there is an early bias towards the statistical occurrence patterns of vowels rather than consonants (Nespor, Peña, & Mehler, 2003), which is in stark contrast to the reverse pattern (i.e., a consonant bias) from late in the first year, and onwards for the rest of the lifespan. The later pattern has its source in vocabulary structure, while the earlier pattern is held to arise from the fact that vowels are the carriers of prosodic structure (Nazzi & Cutler, 2019). Prosody in turn carries talker identity information and emotional information, and these are communicatively useful even in the time before pairing of sound and meaning in a vocabulary has begun. There is scope for much future illumination of this proposal, however.

In second languages, acquisition of prosodic patterning is a protracted process (Mennen, 2004). Whether listeners can apply their L1 prosodic strategies in their L2 may depend, as indeed suggested above, on how they process the conjunction of segmental and suprasegmental information in the nonnative language (Lee & Nusbaum, 1993). Future experiments here could investigate whether there are more subtle ways in which L2 listeners are susceptible to prosodic cues. For instance, in a situation such as we created in the present study, in which the processing of acoustically identical word tokens is potentially affected by manipulations that are solely prosodic, might participants be able to better remember target words from sentences with predicted high stress contexts compared to low stress contexts? Similarly, might listeners show greater influence of word priming for target words in predicted high than in predicted low stress contexts? That is, there may still be processing effects in the L2 situation which have as yet eluded the researcher's grasp.

4. Conclusion

Even though Mandarin has lexical tone, whereby F_0 patterns carry a lexical as well as a sentence-level functional load, Mandarin listeners entrain to preceding intonation at the sentence level to predict upcoming focus, just as had already been observed for English. Acoustic analyses of the present Mandarin stimuli revealed a narrower range of prosodic cues than were shown in the present English stimuli. Despite this, the preceding prosodic context facilitated listeners' prediction of upcoming accents to an equivalent extent in each language. In line with other evidence of failure to exploit prosodic information in L2, however, Mandarin listeners did not display prosodic entrainment in (L2) English. Nonetheless, the fact that both English and Mandarin native processing used entrainment to the same extent, despite the cue range differences, points towards an overall strategy operating in a universal manner. Both concurrent and anticipatory uses of cues to informational salience appear to be options for all listeners, everywhere.

CRediT authorship contribution statement

Martin Ho Kwan Ip:Methodology, Data curation, Formal analysis, Writing - original draft, Writing - review & editing.**Anne Cutler**:Conceptualization, Methodology, Supervision, Writing - review & editing, Funding acquisition.

Acknowledgements

Financial support was provided by the MARCS Institute and the Australian Research Council Centre of Excellence for the Dynamics of Language [CE140100041]. We are grateful to Jason Shaw, Jessie Nixon, Laurence Bruggeman, and Jeremy Zehr for their advice in our statistical analyses, Mark Antoniou and Chris Carignan for technical support and advice, Zhang Yong, Cheng Cheng, and Matthew Stansfield for assistance with participant recruitment, Ma Jiayi for help in translating the written instructions into Chinese, and Bob Ladd, Giuseppina Turco, and Yi Xu for helpful comments on the project. Portions of the data reported here were presented at INTERSPEECH 2017 in Stockholm, Sweden.

Appendix A

Note: Target-bearing words are italicised. The capitalised words are words with the predicted accent in the (a) predicted high and (b) predicted low stress sentences

Experimental sentences

1. (a) I wish he weren't going to a *PARTY* on Monday
   (b) I wish he weren't going to a *party* on MONDAY
2. (a) The old lady thought she saw three *PIXIES* in her garden
   (b) The old lady thought she saw three *pixies* in her GARDEN
3. (a) All the contestants were in a state of PANIC when their names were called out
   (b) All the contestants were in a state of panic when their NAMES were called out
4. (a) Getting an Academy Award was the very *PEAK* of his extremely long career
   (b) Getting an Academy Award was the very *peak* of his EXTREMELY long career
5. (a) Her servants finally found a *PERFECT* way to disguise the stain
   (b) Her servants finally found a *perfect* way to DISGUISE the stain
6. (a) A crowd of activists threw *POWDER* at the mayor's face
   (b) A crowd of activists threw *powder* at the mayor's FACE
7. (a) None of the students could solve the *PUZZLES* the Russians had made
   (b) None of the students could solve the *puzzles* the RUSSIANS had made
8. (a) That summer four years ago I ate roast *PEANUTS* for every meal
   (b) That summer four years ago I ate roast *peanuts* for EVERY meal
9. (a) My friends and I used to meet in the *PARK* every day
   (b) My friends and I used to meet in the *park* every DAY

10. (a) They want to inform my *PARTNER* that I was sent home from work
    (b) They want to inform my *partner* that I was sent HOME from work

11. (a) Most of the jurors find it odd that the millionaire was *PARDONED* after the verdict
    (b) Most of the jurors find it odd that the millionaire was *pardoned* AFTER the verdict

12. (a) The hotel wants to hire more *PORTERS* to deal with the increase in guests
    (b) The hotel wants to hire more *porters* to deal with the increase in GUESTS

13. (a) Our clock no longer works ever since the *PENDULUM* went missing
    (b) Our clock no longer works ever since the *pendulum* went MISSING

14. (a) The surgeons must quickly remove her *PANCREAS* to delay the cancer from advancing
    (b) The surgeons must quickly remove her *pancreas* to delay the CANCER from advancing

15. (a) The Greeks once lived in a society where citizens had the *POWER* to demand their leaders' dismissal
    (b) The Greeks once lived in a society where citizens had the *power* to demand their leaders' DISMISSAL

16. (a) In some convents nuns still use *PADLOCKS* to seal their gates from the outside world
    (b) In some convents nuns still use *padlocks* to seal their GATES from the outside world

17. (a) Down on the farm we were amused to see a *PARROT* who could sing in French
    (b) Down on the farm we were amused to see a *parrot* who could sing in FRENCH

18. (a) Unfortunately the geologist didn't have enough time to *POLISH* all his minerals for the show
    (b) Unfortunately the geologist didn't have enough time to *polish* ALL his minerals for the show

19. (a) The naval officer shook hands with a *PIRATE* who rescued him from the fire
    (b) The naval officer shook hands with a *pirate* who RESCUED him from the fire

20. (a) A child who witnessed the crime said the gunman used his *PENCIL* to scare her away
    (b) A child who witnessed the crime said the gunman used his *pencil* to SCARE her away

21. (a) I was quite shocked to see the Archduke's *POODLES* eating truffles for lunch
    (b) I was quite shocked to see the Archduke's *poodles* eating TRUFFLES for lunch

22. (a) It is sad that the chief commander will *PUNISH* his men for saving the foreigners
    (b) It is sad that the chief commander will *punish* his men for SAVING the foreigners

23. (a) Marine scientists were angry when they discovered *PETROL* inside the whale's eyes
    (b) Marine scientists were angry when they discovered *petrol* inside the whale's EYES

24. (a) These tourists said they would like to *PICNIC* in the desert
    (b) These tourists said they would like to *picnic* in the DESERT

Filler sentences

4 filler sentences with early occurrence of the phoneme target

1. *PARSLEY* is the only thing you should add to the salad
2. In *POLAND* watching movies like "Home Alone" is now a Christmas tradition
3. Kim is *PAINTING* her own face with green and yellow ink for the soccer finale
4. You should not *PONDER* over what colour dress you will wear

4 filler sentences with late occurrence of the phoneme target.

5. The examiner failed us on our driver's license after we told her she was too *PICKY*
6. According to researchers, children under eleven don't understand what a *PARTICLE* is
7. If something goes wrong during the flight the lead stewardess must tell the *PILOTS*.
8. Many seafood lovers are unaware that some of the fish they eat may have *POISON* in their scales.

