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

Martin Ho Kwan Ip¹,b,⁎, Anne Cutler¹,b

¹ The MARCS Institute, Western Sydney University, Australia
² 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, 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, & Bienkowski, 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),
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; Lehiste, 1976; 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: Brunelliére, 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: Kusch, Igalalda, & 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 & Garney, 1995; Friedorf, 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 [main=] 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 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 (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 relationship 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 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 F0 register or
fluctuation of F0 range (e.g., Mandarin: Xu, 1999; Yoloxóchitl Mixtec: DiCanio, Benn, & Garcia, 2018). Similarly, some production studies suggest that prosody plays a dual role in the expression of information structure and lexical tones because features like F0, 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 (Kratosvii, 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; Tswan: 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 prefixed 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 as the Austronesian language Ma’ya (Remijisien, 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 F0 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, & Tayarin, 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 F0 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 F0 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 (Mage = 23.96 years, SD = 8.64 years; 16 females) and 23 native speakers of Mandarin Chinese (Mage = 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 [pʰ] occurring at the start of the target-bearing word's first syllable (e.g., “peanuts” [pʰɪnəts]; “葡萄” grapes [pʰu2 tʰau0]). 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 [pʰu2 tʰau0]) and half had the target on a falling fourth tone (e.g., “騙子” swindler [pʰ³au4 ʂɔ])

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”).
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.

1.2. 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.
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, F₀, 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, F₀) 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 (χ² (1) = 5.95, p = .015 in leave-one-out comparisons; β = 118.42, t = 2.50). However, there was no significant effect of gender (χ² (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 (χ² (1) = 16.67, 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 (χ² (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 (χ² (1) = 0.16, p = .686; β = 6.23, t = 0.42). There was a significant 2-way interaction between trial and language (χ² (1) = 9.08, p = .003; β = −32.62, t = −3.35:...

![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 ≤ .05.](image)
The acoustic results for the preceding duration, F0, 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 F0 range between the predicted high and low stress contexts, *t*(23) = 3.78, *p* = .001. Maximum F0 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 F0, 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 F0, *t*(23) = 2.23, *p* = .036, higher maximum F0, *t*(23) = 3.78, *p* = .001, greater F0 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 F0, F0 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 cross-language 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 F0 and F0 range and pre-target 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
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 (Mage = 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 participant has 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 responses 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 regression analyses (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 (χ²(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) (χ²(1) = 0.97, p = .322; 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 3
Preceding prosody F0 (mean, maximum, and range in Hz) and duration (in milliseconds) three or four syllables before target onset in predicted high versus low stress contexts.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Mean prosodic variables (SD) [Range]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean F0</td>
</tr>
<tr>
<td></td>
<td>High Stress</td>
</tr>
<tr>
<td>English</td>
<td>180.84 ⊹</td>
</tr>
<tr>
<td>(24 sentence pairs)</td>
<td>(15.43)</td>
</tr>
<tr>
<td>Mandarin</td>
<td>200.97</td>
</tr>
<tr>
<td>(24 sentence pairs)</td>
<td>(12.85)</td>
</tr>
</tbody>
</table>

∗ p ≤ .05 significant differences from predicted low stress contexts (two-tailed).
∗∗ p ≤ .001 significant differences from predicted low stress contexts (two-tailed).

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.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Mean Prosinic Variables (SD) [Range]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Intensity</td>
</tr>
<tr>
<td></td>
<td>High Stress</td>
</tr>
<tr>
<td>English</td>
<td>53.63∗∗∗</td>
</tr>
<tr>
<td>(24 sentence pairs)</td>
<td>(2.09)</td>
</tr>
<tr>
<td>Mandarin</td>
<td>54.44</td>
</tr>
<tr>
<td>(24 sentence pairs)</td>
<td>(3.60)</td>
</tr>
</tbody>
</table>

∗∗∗ p ≤ .001 significant differences from predicted low stress contexts (two-tailed).
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: $\chi^2 (1) = 0.37, p = .543$; recognition test scores: $\chi^2 (1) = 2.84, p = .092$). 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 ($r = 0.054, n = 24, p = .801$) 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 ($\chi^2 (1) = 0.37, p = .543$). 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. 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.](image_url)

![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.](image_url)
sentences 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 identity is important for focus location even when the pitch contour is simultaneously contouring the argument that the entrainment strategy is universal. This was despite the fact that the languages differed in where prosodic accents in both English and Mandarin to the same extent, further supporting the argument that the entrainment strategy is universal. This was reflected by evidence of faster RT as a function of the 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 prosodic contexts (in proportions) and length of stay in Australia (in months) in Experiment 2 (top right), and their post-test recognition scores (between 70% to 100%) in Experiment 2 (bottom centred).

