9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Prosodic disambiguation by Chinese EFL learners in a cooperative game task

Yuanyuan Zhang[1]*, Hongwei Ding*[1*]*, Peter Zelchenko*[2]*, Xin Cui*[1]*, Yi Lin*[1]*, Yuqing Zhan*[1]*, Hui Zhang*[1]

[1] Institute of Cross-Linguistic Processing and Cognition, School of Foreign Languages, Shanghai Jiao Tong University, China
[2]School of Media and Design, Shanghai Jiao Tong University, China
yyuan.zhang@sjtu.edu.cn, hwding@sjtu.edu.cn

Abstract

In this study, a game-based production experiment was adopted to examine the prosodic realization of syntactically ambiguous sentences by Chinese learners of English as a foreign language (EFL hereafter). 20 Chinese undergraduates and 10 native speakers of American English participated in this experiment. Subjects followed the guides in pictures and instructed listeners to move objects on the computer screen by using the critical instructions with prepositional attachment sentences (PP-attachment hereafter). In all, 10 pairs of ambiguous PP-attachment sentences that might refer to one situation or another were adopted. It was found that both the native speakers and Chinese EFL learners used pre-boundary lengthening and pause to distinguish the alternative meanings of the ambiguous PP-attachment sentences. While native speakers also showed domain-initial strengthening which may be related to the length of previous phrase, and greater pre-boundary lengthening and longer pause than the learners. In addition, native speakers displayed pitch reset at the prosodic boundary, indicating a pitch declination of the utterances. However, the learners might not consistently use pitch reset at the prosodic boundary.

Index Terms: prosody, prosodic boundary, syntactic ambiguity, English learners

1. Introduction

Prosody plays a crucial role in spoken language comprehension. One important function of prosody is to segment the utterances into smaller phrases. Prosodic phrasing is not isomorphic to syntactic phrasing, rather, it is a reflection of the syntactic structure of the utterance [1]. In the study of the effect of prosodic phrasing on syntactic structure, prosodic disambiguation (i.e., using prosody to resolve syntactic ambiguity) has been of particular interest [2]. Some syntactic ambiguities can be resolved by the placement of prosodic boundaries, especially for bracketing ambiguities [3]. For example, the PP-attachment sentence "*Tap the frog with the flower*" may have two interpretations depending on the situation: that is, to use the flower to tap the frog, or to tap the frog that has a flower [4]. The intended meaning of this ambiguity can be distinguished by bracketing the sentence in different ways.

Prosodic disambiguation has been widely studied in speech production and perception. It has been shown that both speakers and listeners can use prosodic cues to resolve syntactic ambiguity [4–17]. The major acoustic correlates of prosodic phrasing are duration, fundamental frequency (F0), and intensity, but intensity has been less often studied as a signal for prosodic boundaries. According to [18], pre-boundary lengthening, pause insertion, and domain-initial strengthening are the main durational cues relating to prosodic phrasing. In addition, boundary tone, pitch reset, and voice quality changes at the end of the prosodic domain also correlate with prosodic boundaries.

In the study of the role of prosody in sentence processing and language comprehension, prosodic phrasing has also been investigated in second language research. In a study of early Spanish-English bilinguals carried out by [19], it was shown that the bilinguals adopted a greater duration and a less extreme pitch movement compared to the monolingual. In [20], it was found that French-English learners could use prosodic information for interpretation, but their preference for prosodic phrasing was associated with their French proficiency. English-German learners and German-English learners were found not to fully transfer prosodic cues to disambiguate from first language to second language, which was due to their different use of pitch information [21]. In a study of Taiwanese English learners' prosodic disambiguation, it was demonstrated that the advanced English learners employed more distinguishable pause duration difference than the native speakers, while the limited learners could not pause reliably to disambiguate [22]. Prosodic disambiguation by Chinese learners of English was investigated in [23], finding that the learners used pauses more often to disambiguate, and that their intended meaning could not be effectively perceived by the native speakers. It was shown that the prosody of the learners' native language can influence the second language. The above studies also suggest that the learners' use of prosody to disambiguate in speech production is related to their language experience, and that it may be influenced by their native language.