16 filler sentences with no phoneme target

9. Shareholders sometimes take TOO much risk to make themselves rich
10. At the meeting the climatologists told the winery owners that they will NEVER survive if there's no rain
11. His new house is of EXACTLY the same height as the surrounding high rises
12. Anna's colleagues NEARLY fell down the stairs when they were getting off the train

13. After the earthquake our family had to SCAVENGE for food
14. Their new show was not good enough to AMAZE the audience
15. The giant ran towards the garden and DEVOURED all the flowers
16. Several folks from the village were DANCING in the streets
17. Magicians can use their cunning skills to CONTROL the audience's emotions
18. In Congolese culture newlyweds are NOT allowed to smile on their wedding day
19. To get rid of such a massive amount of snow an ELECTRIC shovel is more convenient
20. Construction workers often work in all KINDS of weather conditions
21. The dressmakers at the fashion firm used METAL as material for their couture gowns
22. Quite a few travellers were arrested after COCAINE was found in their luggage
23. Everyone is talking about the HUNTER who lost his way in the woods
24. More than a THOUSAND cars were sold last year even though the economy wasn't so good

Appendix B

Experimental Sentences in Mandarin (with rough IPA transcriptions)

1.
(a) 他们上星期去***爬山***踩了很多野花
(b) 他们上星期去*爬山*踩了**很多**野花

tʰa1 mən2 ʂaŋ4 ɕin1 tɕʰi1 tɕʰy4 **pʰa2 ʂan1** tsʰai3 lə5 **xən3 two1** jɛ3 xwa1
他 们 上 星 期 去 ***爬 山*** 踩 了 **很 多** 野 花

他们 上星期 去 ***爬山*** 踩-了 **多**
3.PL.M last week go ***CLIMB-MOUNTAIN*** tread-PFV **MANY**

野-花
wild-flowers

"They tread on a lot of wild flowers while out mountain-climbing last week"

2.
(a) 他想马上回家因为他的***朋友***想偷他的钱
(b) 他想马上回家因为他的*朋友*想偷他的**钱**

tʰa1 ɕjaŋ2 ma3 ʂaŋ4 xwei2 tɕja1 jin1 wei2 tʰa1 tɤ5 **pʰəŋ2 joʊ4** ɕjaŋ3 tʰou1 tʰa1 tɤ5 **tɕʰjɛn2**
他 想 马 上 回 家 因 为 他 的 ***朋 友*** 想 偷 他 的 **钱**

他 想 马上 回-家 因为 他-的 ***朋友***
3.s.M want immediately return-home because 3.s.M-GEN ***FRIEND***

想 偷 他-的 **钱**
want steal 3.s.M-GEN **MONEY**

"He wants to immediately return home because he suspects that his friend wants to steal his money"

3.
(a) 笑死人了, 这几位游客想穿***皮衣***在沙滩上溜达
(b) 笑死人了, 这几位游客想穿*皮衣*在**沙滩**上溜达

ɕjau4 si3 ɻən2 lɤ5 ʈʂɤ4 tɕi3 wei4 jou2 kʰɤ4 ɕjɛŋ3 ʈʂwaŋ1 **pʰi2 ji1** ʈʂai4 **ʂa1 tʰan1** ʂaŋ1 ljou1 dɤ5
笑 死 人 了, 这 几 位 游 客 想 穿 ***皮 衣*** 在 **沙滩** 上 溜 达

笑-死-人-了 这 几-位 游客 想 穿
laugh-die-people-VOC.PTCL DEM few-CLF tourist want wear

皮-衣 在 **沙滩** 上 溜达
LEATHER-JACKET PREP **BEACH** POST stroll

"How funny! These tourists want to wear their leather jackets while strolling in the beach"

4.

(a) 昨天我看见两个爱人在***苹果树***下偷偷地亲嘴
(b) 昨天我看见两个爱人在*苹果树*下偷偷地**亲嘴**

tswo2 tʰjɛn1 wo3 kʰan4 tɕʰjɛn4 ljaŋ3 kʰɤ4 ai4 ɻən2 ʦai4 **pʰiŋ2 kwo3 ʂu4** ɕja4 tʰou1 tʰou1
昨 天 我 看 见 两 个 爱 人 在 ***苹 果 树*** 下 偷 偷

dɤ5 tɕʰin1 tswei3
地 **亲 嘴**

昨天 我 看-见 两-个 爱人 在 ***苹果-树***
yesterday 1.s see-RES.COMP two-CLF lover PREP ***APPLE-TREE***

下 偷偷-地 **亲嘴**
POST secret-MOD **KISS**

"*Yesterday I saw two lovers secretly kissing under the apple tree*"

5.

(a) 没有人在中国能相信***葡萄***能制造香水
(b) 没有人在中国能相信*葡萄*能制造**香水**

mei2 jou3 ɻən2 tsai4 ʈʂʊŋ1kwo3 nəŋ2 ɕjɑŋ1 ɕin4 **pʰu2 tʰau5** nəŋ2 ʈʂɨ 4 tsau4 ɕjaŋ1 ʂwei3
没 有 人 在 中 国 能 相 信 ***葡 萄*** 能 制 造 **香 水**

没有 人 在 中国 能 相信 ***葡萄*** 能 制造 **香水**
NEG people PREP China can believe ***GRAPES*** can create **PERFUMES**

"*No one in China believes that grapes can be used to make perfumes*"

6.

(a) 我将家里的一套***盘子***送给我的偶像
(b) 我将家里的一套*盘子*送给我的**偶像**

wo3 tɕjaŋ1 tɕja1 li3 tɤ5 ji2 tʰau4 **pʰaŋ2 tsɨ5** sʊŋ4 kei3 wo3 tɤ5 ou3 ɕjaŋ4
我 将 家 里 的 一 套 ***盘 子*** 送 给 我 的 **偶 像**

我 将 家-里-的-一-套" ***盘子*** 送-给
1.s FUT home-POST-GEN-one-CLF ***PLATE*** give-RES.COMP

我-的 **偶像**
1.s-GEN **IDOL**

"*I shall give away my dinnerware as a present to my idol*"

7.

(a) 很多演员认为这***牌子***的鞋已经过时了

(b) 很多演员认为这*牌子*的鞋已经**过时**了

xəŋ3 twɔ1 jan3 ɥɛn2 ʐən4 wei2 ʈʂɤ4 **pʰai2 tsɨ5** tɤ5 ɕje2 ji2 tɕiŋ1 kwo4 ʂɨ2 lɤ5

很 多 演 员 认 为 这 ***牌 子*** 的 鞋 已 经 **过 时** 了

很多 演员 认为 这 ***牌子***-的 鞋 已经

Many actors think DEM ***BRAND***-GEN shoe already

过时-了

OUTDATED-PRF.PTCL

"*A lot of actors think that the shoes made by this brand are no longer in fashion*"

8.

(a) 听说村里那个长得像***螃蟹***的男孩要结婚

(b) 听说村里那个长得像*螃蟹*的男孩要**结婚**

tʰin1 ʂwo1 tsʰun1 li3 na4 kɤ5 ʈʂaŋ3 dɤ5 ɕjaŋ4 **pʰaŋ2 ɕɛ4** tɤ5 nan2 xai2 jau4 tɕjɛ2 xwən1

听 说 村 里 那 个 长 得 像 ***螃 蟹*** 的 男 孩 要 **结 婚**

听-说 明天 村-里 那-个 长-得 像

heard-say tomorrow village-POST ART-CLF look.V-COMP like

螃蟹-的 男孩 要 **结婚**

CRAB-MOD boy AUX.FUT **MARRY**

"*It has been rumoured that boy from the village who looks like a crab will get married tomorrow*"

9.

(a) 你可以看见他肚子***膨胀***得越来越大

(b) 你可以看见他肚子*膨胀*得越来越大

ni3 kɤ2 yi3 kʰan4 tɕʰjɛn4 tʰa1 tu4 tsɨ5 **pʰəŋ2 ʈʂaŋ4** tɤ5 ɥe4 lai2 ɥe4 da4

你 可 以 看 见 他 肚 子 ***膨 胀*** 得 越 来 越 **大**

你 可以 看-见 他 肚子 ***膨胀***-得

2.s can see-RES.COMP 3.s.M stomach ***SWOLLEN***-MOD

越来越 **大**

more and more **BIG**

"*You can see that his stomach is getting bigger and bigger*"

10.
(a) 我挺惊讶他会申请那套***便宜***的房子给自己住
(b) 我挺惊讶他会申请那套*便宜*的房子给**自己**住

wo3 tʰiŋ3 tɕiŋ1 ja4 tʰa1 xwei4 ʂəŋ1 tɕʰiŋ3 na4 tʰau4 **pʰjɛn2 ji5** tɤ5
我 挺 惊 讶 他 会 申 请 那 套 ***便 宜*** 的

faŋ2 tsɨ5 kei3 tsɨ4 tɕi3 tʂu4
房 子 给 **自 己** 住

我 挺 惊讶 他" 会 申请 那-套
1.s quite-MOD surprise 3.s.M AUX.FUT apply DEM-CLF

便宜-的 房子 给 **自己** 住
CHEAP-MOD house give **SELF** live

"*I am quite surprised that he will apply to live in that cheap house by himself*"

