

Design and Collection of an L2 English Corpus with a Suprasegmental Focus for Chinese Learners of English

Mingxing Li, Shuang Zhang, Kun Li, Alissa Harrison, Wai-Kit Lo and Helen Meng

Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong,
Hong Kong SAR of China

{mxli, zhangs, kli, alissa, wklo, hmmeng}@se.cuhk.edu.hk

ABSTRACT

English and Chinese have stark differences in both segmental and suprasegmental features. In particular, for Chinese learners of English as a secondary language (L2), there are expected deviations in the production of English prosody as a result of transfer of linguistic characteristics from their mother tongue. We have designed, collected and annotated a corpus to elicit suprasegmental information from over 200 Chinese learners of English, focusing on *lexical stress, utterance level stress, intonation, reduction of function words*, as well as *prosodic disambiguation*. The corpus provides empirical data for studies in pedagogy and computer-aided language learning (CALL), etc.

Keywords: suprasegmental corpus, annotation, L2 English, Chinese learner, Mandarin, Cantonese

1. INTRODUCTION

Understanding and speaking with proper prosody is important to language learning. Previous studies have shown that suprasegmental features have a stronger effect than segmental features on expert judgments of the proficiency of second language (L2) learners. It is common for L2 speakers to carry segmental and suprasegmental properties from their primary language (L1) to the target L2. While language transfer effects add color to L2 speech, in some cases they may also hinder communication. To facilitate the study of such effects in L2 English spoken by Chinese learners, we have designed and collected a corpus with a special focus on suprasegmental features.

There are stark suprasegmental differences between English and Chinese (e.g. Mandarin and Cantonese). To name a few: (i) Chinese is a tonal language with syllables carrying lexical tones, while English is not; (ii) Chinese shows stress patterns only in a small subset of its lexical items,

while stress is a prominent property of English words; (iii) Chinese is generally regarded as a syllable-timed language, while English is stress-timed.

The collection of our L2 English corpus focuses on the following five specific aspects:

- i) Lexical Stress
- ii) Utterance Level Stress
- iii) Intonation
- iv) Reduced / Unreduced Function Words
- v) Prosodic Disambiguation

We believe that the data provides empirical support for L2 prosodic studies, pedagogical studies, and the development of computer-aided language learning (CALL) technologies.

2. THE L2 ENGLISH CORPUS

The corpus includes spoken L2 English data from 217 Chinese speakers: 109 Mandarin (56 female and 53 male) and 108 Cantonese (55 female and 53 male). All speech recordings are prompted. The speakers read the text shown on a computer screen, with contextual information provided wherever necessary.

2.1. Corpus collection

The recording was conducted in a silent room, via a Sennheiser PC156 microphone headset. The data is sampled at 16kHz 16-bit linear PCM resolution. Each of the 217 speakers read 124 prompts (about 1 hour of recorded data). A total of over 25,000 utterances were collected and manually annotated. The details of the annotation procedures are presented in the later sections.

2.2. Corpus design

The prompting materials are designed with reference to the five aspects described in Section 1 above. Details are given in Table 1.

Table 1: Examples from the five parts of the L2 English suprasegmental corpus.

Part 1: Lexical stress
<i>Goal</i> To study how L2 English learners place and produce lexical stress at different positions
<ul style="list-style-type: none"> • 35 target words selected with lexical stresses located in various positions within the words • Target words embedded in a carrier sentence – “I say/said [target word] five/ten times”
<i>Example</i> “I said apartment five times”.
Part 2: Utterance-level stress
<i>Goal</i> To capture how L2 English learners produce a focused word in a sentence
<ul style="list-style-type: none"> • 35 words put in the narrow focus positions • Target words are highlighted (by boldface) and contextual descriptions are given
<i>Example</i> [context] <i>Did Bill lose everything in the robbery?</i> [prompt] “No, his money was taken”.
Part 3: Intonation
<i>Goal</i> To capture how L2 English learners produce intonations as instructed
<ul style="list-style-type: none"> • 29 utterances covering rising intonation, falling intonation and continuation rises • Speakers are instructed to read in rising tone when “↗” is shown. Similarly, “↘” indicates falling tone.
<i>Example</i> “Do you need any money? ↗”
Part 4: Reduced/Unreduced Function Words
<i>Goal</i> To study how L2 English learners read function words in the reduced and unreduced forms
<ul style="list-style-type: none"> • Five sentences covering the same function word in both reduced and unreduced forms
<i>Example</i> “If the birthday party wasn’t for Mary, then who was it for ?”
Part 5: Prosodic Disambiguation
<i>Goal</i> To study how L2 English learners use prosody in semantic disambiguation
<ul style="list-style-type: none"> • Information on possible semantic meanings of an ambiguous sentence is provided • Speakers requested to read the sentence with appropriate prosody to disambiguate between the different semantic meanings
<i>Example pair</i> [context] <i>The song was over.</i> [prompt] “I stopped singing”. [context] <i>I was walking along in the rain and suddenly felt very happy.</i> [prompt] “I stopped, singing”.

