# A Corpus-based Study on Prosodic Grouping and Boundary Tones in Mandarin Learners' English

*Sally Chen, Janice Fon*

Graduate Institute of Linguistics, National Taiwan University, Taiwan
{sallyhfchen, jfon}@ntu.edu.tw

## Abstract

This study investigated the prosodic grouping of L2 spoken English. A set of ten recordings was extracted from an in-progress learner corpus on an English proficiency test. Each recording consisted of two passages read aloud by a learner who had received a grade of 3, the median grade of the test, on a five-point scale. A group of ten native speakers was recruited to serve as control. They were given the same test materials and their readings were recorded under a test scenario similar to that of the L2 learners. Labeling followed the English ToBI convention. Results showed that in terms of break indices, liaisons only occurred among native speakers, and the L2 learners were in general not as fluent in reading the same texts aloud. In addition, L2 learners assigned more tones in their production and matched with the native speakers in assigning positions for boundary tones. However, the consistency of tone types between native and L2 speech was found only for major syntactic boundaries.

**Index Terms**: Prosodic grouping, boundary tone, ToBI, L2

## 1. Introduction

We derive our impression about other people from the way they intonate, and inappropriate intonation patterns may bring up misunderstandings [17, 18]. This is due to the fact that in addition to conveying linguistic information, intonation also plays an essential role in regulating discourse. In second language (L2) speech, inappropriate intonation is commonly observed. However, most L2 studies have been done on the investigation of segments, and several well-known models have been developed. For example, both the speech learning model (SLM) [6, 7] and the perceptual assimilation model (PAM) [3, 4, 5] predict the degree of success listeners have in perceiving nonnative sounds is based on the perceived phonetic distance between L1 and L2 sounds. Compare to that, research regarding suprasegments has been scant. Most related ones concentrated on intonational errors [16].

In terms of suprasegments, how people segment the speech stream and how they address prominence along the way they speak are the most important. Originally established on the phonology and phonetics of English intonation in the 1980's, Tone and Break Indices (ToBI) has been considered one of the most representative systems with regards to prosodic transcriptions, specifically prosodic grouping and tonal assignment [2, 19]. Nowadays, ToBI concepts have been applied to the labeling of prosodic features of various languages, such as German [8, 9], Greek [1], Japanese [21], and Korean [12].

The above applications, however, mainly focused on the language in the native sense. As Gut et al. have put it, "Cross-linguistic descriptions of the intonational system of languages are still few and far between." [10]

In accordance with the few empirical studies investigating non-native intonations, like segments, suprasegments of L2 production were also influenced by L1 transfer [11, 14]. Though rarely L2 speakers might reach the so-called "full attainment"—as the case reported by Mennen, given early exposure through immersion—the majority of the L2 speakers, even advanced learners, were found hard to resist this influence [13, 15, 16]. In the present study, we would like to take the initiative to figure out possible L1 influence on the prosodic features of L2 English production, and to explore the similarities and differences between L1 and L2 intonation.

## 2. Aims of the study

The aims of this study include: 1) whether L2 learners share the same prosodic grouping patterns as native English speakers; 2) whether L2 and native production have a similar accent density; 3) whether L2 learners assign boundary tones in the same way as native speakers.

## 3. Method

### 3.1. Subjects

#### 3.1.1. L2 learners

A set of 10 recordings was extracted from an in-progress learner corpus of General English Proficiency Test (GEPT), a nation-wise English proficiency held annually in Taiwan. These recordings all received a grade of 3, the median score, on a five-point scale.

#### 3.1.2. Native speakers

A group of 10 native speakers was recruited to serve as control. They were all native speakers of American English, as American accent has been the mainstream for learners in Taiwan. These participants were given the same test materials and their readings recorded in a sound-treated booth under a similar test scenario.

### 3.2. Materials

Two texts consist of 165 words in total were used. Sentences in both texts were all statements with punctuations clearly marked in the test sheet. The test sheets were given to the test takers/native speakers 1 minute prior to the recording. They were instructed to read at a comfortable speed. All readings had to be finished within a limit of 2 minutes.

### 3.3. Labeling

The labeling followed English ToBI convention. The two authors independently judged the data and both agreed on the labels.

