



## Mandarin and English Adults' Cue-weighting of Lexical Stress

Zhen Zeng<sup>1</sup>, Karen Mattock<sup>1</sup>, Liquan Liu<sup>1,2</sup>, Varghese Peter<sup>1</sup>, Alba Tuninetti<sup>1,3</sup>, Feng-Ming Tsao<sup>4</sup>

<sup>1</sup>Western Sydney University, Australia

<sup>2</sup>University of Oslo, Norway

<sup>3</sup>Bilkent University, Turkey

<sup>4</sup>National Taiwan University, Taiwan

j.zeng@westernsydney.edu.au, k.mattock@westernsydney.edu.au,  
l.liu@westernsydney.edu.au, v.peter@westernsydney.edu.au,  
alba.tuninetti@bilkent.edu.tr, tsaosph@ntu.edu.tw

### Abstract

Listeners segment speech based on the rhythm of their native language(s) (e.g., stress- vs. syllable-timed, tone vs. non-tone) [1,2]. In English, the perception of speech rhythm relies on analyzing auditory cues pertinent to lexical stress, including pitch, duration and intensity [3]. Focusing on cross-linguistic impact on English lexical stress cue processing, the present study aims to explore English stress cue-weighting by Mandarin-speaking adults (with English adults as control), using an MMN multi-feature paradigm.

Preliminary ERP data revealed cross-linguistic perceptual differences to pitch and duration cues, but not to intensity cues in the bisyllabic non-word /dede/. Specifically, while English adults were similarly sensitive to pitch change at the initial and final syllable of the non-word, they were more sensitive to the duration change at the initial syllable. Comparatively, Mandarin adults were similarly sensitive to duration change at each position, but more sensitive to pitch at the final syllable. Lastly, both the Mandarin group and the English group were more sensitive to the intensity sound change at the second syllable. Possible explanations for these findings are discussed.

**Index Terms:** stress perception, cue weighting, event-related potentials, multi-feature paradigm, Mandarin adults

### 1. Introduction

Speech rhythm exists at different prosodic hierarchical levels [4-6]. For stress-timed languages, speech rhythm can be perceived through lexical stress at the word level [7]. Lexical stress is signaled by cues such as pitch, intensity and duration [3]. While listeners use lexical stress to segment/group information, the extent to which they rely on these cues is dependent on the rhythmic nature of their native language(s) (e.g., stress- vs. syllable-timed) [1,2]. Moreover, iambic-trochaic law (ITL) [8] – a domain-general law of sound grouping – posits that sounds varying in pitch or intensity are grouped trochaically, while those varying in duration are grouped iambically [9]. Studies examining language-specific vs. ITL influence have been largely conducted with European languages, and mixed results have emerged over the past decades in retesting ITL using differing stimuli and methods, in differing language groups as well as at different developmental stages (see [10]). Thus, the mechanism underlying the perceptual weighting of these cues in speech is not yet fully understood, especially in tone language speakers.

This study examined ITL vs language-specific influence in speech cue-weighting in Mandarin adults and English adults' weighting of pitch, intensity and duration cues. We focused on the cross-lingual effect of Mandarin on English cue-weighting, due to the multi-faceted linguistic constraints of Mandarin. Firstly, a behavioural study using a forced-choice paradigm reported that native English speakers use pitch (F0) duration and intensity in English stress perception of bisyllabic non-words, but only F0 was found to be a decisive cue for stress perception in Mandarin learners of English. The author suggests that the Mandarin listeners are 'deaf' to duration and intensity contrasts [11]. This finding was also replicated in other studies using similar behavioural paradigms [12,13]. Secondly, acoustic analysis has revealed that Mandarin can be classified as a syllable-timed language [14], and Taiwan Mandarin has no stress at the word level [15]. These studies lead to the hypothesis that although (Taiwan) Mandarin adults are sensitive to pitch cues, they do not have a preference in terms of syllable positions when processing stress. Third, there is no consistent evidence that Taiwan Mandarin has a trochaic or iambic speech pattern at the word level, compared to English which is a trochaic language [16]. Studies looking at Mandarin speakers' stress perception from a neural point of view have been scarce. In an electroencephalography (EEG) study, [17] found that English adults track intensity for non-word pair "noCTiCity" and "NOcticity" better than Mandarin speakers learning English as a second language. This evidence contrasts with another EEG study, which found that Cantonese (syllable-timed and tonal) school children who were learning English had larger mismatch negativity (MMN) responses to stress cues compared to English monolingual children using a real word pair MOther and toDay [18].