11.
(a) 没想到她干女儿的***脾气***能让她得癌症
(b) 没想到她干女儿的*脾气*能让她得**癌症**

mei2 ɕjaŋ3 dau4 tʰa1 kan1 ny3 ɚ2 tɤ5 **pʰi2 tɕʰi4** nəŋ2 ɻan4 tʰa1 tɤ3 ai2 tsən4
没 想 到 她 干 女 儿 的 ***脾 氣*** 能 让 她 得 **癌 症**

没 想-到 她 干-女儿-的 ***脾氣***
NEG think-RES.COMP 3.s.F adopted-daughter-GEN ***TEMPER***

能 让 她 得 **癌症**
can CAUS.V 3.s.F contract **CANCER**

"*Nobody would have thought that her adopted daughter's temper led her to have cancer*"

12.
(a) 身体虚弱的年轻人需要吃***排骨***来增加营养
(b) 身体虚弱的年轻人需要吃*排骨*来增加**营养**

ʂən1 tʰi3 ɕy1 ɻwo4 tɤ5 njɛn2 tɕʰiŋ1 ɻən2 ɕu1 jau4 tʂʰɨ1 **pʰai 2 ku3** lai2 tsəŋ1 tɕja1 jiŋ2 jaŋ3
身 体 虚 弱 的 年 轻 人 需 要 吃 ***排 骨*** 来 增 加 **营 养**

身体-虚弱-的 年轻-人 需要 吃 ***排骨*** 来 增加 **营养**
body-weak-COMP young-people need eat ***RIBS*** CNJ add **NUTRIENTS**.

"*Young people who are physically weak need to eat some ribs to gain more nutrients*"

13.
(a) 这些狗仔队能***破坏***总统的名声
(b) 这些狗仔队能*破坏***总统**的名声

ʈʂɤ4 ɕjɛ1 gou2 tsai3 twei4 nəŋ2 **pʰwo4 xwai4** tsʊŋ2 tʰʊŋ3 tɤ5 miŋ2 ʂəŋ1
这 些 狗 仔 队 能 ***破 坏*** **总 统** 的 名 声

这-些 狗仔队 能 ***破坏*** **总统**-的 名声
DEM-CLF paparazzi can ***RUIN*** **PRESIDENT**-GEN reputation

"*These paparazzi can ruin the president's reputation*"

14.
(a) 红楼梦里的姑娘长得***漂亮***因为她们吃过仙丹
(b) 红楼梦里的姑娘长得*漂亮*因为她们吃过**仙丹**

xʊŋ2 lou2 məŋ4 li3 tɤ5 ku1 njaŋ5 ʈʂaŋ3 tɤ5 **pʰau 4 ljɛn5** jin1 wei2 tʰa1 mən2
红 楼 梦 里 的 姑 娘 长 得 ***漂 亮*** 因 为 她 们

ʈʂʰɨ2 kwo4 ɕjɛn1 tan1
吃 过 **仙 丹**

红楼梦-里-的 姑娘 长-得 *漂亮*
Dream Red Mansion-POST-GEN maiden look.V-MOD ***BEAUTIFUL***

因为 她们 吃-过 **仙丹**
because 3.s.F.PLR eat-ASP.PTCL **DIVINE PILL**

"*The maidens from the novel "Dream of the Red Chamber" were beautiful because they once swallowed a divine pill*"

15.
(a) 我家大姐的行为像***叛徒***因为她取笑我们家的秘方
(b) 我家大姐的行为像*叛徒*因为她**取笑**我们家的秘方

wo3 tɕja1 da4 tɕjɛ3 tɤ5 ɕiŋ2 wei2 ɕjaŋ4 **pʰan4 tʰu2** jin1 wei2 tʰa1 tɕʰu3 ɕjau4
我 家 大 姐 的 行 为 像 *叛 徒* 因 为 她 **取 笑**

wo3 mən4 tɕja1 dɤ5 mi4 faŋ1
我 们 家 的 秘 方

我 家 大姐-的 行为 像 ***叛徒*** 因为 她 **取笑**
1.s family sister-GEN behaviour like ***TRAITOR*** because 2.s.F **MOCK**

我们-家-的 秘方
1.PL-family-GEN secret recipe

"*Our oldest sister acts like a traitor because she made a mockery of our family recipe*"

16.
(a) 时代杂志的分析家预测音乐会的*票*将要跌下来
(b) 时代杂志的分析家预测音乐会的*票*将要**跌**下来

ʂɨ2 tai4 tsa2 ʈʂɨ4 tɤ5 fən3 ɕi1 tɕja1 jy4 tsʰɤ4 jin1 ɥe4 xwei4 tɤ5 **pʰjau4** tɕjaŋ1 jau4 tjɛ2
时 代 杂 志 的 分 析 家 预 测 音 乐 会 的 ***票*** 将 要 **跌**

ɕja4 lai2
下 来

时代-杂志-的	分析家	预测	音乐会-的	***票***	将要
Times-Magazine-GEN	analysts	predict	concert-GEN	***TICKET***	shall.FUT

跌	下来
DOWN	CMPD.DIRECT.COMP

"*Analysts from the Times Magazine predict that the price of the concert will go down*"

17.
(a) 餐厅经理听到***炮响***都吓呆了
(b) 餐厅经理听到***炮***响都吓呆了

tsʰan1 tʰin1 tɕin1 li3 tʰiŋ1 tau4 **pʰau4 ɕjaŋ3** tou1 ɕja4 tai1 lɤ5
餐 厅 经 理 听 到 ***炮 响*** 都 吓 呆 了

餐厅-经理	听-到	***炮-响***	都
restaurant-manager	hear-RES.COMP	***CANNON-SOUND***	ADV

吓-呆-了
SCARE-STIFF-PRF

"*The restaurant manager was scared stiff after he heard a blast from the cannon*"

18.
(a) 有些护士喜欢向婴儿的*屁股*打针
(b) 有些护士喜欢向婴儿的*屁股*打**针**

jou3 ɕjɛ1 xu4 ʂɨ5 ɕi3 xwaŋ1 ɕjaŋ4 jin1 ɚ2 tɤ5 **pʰi4 ku5** ta3 tʂəŋ1
有 些 护 士 喜 欢 向 婴 儿 的 ***屁股*** 打 **針**

有-些	护士	喜欢	向	婴儿-的	***屁股***
EXIST.V-CLF	nurse	like	DIRECT.PREP	infant-GEN	***GLUTE***

打-**针**
apply-**INJECTION**

"*Some nurses prefer performing glute injections on toddlers*"

19.

(a) 李先生逛超市时看见一位***胖子***买红豆

(b) 李先生逛超市时看见一位*胖子*买**红豆**

li3 ɕjɛn1 ʂən1 kwaŋ4 ʈʂʰau1 ʂɨ4 ʂɨ2 kʰan4 tɕjɛn4 ji2 wei4 **pʰaŋ4 tsi5** mai3 xuŋ2 tou4

李 先 生 逛 超 市 时 看 见 一 位 ***胖 子*** 买 **紅 豆**

李先生	逛-超市-时	看-见	一-位	***胖子***
Mr. Li	stroll-market-when.CNJ	see-RES.COMP	one-CLF	***OBESE PERSON***

买	紅-豆
buy	**RED-BEAN**

"*While doing grocery shopping, Mr Li saw an obese guy buying red adzuki beans*"

20.

(a) 在罗马有三个***骗子***往我们的方向走

(b) 在罗马有三个*骗子*往**我们**的方向走

tsai4 lwo2 ma3 jou3 san1 kɤ4 **pʰjɛn4 tsi5** waŋ3 wo3 mən2 tɤ5 faŋ1 ɕjaŋ4 tsou3

在 罗 马 有 三 个 ***骗 子*** 往 **我 们** 的 方 向 走

在-罗马	有	三-个	***骗子***	往	**我们**-的-方向	走
PREP-Rome	EXIST.V	three-CLF	***SWINDLER***	PREP	**1.PL-GEN**-direction	go

"*When we were in Rome, three swindlers were walking in our direction*"

21.