Following the convention, in addition to the "words" tier, we also labeled two extra tiers: the "breaks" tier and the "tones" tier. As shown in Figure 1, the first tier is the "tones" tier, the second "words," and the third "breaks."

In the "tones" tier, single tones H and L are used alone or as elements of bi-tonal combinations of either pitch accents (ended with an asterisk *) or boundary tones (hyphen – marks intermediate phrasal boundaries and percent % marks intonational phrasal boundaries).

In terms of the "breaks" tier, break indices 4 (BI4s) are labeled for intonational phrasal boundaries, BI3s for intermediate phrasal boundaries, BI1s for prosodic word boundaries, and B0s for liaisons, or merges between any adjacent words. In addition, numbers followed by the letter *p* are also used to mark different levels of disfluency: 1p is used for truncation, 2p for any hesitation or prolongation without pitch reset of the following word, and 3p is used while the hesitation or prolongation is followed with a pitch reset. BI2s are used for any mismatch between the breaks tier and the tones tier.
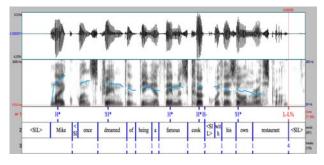


Figure 1: *Demonstration of the ToBI labeling.*

### 3.4. Analyses

For the current stage, three main analyses have been conducted on half of the data (5 people from each group). First, we examined the BI0 labels and labels that mark different levels of disfluency (BI-p) in native and L2 speech. Average number of tokens were calculated and compared.

The second analysis, accent density, was computed with the total number of pitch accents divided by the total number of words in the production for both groups. Given that native speakers and L2 learners were reading the same texts, higher accent density indicates more accents were assigned in the production.

Finally, we investigated the positions and tone types in which BI4 labels were given in the "breaks" tier. For position, major and minor syntactic boundaries, as well as regular word boundaries were examined. For tone types, distribution of different boundary tone combinations was calculated to examine the distributional pattern of the two groups.

## 4. Results

Results will be discussed following the order of the analyses as aforementioned.

### 4.1. BI0 and BI-p

Our impressionistic observation tells us that native speakers usually do not have a clear boundary for function words (*e.g.*, *and* is frequently pronounced as syllabic *n*) [20]. In addition, they might also merge the boundary between two adjacent words sharing the same phone in final and initial positions (e.g., *He is so*….). These phenomena were all clearly observed in our data via ToBI labeling.

Compared with spontaneous speech, read speech usually implies a more clear enunciation. However, as shown in Figure 2, the native speakers still showed an average of 5 liaisons in their read-aloud production, which was significantly different from L2 learners' production.[1]
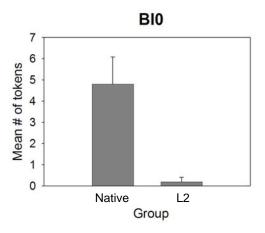


Figure 2: *Mean number of BI0 tokens.*

In terms of disfluency labels, BI-p, as expected, nonnative speech is not as fluent as native speech. In Figure 3, as compared to native speakers' production, the L2 group's production received more disfluency labels (-p) in terms of break indices.
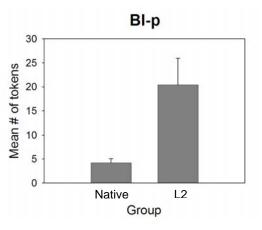


Figure 3: *Mean number of disfluency labels.*

### 4.2. Accent Density

As aforementioned, accent density was computed by dividing the number of pitch accents by the number of words

---

[1] All error bars in the present study stand for standard error.

in the production for each speaker of the L2 and native groups, as shown by the following formula:

$$\frac{\#\,Accent}{\#\,Word}$$

As shown in Figure 4, L2 learners had a significantly higher accent density ($M = 0.72$, $SD = 0.4$) as compared to that of the native speakers ($M = 0.54$, $SD = 0.3$). To put it in a simple way, about two thirds of the words L2 learners read were assigned with pitch accents, while only half of the words were read with accents among native speakers.