3. CORPUS ANNOTATION

The L2 corpus has both segmental and suprasegmental annotations – phonemes with stress marks, words, pitch accents, edge tones

(phrase accents and boundary tones), intonation patterns, narrow focus, reduced / unreduced function words, etc.

We performed forced alignment, using a home-grown automatic speech recognizer, between each recording and the canonical pronunciations in the reference word transcription. This generates a phonemic segmentation of the utterance, which provides the basis for annotation. A linguistically trained annotator then modifies the phoneme labels according to perception and maintains consistency by referring to defined guidelines. The Praat 5.1.44. platform is used for annotation.

3.1. Segmental labeling

We provide annotation of phonemes and words to all five parts of the corpus. Pronunciations in DARPABET are obtained by dictionary lookup from CMUDict 0.7. Should the perceived sound fall outside of DARPABET’s phoneme set, the closest phoneme is used for labeling.

3.2. Suprasegmental labeling

At the suprasegmental level, we provide both general (e.g., pitch accents and breaks) and specific (e.g., lexical stress, focus, reduced function word) annotations. Details are presented in this section.

3.2.1. Lexical stress

For lexical stress, annotation is based on the perception of pitch, duration and intensity. Vowels are labeled with “1” for primary stress and “2” for secondary stress, e.g., the word “motorcyclist” may be read as “/m ow2 t ax s ay1 k l ih s t/”.

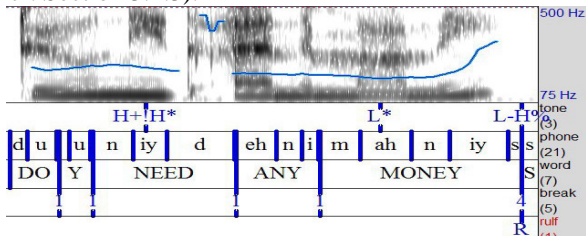
3.2.2. Tones and break indices

The ToBI annotation system [4,5,6] is used to describe the intonation patterns of the L2 English utterances. Eight labels **H***, **L***, **L+H***, **L*+H**, **H+!H***, **!H***, **L+!H***, **L*+!H** are used for pitch accents (see Figure 1). For cases where there is difficulty in deciding if a pitch accent exists, the label “*?” is used.

In order to capture the pitch movements at the end of an intonational phrase (IP), we adopt two types of phrase accents (**L-** and **H-**) and four types of edge tones to describe the intonational contour after the last pitch accent of the IP, i.e. **L-L%** (the falling to low contour as in the standard declarative contour), **L-H%** (as in ‘continuation rise’), **H-H%** (as in the canonical contour for yes-no questions, usually a high value for H%), and **H-L%** (a final ‘plateau’). Examples of the ToBI annotation are shown in the first tier of labels in Figure 1.

For annotating prosodic breaks, we used the ToBI break indices – **0**, **1**, **2**, **3**, **4**, **1-**, **2-**, **3-**, **1p**, **2p**, **3p**. Figure 1 also shows examples of break indices in the fourth tier of labels.

Figure 1: Examples from the ‘Intonation’ part of the corpus, showing pitch accents, phrase accents and edge tones (first tier of labels), phones and words (second and third tiers respectively), break indices (fourth tier), and RULF annotations (the bottom tier, cf. Section 3.2.3).