(a) 住在山里的那位小伙子买了一个***铺头***在兰州

(b) 住在山里的那位小伙子买了一个*铺头*在**兰州**

ʈʂu4 tsai4 ʂan1 li3 tɤ5 na4 wei4 ɕjau2 xwo3 tsɨ5 mai3 lɤ5 ji2 kɤ4 **pʰu4 tʰou2**

住 在 山 里 的 那 位 小 伙 子 买 了 一 个 ***铺 头***

tsai4 lan2 ʈʂou1

在 **兰 州**

住-在	山-里-的	那-位	小伙子	买-了	一-个
live-POST	mountain-POST-GEN	DEM-CLF	young man	buy-PRF	one-CLF

铺头	在-**兰州**
STORE	PREP-**LANZHOU**

"*The young man who lived in the mountains bought a retail shop in Lanzhou*"

22.
(a) 小学生听到喊叫后***怕***得猫在桌子底下
(b) 小学生听到喊叫后*怕*得猫在**桌子**底下

ɕjau3 ɕye2 ʂəŋ1 tʰiŋ1 tau4 xan3 tɕjau4 xou4 **pʰa4** tɤ5 mau1 tsai4 tswo1 tsɨ5 ti3 ɕja4
小 学 生 听 到 喊 叫 后 ***怕*** 得 猫 在 **桌 子** 底下

小学生 听-到 喊叫-后 ***怕***-得
Primary-school students hear-RES.COMP scream-after.CNJ ***SCARED***-MOD
hide

猫-在 **桌子** 底-下
hide-POST **TABLE** under-DIRECT.COMP

"*After hearing someone screaming, the primary-school students were so scared that they hid under the table*"

23.
(a) 老奶奶每天一个人站在门前***盼望***她儿子从战争回家
(b) 老奶奶每天一个人站在门前*盼望*她儿子从**战争**回家

lau2 nai3 nai5 mei3 tʰjɛn1 ji2 kɤ4 ɻən2 ʈʂan4 tsai4 mən2 tɕʰjɛn2 **pʰan4 wɑŋ4** tʰa1 ɚ2 tsɨ5 tsʰʊŋ2
老 奶 奶 每 天 一 个 人 站 在 门 前 ***盼望*** 她 儿 子 从

ʈʂan4 ʈʂəŋ1 xwei2 tɕja1
战 争 回 家

老奶奶 每-天 一个人 站-在 门-前 ***盼望***' 她-儿子
Old lady every-day alone stand-POST door-front ***YEARN*** 2.s.F-son

从-**战争** 回-家
from.PREP-**WAR** return-home

"*Every day, the old lady stood in front of her doorstep and yearned for her son's return from war*"

24.
(a) 我很惊愕她开车***碰撞***了一只大象
(b) 我很惊愕她开车*碰撞*了一只**大象**

wo2 xən3 tɕiŋ1 ɤ4 tʰa1 kʰai1 ʈʂʰɤ1 **pʰəŋ4 ʈʂʰwaŋ4** lɤ5 ji4 ʈʂɨ1 ta4 ɕjaŋ4
我 很 惊 愕 她 开 车 ***碰 撞*** 了 一 只 **大象**

我 很-惊愕 她 开车 ***碰撞***-了 一-只 **大象**
1.s. very.INT-horrified 3.s.F drive ***COLLIDE***-PRF one-CLF **ELEPHANT**

"*I am very horrified that she collided with an elephant while she was driving*"

Filler Sentences

4 filler sentences with early occurrence of the phoneme target and their translations in English

1.
她想***陪着***她的母親去澳大利亚參加婚礼
"She wants to accompany her mother when they go to Australia for the wedding ceremony"

2.
啤酒有時能候让嗓子难受
"Beer can sometimes make your throat feel uncomfortable"

3.
在***派出所***我遇到了很多人
"At the police station I encountered a lot of people"

4.
我很***佩服***很多非常勇敢的哲学家
"I really admire many philosophers who are very brave"

4 filler sentences with late occurrence of the phoneme target

5.
這倆位园丁花三天三夜設計一個很华丽的***盆景***
"The two gardeners spent three days and nights designing a very beautiful bonsai tree"

6.
研究地理的工程師喜歡在松树的***旁边***休息
"The engineers who do research in geology prefer to take a rest next to the pine tree"

7.
我们的手上粘滿了肥皂***泡沫***
"Our hands are filled with soap bubbles"

8.
地震災民現在对食物要求很迫切
"The earthquake victims currently have a very *urgent* need for food"

16 filler sentences with no phoneme target

9.
真的沒看出來他的艺术眼光有**那麼**差
"*I have never really noticed that his taste for art can be that bad*"

10.
大公司的会计師**老是**埋怨他們公司的经济困难
"*Accountants from big companies are always complaining about their company's financial problems*"

11.
我在俄罗斯下火車**差点儿**跌倒了 (note: focused word is disyllabic)
"*I almost fell down when I was getting off the train in Russia*"

12.
这塔**正好**跟附近的楼一样高
"*This tower is of exactly the same height as the surrounding buildings*"

13.
药剂师知道怎么**混和**中药和其他的香料来提高药的味道
"*Pharmacists know how to mix Chinese herbal medicine with other ingredients to enhance the medicine's flavour*"

14.
安娜卡列尼娜曾经和渥伦斯基**说过**她的生命很痛苦
"*Anna Karenina did tell Vronsky that she has suffered a lot in her life*"

15.
调查人员发现房间的温度很**热**
"*The investigator discovered that the room's temperature was very hot*"

16.
机场海关**没有**没收走私的伪造手袋
"*At the airport the customs officers did not confiscate the smugglers' counterfeit handbags*"

17.
工会说建筑工人总是在**危险**的环境工作
"*The unions said the construction workers' are working under very dangerous conditions*"

18.
快乐的夫妇从来没有在公共场合吵过架
"*Couples who are happily married would never quarrel in public*"

生存在城市的麻雀经常喜欢从垃圾桶掏食物吃
"*Finches living in big cities often like to scavenge for food from trash cans*"

20.
我知道有些人喜欢在**酒店**开会
"*I know there are people who prefer setting up conferences in hotels*"

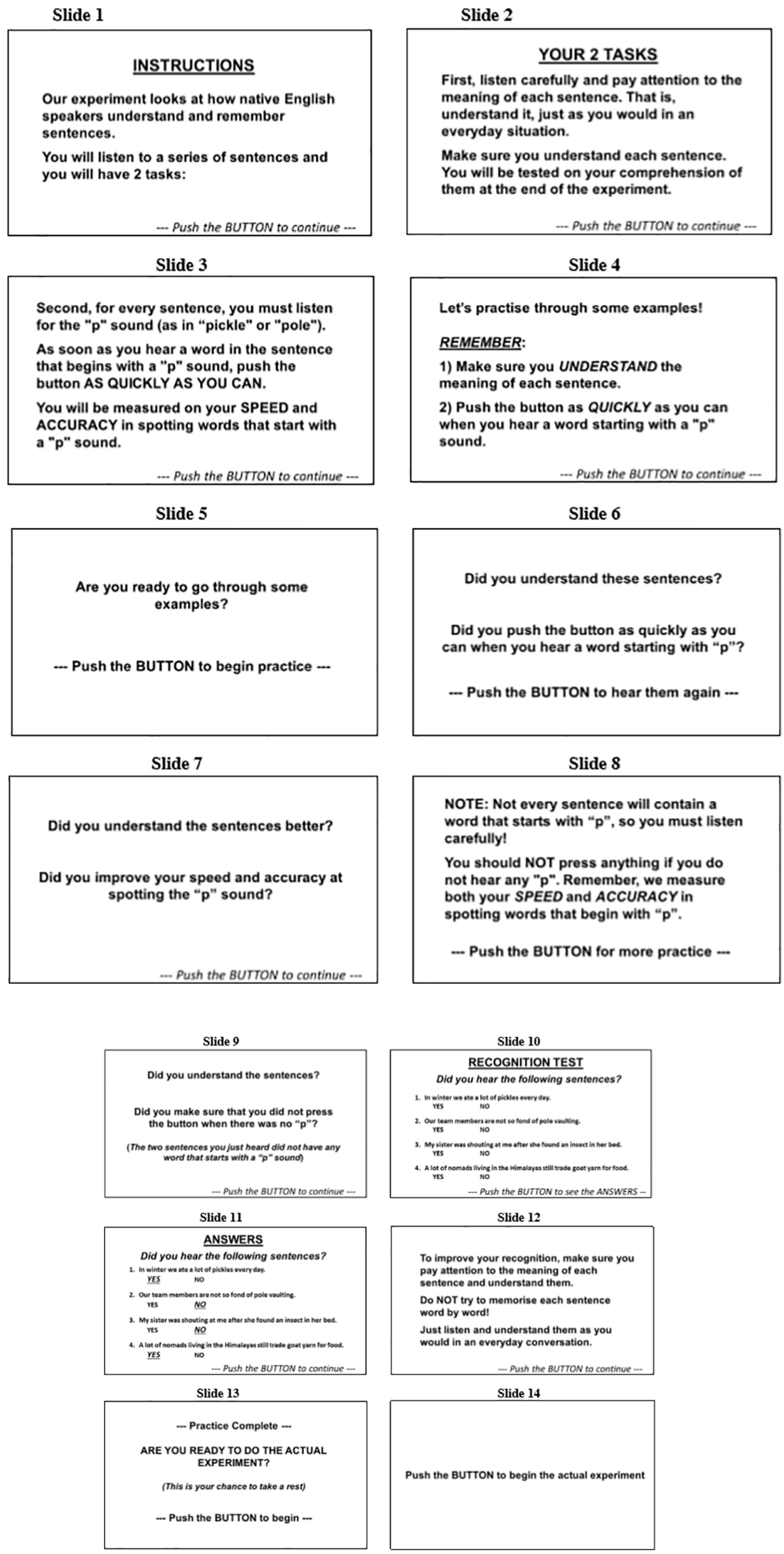
21.
所有的律师同意马路的卫生是**清洁工**的责任
"*All the lawyers unanimously agree that the hygiene in our streets is the cleaners' responsibility*"

