# A COMPARATIVE STUDY ON ACCENTUATION IMPLEMENTATION OF CHINESE EFL LEARNERS VS. AMERICAN NATIVE SPEAKERS

*Xia WANG[1], Yuan JIA[2], Aijun LI[2]*

[1]Nokia Research Growth Economies Lab, Beijing, China
[2]Institute of Linguistics, Chinese Academy of Social Sciences, China

## ABSTRACT

This paper investigates how Chinese EFL (English as foreign language) learners produce accentuation when speaking English. The study focuses on prosodic research of Chinese EFL Learners' English vs. native English through comparative evaluation of phonological pattern and accent related prosodic parameters. The research results show that the average length of intermediate phrases and intonational phrase is smaller in Chinese EFL learners' English than that in native English; the better the Chinese learner's English is, the closer the partition of intermediate/intonational phrases and the accent pattern are in his/her speech to those of native speakers; Chinese speakers tend to use pitch range amplification mechanism to realize accentuation rather than durational lengthening due to the negative language transfer from their native language, Chinese.

***Index Terms—*** Accentuation, pitch, duration, English as Foreigner Language (EFL), Prosody, Intermediate phrase, Intonational phrase

## 1. INTRODUCTION

Metrically, English is a stress language, while Chinese is a tone language. Among the three factors of pitch, length and loudness, pitch is considered to be the most efficacious and important cue to the perception of English stress [1]. In Chinese, pitch also plays an essential role in the production and perception of contrastive stress, weak stress and normal stress [2]. Xu [3] investigated the intonation realizations of statements and questions in American English by examining their interaction with focus and word stress. He pointed out that focus expands the pitch range of the focused syllable, and compresses that of the post-focus syllables, while leaving that of pre-focus syllables largely unaffected. As for the expansion of pitch range of the focused syllable, it can be either lowering the L target tone or raising the H target tone. Jia [4] pointed out that for Mandarin statements, the phonetic realization of the intonational stress is to enlarge the pitch range of the under-focus position, and compress the pitch range of the post-focus syllables. The enlargement is realized through raising the H target tone, while leaving L tone basically unchanged. Together with other research [5], it was proven that the intonation stress in Mandarin is realized mainly through the H tone. The differences of prosodic characteristics in English and Chinese demonstrated that the perception of prominence distribution pattern by the Chinese EFL learners might be different from the Americans.

In [6], I did extensive study on Chinese EFL learners' English prosody acquisition through comparative study on a parallel database. Overall, the speech rate of Chinese learners was much slower than that of native English speakers. Chinese learners' English had a lot more intonational phrases, i.e. each intonational phrase contained less syllables, while the number of sentence stress or accentuation was far more than that of Native American English speakers in this parallel database. In the native speaker's utterances, content words carried sentence stress, while in those of Chinese Learners, non-content words such as pronouns and prepositions often carried sentence stress too.

This paper focuses on how Chinese EFL learners produce accent in intermediate phrases and intonational phrases, in comparison with Native American English speakers, especially on the following aspects:

- Similarities and differences on phonological representation
- Similarities and differences on positions of accentuations
- Similarities and differences on duration of accentuated vowels
- Similarities and differences on pitch of accentuated units
- Relationship between English proficiency level and the prosodic parameter similarity to native speakers, as well as the impact of their native language, Chinese, on their English accent production.

## 2. METHOD

The material used in this study is a subset of CELSCOM (Chinese EFL Learners' Speech Corpus with Multi-accents) corpus [7]. The corpus is designed for comparative studies of phonetic aspects among Chinese dialect, regional Mandarin, standard Mandarin, EFL (English as foreign language) learners' English and native English, which will benefit speech recognition and synthesis, pronunciation evaluation and language teaching as well.

The materials used in this study contain 150 sentences, including imperatives, yes-no questions and statements. Each sentence was uttered twice by 19 speakers, in which 7 (6 male, 1 female) are American native speakers, mostly from western United States and 12 (5 male, 7 female) are Chinese EFL learners, from northern China. All of them had no self-reported speech or hearing disorders.