3.2.3. Intonation pattern in RULF

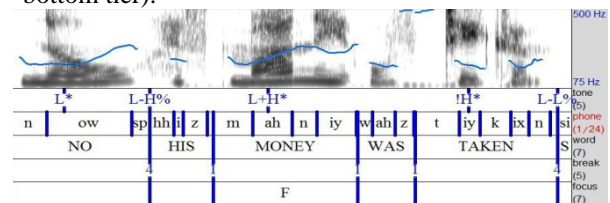
With reference to [7], we devised a coarse labeling scheme (RULF) for the overall intonation pattern of an IP. This RULF label is provided in the parts of the corpus on ‘Intonation’ and ‘Prosodic Disambiguation’. RULF is included as a coarse descriptive annotation of the intonation pattern. We plan to use it in for simplified evaluation of the production of intonation by L2 speakers. We annotate the utterances based on the intonational contour from the final pitch accent to the end of the IP (see bottom tier in Figure 1), according to the following procedure:

First, we decide whether the perceived intonation is rising (denoted by **R**) or falling (denoted by **F**). Should it be difficult to make such a distinction, we back off to decide between a high level of perceived intonation (denoted by **U**, for ‘upper’), versus a low level (denoted by **L**). Should this distinction be unclear, we use the label ‘?’ to indicate the uncertainty. Annotation using this coarse scheme is performed independently of the ToBI tone annotation.

3.2.4. Utterance-level stress (Narrow focus)

For the part of the corpus on ‘Utterance-level Stress’, particular words carrying narrow focus are marked with an ‘F’ label, to signify their prominence and discourse function in relation to the given context [8] (see Figure 2). In case of ambiguity, the label ‘?’ is used.

Figure 2: Annotation of narrow focus in the part of the corpus for ‘Utterance-level Stress’ (shown in the bottom tier).



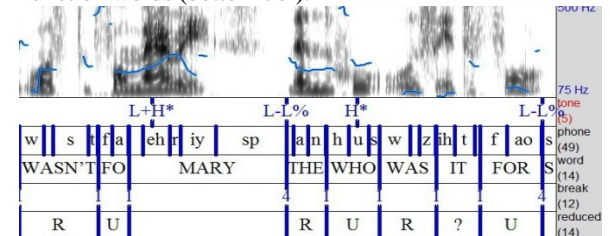
3.2.5. Reduced / unreduced function words

For the part of the corpus on ‘Reduced/Unreduced Function Words’, the notations used are:

- **R**: for words read in reduced (weak) forms [9]
- **U**: for words read in unreduced (strong) forms
- **?**: for uncertainty regarding **R** versus **U**

Judgment on a function word is made according to its vowel quality and its relative intensity with neighboring words. (see Figure 3).

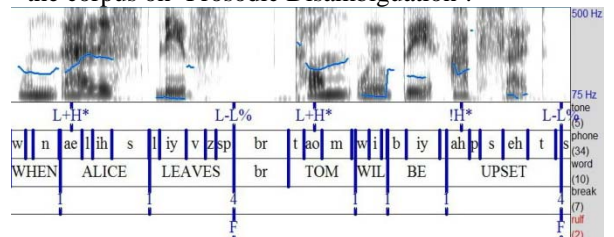
Figure 3: Annotation of ‘Reduced’ versus ‘Unreduced’ function words (bottom tier).



3.2.6. Additional pause label

Utterances in part of the corpus on ‘Prosodic Disambiguation’ are also annotated with an additional symbol ‘br’ in the phone and word tiers (see the second and third tiers in Figure 4). This symbol serves to indicate the use of pausing in semantic disambiguation.

Figure 4: Annotation of an utterance in the part of the corpus on ‘Prosodic Disambiguation’.



3.3. Summary of Annotations

Table 2 summarizes the annotations made available to each of the five parts in the L2 corpus, according to the design goal.

Table 2: Available annotation in each of the five parts in the L2 English corpus.