22.
厉害害的魔术师能用他的**手法**来影响其他人的心情
"*Skilled magicians can use his legerdemain to influence other people's mood*"

23.
没见过一个**模特**有那么多的学问
"*I have never met a model who is that knowledgeable*"

24.
艺术馆失踪了**一千**幅画因为晚上值勤的员工光顾看电视
"*A thousand paintings are missing at the art gallery because the night staff were watching TV*"

Appendix C


Slide 1
INSTRUCTIONS
Our experiment looks at how native English speakers understand and remember sentences.
You will listen to a series of sentences and you will have 2 tasks:
--- Push the BUTTON to continue ---
Slide 2
YOUR 2 TASKS
First, listen carefully and pay attention to the meaning of each sentence. That is, understand it, just as you would in an everyday situation.
Make sure you understand each sentence. You will be tested on your comprehension of them at the end of the experiment.
--- Push the BUTTON to continue ---
Slide 3
Second, for every sentence, you must listen for the "p" sound (as in "pickle" or "pole").
As soon as you hear a word in the sentence that begins with a "p" sound, push the button AS QUICKLY AS YOU CAN.
You will be measured on your SPEED and ACCURACY in spotting words that start with a "p" sound.
--- Push the BUTTON to continue ---
Slide 4
Let's practise through some examples!
REMEMBER:
1) Make sure you UNDERSTAND the meaning of each sentence.
2) Push the button as QUICKLY as you can when you hear a word starting with a "p" sound.
--- Push the BUTTON to continue ---
Slide 5
Are you ready to go through some examples?
--- Push the BUTTON to begin practice ---
Slide 6
Did you understand these sentences?
Did you push the button as quickly as you can when you hear a word starting with "p"?
--- Push the BUTTON to hear them again ---
Slide 7
Did you understand the sentences better?
Did you improve your speed and accuracy at spotting the "p" sound?
--- Push the BUTTON to continue ---
Slide 8
NOTE: Not every sentence will contain a word that starts with "p", so you must listen carefully!
You should NOT press anything if you do not hear any "p". Remember, we measure both your SPEED and ACCURACY in spotting words that begin with "p".
--- Push the BUTTON for more practice ---
Slide 9
Did you understand the sentences?
Did you make sure that you did not press the button when there was no "p"?
(The two sentences you just heard did not have any word that starts with a "p" sound)
--- Push the BUTTON to continue ---
Slide 10
RECOGNITION TEST
Did you hear the following sentences?
1. In winter we ate a lot of pickles every day.
YES NO
2. Our team members are not so fond of pole vaulting.
YES NO
3. My sister was shouting at me after she found an insect in her bed.
YES NO
4. A lot of nomads living in the Himalayas still trade goat yarn for food.
YES NO
--- Push the BUTTON to see the ANSWERS --
Slide 11
ANSWERS
Did you hear the following sentences?
1. In winter we ate a lot of pickles every day.
YES NO
2. Our team members are not so fond of pole vaulting.
YES NO
3. My sister was shouting at me after she found an insect in her bed.
YES NO
4. A lot of nomads living in the Himalayas still trade goat yarn for food.
YES NO
--- Push the BUTTON to continue ---
Slide 12
To improve your recognition, make sure you pay attention to the meaning of each sentence and understand them.
Do NOT try to memorise each sentence word by word!
Just listen and understand them as you would in an everyday conversation.
--- Push the BUTTON to continue ---
Slide 13
--- Practice Complete ---
ARE YOU READY TO DO THE ACTUAL EXPERIMENT?
(This is your chance to take a rest)
--- Push the BUTTON to begin ---
Slide 14
Push the BUTTON to begin the actual experiment

Appendix D

Appendix E

Recognition test

Did you hear the following sentences? Please circle your response.

1) The very peak of his acting career was not when he received the Golden Globe's award.
YES NO

2) After the earthquake, our family had to scavenge for food.
YES NO

3) That summer four years ago, I ate roast peanuts for every meal.
YES NO

4) Most of the jurors find it odd that the millionaire was pardoned after the verdict
YES NO

5) No one in the farm was surprised to see the parrot when it sang in German.
YES NO

6) Down on the farm we were amused to see a parrot who could sing in French.
YES NO

7) The porter stole a tourist's suitcase while he was working in the lobby.
YES NO

8) Three fairies appeared in my grandmother's backyard yesterday.
YES NO

9) Magicians can use their cunning skills to control the audience's emotions.
YES NO

10) Everyone is talking about the hunter who lost his way in the woods.
YES NO

11) The teacher called her partner and told him that their daughter was sent home from school.
YES NO

12) The giant ran towards the gate and devoured all the flowers.
YES NO

13) The countess's dogs are very spoiled because they eat caviar every morning.
YES NO

14) Most of the farmers in the village say they like to dance when they hear music.
YES NO

15) Unfortunately the geologist didn't have enough time to polish all his minerals for the show.
YES NO

16) Several of my friends from Wall Street are now in danger of losing their wealth.
YES NO

17) Some students always party, even when they should be revising for the exams.
YES NO

18) The soldiers couldn't break the code the foreigners had used.
YES NO

19) All the contestants were in a state of panic when their names were called out.
YES NO

20) The dressmakers at the fashion firm used metal as material for their couture gowns.
YES NO

Appendix F

识别测试

您有没有听到下面的这些句子？请在答案上画圈

1 *我认为这牌子的衣服还是太土了*
有 沒有

2 我在俄罗斯下火车差点儿跌倒了
有 沒有

3 没有人在中国能相信葡萄能制造香水
有 沒有

4 我挺惊讶他会申请那套便宜的房子给自己住
有 沒有

5 大家都很高兴因为那个长得像螃蟹的女孩要结婚
有 沒有

6 听说村里那个长得像螃蟹的男孩要结婚
有 沒有

7 我对我的朋友很失望因为他们现在都很贪财
有 沒有

8 这些游客在市场买了很多西瓜
有 沒有

9 厉害的魔术师能用他的手法来影响其他人的心情
有 沒有

10 机场海关沒有沒收走私的伪造手袋
有 沒有

11 很多人喜欢用大盘子吃意粉
有 沒有

12 所有的律师同意马路的卫生是清洁工的责任

有 沒有

13 我的大哥在香港岛买了一套很小的公寓

有 沒有

14 我的同事经常说我应该讲话大点声

有 沒有

15 昨天我看见俩位爱人在苹果树下偷偷地亲嘴

有 沒有

16 有些老人依靠卖奶粉来生存

有 沒有

17 我的家人喜欢在加拿大爬山多过游泳

有 沒有

18 今天我把我的两千块杯子送给了我的最崇拜的歌星

有 沒有

19 我家大姐的行为像叛徒因为她取笑我们家的秘方

有 沒有

20 艺术馆失踪了一千幅画因为晚上值勤的员工光顾看电视

有 沒有

Appendix G

Experiment 1

Random effects for linear mixed-effects model analyses for RT (Box-Cox converted). Analyses were based on 1088 datapoints (46 participants and 24 items).