The use of prosody in Chinese has been examined in speech production and perception. It was demonstrated in [24] and [25] that pause insertion, F0 lowering, intensity reduction, pre-boundary lengthening, and laryngealization at the prosodic boundary were the acoustic correlates to prosodic disambiguation. It has been proposed that pause is the primary prosodic cue and pre-boundary lengthening is less important. A comparison between the prosodic cues used in English and Chinese was conducted in [26], finding that both groups of speakers utilized pause insertion, pre-boundary lengthening and pitch changes to signal prosodic boundary, but their use of pitch changes differed. It has been proposed that pitch slope in English is more effective, while in Chinese pitch reset is more effective. English is an intonation language, while Chinese is a tonal language, in which pitch is used to distinguish lexical items at the word level, and signal variations in intonation at the sentence level [26]. In addition, it was proposed that the prosodic cues indicating prosodic boundaries may be language

*Corresponding author

 10.21437/SpeechProsody.2018-198

specific [25]. For this reason, it is assumed that the Chinese learners of English may employ prosodic cues differently from native speakers.

As indicated in recent studies carried out by [4] and [11], prosodic disambiguation in English has yielded mixed results regarding whether speakers could employ reliable prosodic cues to resolve syntactic ambiguity. Thus, it can be inferred that speakers' consistency in using prosodic cues to disambiguate may depend on the discourse situation and on their awareness of the ambiguity and the need to clarify [21]. A game-based conversation task was adopted in [11] to elicit target utterances, finding that both speakers and listeners could reliably use prosodic cues to disambiguate, even when they didn't notice the ambiguity. Thus, this study will also adopt a cooperative game task to ensure that all of the target utterances produced by the speaker are more natural and that their prosodic cues to disambiguate are effective for our analyses.

2. Method

In this study, a cooperative role-play game task was adopted to elicit the target utterances. The subjects served as the speaker (hereafter called the *director*) and the confederate as the listener (hereafter called the *mover*).

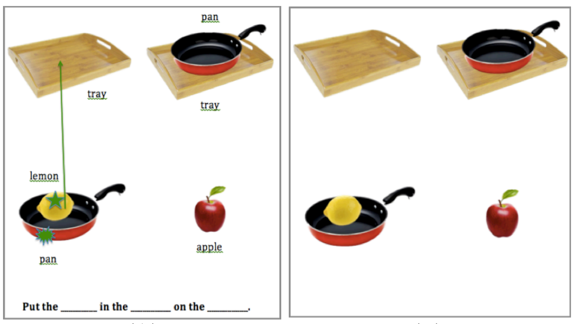
2.1. Production experiment

2.1.1. Participants

Directors were divided into two groups, i.e., the native speakers and EFL learners. 10 native speakers of American English (5 males and 5 females) and 20 Chinese EFL learners (13 males and 7 females) were recruited. All EFL learners were undergraduate students from Shanghai, whose native language was Mandarin Chinese. All had been learning English for approximately 10 years. The native speakers were also undergraduate students, who came to China for short-term international exchange. All subjects were 18 to 22 years old. In addition, two master students (1 male and 1 female), and one American-born doctorial student (male) acted as the confederate movers. No participant reported any speech or hearing disorders.

2.1.2. Materials

The target materials consisted of 10 pairs of PP-attachment sentences. These sentences were presented to the director as instructions for each display, and to the mover as manipulable objects on the computer screen, as illustrated in Figure 1. All of the displays for the mover were programmed into the computer.



(A) (B)

Figure 1: *Schematic sample set of the display presented to the instructor (A) and to the mover (B).*

Figure 1 (A) is the display as presented on a hand-held tablet computer to the director. It consisted of four objects: (1) the target object to be moved as indicated by the green star; (2) the destination object where to move the target object indicated by the green arrow; (3) the destination object if the PP-attachment sentence was interpreted incorrectly by the mover; and, (4) the distractor. Which object the director should indicate when giving instructions was marked by an asterisk. Figure 1 (B) is the display presented to the mover on the computer screen, and each of the objects could be moved randomly.

Each object's name was given beside the object in 16-point Times New Roman type, so that the directors would be consistent in naming the objects as well as to be more fluent in their speech. The sentence structure, "*Put the __ in the __*." was also given at the bottom of the picture to ensure that the target utterances could be elicited consistently. To avoid the mechanical production of target structures, filler trials were also inserted between each target trial. Giving the object names and images on screen above, with the fill-in sentence patterns elsewhere below, acted as an internal distractor activity and seemed to serve more spontaneous production while maintaining consistency. The filler trials had a variety of structures and the directors were given the freedom to speak them in any way they preferred.