### 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 $\chi^2$ values and corresponding p-values come from leave-one-out model comparisons. Analyses were based on 548 datapoints from 24 participants and 24 items.

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>$\beta$</th>
<th>SE</th>
<th>$\chi^2$ (1)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>604.01</td>
<td>32.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prosodic context</td>
<td>1.14</td>
<td>0.96</td>
<td>0.322</td>
<td></td>
</tr>
<tr>
<td>Participation in Experiment 1</td>
<td>−84.46</td>
<td>57.38</td>
<td>2.47</td>
<td>0.116</td>
</tr>
<tr>
<td>Length of stay</td>
<td>−16.18</td>
<td>40.77</td>
<td>0.37</td>
<td>0.543</td>
</tr>
<tr>
<td>Post-test recognition scores</td>
<td>383.29</td>
<td>242.04</td>
<td>2.84</td>
<td>0.092</td>
</tr>
<tr>
<td>Preceding duration × Prosodic context</td>
<td>−158.15</td>
<td>114.61</td>
<td>2.27</td>
<td>0.019</td>
</tr>
<tr>
<td>Preceding interval duration × Prosodic context</td>
<td>−327.37</td>
<td>209.13</td>
<td>2.95</td>
<td>0.040</td>
</tr>
<tr>
<td>Mean F0 × Prosodic context</td>
<td>114.36</td>
<td>376.78</td>
<td>1.32</td>
<td>0.076</td>
</tr>
<tr>
<td>Maximum F0 × Prosodic context</td>
<td>87.05</td>
<td>289.40</td>
<td>2.01</td>
<td>0.050</td>
</tr>
<tr>
<td>F0 Range × Prosodic context</td>
<td>−41.79</td>
<td>88.29</td>
<td>1.99</td>
<td>0.054</td>
</tr>
<tr>
<td>Mean Intensity × Prosodic context</td>
<td>485.44</td>
<td>815.48</td>
<td>5.55</td>
<td>0.136</td>
</tr>
<tr>
<td>Maximum Intensity × Prosodic context</td>
<td>449.00</td>
<td>1078.50</td>
<td>4.80</td>
<td>0.187</td>
</tr>
<tr>
<td>Intensity range × Prosodic context</td>
<td>−19.46</td>
<td>129.02</td>
<td>4.23</td>
<td>0.038</td>
</tr>
</tbody>
</table>

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).
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 detect 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 detect 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 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 F0 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. Vanlangeck-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 a 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; Safran et al., 1996; Vallabha,
Conclusion

searcher’s grasp.

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 contexts compared to 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 F0 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.
10. (a) They want to inform my \textbf{PARTNER} that I was sent home from work
(b) They want to inform my \textbf{partner} that I was sent \textbf{HOME} from work

11. (a) Most of the jurors find it odd that the millionaire was \textbf{PARDONED} after the verdict
(b) Most of the jurors find it odd that the millionaire was \textbf{pardoned \textit{AFTER}} the verdict

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

13. (a) Our clock no longer works ever since the \textbf{PENDULUM} went missing
(b) Our clock no longer works ever since the \textit{pendulum} went \textbf{MISSING}

14. (a) The surgeons must quickly remove her \textbf{PANCREAS} to delay the cancer from advancing
(b) The surgeons must quickly remove her \textit{pancreas} to delay the \textbf{CANCER} from advancing

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

16. (a) In some convents nuns still use \textbf{PADLOCKS} to seal their gates from the outside world
(b) In some convents nuns still use \textit{padlocks} to seal their \textbf{GATES} from the outside world

17. (a) Down on the farm we were amused to see a \textbf{PARROT} who could sing in French
(b) Down on the farm we were amused to see a \textit{parrot} who could sing in \textbf{FRENCH}

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

19. (a) The naval officer shook hands with a \textbf{PIRATE} who rescued him from the fire
(b) The naval officer shook hands with a \textit{pirate} who \textbf{RESCUED} him from the fire