To expand the neural evidence of the cross-linguistic effect of Mandarin on English in terms of cue-weighting with better controlled stimuli, this study used a symmetrical bisyllabic non-word /dede/ and adopted an MMN multi-feature paradigm, as used in [18]. The MMN response, as an automatic neural change-detection response and a component of event-related potentials (ERPs), has been used to investigate short and long-term memory traces and discrimination accuracy of speech sounds (for a review, see [19]). Accurate stimulus discrimination is associated with large MMN amplitudes and inaccurate discrimination with small ones [20]. In optimizing the recording procedure, [21] developed a faster paradigm, the so-called multi-feature paradigm (Optimum-1), to facilitate multiple-deviant discrimination, compared to the traditional oddball paradigm where only one deviant is included for one recording session.

According to the previous research (mainly behavioural), we expect cross-linguistic differences in duration cue perception, such that native English (stress-timed) speakers have significantly different MMN amplitudes across syllable positions to duration changes, whereas Mandarin (syllable-timed) speakers have a lack of difference in MMN amplitudes.

## 2. Methods

### 2.1. Participants

The current sample consisted of Mandarin adult listeners ( $N = 10$ ) recruited from National Taiwan University and Australian English listeners ( $N = 9$ ) recruited from Western Sydney University. According to the Language Experience and Proficiency Questionnaire (LEAP-Q) [22] completed, all Mandarin adults had L2 English classes in school but were not fluent in English and had never lived in an English-speaking country at the time of testing. While their main language exposure was to Mandarin, some of these adults were also exposed to a spoken Chinese dialect (Fujian = 2, Taiwanese = 3). Their exposure to the dialect ranged from zero to 30% as measured by the LEAP-Q. All English adult participants were monolingual, although some ( $N = 2$ ) of their parents speak another language other than English (Egyptian). None of the adult participants were musicians or reported receiving music training.

### 2.2. Stimuli

In order to ensure that there would be no lexicality effect in the MMN paradigm (e.g., [23]), we selected a stimulus that was not a word in either English or Mandarin Chinese. The single syllable /de/ with a neutral pitch contour was produced by a monolingual Australian English female speaker and recorded using Adobe Audition in a sound-proof booth in MARCS Institute Phonetics Lab. The syllable was then repeated to form a bisyllabic non-word (in both English and Mandarin) /dede/ with duration of 380ms, intensity = 65dB, and of F0 at 190Hz; this bisyllabic nonword served as the ‘standard’ stimulus in the MMN paradigm. This allowed us to control the acoustic features between the first and the second syllable. Stress deviants were manipulated in Praat software [24] with pitch, intensity and duration changes at either initial or final syllable position (see Table 1), with pitch deviant 10% higher F0, intensity deviant 6 dB louder, and duration deviant 33% longer than the standard.

Table 1: *Manipulation of Deviants*

	Pitch	Duration	Intensity
Initial	[~dede]	[de:de]	['dede]
Final	[de~de]	[dede:]	[de'de]

### 2.3. Experimental Design and Procedure

#### 2.3.1. EEG paradigm

An MMN multi-feature paradigm was used to facilitate recording among multiple speech stress cue-weighting stimuli. During the experimental session, the standard sounds were

alternating with each of the six randomized deviants (in the oddball block). A control block consisting of the six deviants was also included, with each of the six deviants repeating for 100 trials and the presenting order of the six deviants randomised. The standard/deviant probability ratio was 50%/50% [21]. There were 1800 stimuli in total, with 600 deviants for the control block, 600 standards and 600 deviants for the oddball block. The stimuli were presented with an inter-stimulus interval of 500ms at a constant intensity of 65dB SPL. The total duration of the experiment was 30 minutes.

#### 2.3.2. EEG Recording

Participants were seated in a sound-attenuated room  $\sim 0.5$  m from a Genelec 8010A speaker where the sound stimuli were played. While listening to the stimuli presented through Presentation software (Neurobehavioral Systems, Inc.), participants were fitted with a 32-channel EEG cap (ActiCAP Slim, Brain Products) attached with gel and electrodes. The continuous EEG signal was recorded at a sampling rate of 500Hz with the reference electrode at FCz. Electrode impedances were kept below 50k $\Omega$  at the start of the recording. Participants’ EEG was recorded using LiveAmp amplifier and BrainVision Recorder. To maintain engagement during the study, a movie was played silently on a screen next to the speaker with subtitles in their native language.