### 2.1. Recording

The speech materials were recorded in quite meeting room environment, at 16 kHz sampling rate and 16 bit precision, using Shure Beta53 microphone and FireFace 800 sound card.

### 2.2. Annotation

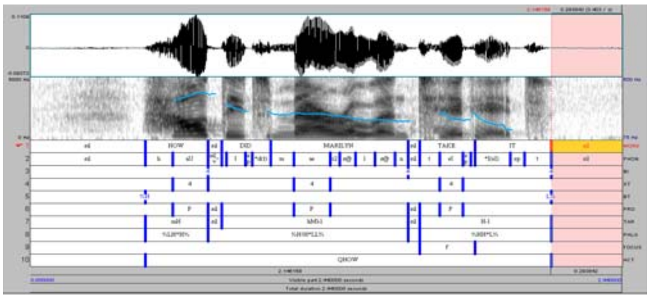

Figure 1. Annotation example

As shown in Figure 1, we adopt a combined ToBI and IViE system for annotation, to cover both segment annotation as well as prosody annotation.

Layer 1 WORD and Layer 2 PHON are for segmental features. WORD is the real pronunciation in word level; PHON is the real pronunciation in phoneme level. Layers 3 to Layer 8 are for prosody annotation. Layer 3 BI is Break Index, as in ToBI, in which "3" marks an intermediate phrase and "4" marks a intonational phrase. Layer 4 ST is Stress Tier, in which "3" marks the accentuated position in the intermediate phrase and "4" marks the accentuated position in the intonational phrase. Layer 5 BT is for Boundary Tone, using "H" and "L" to indicate the boundary tones. Layer 6 PRO is for prominence annotation, same as in IViE, using "P" to mark prominent syllables. Layer 7 TAR is for Target annotation, same as in IViE, to annotate the intonation target and its movement. Layer 8 PHLG is for phonologcal annotation, same as IViE. Layer 9 FOCUS is for sentence focus. Layer 10 ACT is to annotate the sentence functions.

### 2.3. Evaluation of oral English proficiency level

The oral English proficiency levels of Chinese speakers are evaluated by some American teachers from the United States, working or visiting China. The evaluation is mainly based on the learners' intonation. The American teachers were asked to rank the Chinese EFL Learners according to their oral English proficiency levels (see Table 1). In the following tables and figures, 'A' represents American and 'C' symbolizes Chinese; 'F' represents female and 'M' represents male, respectively; For example, CF01 indicates that this is a Chinese female speaker and her No. is 01, AM01 indicates that is an American male speaker with the No. of 01. The level of American Speakers' proficiency level is defined as 5. The Chinese EFL learners' are categorized into 3 levels, i.e. 2, 3, 4, in which 2 is least proficient and 4 is most proficient.

Table 1. Chinese EFL learners' oral English proficiency

| Speaker | Proficiency Level |
|---|---|
| CF03 | 4 |
| CF04 | 4 |
| CF06 | 4 |
| CF07 | 4 |
| CF01 | 3 |
| CF05 | 3 |
| CM09 | 3 |
| CM10 | 3 |
| CF02 | 2 |
| CM08 | 2 |
| CM11 | 2 |
| CM12 | 2 |

## 3. EXPERIMENT RESULTS

The study focuses on the similarities and differences of accentuation in Chinese EFL learners vs. Native American speakers. The research goal is approached through the examination on the following aspects: i)length of prosodic phrases; ii) positions of ST3 and ST4 boundaries; iii) positions of accentuation; iv) durational lengthening effect of accentuated vowels; v)Pitch range enlargement effect of accentuated vowels; and vi) prosodic parameters vs. English proficiency level.

### 3.1. Length of Prosodic Phrases

According to [6], there is clearly difference on the length of intermediate phrase and intonational phrase between Chinese EFL learners and Native American speakers.

To simplify, in the present paper, we treat all the vowels the same way, no matter it is monophthong or diphthong, long vowel or short vowel. The length of a prosodic phrase or a sentence is defined as the number of vowels in the phrase or the sentence.