20. (a) A child who witnessed the crime said the gunman used his \textbf{PENCIL} to scare her away
(b) A child who witnessed the crime said the gunman used his \textit{pencil} to \textbf{SCARE} her away

21. (a) I was quite shocked to see the Archduke's \textbf{POODLES} eating truffles for lunch
(b) I was quite shocked to see the Archduke's \textit{poodles} eating \textbf{TRUFFLES} for lunch

22. (a) It is sad that the chief commander will \textbf{PUNISH} his men for saving the foreigners
(b) It is sad that the chief commander will \textit{punish} his men for \textbf{SAVING} the foreigners

23. (a) Marine scientists were angry when they discovered \textbf{PETROL} inside the whale's eyes
(b) Marine scientists were angry when they discovered \textit{petrol} inside the whale's \textbf{EYES}

24. (a) These tourists said they would like to \textbf{PICNIC} in the desert
(b) These tourists said they would like to \textit{picnic} in the \textbf{DESERT}

\textbf{Filler sentences}

4 filler sentences with early occurrence of the phoneme target

1. \textit{PARSLEY} is the only thing you should add to the salad
2. In \textbf{POLAND} watching movies like “Home Alone” is now a Christmas tradition
3. Kim is \textbf{PAINTING} her own face with green and yellow ink for the soccer finale
4. You should not \textbf{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 \textbf{PICKY}
6. According to researchers, children under eleven don't understand what a \textbf{PARTICLE} is
7. If something goes wrong during the flight the lead stewardess must tell the \textbf{PILOTS}.
8. Many seafood lovers are unaware that some of the fish they eat may have \textbf{POISON} in their scales.

16 filler sentences with no phoneme target

9. Shareholders sometimes take \textbf{TOO} much risk to make themselves rich
10. At the meeting the climatologists told the winery owners that they will \textbf{NEVER} survive if there's no rain
11. His new house is of \textbf{EXACTLY} the same height as the surrounding high rises
12. Anna's colleagues \textbf{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
   他们 上星期 去 CLIMB-MOUNTAIN 踩了 many wild-flowers
   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
   他 想 马上 回家 因为 他的 FRIEND 想 偷 他 的 MONEY
   想 偷 他 的 MONEY 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
   笑死人 了 这 几 位 游客 想 穿 LEATHER-JACKET 在 BEACH 上 溜达
   笑死人了 这 几 位 游客 想 穿 leaCE-misc 在 海滩 上 巡 游
   laugh-die-people-VOc.PT CL 3.PL.M tourist wear leather-JACKET in 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
day two see:RES.CLP 1s lover PREP APPLE TREE
下偷偷地
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
no one see:RES.CLP 1s grape can create perfume
没有人在中国能 相信 葡萄 能制造 香水
NEG people 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
1s PUT home one CLP PLATE give:RES.CLP 1s GEN IDOL

“I shall give away my dinnerware as a present to my idol”
7. (a) Many actors think the shoes are already outdated.
(b) Many actors think the shoes are already outdated.

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

8. (a) It has been rumoured that boy from the village who looks like a crab will get married tomorrow.
(b) It has been rumoured that boy from the village who looks like a crab will get married tomorrow.

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

9. (a) You can see his stomach is getting越来越大.
(b) You can see his stomach is getting越来越大.

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

10. (a) You can see that his stomach is getting越来越大.
(b) You can see that his stomach is getting越来越大.

"You can see that his stomach is getting bigger and bigger."
10. (a) I’m surprised he would apply for 那 套 廉 价 的 房 子 给 自 己 住
    (b) 我 想 他 会 申 请 那 套 廉 价 房 子 给 自 己 住

    我 想 他 会 申请 那 套 廉 价 的 房 子 给 自 己 住
    I am quite surprised that he will apply to live in that cheap house by himself

11. (a) 没 想 到 她 女 儿 的 气 愤 能 让 她 患 上 癌 症
    (b) 没 想 到 她 女 儿 的 气 愤 能 让 她 患 上 癌 症

    没 想 到 她 女 儿 的 气 愤 能 让 她 患 上 癌 症
    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