#### 2.3.3. EEG Analysis

The EEG was analysed using Fieldtrip Toolbox [25] in MATLAB 2019a firstly and then band-pass filtered between 0.1-30Hz and divided into epochs between -100 and 800ms relative to sound onsets. Epochs were then baseline corrected between -100 and 0ms. The EEG was then subjected to Independent Component Analysis. Components with stereotypical features of eye blinks and eye movements were removed from the EEG. The EEG signal was then re-referenced to the average of the mastoids. Trials exceeding  $\pm 100\mu\text{V}$  were removed and then averaged separately for deviant and control to obtain the ERP waves. Difference waves were computed by subtracting the ERP for the control stimulus from the deviant stimulus. In this way, ERPs to physically identical stimuli were compared for the calculation of MMN responses, as it reflects the brain response to a change as opposed to ERP effects due to physical differences between standard and deviant [26]. Individual ERP waves were averaged to create grand-averaged ERPs.

#### 2.3.4. Statistical Analysis

The presence of an MMN response was tested using nonparametric cluster-based permutation statistics [26]. First, a series of t-tests were computed at each electrode and each time point, comparing the deviant and control waveforms. From this, clusters were formed by combining the sampling points where a significant effect was obtained ( $p < .05$ , two-tailed) based on temporal and spatial adjacency and polarity of the effect. Cluster-level statistics were then calculated by adding together all the t values within the cluster. To control for Type I errors, a permutation approach was used where the condition labels were randomly swapped, and the t-tests were repeated 2000 times to generate a data-driven null hypothesis distribution. The cluster-level statistics from the first step was considered significant if it fell in the top 2.5 or bottom 2.5 percentile of the distribution.

### 3. Results

The control, deviant and deviant-control difference waves for the two language groups are shown in Figures 1 and 2. The difference waves showed negative peaks between 0 - 500ms for cue position at the first syllable and between 200 - 700ms for cue position at the second syllable for both language groups. These responses were confirmed by the cluster-based permutation tests. Since the latency and scalp location matched the expected MMN response latencies (100 - 300ms after the onset of sound change) and location (frontocentral), these responses were classified as MMN responses. This paper will not discuss positive and negative peaks outside of the MMN latency ranges as it is beyond the scope of this paper. Table 2 shows the results of the analysis.

Table 2: Results of the cluster-based permutation statistics

Group	Cue Type	Cluster Type	Time Window (ms)
English	Pitch1	+	174 - 228
	Pitch2	+	356 - 444
	Intensity1	N/A	
	Intensity2	+	336 - 406
	Duration1	+	394 - 458
	Duration2	N/A	
Mandarin	Pitch1	+	162 - 266
	Pitch2	+	350 - 454
	Intensity1	N/A	
	Intensity2	+	356 - 396
	Duration1	+	430 - 494
	Duration2	+	390 - 452

According to the results of the cluster permutation test (as shown in Table 2), whereas English adults had significant MMN responses to the intensity at the second syllable (I2), the duration at the first syllable position (D1), and both positions for pitch cues (P1 and P2), Mandarin adults had significant MMN responses to all conditions except for I2. When the same cue change was significant at both syllabic positions, the significant MMN amplitudes at the two syllabic positions were compared using paired *t*-tests. Here, the MMN amplitude values were calculated by averaging 40ms around the negative peak of the waves, from five electrodes in the frontocentral region (Fz, FC1, Cz, FC2 and FCz). We found that Mandarin adults showed a significantly larger MMN amplitude to P2 compared to the P1,  $t(9) = 4.740, p < .0001$ , whereas amplitude to D2 was not significantly larger than the amplitude to D1,  $t(9) = .685, p = .510$ . For English adults, the average amplitude to P1 was not significantly different compared to P2,  $t(8) = 1.633, p = .154$ .

In short, the results showed that Mandarin adults are more sensitive to acoustic changes in pitch and intensity at the second syllable position, and similarly sensitive to acoustic

change in duration at the initial syllable position and at the final syllable position. Comparatively, English adults showed more sensitivity to a duration acoustic change at the initial syllable position and showed similar sensitivity to pitch in both syllable positions. Interestingly, both language groups show more sensitivity to intensity change at the second syllable.