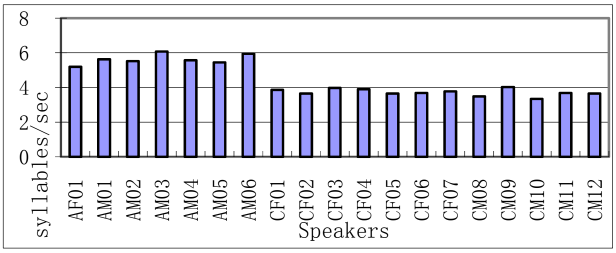

Figure 2. Statistical results of overall speech rate

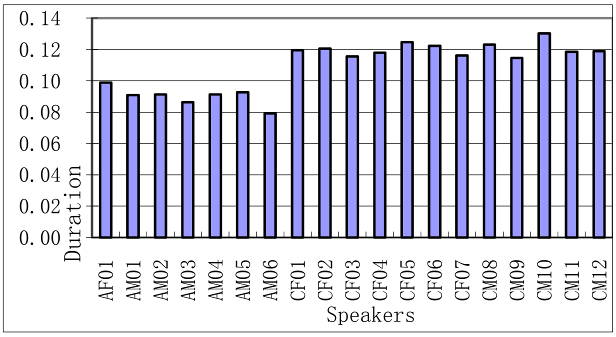

Figure 3. Statistical results of vowel length

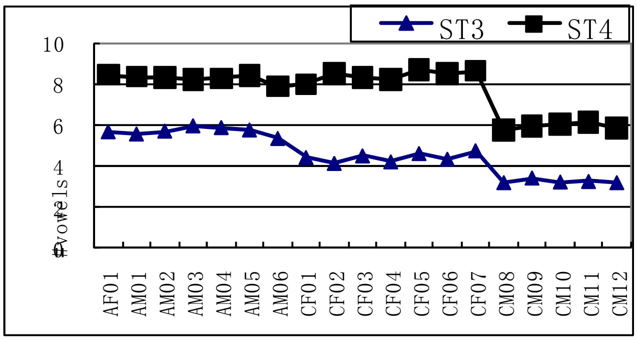

Figure 4. Length of ST3 and ST4

Figure 2 shows the average speech rate for each speaker. It is obvious that Native American speakers show faster speech, i.e. over 5 syllables per second, which is faster than Chinese Learners. Chinese EFL learners speak slower, less than 4 syllables per second, some below 3.5 syllables per second. There is no correlation between speech rate and English proficiency level.

Figure3 gives the average length of vowels. It is easy to understand that because Chinese EFL learners exhibit slower speech rate, so that their vowel length is longer than Native American speakers.

Figure 4 gives the statistics of average length of intermediate phrases (ST3) and intonational phrases (ST4) counted by number of syllables or vowels. This figure shows that Chinese EFL learners show more pauses and tend to break down the prosodic phrases into smaller units. For intonational phrases, Chinese female learners exhibited good learning capability. This figure indicates that intonational phrases manifest better stability. However, less proficient speakers (Chinese male speakers) exhibit much smaller intonational phrases in length. They tend to make more pauses in one sentence. Proficient learners have no acquisition problems for intonational phrases. While the intermediate phrase (ST3) is more flexible and Chinese EFL learners have difficulty on the acquisition of ST3 phrase boundaries.

### 3.2. Positions of ST3 and ST4 boundaries

We first analyze the position of prosodic boundaries of intermediate phrases (ST3) and intonational phrases (ST4) in the parallel database. For a sentence k with N syllables, we define the prosodic boundary vector as following:

$A_k = (a_1, a_2, .\Lambda\ , a_N)$, in which

$$a_n = \begin{cases} 0 & non-phraseboundary \\ 1 & IntermediatePhraseBoundary \\ 2 & IntonationPhraseBoundary \end{cases} \quad n = 1, \Lambda\ , N$$

For example, a sentence with 6 syllables could have an phrase boundary vector as A=(0, 0, 0, 1, 0, 2), indicating that there is an intermediate phrase (ST3) boundary at the fourth syllable and an intonational phrase (ST4) boundary at the last syllable.