    身 体 虚 弱 的 年 轻 人 需 要 吃 排 骨 来 增 加 营 养
    Young people who are physically weak need to eat some ribs to gain more nutrients
13. (a) These dog teams are 腐败 案件 的 名声 语文
   (b) These dog teams are 腐败 案件 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
   红楼梦里的姑娘长得漂亮 因为她们吃过仙丹
   Dream Red Mansion POST-GEN maiden look V-MOD BEAUTIFUL
   因为她们吃过仙丹 because 3.S.F.PL eat ASP.PTCP 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 魔鬼 因为她取笑我们家的秘密
   (b) 我家大姐的行为像 魔鬼 因为她取笑我们家的秘密
   我家大姐的行为像 魔鬼 因为她取笑 we家的秘密
   1.S.F.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

*Times*-Magazine-GEN analysts predict concert-GEN ticket shall.FUT
down CMPD.DIRECT.COMP

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

17. (a) 餐厅经理听到都吓呆了
(b) 餐厅经理听到都吓呆了

restaurant-manager hear-RES.COMP CANNON-SOUND ADV

down 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屁打针
(b) 有些护士喜欢向婴儿的屁屁打针

exist.v-clf nurse like direct.prep infant-gen glutte

apply INJECTION

“Some nurses prefer performing glute injections on toddlers”
19.
(a) 李先生逛超市时看见一位小朋友
(b) 李先生逛超市时看见一位小朋友

李先生逛超市时看见一位小朋友
Mr. Li saw a child in the supermarket.

“While doing grocery shopping, Mr. Li saw a child.”

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

在罗马有三个骗子往我们的方向走
In Rome, there were three swindlers walking in our direction.

“When we were in Rome, three swindlers were walking in our direction.”

21.
(a) 在山里的那位小伙子买了一个馍在兰州
(b) 在山里的那位小伙子买了一个馍在兰州

在山里的那位小伙子买了一个馍在兰州
The young man who lived in the mountains bought a retail shop in Lanzhou.

“The young man who lived in the mountains bought a retail shop in Lanzhou.”
22. (a) 小学生听到喊叫后躲猫猫在桌子底下
(b) 小学生听到喊叫后躲猫猫在桌子底下

小学生听到喊叫后躲猫猫在桌子底下

Primary-school students hear-RES.COMP scream-after.CNJ SCARED-MOD hide

hide-POST TABLE under-DIRECT.COM

"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窥视儿子从战争回家
(b) 老奶奶每天一个人站在门前窥视儿子从战争回家

老奶奶每天一个人站在门前窥视儿子从战争回家

Old lady every-day alone stand-POST door-front VISOR 2.s.F-scm

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) 我很惊讶她开车撞了一只大象
(b) 我很惊讶她开车撞了一只大象

我很惊讶她开车撞了一只大象

I am very 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. 她想陪伴她的母親去澳大利亞參加婚禮
   “She wants to accompany her mother when they go to Australia for the wedding ceremony.”

2. 這時有時能讓人喉嚨難受
   “This can sometimes make your throat feel uncomfortable.”

3. 在警察局我遇到了很多人
   “At the police station I encountered a lot of people.”

4. 我很崇拜很多非常勇敢的哲學家
   “I really admire many philosophers who are very brave.”

4 filler sentences with late occurrence of the phoneme target

5. 這位園丁花三天三夜設計一個很華麗的盆栽
   “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樹旁休息
   “The engineers who do research in geology prefer to take a rest next to the pine tree.”

7. 我們的手上沾滿了肥皂泡泡
   “Our hands are filled with soap bubbles.”

8. 地震災民現在對食物要求很迫切
   “The earthquake victims currently have a very urgent need for food.”
16 filler sentences with no phoneme target

9. 
真的没看出来他的艺术眼光有那么差
“I have never really noticed that his taste for art can be that bad”

10. 
大公司会计员总是抱怨他们公司的经济困难
“Accountants from big companies are always complaining about their company’s financial problems”

11. 
我在俄罗斯火车站差点跌倒了 (note: focused word is disyllabic)
“I almost fell down when I was getting off the train in Rustia”

12. 
这座塔和附近的楼一样高
“This tower is of exactly the same height as the surrounding buildings”

13. 
药剂师知道怎么混合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. 
安娜卡列尼娜曾经听说罗曼斯科说过她的生命很痛苦
“Anna Karenina did tell Yronsky that she has suffered a lot in her life”

15. 
调查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假货手袋
“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下工作
“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找食物吃
“Pinches 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扫工的责任
“All the lawyers unanimously agree that the hygiene in our streets is the cleaners’ responsibility”

22. 
厉害的魔术师能用他的手法来影响其他人的感情
“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班的员工光顾看电视
“A thousand paintings are missing at the art gallery because the night staff were watching TV”
Appendix C

**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

**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 some weeks a girl of paleComplexity 1081
2. The young man is lost on paleComplexity 1081
3. We give you a chance of live in the house Complex 1082
4. A lot of money in the money box before going for food.