This significant difference might be contributed to the fundamental difference regarding rhythmic categorization of Mandarin and English, native language of the two groups of participants in this study.
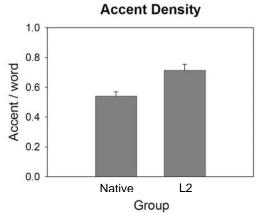


Figure 4: *Accent density of native and L2 speech.*

## 4.3. BI4 positions & accent types

### 4.3.1. BI4 positions

Boundaries were divided into three types in accordance with their positions in the syntactic structure: sentential (major syntactic boundaries), phrasal (minor syntactic boundaries) and regular word boundaries. Those marked as break index 4s were observed for their distributions.

In Figure 5, the data were plotted on the basis of subjects of each group. Thus, for the current stage, the maximal token for each boundary type was five. As observed, for both native speakers and L2 learners, BI4s were mainly assigned to major syntactic boundaries. In addition, native speakers assigned a slightly larger number of BI4s at phrasal boundaries, while L2 learners addressed more BI4s on regular word boundaries, which by default should be assigned with BI1s.
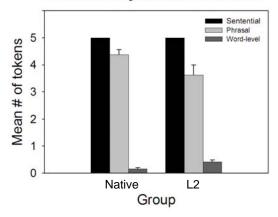


Figure 5: *Mean number of BI4 tokens placed at different levels of syntactic boundaries.*

Figure 6 (on the next page) illustrates a typical distribution of BI4 labels at different syntactic boundaries in the two groups of speakers. Native speakers showed a consistent tendency in assigning BI4s to sentential or phrasal boundaries, while L2 learners had a more "scattered" pattern. BI4s were also assigned to various word boundaries in addition to the regular occurrence on major and minor syntactic boundaries.

### 4.3.2. Accent types

With regards to the accent types assigned to these BI4 boundaries, intriguing results have been observed.

As shown in Table 1, for BI4 labels assigned to sentential boundaries, both native and nonnative speakers consistently placed L-L% tones, default tones for sentence-final boundary tones. For those assigned to a phrasal boundary, native speakers used the default boundary tone type in English, L-H%, for continuous intonation. However, L2 learners tended to use H-L% instead. On the rightmost column of the table, we can see that L2 speakers in general assigned more than twice as many BI4s as in regular word boundaries, and even in these cases, they still preferred assigning H-L% than L-H% as the boundary tone.

Table 1. *Distribution of accent types of BI4s at different levels of syntactic boundaries.*

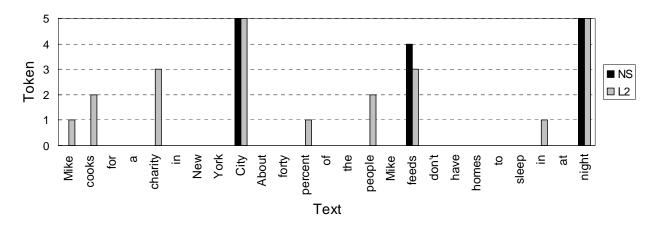|  | Sentential | | Phrasal | | Word-level | |
|---|---|---|---|---|---|---|
|  | NS | L2 | NS | L2 | NS | L2 |
| L-L% | 60 | 63 | 12 | 9 | 11 | 8 |
| L-H% |  |  | 16 | 6 | 9 | 5 |
| H-L% |  |  | 7 | 12 | 2 | 33 |
| H-H% |  |  |  | 2 | 1 | 14 |
| Total | 60 | 63 | 35 | 29 | 23 | 60 |

Figure 6: *Illustration of BI4s at different levels of syntactic boundaries.*

## 5. Conclusion

In the present study we not only support our impressionistic observations on native and L2 English read speech, but showed similarities and differences of the two types of speech on accent density as well as boundary position and tone types.

For the first part, native speakers were found to place liaisons even in read speech. L2 production was found as less fluent with significantly more disfluency labels marked for their speech. Their accent density was also significantly higher than that of the native speakers', indicating a negative L1 transfer.

For the second part, we may conclude that L2 learners shared the native prosodic grouping patterns on major syntactic boundaries as native speakers. Tone types were found consistent for both groups at sentential boundaries; however, the two groups preferred different tone types when assigning smaller boundaries.

## 6. References

[1] Arvaniti, A. and Baltazani, M., "Intonational analysis and prosodic annotation of Greek", in S.-A. Jun [Ed.], Prosodic Typology: The Phonology of Intonation and Phrasing, 84-117, Oxford University Press, 2005.