In order to estimate the similarity of positions of phrase boundary among different speakers, we use the following e-index distance [Dang] to calculate the distance between 2 speakers. For the same sentence k, the distance of the phrase boundary vector by 2 speakers numbered as *i* and *j* is defined as:

$$S(a_i, a_j) = e \frac{\|a_i - a_j\|}{\|a_i\| \|a_j\|}$$

The bigger $S(a_i, a_j)$ is, the more similar the 2 speakers produce the same sentence regarding to the positions of ST3 and ST4 phrase boundaries.

Then we apply Mutli-Dimensional Scaling to the similarity matrix among 19 speakers for all the sentences and get the 3-D space distribution of all speakers as in Figure 5.

In the 3-D space, it is obvious that 19 speakers were categorized into 3 groups: Native American (AF01, AM01-05), Chinese Female (CF01-07) and Chinese Male (DM08-12). AM06 is different from other American speakers because this speaker comes from Southern East, not like the others who come from mid-west. The distance from one group to another is significantly different from each other. This shows from the places of ST3 and ST4 boundaries, Chinese female has better proficiency, which is in line with the proficiency evaluation. Native speakers show clear consistency on the place of ST3 and ST4 boundaries. Chinese male speakers are scattered in the space because of lower proficiency level and lacking of consistency.

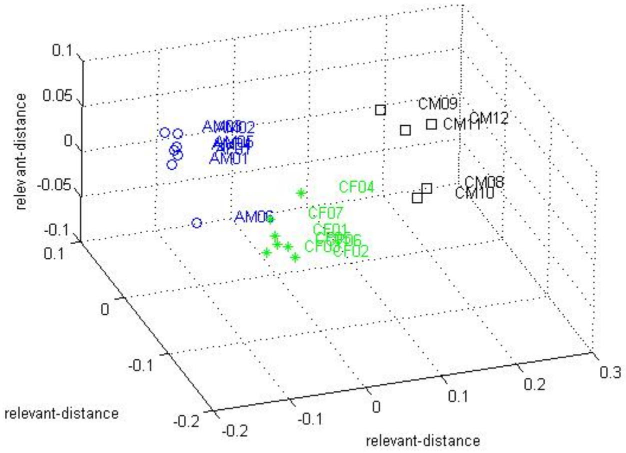

Figure 5. Positions of ST3 and ST4 boundary similarity distribution

### 3.3 Positions of accentuation

Now we come to the position of accentuation in the parallel database. For a sentence k with N syllables, we define the accent vector as following, similar to the phrase boundary vector:

$B_k = (b_1, b_2, .\Lambda\ , b_N)$, in which

$$b_n = \begin{cases} 0 & non-accent \\ 1 & IntermediatePhraseAccent \quad n = 1, \Lambda\ , N \\ 2 & IntonationPhraseAccent \end{cases}$$

For example, a sentence with 6 syllables could have an accent vector as B=(0, 0, 0, 1, 0, 2), indicating that there is an intermediate phrase accent at the fourth syllable and an intonational phrase accent at the last syllable.

In order to estimate the similarity of accent positions among different speakers, we use the following e-index distance [8] to calculate the distance between 2 speakers. For the same sentence k, the distance of the accent vector by 2 speakers numbered as *i* and *j* is defined as:

$$S(b_i, b_j) = e\frac{\|b_i - b_j\|}{\|b_i\|\|b_j\|}$$

The bigger $S(b_i, b_j)$ is, the more similar the 2 speakers produce the same sentence regarding to the accentuation pattern.