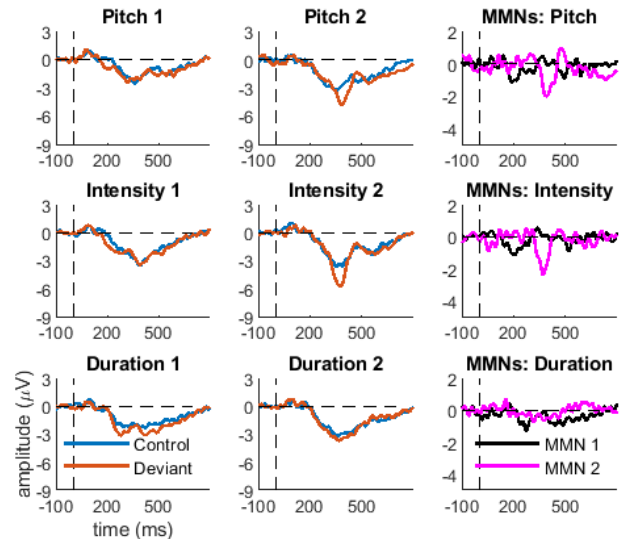


Figure 1: Control, deviant, and difference waves at frontocentral electrodes in English-speaking adults.

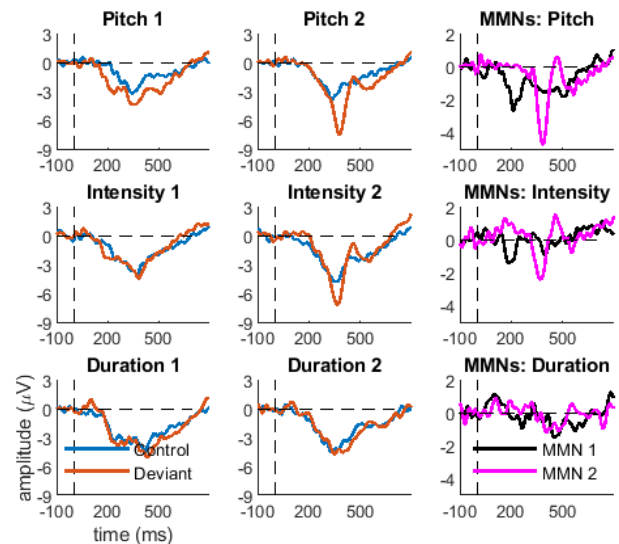


Figure 2: Control, deviant, and difference waves at frontocentral electrodes in Mandarin-speaking adults.

### 4. Discussion

To examine the cross-linguistic impact of Mandarin on English lexical stress perception, sensitivity to stress cues at the first syllable position and the second syllable position was measured using an MMN multi-feature paradigm in Mandarin-speaking adults and English-speaking adults. Preliminary ERP

data have revealed cross-linguistic perceptual differences to pitch and duration cues, but not to intensity cues in the bisyllabic non-word /dede/.

English-speaking adults had significant MMN responses to pitch deviants at both syllabic positions. However, these were not significantly different from each other, suggesting comparable sensitivity to this cue across the two syllables. We hypothesize that although pitch incremental change, or pitch leap, is not essential for stress discrimination for their native language, the cue itself is a perceptually salient one. This is in line with behavioural findings showing that non-tone language speakers are sensitive to lexical tones though they perceive them in a psycho-acoustic fashion [28, 29]. Moreover, English listeners were more sensitive to the duration sound change at the first syllable, and more sensitive to the intensity sound change at the second syllable. As MMN responses typically represent the brain response to violations of auditory memory traces stored in the brain [19], native English speakers were likely using the iambic-trochaic law to group bisyllabic speech sounds as short-long and strong-weak in duration and intensity cues, respectively [30]. Thus, when deviants were against this usual perceptual pattern, significant MMN responses were elicited.