Parts of Corpus \ Annotation	Lexical Stress	Utterance-level Stress	Intonation	Reduced / Unreduced Function Words	Prosodic Disambiguation
Segments (DARPBET& words)	*	*	*	*	*
Lexical Stresses	*				
Coarse Intonation (RULF)			*		*
Narrow Focus		*			
Reduced / Unreduced				*	
Tone and Break Index (ToBI)		*	*	*	*

4. SALIENT SUPRASEGMENTAL ERRORS OBSERVED FROM THE L2 LEARNERS

4.1. Lexical stress

According to the annotated corpus, L2 learners made lexical stress errors on certain words. For example, “MISUNDERSTAND” may be read as “MISUNDERSTAND” (63%, 128 out of 201 tokens without deleted/added syllables); “REFRIGERATOR” as “REFRIGERATOR” (31%, 64 out of 206 tokens), etc.

Speakers generally made frequent errors only for a limited number of words in the corpus. This implies that they are capable of producing lexical stress, and errors may be due to insufficient or incorrect knowledge for certain words.

4.2. Pitch accents

We also observe that Chinese learners generally produce more accented syllables in utterances than native American English speakers. For example, they may produce pitch accents for consecutive syllables, instead of specific syllables such as those underlined in: “I DIDN’T MISUNDERSTAND THE QUESTION”. This may be attributed to language transfer effects from the learners’ L1. Chinese is a syllable-timed language, with every syllable taking up roughly the same amount of time 0. As they carry this prosodic feature to a stress-timed L2 like English, Chinese learners tend to use consecutive syllables as rhythmic units, rather than the stressed syllables in English.

4.3. Edge tones

Chinese learners tend to use **L-H%** for yes-no questions instead of **H-H%**. For example, the sentence “DO YOU NEED ANY MONEY?” usually ends with **L-H%** (92%, 199/217). That is, a low-rising pitch contour is produced instead of a high-rising contour for the edge tone of the IP.

Conversely, L2 learners sometimes use **H-H%** for continuation rise instead of **L-H%**. For example, “HE BOUGHT STRAWBERRIES, PINEAPPLES, BANANAS...” may be read as “STRAWBERRIES **H-H%** (17%, 30/217), PINEAPPLES **H-H%** (11%, 23/217)...”.

5. CONCLUSIONS AND FUTURE WORK

We have designed and compiled an L2 English corpus recorded from over 200 Chinese learners, with a focus on suprasegmental features. From the annotated data, we observed some common prosodic errors (e.g., in lexical stress, pitch accents, and edge tones) among the Chinese learners. We believe that this corpus will be very useful in supporting investigations in language transfer phenomena relating to prosody.

6. ACKNOWLEDGMENTS

This work is affiliated with the CUHK MoE-Microsoft Key Laboratory of Human-centric Computing and Interface Technologies. This project is partially supported by a General Research Fund (CUHK416108) from the Hong Kong SAR Government Research Grants Council. We also thank Professor Chiuyu Tseng of Academia Sinica and Dr. Tanya Viselgia of Ming Chuan University for helpful discussions.

7. REFERENCES

- [1] Anderson-Hsieh, J. *et al.* 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning*, 42:4.
- [2] Meng, H. *et al.* 2009. Studying L2 suprasegmental features in Asian Englishes: a position paper. *Interspeech 2009*, 1715-1718.
- [3] Mok, P. 2008. On the syllable-timing of Cantonese and Beijing Mandarin. In *Proc. 8th Phonetics Conference of China (PCC 2008) and the International Symposium on Phonetic Frontiers (ISPF 2008)*. Beijing.
- [4] Beckman, M. E., Elam, G. A. 1993/1997, Guidelines for ToBI Labelling. http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/singer_tobi.html.
- [5] Beckman, M. E. *et al.* 2005. The original ToBI system and the evolution of the ToBI framework. in Jun, S.-A. (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford Univ. Press.
- [6] Veilleux, N. *et al.* 2006. Transcribing prosodic structure of spoken utterances with ToBI. MIT Opencourseware.
- [7] Huang, X. *et al.* 2001. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Prentice Hall.
- [8] Gussenhoven, C. 2007. Types of focus in English. In Lee, C., *et al.* (eds.) *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*, Springer. 83-100.
- [9] Roach, P. 1991. *English Phonetics and Phonology: A Practical Course* (2nd edition). Cambridge Univ. Press.