Random effects		Variance	SD
Participant	(Intercept)	3.743e-07	0.0006
	Language	5.030e-06	2.243e-03
	Gender	7.706e-06	0.0028
	Prosodic context	2.694e-05	5.190e-03
	Language × Prosodic context	2.663e-05	5.161e-03
	Trial	2.690e-05	5.186e-03
	Trial × Prosodic context	2.697e-05	5.193e-03
	Trial × Language	2.773e-05	5.266e-03
Prosodic context | Participant	(Intercept)	2.274e-06	0.0015
	Language	2.314e-06	1.521e-03
	Gender	4.805e-06	0.0022
	Prosodic context	1.172e-06	1.083e-03
	Language × Prosodic context	1.113e-06	1.055e-03
	Trial	1.164e-06	1.079e-03
	Trial × Prosodic context	1.181e-06	1.087e-03
	Trial × Language	1.301e-06	1.140e-03
Item	(Intercept)	9.546e-07	0.0010
	Language	1.336e-09	3.655e-05
	Gender	1.292e-06	0.0011
	Prosodic context	1.400e-06	1.183e-03
	Language × Prosodic context	0.000e+00	0.000e+00
	Trial	1.241–06	1.114e-03
	Trial × Prosodic context	0.000e+00	0.000e+00
	Trial × Language	2.523e-06	1.588e-03

(continued on next page)

Experiment 1 (*continued*)

Random effects		Variance	SD
Prosodic context | Item	(Intercept)	8.867e-06	0.0030
	Language	8.995e-06	2.999e-03
	Gender	3.710e-06	0.0019
	Prosodic Context	6.606e-07	8.128e-04
	Language × Prosodic context	1.509e-06	1.229e-03
	Trial	7.911e-06	8.894e-04
	Trial × Prosodic context	1.307e-06	1.143e-03
	Trial × Language	5.306e-08	2.303e-04

Appendix H

Experiment 1

Fixed effects for linear mixed-effects model analyses for RT (Box-Cox converted). Analyses were based on 1088 datapoints (46 participants and 24 items).

Fixed effects	β	SE
(Intercept)	1.623	8.171e-04
Language	0.005	0.0020
Gender	2.147e-03	1.361e-03
Prosodic context	3.101e-03	6.713e-04
Language × Prosodic context	−0.002	0.0021
Trial	3.736e-03	8.934e-04
Trial × Prosodic context	−0.0005	0.0008
Trial X Language	−0.0029	0.0009
English listeners		
(Intercept)	1.90	0.0018
Prosodic context	0.0068	0.0019
Preceding duration × Prosodic context	0.0086	0.0065
Pretarget interval duration × Prosodic Context	−0.0195	0.0107
Mean F0 × Prosodic context	0.0172	0.0208
Maximum F0 × Prosodic context	0.0180	0.0146
F0 range × Prosodic context	6..735e-03	4.771e-03
Mean intensity × Prosodic context	7.689e-03	4.463e-02
Maximum intensity × Prosodic context	7.456e-03	5.712e-02
Intensity range × Prosodic context	−0.0075	0.0074
Mandarin listeners		
(Interval)	1.409	0.006
Prosodic context	1.340e-03	5.005e-04
Preceding duration × Prosodic context	−1.638e-03	2.644e-03
Pretarget interval duration × Prosodic context	1.140e-03	1.790e-03
Mean F0 × Prosodic context	0.0084	0.0054
Maximum F0 × Prosodic context	0.0076	0.0075
F0 range × Prosodic context	1.609e-04	1.561e-03
Mean intensity × Prosodic context	0.0163	0.0071
Maximum intensity × Prosodic context	0.0191	0.0074
Intensity range × Prosodic context	5.180e-04	1.714e-03

Appendix I

Experiment 2

Random participant effects for linear mixed-effects model analyses for RT (Box-Cox converted). Analyses were based 548 datapoints (46 participants and 24 items).

Random effects		Variance	SD
Participant	(Intercept)	1.157e-05	0.0034
	Prosodic context	1.402e-05	3.745e-03
	Participation in experiment 1	1.092e-05	3.305e-03
	Length of stay	4.111e-08	0.0002
	Post-test recognition scores	1.638e-05	4.047e-03
	Preceding duration × Prosodic context	2.572e-06	0.0016
	Pretarget interval duration × Prosodic context	4.363e-05	0.0066
	Mean F0 × Prosodic context	6.539e-06	0.0026
	Maximum F0 × Prosodic context	2.861e-05	5.349e-03
	F0 range × Prosodic context	7.584e-06	2.754e-03
	Mean intensity × Prosodic context	8.007e-06	2.830e-03
	Maximum intensity × Prosodic context	1.670e-05	4.087e-03
	Intensity range × Prosodic context	1.660e-05	4.074e-03
Prosodic Context | Participant	(Intercept)	1.453e-05	0.0038
	Prosodic context	1.333e-05	3.651e-03

(*continued on next page*)

Experiment 2 (*continued*)

Random effects		Variance	SD
	Participation in Experiment 1	1.449e-05	3.807e-03
	Length of stay	1.453e-05	0.00381
	Post-test recognition scores	1.443e-05	3.799e-03
	Preceding duration × Prosodic context	1.336e-05	3.656e-03
	Pretarget interval duration × Prosodic context	1.337e-05	0.0036
	Mean F0 × Prosodic context	1.347e-05	0.0037
	Maximum F0 × Prosodic context	1.352e-05	3.677e-03
	F0 Range × Prosodic context	1.335e-05	3.654e-03
	Mean intensity × Prosodic context	1.341e-05	3.662e-03
	Maximum intensity × Prosodic context	1.327e-05	3.643e-03
	Intensity range × Prosodic context	1.348e-05	3.671e-03

Appendix J

Experiment 2
Random item effects for linear mixed-effects model analyses. Analyses were based on 548 datapoints (24 participants and 24 items).

Random effects		Variance	SD
Item	(Intercept)	1.191e-07	0.0035
	Prosodic context	1.427e-10	1.195e-05
	Participation in Experiment 1	8.207e-09	9.059e-05
	Length of stay	1.179e-08	0.0001
	Post-test recognition scores	1.638e-05	4.047e-03
	Preceding duration × Prosodic context	1.618e-10	1.272e-05
	Pretarget interval duration × Prosodic context	0.000e+00	0.0000
	Mean F0 × Prosodic context	7.427e-08	0.0003
	Maximum F0 × Prosodic context	3.441e-09	5.866e-05
	F0 Range × Prosodic context	3.585e-13	5.988e-07
	Mean intensity × Prosodic context	5.093e-11	7.136e-06
	Maximum intensity × Prosodic context	1.519e-11	3.897e-06
Prosodic context | Item	Intensity range × Prosodic context	2.615e-11	5.425e-03
	(Intercept)	5.288e-06	0.0023
	Prosodic context	4.947e-06	2.224e-03
	Participation in Experiment 1	5.192e-06	2.279e-03
	Length of stay	5.214e-06	0.0023
	Post-test recognition scores	5.147e-05	2.269e-03
	Preceding duration × Prosodic context	4.154e-06	2.038e-03
	Pretarget interval duration × Prosodic context	6.003e-06	0.0025
	Mean F0 × Prosodic context	5.128e-06	0.0026
	Maximum F0 × Prosodic context	5.329e-06	2.308e-03
	F0 Range × Prosodic context	5.392e-06	2.322e-03
	Mean intensity × Prosodic context	4.966e-06	2.228e-03
	Maximum intensity × Prosodic context	5.592e-06	2.365e-03
	Intensity range × Prosodic context	4.376e-06	2.092e-03

Appendix K

Experiment 2
Fixed effects for linear mixed-effects model analyses for RT (Box-Cox converted). Analyses were based on 548 datapoints (24 participants and 24 items).

Fixed effects	β	SE
(Intercept)	1.629	0.0011
Prosodic context	0.0011	0.0011
Participation in Experiment 1	−0.0030	0.0019
Length of stay	−0.0008	0.0013
Post-test recognition scores	0.0139	0.0079
Preceding duration × Prosodic context	−0.0034	0.0032
Pretarget interval duration × Prosodic context	−0.0062	0.0060
Mean F0 × Prosodic context	0.0056	0.0106
Maximum F0 × Prosodic context	0.0048	0.0084
F0 Range × Prosodic context	8.186e-03	2.503e-03
Mean Intensity × Prosodic context	0.0046	0.0229
Maximum Intensity × Prosodic context	0.0144	0.0307
Intensity range × Prosodic context	0.0021	0.0035

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104311.

References

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition, 6*, 81–96. https://doi.org/10.1017/S1366728903001056.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1.

Baese-Berk, M., Heffner, C., Dilley, L., Pitt, M., Morrill, T., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science, 25*, 155–1546. https://doi.org/10.1177/0956797614533705.

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1563–1571. https://doi.org/10.1037/a0032186.

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

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30*, 644–663. https://doi.org/10.1016/0749-596X(91)90030-N.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.). *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 167–200). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro, & O.-S. Bohn (Eds.). *Second language speech learning* (pp. 13–34). Amsterdam: John Benjamins.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech, 38*, 365–391. https://doi.org/10.1177/002383099503800403.