2.1.3. Procedure

The experiment was performed in the professional recording studio at the School of Foreign Languages in Shanghai Jiao Tong University. The speech was digitized into a computer (Macbook Pro) by 16-bit, 44.1kHz sampling rate, using audio sound (AVID Mbox Mini). The mover took a seat in front of the computer, and the director was seated a little behind and to the side of the mover, so that they could exchange information. Moreover, the director could only use speech to give instructions, instead of eye gaze or gestures. Before beginning the game, participants were told that the aim of the experiment was for the director to use the given instructing statements to get the mover to perform the intended actions as indicated in pictures on the tablet computer. Confederate movers were told to perform the intended actions after they truly understood the instructions given by the directors. If there was any doubt, a mover was to ask for clarification, until the target utterances from the director were effective. For the native group, the native confederate acted as the mover; for the learner group, the Chinese student acted as the mover.

Before each trial, directors were told to briefly describe the objects in each display and their position to the mover before delivering the target sentence, so that they would be more familiar with the objects and more fluent in the target utterances. During the experiment, the director was to observe the actions performed by the mover. The director was to tell the mover if his or her action was right or wrong. If the mover performed incorrectly, the director was to repeat the instructions until the mover performed correctly. Before the experiment, eight practice trials were presented to the participants, so as to familiarize the participants with the task of the experiment

2.1.4. Acoustic measurement

Since in this study the utterances were elicited spontaneously, the directors occasionally made mistakes, repetitions, hesitations, coughs, giggles, and other disfluency. Only utterances that were correctly disambiguated and fluent were chosen. Two acoustic correlates (duration and F0) were analyzed to investigate the prosodic cues employed by the Chinese learners and the native speakers.

The duration data extraction was done with a Praat script ProsodyPro [27]. To measure duration correlates, measurements of pre-boundary lengthening, pause, as well as their combination were analyzed. To normalize the individual differences in speaking rates, the duration ratio was derived by dividing the duration of the critical items from that of the entire utterance. From this, a pre-boundary duration ratio and pause ratio were obtained. For the F0 data analyses, pitch reset was analyzed to signal the pitch changes at the critical location. The F0 data extraction was done with VoiceSauce using the STRAIGHT algorithm [28]. After F0 data was extracted, pitch contours for the critical pre-boundary and post-boundary syllables were drawn manually. To ensure a smooth pitch contour, observed errors were manually corrected by calculating the F0 in the narrow-bandwidth Praat window. Following this step, the F0 values were converted from hertz (Hz) to semitones (st) using the following formula:

$$f_{st} = 12\times log_2(f0/f0_{ref}) \quad (1)$$

in which f0 refers to the measured pitch values, $f0_{ref}$ is assumed to be 1 Hz.

In the calculation of pitch reset, this study followed the method in [26], taking the difference between the minimum F0 of the pre-boundary syllable and that of the post-boundary syllable. A script was edited to extract the maximum F0 and minimum F0.

2.2. Results

2.2.1. Duration

Figure 2 displays the mean duration ratios of the pre-boundary syllable, pause, and post-boundary syllable. The critical sentence structure in this study is composed of two prepositional phrases, i.e., "__ *in* __ *on* __". Thus, there might be two alternative prosodic phrasings or both: "__ # *in* __ *on* __" (context A), or "__*in*__ # *on*__" (context B). Thus, Pre1 refers to the first pre-boundary syllable, P1 the first pause, Prep1 the first preposition, Pre2 the second pre-boundary syllable, P2 the second pause, and Prep2 the second preposition.

It may be observed that compared with context B, Pre1 is lengthened and P1 is inserted in both groups in context A, and P2 of the learners is much greater than that of the native. In Context B, P1 of the learners is much shorter than that of the native. In addition, P2 is inserted and Pre2 is lengthened in both groups. Meanwhile, a lengthening could be observed on Prep2 among the native.

Mixed between-within subjects ANOVA with context as within-subjects variable and group as between-subjects factor were conducted to test the significance of durational changes. The results show that the changes are significant across the two contexts ($p < 0.001$), and significant differences also exist between the two groups ($p < 0.001$). To determine which item in the utterance caused the significant differences between the two groups, post-hoc pairwise comparisons were conducted. The results in Table 1 (N: Native, C: Chinese) show that for the native, significant differences exist in Pre1, P1, Pre2, P2, and Prep2 between context A and context B, while for the