- Push the BUTTON to see the ANSERS

**Slide 11**

**Answers**

Did you hear the following sentences?

1. In some weeks a girl of paleComplexity 1081
2. The young man is lost on paleComplexity 1081
3. We give you a chance of live in the house Complex 1082
4. A lot of money in the money box before going for food.

- 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 remember 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

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
**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. 我认为这件裤子的衣服还是太大了
   有  没有

2. 我在俄罗斯下火车差点儿跌倒了
   有  没有

3. 没有人在中国能相信葡萄能制造香水
   有  没有

4. 我很惊讶他会申请那套便宜的房子给自己住
   有  没有

5. 大家都很高兴因为那个长得像螃蟹的女孩要结婚
   有  没有

6. 听说村里那个长得像螃蟹的男孩要结婚
   有  没有

7. 我对我的朋友很失望因为他们现在都很有钱
   有  没有

8. 这些游客在市场上买了很多西瓜
   有  没有

9. 厉害的魔术师能用他的手法来影响其他人的心情
   有  没有

10. 机场海关没有没收走私的仿造手袋
    有  没有

11. 很多人喜欢用大盘子吃意粉
    有  没有
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).

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

(continued on next page)
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).

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>(\beta)</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.623</td>
<td>8.17e-04</td>
</tr>
<tr>
<td>Language</td>
<td>0.005</td>
<td>0.0020</td>
</tr>
<tr>
<td>Gender</td>
<td>2.147e-03</td>
<td>1.361e-03</td>
</tr>
<tr>
<td>Prosodic context</td>
<td>3.101e-03</td>
<td>6.713e-04</td>
</tr>
<tr>
<td>Language (\times) Prosodic context</td>
<td>-0.002</td>
<td>0.0021</td>
</tr>
<tr>
<td>Trial</td>
<td>3.736e-03</td>
<td>8.934e-04</td>
</tr>
<tr>
<td>Trial (\times) Prosodic context</td>
<td>-0.0005</td>
<td>0.0008</td>
</tr>
<tr>
<td>Trial (\times) Language</td>
<td>-0.0029</td>
<td>0.0009</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English listeners</th>
<th>(\beta)</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.90</td>
<td>0.0018</td>
</tr>
<tr>
<td>Prosodic context</td>
<td>0.0068</td>
<td>0.0019</td>
</tr>
<tr>
<td>Preceding duration (\times) Prosodic context</td>
<td>0.0086</td>
<td>0.0065</td>
</tr>
<tr>
<td>Pretarget interval duration (\times) Prosodic context</td>
<td>-0.0195</td>
<td>0.0107</td>
</tr>
<tr>
<td>Mean F0 (\times) Prosodic context</td>
<td>0.0172</td>
<td>0.0208</td>
</tr>
<tr>
<td>Maximum F0 (\times) Prosodic context</td>
<td>0.0180</td>
<td>0.0146</td>
</tr>
<tr>
<td>F0 range (\times) Prosodic context</td>
<td>6.735e-03</td>
<td>4.771e-03</td>
</tr>
<tr>
<td>Mean intensity (\times) Prosodic context</td>
<td>7.689e-03</td>
<td>4.463e-02</td>
</tr>
<tr>
<td>Maximum intensity (\times) Prosodic context</td>
<td>7.456e-03</td>
<td>5.712e-02</td>
</tr>
<tr>
<td>Intensity range (\times) Prosodic context</td>
<td>-0.0075</td>
<td>0.0074</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mandarin listeners</th>
<th>(\beta)</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Interval)</td>
<td>1.409</td>
<td>0.006</td>
</tr>
<tr>
<td>Prosodic context</td>
<td>1.340e-03</td>
<td>5.005e-04</td>
</tr>
<tr>
<td>Preceding duration (\times) Prosodic context</td>
<td>-1.638e-03</td>
<td>2.644e-03</td>
</tr>
<tr>
<td>Pretarget interval duration (\times) Prosodic context</td>
<td>1.140e-03</td>
<td>1.790e-03</td>
</tr>
<tr>
<td>Mean F0 (\times) Prosodic context</td>
<td>0.0084</td>
<td>0.0054</td>
</tr>
<tr>
<td>Maximum F0 (\times) Prosodic context</td>
<td>0.0076</td>
<td>0.0075</td>
</tr>
<tr>
<td>F0 range (\times) Prosodic context</td>
<td>1.609e-04</td>
<td>1.561e-03</td>
</tr>
<tr>
<td>Mean intensity (\times) Prosodic context</td>
<td>0.0163</td>
<td>0.0071</td>
</tr>
<tr>
<td>Maximum intensity (\times) Prosodic context</td>
<td>0.0191</td>
<td>0.0074</td>
</tr>
<tr>
<td>Intensity range (\times) Prosodic context</td>
<td>5.160e-04</td>
<td>1.714e-03</td>
</tr>
</tbody>
</table>