Birch, S., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language, 34*, 232–267. https://doi.org/10.1006/jmla.1995.1011.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics, 18*, 37–49.

Blutner, R., & Sommer, R. (1988). Sentence processing and lexical access: The influence of the focus-identifying task. *Journal of Memory and Language, 27*, 359–367. https://doi.org/10.1016/0749-596X(88)90061-7.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. retrieved 8 September 2018 from http://www.praat.org/Version 6.0.43, 2018.

Box, G., & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological), 26*, 211–252.

Braun, B., Asano, Y., & Dehé, N. (2018). When (not) to look for contrastive alternatives: The role of pitch accent type and additive particles. *Language and Speech, 62*, 751–778. https://doi.org/10.1177/0023830918814279.

Braun, B., & Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes, 25*, 1024–1043. https://doi.org/10.1080/01690960903036836.

Breen, M., Dilley, L. C., Devin McAuley, J., & Sanders, L. D. (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech. *Language, Cognition and Neuroscience, 29*, 1131–1146. https://doi.org/10.1080/23273798.2014.894642.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*, 93–125. https://doi.org/10.1016/S0010-0277(96)00719-6.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychometric Bulletin and Review, 18*, 1189–1196. https://doi.org/10.3758/s13423-011-0167-9.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 306–323. https://doi.org/10.1037/a0038689.

Brunellière, A., Auran, C., & Delrue, L. (2019). Does the prosodic emphasis of sentential context cause deeper lexical-semantic processing? *Language, Cognition and Neuroscience. 34*(1), 29–42. https://doi.org/10.1080/23273798.2018.1499945.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. N. Li (Ed.). *Subject and topic* (pp. 27–55). New York: Academic Press.

Chávez-Peón, M. (2010). *The interaction of metrical structure, tone, and phonation types in Quiaviní Zapotec. Ph.D. dissertation*Vancouver, Canada: University of British Columbia.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in standard Chinese. *Journal of Phonetics, 36*, 724–746. https://doi.org/10.1016/j.wocn.2008.06.003.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*, 221–268. https://doi.org/10.1080/016909698386528.

Connell, B. (2017). Tone and intonation in Mambila. In L. J. Downing, & A. Rialland (Eds.). *Intonation in African tone languages* (pp. 132–166). Berlin: De Gruyter.

Connell, K., Hüls, S., Martínez-García, M. T., Qin, Z., Shin, S., Yan, H., & Tremblay, A. (2018). English learners' use of segmental and suprasegmental cues to stress in lexical access: An eye-tracking study. *Language Learning,* 1–34. https://doi.org/10.1111/lang.12288.

Cutler, A. (1976). Phoneme monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics, 20*, 55–60. https://doi.org/10.3758/BF03198706.

Cutler, A., & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics, 29*, 217–224. https://doi.org/10.3758/BF03207288.

Cutler, A., Demuth, K., & McQueen, J. M. (2002). Universality versus language-specificity in listening to running speech. *Psychological Science, 13*, 258–262. https://doi.org/10.1111/1467-9280.00447.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition, 7*, 49–59. https://doi.org/10.1016/0010-0277(79)90010-6.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech, 20*, 1–10. https://doi.org/10.1177/002383097702000101.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*, 292–314. https://doi.org/10.1016/S0749-596X(02)00001-3.

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. https://doi.org/10.1037/0096-1523.28.1.218.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics, 32*, 493–516. https://doi.org/10.1016/j.wocn.2004.05.002.

DiCanio, C., Benn, J., & García, R. C. (2018). The phonetics of information structure in Yoloxóchitl Mixtec. *Journal of Phonetics, 68*, 50–68. https://doi.org/10.1016/j.wocn.2018.03.001.

DiCanio, C., & Hatcher, R. (2018). On the non-universality of intonation: Evidence from Triqui. *Journal of the Acoustical Society of America, 144*. https://doi.org/10.1121/1.5068494.

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language, 63*, 274–294. https://doi.org/10.1016/j.jml.2010.06.003.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*, 294–311. https://doi.org/10.1016/j.jml.2008.06.006.

Dilley, L. C., Morrill, T., & Banzina, E. (2013). New tests of the distal speech rate effect: Examining cross-linguistic generalizability. *Frontiers in Language Sciences, 4*, 1–13. https://doi.org/10.3389/fpsyg.2013.01002.

Dilley, L. C., & Pitt, M. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*, 1664–1670. https://doi.org/10.1177/0956797610384743.

Dwyer, M. (2004). More is better: The impact of study abroad program duration. *Frontiers: The Interdisciplinary Journal of Study Abroad, 10*, 151–163.

Félix-Brasdefer, J. (2004). Interlanguage refusals: Linguistic politeness and length of residence in the target community. *Language Learning, 54*, 587–653.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.). *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Fowler, C. A., & Housum, J. (1987). Talker's signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504. https://doi.org/10.1016/0749-596X(87)90136-7.

Fraundorf, S., Watson, D., & Benjamin, A. (2010). Recognition memory reveals just how contrastive contrastive accenting really is. *Journal of Memory & Language, 63*, 367–386. https://doi.org/10.1016/j.jml.2010.06.004.

Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America, 104*, 505–510. https://doi.org/10.1121/1.413004.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., ... Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language, 84*, 318–336. https://doi.org/10.1016/S0093-934X(02)00505-9.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*, 613–656. https://doi.org/10.1080/016909697386646.

Gómez, D. M., Berent, I., Benavides-Varela, S., Bion, R. A. H., Cattarossi, L., Nespor, M., & Mehler, J. (2014). *Language. Universals at birth. Proceedings of the National Academy of. Sciences. 111*, 5341–5837. https://doi.org/10.1073/pnas.1318261111.

Gotzner, N., Spalek, K., & Wartenburger, I. (2013). How pitch accents and focus particles affect the recognition of contextual alternatives. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.). *Proceedings of the 35th annual meeting of the cognitive science society* (pp. 2434–2440). Austin, TX: Cognitive Science Society.

Grassmann, S., & Tomasello, M. (2007). Two-year-olds use primary sentence accent to learn new words. *Journal of Child Language, 34*, 677–687.

Grassmann, S., & Tomasello, M. (2010). Prosodic stress on a word directs 24-month-olds' attention to a contextually new referent. *Journal of Pragmatics, 42*, 3098–3105.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics, 19*, 377–419. https://doi.org/10.1017/S0022226700007799.

Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. In B. Yuan, T. Huang, & X. Tang (Vol. Eds.), *Proceedings of The Sixth International Conference on Spoken Language Processing. Vol. I. Proceedings of The Sixth International Conference on Spoken Language Processing* (pp. 91–94). Beijing, China: China Military Friendship Publish.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies.* Chicago, IL: Chicago University Press.

Heffner, C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes, 28*, 1275–1302. https://doi.org/10.1080/01690965.2012.672229.

Howell, P. (1993). Cue trading in the production and perception of vowel stress. *Journal of the Acoustical Society of America, 94*, 2063–2073. https://doi.org/10.1121/1.407479.

Hsu, C.-H., Evans, J. P., & Lee, C.-Y. (2015). Brain responses to spoken F0 changes: Is H special? *Journal of Phonetics, 51*, 82–92. https://doi.org/10.1016/j.wocn.2015.02.003.

Ife, A., Vives, G., & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics, 4*, 55–84.

Ip, M. H. K., & Cutler, A. (2016). Cross-language data on five types of prosodic focus. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.). *Proceedings of Speech Prosody 2016* (pp. 330–334). . Boston, USA 10.21437/SpeechProsody.2016-68.

Ip, M. H. K., & Cutler, A. (2018). Cue equivalence in prosodic entrainment for focus detection. In J. Epps, J. Wolfe, J. Smith, & C. Jones (Eds.). *Proceedings of the 17th Australasian international conference on speech science and technology* (pp. 153–156). Canberra, Australia: ASSTA. Retrieved from https://pure.mpg.de/rest/items/item_2639790_4/component/file_3012564/content.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama, & K. Maekawa (Eds.). *Casual speech: Data and analysis* (pp. 29–54). Tokyo, Japan: The National Institute for Japanese Language.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language. 32*, 402–420. https://doi.org/10.1006/jmla.1993.1022.

Karlsson, A., House, D., Svantesson, J., & Tayanin, D. (2010). Influence of lexical tones on intonation in Kammu. In W. Hess (Ed.). *Proceedings of the 12th Annual Conference of the International Speech Communication Association* (pp. 1740–1743). Japan: Makuhari.