For Mandarin-speaking adults, significant MMN responses were elicited to duration cues at both syllable positions with similar MMN amplitude values, indicating comparable sensitivity to this cue in the two syllable positions. These speak to evidences that Mandarin is a syllable-timed language [14], and Taiwan Mandarin has few durational contrasts at the word level [15]. Significant MMN responses were also elicited to pitch cues at both syllabic positions (although with different amplitude values across syllables). It seems that Mandarin speakers over-relied on pitch cues for English stress perception [11-13], due to the fact that pitch changes are lexical in Mandarin and elicit stronger responses than non-lexical prosodic cues. Mandarin adults were also more sensitive to pitch and intensity cues at the second syllable position, with significant MMN responses to pitch cues at both positions and intensity cues at the second syllable position. It is possible that Mandarin adults were grouping pitch and intensity cues following the ITL like their English-speaking peers, indicating a domain-general use for the ITL. Alternatively, as the sample of Mandarin adults also had some level of English experience, it is not impossible that their responses to the examined cues were somewhat acquired from or influenced by English experience.

Another possibility is that the MMN amplitudes are representative of the main structure of the language. It could be that there are more low-high tone occurrences in Mandarin [31], and pitch and intensity are part of this feature [32]. Thus, the listeners collapsed pitch/intensity cues into low-high tone occurrences. Some previous studies also found that adult MMN responses to stress cue changes are representative of the main pattern of the native language. For example, [17] found that English adults had larger MMN responses to pitch and intensity cues of English lexical stress, and neurally tracked the stimuli sound envelope better than the Mandarin group. Therefore, to disentangle nature from nurture and novelty vs. familiarity effect, it would be useful to examine the questions in monolingual groups of the two languages respectively at a younger age.

## 5. Conclusions

The present study shows that the cross-linguistic perceptual differences to English lexical stress cues between native Mandarin-speaking and native English-speaking adults are characterized of both the rhythmic nature of their languages and ITL. In particular, this is driven by their different perceptual patterns to duration and pitch cues. Importantly, we found that the perceptual pattern to a duration sound change was different to that of pitch and intensity cues in Mandarin adults, a result not seen in previous research designs. While these preliminary results do not compare the language groups' responses to each other or compare absolute monolingual English speakers to absolute monolingual Mandarin speakers, future work should aim to compare responses across languages and specific age groups to further clarify the role of cues in stress perception.

## 6. Acknowledgements

This project is jointly funded by HDR funds from Western Sydney University and a 2019 Research Grant from the Australian Linguistic Society.

## 7. References

- [1] L. Goyet, S. de Schonen, and T. Nazzi "Words and syllables in fluent speech segmentation by French-learning infants: An ERP study". *Brain Research*, 1332, 75–89, 2010 <https://doi.org/10.1016/j.brainres.2010.03.047>
- [2] D. Houston, L. Santelmann, and P. Jusczyk "English-learning infants' segmentation of trisyllabic words from fluent speech". *Language and Cognitive Processes*, 19(1), 97–136, 2004. <https://doi.org/10.1080/01690960344000143>
- [3] C. Adams, and R. R. Munro "In search of the acoustic correlates of stress: fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English". *Phonetica*, 35(3), 125–156, 1978. <https://doi.org/10.1159/000259926>
- [4] D. B. Fry "Duration and intensity as physical correlates of linguistic stress". *The Journal of the Acoustical Society of America*, 27(4), 765–768, 1955.
- [5] D. B. Fry. Experiments in the perception of stress. *Language and Speech*, 1(2), 126–152, 1958.
- [6] M. Nespors, M. Shukla, R. van de Vijver, C. Avesani, H. Schraudolf, & C. Donati "Different phrasal prominence realizations in VO and OV languages". *Lingue e Linguaggio*, 7(2), 139–168, 2008.
- [7] P.W. Jusczyk, D. M. Houston, and M. Newsome "The Beginnings of Word Segmentation in English-Learning Infants". *Cognitive Psychology*, 39(3), 159–207, 1999. <https://doi.org/10.1006/cogp.1999.0716>
- [8] B. Hayes "Metrical Stress Theory: Principles and case studies". *Chicago: University of Chicago Press*, 1995.
- [9] A. Bhatara, N. Boll-Avetisyan, A. Unger, T. Nazzi, & B. Höhle. "Native language affects rhythmic grouping of speech". *The Journal of the Acoustical Society of America*, 134(5), 3828–3843, 2013.
- [10] M. J. Crowhurst "The iambic/trochaic law: Nature or nurture?". *Language and Linguistics Compass*, 14(1), e12360, 2020.
- [11] Q. Wang "L2 stress perception: the reliance on different acoustic cues". In *Proc. 4th Conf. on Speech Prosody, Campinas, Brazil*, 2008, pp. 635–638.
- [12] S. C. Ou "Taiwanese EFL learners' perception of English word stress". *Concentric: Studies in Linguistics*, 36(1), 1–23, 2010.