Appendix I

Experiment 2
Random participant effects for linear mixed-effects model analyses for RT (Box-Cox converted). Analyses were based on 548 datapoints (46 participants and 24 items).

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant</td>
<td>1.157e-05</td>
<td>0.0034</td>
</tr>
<tr>
<td>Prosodic context</td>
<td>1.402e-05</td>
<td>3.745e-03</td>
</tr>
<tr>
<td>Participation in experiment 1</td>
<td>1.092e-05</td>
<td>3.305e-03</td>
</tr>
<tr>
<td>Length of stay</td>
<td>4.111e-08</td>
<td>0.0002</td>
</tr>
<tr>
<td>Post-test recognition scores</td>
<td>1.638e-05</td>
<td>4.047e-03</td>
</tr>
<tr>
<td>Preceding duration (\times) Prosodic context</td>
<td>2.572e-06</td>
<td>0.0016</td>
</tr>
<tr>
<td>Pretarget interval duration (\times) Prosodic context</td>
<td>4.363e-05</td>
<td>0.0066</td>
</tr>
<tr>
<td>Mean F0 (\times) Prosodic context</td>
<td>6.539e-06</td>
<td>0.0025</td>
</tr>
<tr>
<td>Maximum F0 (\times) Prosodic context</td>
<td>2.861e-05</td>
<td>5.349e-03</td>
</tr>
<tr>
<td>F0 range (\times) Prosodic context</td>
<td>7.584e-06</td>
<td>2.754e-03</td>
</tr>
<tr>
<td>Mean intensity (\times) Prosodic context</td>
<td>8.007e-06</td>
<td>2.830e-03</td>
</tr>
<tr>
<td>Maximum intensity (\times) Prosodic context</td>
<td>1.670e-05</td>
<td>4.087e-03</td>
</tr>
<tr>
<td>Intensity range (\times) Prosodic context</td>
<td>1.660e-05</td>
<td>4.074e-03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mandarin listeners</th>
<th>(\beta)</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Interval)</td>
<td>1.409</td>
<td>0.006</td>
</tr>
<tr>
<td>Prosodic context</td>
<td>1.340e-03</td>
<td>5.005e-04</td>
</tr>
<tr>
<td>Preceding duration (\times) Prosodic context</td>
<td>-1.638e-03</td>
<td>2.644e-03</td>
</tr>
<tr>
<td>Pretarget interval duration (\times) Prosodic context</td>
<td>1.140e-03</td>
<td>1.790e-03</td>
</tr>
<tr>
<td>Mean F0 (\times) Prosodic context</td>
<td>0.0084</td>
<td>0.0054</td>
</tr>
<tr>
<td>Maximum F0 (\times) Prosodic context</td>
<td>0.0076</td>
<td>0.0075</td>
</tr>
<tr>
<td>F0 range (\times) Prosodic context</td>
<td>1.609e-04</td>
<td>1.561e-03</td>
</tr>
<tr>
<td>Mean intensity (\times) Prosodic context</td>
<td>0.0163</td>
<td>0.0071</td>
</tr>
<tr>
<td>Maximum intensity (\times) Prosodic context</td>
<td>0.0191</td>
<td>0.0074</td>
</tr>
<tr>
<td>Intensity range (\times) Prosodic context</td>
<td>5.160e-04</td>
<td>1.714e-03</td>
</tr>
</tbody>
</table>

(continued on next page)
## Experiment 2 (continued)

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

## Appendix J

### Experiment 2

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

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

## 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).

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