[2] Beckman, M. E., Hirschberg, J. and Shattuck-Hufnagel, S., "The original ToBI system and the evolution of the ToBI framework", in S.-A. Jun [Ed] Prosodic Typology: The Phonology of Intonation and Phrasing, 9-54, Oxford University Press, 2005.

[3] Best, C., "A direct realist view of cross-language speech perception", in W. Strange [Ed], Speech Perception and Linguistic Experience: Issues in Cross-language Research, York Press, 171-204, 1995.

[4] Best, C. T., McRoberts, G. W. and Goodell, E., "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system", Journal of Acoustical Society of America, 109(2): 775-792, 2001.

[5] Best, C. T., McRoberts, G. W. and Sithole, N., "The phonological basis of perceptual loss for non-native contrasts: maintenance of discrimination among Zulu clicks by English speaking adults and infants", Journal of Experimental Psychology, Human Perception Performance 14: 345-360, 1988.

[6] Flege, J. E., "Second Language Speech Learning: Theory, findings, and problems", in W. Strange [Ed] Speech Perception and Linguistic Experience: Issues in Cross-language Research, York Press, 233-277, 1995.

[7] Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H. and Tsukada, K., "Degree of foreign accent in English sentences produced by Korean children and adults", *Journal of Phonetics* 34: 153-175, 2006.

[8] Grice, M., Baumann, S. and Benzmuller, R., "German intonation in autosegmental-metrical phonology", in S.-A. Jun [Ed.], Prosodic Typology: The Phonology of Intonation and Phrasing, 55-83, Oxford University Press, 2005.

[9] Grice, M., Reyelt, M., Benzmuller, R., Mayer, J. and Batliner, A., "Consistency in transcription and labeling of German intonation with GToBI", Proceedings of the 1996 International Conference on Spoken Language Processing. Citation Delaware, 1716-1719. 1996.

[10] Gut, U., Trouvain, J. and Barry, W., "Bridging research on phonetic descriptions with knowledge from teaching practice-the case of prosody in non-native speech", in J. Trouvain and U. Gut [Eds], Nonnative Prosody: Phonetic Description and Teaching Practice, Mouton de Gruyter, 3-21, 2007.

[11] Jilka, M. 2007., "Different manifestations and perceptions of foreign accents of foreign accent in intonation", in J. Trouvain and U. Gut [Eds], Nonnative Prosody: Phonetic Description and Teaching Practice, Mouton de Gruyter, 77-96, 2007.

[12] Jun, S.-A., "Korean intonational phonology and prosodic transcription", in S.-A. Jun [Ed.], Prosodic Typology: The Phonology of Intonation and Phrasing, 201-229, Oxford University Press, 2005.

[13] Ladd, D. R., Intonational Phonology. Cambridge University Press. 1996.

[14] Ladd, D. R., Mennen, I. and Schepman, A., "Phonological conditioning of peak alignment in rising pitch accents in Dutch", Journal of Acoustical Society of America, 107(5) Pt. 1: 2685-2696, 2000.

[15] Mennen, I., "Second language acquisition of intonation: the case of peak alignment", Chicago Linguistic Society 34 (2). 1998.

[16] Mennen, I., "Bi-directional interference in the intonation of Dutch speakers of Greek", Journal of Phonetics 32: 543-563, 2004.

[17] Mennen, I., "Phonetic and phonological influences in non-native intonation: an overview for language teachers", QMUC Speech Science Research Centre Working Paper 9: 1-18, 2006.

[18] Mennen, I., "Phonological and phonetic influences in non-native intonation", in J. Trouvain and U. Gut [Eds], Nonnative Prosody: Phonetic Description and Teaching Practice, Mouton de Gruyter, 53-76. 2007.

[19] Pierrehumbert, J., The Phonology and Phonetics of English Intonation. Ph.D. dissertation, MIT, 1980.

[20] Raymond, W. D., Makashay, M. J., Dautricourt, R., Johnson, K., Hume, E. and Pitt, M., "Variation in conversation: An introduction to the Buckeye Speech Corpus", poster session presented at the annual meeting of the Acoustical Society of America, Ft. Lauderdale, FL, 2001.

[21] Venditti, J. J. "The J_ToBI model of Japanese intonation", in S.-A. Jun [Ed.], Prosodic Typology: The Phonology of Intonation and Phrasing, 172-200, Oxford University Press, 2005.