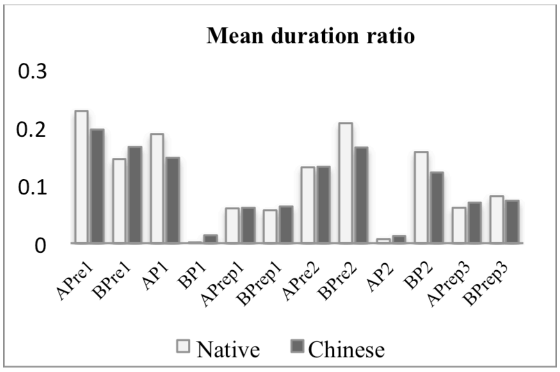


Figure 2: *Mean duration ratios averaged across 10 sentences by native speakers and Chinese learners in contexts A and B.*

learners, they are Pre1, P1 and P2. In context A, Pre1 and P1 of the native are significantly longer than those of the learners, while Prep2 is shorter than that of the learners. In context B, Prep1 of the learners are much longer than those of the native. In addition, Pre2, P2 and Prep2 of the native are significantly longer than those of the learners.

Three indications can be made from the above results: (1) the native display greater durational changes at the prosodic boundary: that is, greater pre-boundary lengthening and longer pause, than the learners; (2) the learners employ pauses more often to interpret the alternative meanings of PP-attachment; and, (3) the significant lengthening effect on Prep2 in context B by the native speakers and its significant differences between the native and the learners indicate that native speakers' prosodic cues may be influenced by the length of the phrase, and there is an initial strengthening, while that of the learners not.

Table 1: *Post-hoc pairwise comparisons for duration ratio.*

Context	Item	Sig.	Group	Item	Sig.
NA * NB	Pre1	s.	NA * CA	Pre1	s.
	P1	n.s.		P1	s.
	Prep1	n.s.		Prep1	n.s.
	Pre2	s.		Pre2	n.s.
	P2	s.		P2	n.s.
	Prep2	s.		Prep2	s.
CA * CB	Pre1	s.	NB * CB	Pre1	n.s.
	P1	s.		P1	n.s.
	Prep1	n.s.		Prep1	s.
	Pre2	n.s.		Pre2	s.
	P2	s.		P2	s.
	Prep2	n.s.		Prep2	s.

Note: sig. is the abbreviation of significance, s. is significant, and n. s. is not significant.

2.2.2. F0

In this study, pitch reset was measured to signal pitch changes at the prosodic boundary. From Figure 3, in which the positive value indicates a pitch reset at the boundary, while no reset is observed when the value is negative, the native speakers are supposed to display a pitch reset at the prosodic boundary in both contexts A and B. Whereas the learners exhibit a reset at the second boundary in context A, and at the first boundary in context B.

Two-way ANOVA show significant differences in pitch reset at the first boundary ($p < 0.05$) and the second boundary ($p < 0.005$) between the two groups. While for the native, the factor of context is not significant at the first boundary.

Post-hoc analysis was conducted to test the significance of pitch reset between the two groups. The reset by the native is shown to be significantly greater than that of the learners at the first boundary ($p < 0.001$) in context A, and at the second boundary ($p < 0.05$) in context B. This suggests that Chinese learners use pitch reset at the first boundary in context B, and at the second boundary in context A. While, the native use pitch reset at both the first and the second boundary in context A and context B, and no significance is observed between resets at the two boundaries, indicating that the use of pitch reset is not influenced by the phrase length.

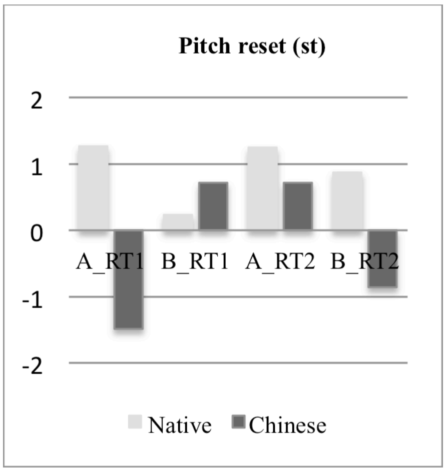


Figure 3: *Mean values of pitch reset.*

3. Discussion

This study adopted a cooperative game task to investigate the Chinese EFL learners' production of prosodic cues to resolve syntactic ambiguity. Both native speakers and learners were shown to use prosodic cues to distinguish the alternative meanings of the ambiguous PP-attachment sentences in this task. Acoustic analyses were conducted to find the prosodic cues employed by the native speakers and the learners, to discover any differences between these two groups.