Kember, H., Choi, J., Yu, J., & Cutler, A. (2019). The processing of linguistic prominence. *Language and Speech.*. https://doi.org/10.1177/0023830919880217.

Kleinschmidt, D. & Jaeger, T. F. (2011) A Bayesian belief updating model of phonetic recalibration and selective adaptation. In *Proceedings of the Cognitive Modeling and Computational Linguistics Workshop* (pp. 10-19). ACL, Portland, OR, June 23rd, 10–19.

Kratochvil, P. (1998). Intonation in Beijing Chinese. In D. Hirst, & A. Di Cristo (Eds.). *Intonation systems: A survey of twenty languages* (pp. 417–431). Cambridge, UK: Cambridge University Press.

Kügler, F. (2017). Tone and intonation in Akan. In L. J. Downing, & A. Rialland (Eds.). *Intonation in African tone languages* (pp. 89–129). Berlin: Mouton de Gruyter.

Kuijpers, C., & van Donselaar, W. (1998). The influence of rhythmic context on schwa epenthesis and schwa deletion in Dutch. *Language and Speech, 41*, 87–108. https://doi.org/10.1177/002383099804100105.

Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience, 33*, 992–1004. https://doi.org/10.1080/23273798.2018.1435894.

Lai, W., & Dilley, L. C. (2016). Cross-linguistic generalization of the distal rate effect: Speech rate in context affects whether listeners hear a function word in Chinese Mandarin. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.). *Proceedings of Speech Prosody 2016* (pp. 1124–1128). . https://doi.org/10.21437/SpeechProsody.2016-231 (Boston, USA).

Laniran, Y. O., & Clements, G. N. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics, 31*, 203–250. https://doi.org/10.1016/S0095-4470(02)00098-0.

Lee, A., Chiu, F., & Xu, Y. (2017). Focus perception in Japanese: Effects of focus location and accent condition. *Proceedings of Meetings on Acoustics, 29*, 60007. https://doi.org/10.1121/2.0000441.

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics, 53*, 157–165. https://doi.org/10.1080/01690960701728261.

Lee, Y.-C., Wang, T., & Liberman, M. (2016). Production and perception of tone 3 focus in Mandarin Chinese. *Frontiers in Psychology, 7*, 1–13. https://doi.org/10.3389/fpsyg.2016.01058.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT Press.

Li, X.-Q., & Ren, G.-Q. (2012). How and when accentuation influences temporally selective attention and subsequent semantic processing during on-line spoken language comprehension: An ERP study. *Neuropsychologia, 50*, 1882–1894. https://doi.org/10.1016/j.neuropsychologia.2012.04.013.

Lieberman, P. (1963). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America, 32*, 451–454. https://doi.org/10.1121/1.1908095.

Lin, C. Y., Wang, M. I. N., Idsardi, W. J., & Xu, Y. I. (2014). Stress processing in Mandarin and Korean second language learners of English. *Bilingualism: Language and Cognition, 17*, 316–346. https://doi.org/10.1017/s1366728913000333.

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica, 62*, 70–87. https://doi.org/10.1159/000090090.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech, 104*, 109–138. https://doi.org/10.1177/00238309040470020101.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyze reaction time data. *Frontiers in Psychology, 30*, 1171. https://doi.org/10.3389/fpsyg. 2015.01171.

Mattys, S., & Samuel, A. G. (2000). Implications of stress-pattern differences in spoken-word recognition. *Journal of Memory and Language, 42*(4), 571–596. https://doi.org/10.1006/jmla.1999.2696.

Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 960–977. https://doi.org/10.1037/0096-1523.33.4.960.

McAllister, J. (1991). The processing of lexically stressed syllables in read and spontaneous speech. *Language and Speech, 34*, 1–26. https://doi.org/10.1177/002383099103400101.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21–46. https://doi.org/10.1006/jmla.1998.2568.

Mennen, I. (2004). Bi-directional interference in the intonation of. Dutch speakers of Greek. *Journal of Phonetics, 32*, 54–563. https://doi.org/10.1016/j.wocn.2004.02.002.

Morrill, T. H., Dilley, L. C., McAuley, J., & Pitt, M. A. (2014). Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition, 131*, 69–74. https://doi.org/10.1016/j.cognition.2013.12.006.

Nazzi, T., & Cutler, A. (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics, 5*, 25–47. https://doi.org/10.1146/annurev-linguistics-011718-011919.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue Linguaggio, 2*, 203–230.

Nolan, F. (2006). Intonation. In B. Aarts, & A. McMahon (Eds.). *Handbook of English linguistics* (pp. 433–457). Oxford, UK: Blackwell Publishing Ltd.

Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology, 53*, 146–193. https://doi.org/10.1016/j.cogpsych.2006.03.001.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299–325. https://doi.org/10.1017/S0140525X00003241.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*, 191–243. https://doi.org/10.1006/cogp.1997.0671.

Ouyang, I. C., & Kaiser, E. (2013). Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience, 30*, 57–72. https://doi.org/10.1080/01690965.2013. 805795.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal, 84*, 372–389. https://doi.org/10.1111/0026-7902.00075.

Pierrehumbert, J. (1999). Prosody and intonation. In R. A. Wilson, & F. C. Keil (Eds.). *MIT encyclopedia of cognitive science* (pp. 679–682). Cambridge, MA: MIT Press.

Pisoni, D. B., & Luce, P. A. (1987). *The Psychophysics of Speech Perception (pp 155–172) NATO ASI series book series (ASID, volume 39)*.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532. https://doi.org/10.1037/0033-2909.114.3.510.

Reed, M., & Michaud, C. (2014). Intonation in research & practice: The importance of metacognition. In M. Reed, & J. M. Levis (Eds.). *The handbook of English pronunciation.* New York: Wiley-Blackwell.

Remijsen, B. (2002). Lexically contrastive accent and lexical tone in Ma'ya. In C. Gussenhoven, & N. Warner (Eds.). *Laboratory phonology 7* (pp. 585–614). Berlin: Mouton de Gruyter.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-olds. *Science, 274*, 1926–1928. https://doi.org/10.1126/science.274.5294.1926.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*, 348–351. https://doi.org/10.1111/1467-9280.00364.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide.* Pittsburgh: Psychology Software Tools Inc.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowsky, M. (1982). Automatic access of the meaning of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*, 489–537. https://doi.org/10.1016/0010-0285(82)90017-2.

Shaw, J. A., Best, C. T., Docherty, G., Evans, B. G., Foulkes, P., Hay, J., & Mulak, K. E. (2018). Resilience of English vowel across regional accent variation. *Laboratory Phonology, 9*, 1–36. https://doi.org/10.5334/labphon.87.

Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*, 2471–2485. https://doi.org/10.1121/1.417955.

Tremblay, A., Broersma, M., & Coughlin, C. E. (2017). The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in a second language. *Bilingualism: Language and Cognition, 21*, 1–13. https://doi.org/10.1017/S136672891700030X.

Vallabha, G., McClelland, J., Pons, F., Werker, J., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*, 13273-13278.

Vanlancker-Sidtis, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. *Applied PsychoLinguistics, 24*, 45–57.

https://doi.org/10.1017/S0142716403000031.
Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374–408. https://doi.org/10.1006/jmla.1998.2618.
Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*, 25–47. https://doi.org/10.1159/000261901.
Xu, Y. (1999). Effect of tone and focus on the alignment of f0 contours. *Journal of Phonetics, 27*, 55–105. https://doi.org/10.1006/jpho.1999.0086.
Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication. 46*, 220–251. https://doi.org/10.1016/j.specom.2005.02.014.
Xu, Y., Chen, S.-W., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *Linguistic Review, 29*, 131–147. https://doi.org/10.1515/tlr-2012-0006.
Yan, M., & Calhoun, S. (2019). Priming effects of focus in Mandarin Chinese. *Frontiers in Psychology, 10*, 1985. https://doi.org/10.3389/fpsyg.2019.01985.
Yuan, J. (2004). *Intonation in Mandarin Chinese: Acoustics, perception, and computational Modeling.* Ph.D. thesisCornell University.
Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *Journal of the Acoustical Society of America, 130*, 4063–4069. https://doi.org/10.1121/1.3651.
Zerbian, S. (2017). Sentence intonation in Tswana (Sotho-Tswana group). In L. J. Downing, & A. Rialland (Eds.). *Intonation in African tone languages* (pp. 89–129). Berlin: Mouton de Gruyter.