- [13] Y. Y. Vickie, and J. E. Andruski “A cross-language study of perception of lexical stress”. *English Journal of psycholinguistic research*, 39(4), 323-344, 2010.
- [14] P. Mok “On the syllable-timing of Cantonese and Beijing Mandarin”. *Chinese Journal of Phonetics*, 2, 148-154, 2009.
- [15] Z. Qin, Y. Chien, and A. Tremblay “Processing of word-level stress by Mandarin-speaking second language learners of English”. *Applied Psycholinguistics*, 38(3), 541-570, 2017. doi:10.1017/S0142716416000321
- [16] A. Cutler and D. M. Carter, “The predominance of strong initial syllables in the English vocabulary”. *Computer Speech and Language*, 2, 133-142, 1987. doi: 10.1016/0885-2308(87)90004-0.
- [17] W. L. Chung, and G. M. Bidelman “Cortical encoding and neurophysiological tracking of intensity and pitch cues signaling English stress patterns in native and nonnative speakers”. *Brain and language*, 155, 49-57, 2016.
- [18] X. Tong, C. McBride, J. Zhang, K. K. Chung, C. Y. Lee, L. Shuai, and X. Tong “Neural correlates of acoustic cues of English lexical stress in Cantonese-speaking children”. *Brain and language*, 138, 61-70, 2014.
- [19] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho “The mismatch negativity (MMN) in basic research of central auditory processing: a review”. *Clinical neurophysiology*, 118(12), 2544-2590, 2007.
- [20] T. Kujala, and R. Näätänen “The adaptive brain: a neurophysiological perspective”. *Progress in neurobiology*, 91(1), 55-67, 2010.
- [21] R. Näätänen, S. Pakarinen, T. Rinne, and R. Takegata, “The mismatch negativity (MMN): towards the optimal paradigm”. *Clinical Neurophysiology*, 115(1), 140-144, 2004. <https://doi.org/10.1016/j.clinph.2003.04.001>
- [22] V. Marian, H. K. Blumenfeld, and M. Kaushanskaya “The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals”. *Journal of Speech, Language, and Hearing Research*, 2007.
- [23] Y. Shtyrov, & F. Pulvermüller. “Neurophysiological evidence of memory traces for words in the human brain”. *Neuroreport*, 13(4), 521-525, 2002.
- [24] P. Boersma, and D. Weenink *Praat: doing phonetics by computer [Computer program]. Version 5.3. 56. Retrieved September 15, 2013.*
- [25] R. Oostenveld, P. Fries, E. Maris, J.M. Schoffelen “FieldTrip: Open source software for advanced analysis of MEG, EEG and invasive electrophysiological data”. *Comput. Intell. Neurosci.* 156869, 2011.
- [26] V. Peter, G. McArthur, and W. F. Thompson “Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN)”. *Neuroscience letters*, 482(1), 71-75, 2010.
- [27] E. Maris and R. Oostenveld “Nonparametric statistical testing of EEG- and MEG- data”. *J. Neurosci. Meth.* 164, 177-190, 2007.
- [28] P. A. Hallé, Y. C., Chang, & C. T. Best “Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners”. *Journal of phonetics*, 32(3), 395-421, 2004.
- [29] L. Liu, A. Chen, & R. Kager “Perception of tones in Mandarin and Dutch adult listeners”. *Language and Linguistics*, 18(4), 622-646, 2017.
- [30] N. Abboub, N. Boll-Avetisyan, A. Bhatara, B. Höhle, and T. Nazzi “An exploration of rhythmic grouping of speech sequences by French-and German-learning infants”. *Frontiers in Human Neuroscience*, 10, 292, 2016. <https://doi.org/10.3389/fnhum.2016.00292>
- [31] R. S. Lavin “Conceptual Foundations of a Prosodic Model for Mandarin”. *Journal of Cognitive Science*, 3(1), 65-83, 2002.
- [32] D. Burnham, and K. Mattock “The perception of tones and phones”. in *Language experience in second language speech learning: In honor of James Emil Flege*, 259-280, 2007.