For the duration correlates, pre-boundary lengthening and pause were analyzed. Both the native and the learners used pre-boundary lengthening and pause to indicate the phrasing of the utterances. But differences were observed between the two groups. First, the native showed greater duration changes on the pre-boundary items and pauses; second, mixed behavior in pitch reset was found among the learners; and third, the native speakers displayed a significant lengthening effect on Prep2 in context B. It's suggested that the learners might pause more often, resulting in more prosodic boundaries. In such a case, the learners presumably use relative prosodic boundary strength to resolve syntactic ambiguity: that is, they may employ two pauses in the utterances in both contexts A and B. The native speakers' lengthening of Prep2 in context B may result from different phrase lengths of the structures, e.g., "*Put the book in the box on the mat*" is composed of two phrases with three syllables and six syllables respectively in context A (i.e., "*put the book* / *in the box on the mat*"); while six syllables and three syllables respectively in context B (i.e., "*put the book in the box* / *on the mat*"). The native speakers, it was proposed, display domain-initial strengthening after a long phrase, while the leaners do not.

For the pitch correlate, pitch reset was calculated and analyzed. The native speakers employed pitch reset at the first and at the second boundary in contexts A and B. This may indicate that there is a pitch declination across the critical part in the utterance. In contrast, the learners' use of pitch reset was mixed. They showed pitch reset at the second boundary in context A and at the first boundary in context B, which may result from the fact that the learner group used pauses more often, hence producing more prosodic boundaries in contexts A and B. As the prosodic boundary is not consistent cross the contexts, their average use of pitch reset might also be different.

In conclusion, both native speakers and Chinese EFL learners use pre-boundary lengthening and pause to resolve syntactic ambiguity. However, the native also showed domain-initial strengthening immediately after the boundary, and employed greater durational changes on the pre-boundary item and pauses compared with the learners. Besides, they also displayed pitch reset at the first and at the second boundary in both contexts A and B, indicating a pitch declination across the critical part of the utterance. On the other hand, the Chinese EFL learners showed mixed pitch reset behaviours, indicating that the learners employed relative boundary strength to distinguish the alternative meanings of the PP-attachment sentences.

This study employed a game-based task to elicit the target utterances, so that more natural and effective utterances could be obtained for the acoustic analyses. However, in the analysis of pitch changes, the mixed results in pitch reset and durational changes suggest that the learners tend to use relative prosodic boundary strength more often to resolve syntactic ambiguity in PP-attachment: that is, to pause at the two boundaries in both contexts. However, only pitch reset was measured in this study. Other pitch changes, such as pitch slope, should also be included in the acoustic analysis, as it has been found that Chinese speakers and English speakers employed these two cues differently [26].

4. Acknowledgements

We wish to express our sincere thanks to Pengfei Shao and Chunlei Deng for their help in annotation and data extraction. Our thanks also go to the participants. The work was sponsored by the Interdisciplinary Program of Shanghai Jiao Tong University (14JCZ03) and the major project of national social science foundation of China (15ZDB103).

5. References

[1] H. Truckenbrodt, "The syntax - phonology interface," Cambridge Handb. Phonol., pp. 435–456., 2007.

[2] H. N. Nagel, L. P. Shapiro, B. Tuller, and R. Nawy, "Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing," J. Psycholinguist. Res., vol. 25, no. 2, pp. 319–344, 1996.

[3] I. Lehiste, “Role of duration in disambiguating syntactically ambiguous sentences,” J. Acoust. Soc. Am., vol. 60, p. 1199, 1976.
[4] J. Snedeker and J. Trueswell, “Using prosody to avoid ambiguity: Effects of speaker awareness and referential context,” J. Mem. Lang., vol. 48, no. 1, pp. 103–130, 2003.
[5] I. Lehiste, “Phonetic Disambiguation of Syntactic Ambiguity,” J. Acoust. Soc. Am., vol. 53, no. 1, pp. 380–380, 1973.
[6] P. J. Price, M. Ostendorf, and M. Park, “The use of prosody in syntactic disambiguation,” J. Acoust. Soc. Am., vol. 90, no. December 1991, pp. 2296–2970, 2015.
[7] F. Ferreira, “Creation of prosody during sentence production,” Psychol. Rev., vol. 100, no. 2, p. 233, 1993.
[8] D. W. Allbritton, G. McKoon, and R. Ratcliff, “Reliability of prosodic cues for resolving syntactic ambiguity,” J. Exp. Psychol. Learn. Mem. Cogn., vol. 22, no. 3, pp. 714–735, 1996.
[9] J. E. Fox Tree, P. J. A. Meijer, J. E. Tree, and P. J. A. Meijer, “Untrained speakers’ use of prosody in syntactic disambiguation and listeners’ interpretations,” Psychol. Res., vol. 63, no. 1, pp. 1–13, 2000.
[10] J. Schafer, S. R. Speer, P. Warren, and S. D. White, “Intonational disambiguation in sentence production and comprehension,” J. Psycholinguist. Res., vol. 29, no. 2, pp. 169–82, 2000.
[11] T. Kraljic and S. E. Brennan, “Prosodic disambiguation of syntactic structure: For the speaker or for the addressee?” Cogn. Psychol., vol. 50, no. 2, pp. 194–231, 2005.
[12] L. A. Streeter, “Acoustic determinants of phrase boundary perception,” J. Acoust. Soc. Am., vol. 64, no. November 1976, pp. 1582–1592, 1978.
[13] C. M. Beach, “The Interpretation of Prosodic Patterns at Points of Syntactic Structure Ambiguity : Evidence for Cue Trading Relations,” vol. 663, pp. 644–663, 1991.
[14] P. Warren, A. J. Schafer, S. R. Speer, and S. D. White, “Prosodic Resolution of Prepositional Phrase Ambiguity in Ambiguous and Unambiguous Situations,” Univ. Calif. Work. Pap. Phonetics, vol. 99, no. Dec, pp. 5–33, 2000.
[15] M. M. Kjelgaard and S. R. Speer, “Prosodic Facilitation and Interference in the Resolution of Temporary Syntactic Closure Ambiguity,” J. Mem. Lang., vol. 40, no. 2, pp. 153–194, 1999.
[16] J. P. Walker, K. Fongemie, and T. Daigle, “Prosodic Facilitation in the Resolution of Syntactic Ambiguities in Subjects with Left and Right Hemisphere Damage,” Brain Lang., vol. 78, no. 2, pp. 169–196, 2001.
[17] K. Steinhauer, K. Alter, and A. D. Friederici, “Brain potentials indicate immediate use of prosodic cues in natural speech processing,” Nat. Neurosci., vol. 2, no. 2, pp. 191–196, 1999.
[18] M. Wagner and D. G. Watson, “Experimental and theoretical advances in prosody: A review,” Lang. Cogn. Process., vol. 25, no. 7–9, pp. 905–945, 2010.
[19] E. M. Fernández, “The prosody produced by Spanish-English bilinguals: A preliminary investigation and implications for sentence processing,” Rev. da ABRALIN, vol. 4, pp. 109–141, 2005.
[20] A. L. Fultz, “The Use of Prosody for Disambiguation in English-French Interlanguage,” Proc. 9th Gener. Approaches to Second Lang. Acquis. Conf. (GASLA 2007), no. Gasla 2007, pp. 130–139, 2008.
[21] M. G. O’Brien, C. N. Jackson, and C. E. Gardner, “Cross-linguistic differences in prosodic cues to syntactic disambiguation in German and English,” Appl. Psycholinguist., vol. 35, no. 1, pp. 27–70, 2014.
[22] P. L. Yang, “English language proficiency and production of prosodic disambiguation: a preliminary study of Taiwanese English learners,” English Teach. Learn., vol. 34, no. 3, pp. 51–84, 2010.
[23] Y. Zhang and H. Ding, “A Preliminary Study of Prosodic Disambiguation by Chinese EFL Learners,” in *Interspeech*, 2017, pp. 374–378.
[24] X. S. Shen, “A pilot study on the relation between the temporal and syntactic structures in Mandarin,” J. Int. Phon. Assoc., vol. 22, no. 1–2, pp. 35–43, 1992.
[25] X. S. Shen, “the use of prosody in disambiguation in mandarin,” Phonetica, vol. 50, pp. 261–271, 1993.
[26] X. Zhang, “A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese,” 2012.
[27] Y. Xu, “ProsodyPro - A tool for large-scale systematic prosody analysis,” Proc. Tools Resour. Anal. Speech Prosody, pp. 7–10, 2013.
[28] K. Y. Shue, Y.-L., P. Keating , C. Vicenik, “VoiceSauce: A program for voice analysis,” in Proceedings of the ICPhS XVII, 2011, pp. 1846–1849.