Then we apply Mutli-Dimensional Scaling to the similarity matrix among 19 speakers for all the sentences and obtain the 3-D special distribution of all speakers as in Figure 6:

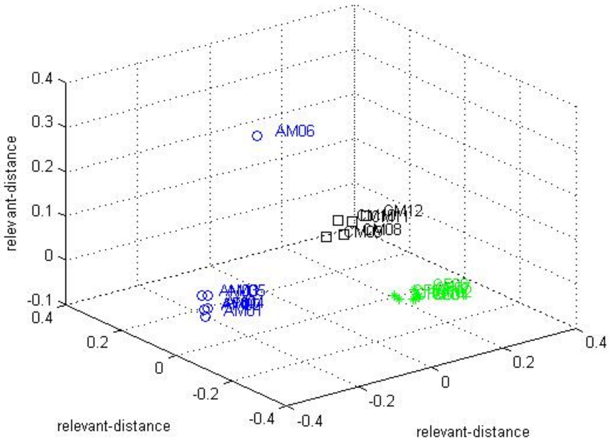

Figure 6. Accentuation position similarity distribution

Obviously from the distribution map, the 19 speakers were categorized into 3 groups: Native American (AF01, AM01-05), Chinese Female (CF01-07) and Chinese Male (DM08-12). AM06 is different from other American speakers due to the f that this speaker comes from Southern East, not like the others who come from mid-west. The distance from one group to another is significant. This shows that the place of accentuation is difficult to learn for EFL learners. Native speakers show clear consistency on the place of accentuation.

### 3.4. Durational lengthening effect of accentuated vowels

The most important acoustic parameters of accentuation are duration and pitch. We will investigate these two parameters in more detail in this section, with the emphasis on the intermediate phrase and intonational phrase.

In all duration analysis, we use normalized duration instead of original duration to get rid of the impact of different speech rate from different speakers.

First, we normalize the phoneme duration using Z-score algorithm, where,

DurZ is defined as the average duration of all the vowels except that those vowels carrying intermediate & accents of intonational phrase.

DurZST3 is defined as the average duration of all the vowels carrying accents of intermediate phrase.

DurZST4 is defined as the average duration of all the vowels carrying accent of intonational phrase.

The normalized duration of Z-score reflects the overall lengthy effect of accentuated vowel and the compression of the non-accentuated vowel. Figure 7 shows that for both Chinese and American speakers, we could observe the vowel lengthy effect on accents, and the higher level the accent is, the lengthy the accentuated vowel is. The native speakers demonstrate bigger durational lengthening effect on accents than Chinese. In Figure 7, A indicates the average duration for all American speakers, while C indicates the average duration for all Chinese ELF learners.

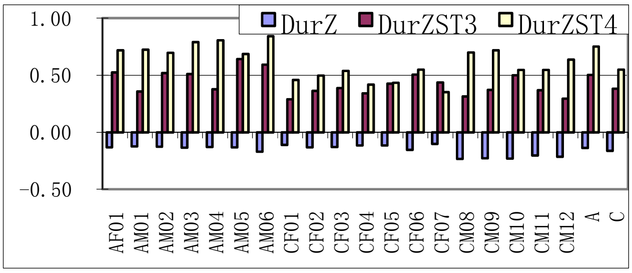

Figure 7. Z-Score normalized duration of ST3 and ST4 vowels

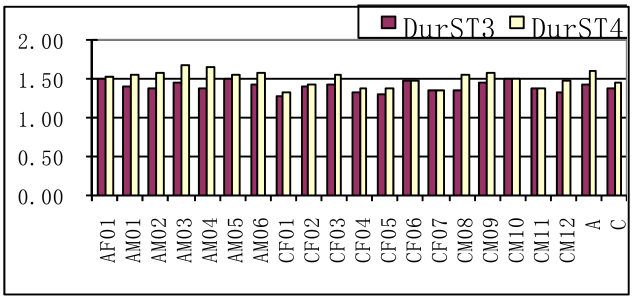

Figure 8. Local normalized duration of ST3 and ST4

We also apply normalization in the local context within the prosodic phrase, defined as:

$$NormD = \frac{Dur}{meanDur}$$

in which, Dur indicates that the duration of the vowel on accentuation in the prosodic phrase, either intermediate phrase or intonational phrase; meander is the average duration of all the other vowels except for the accentuated one in the same phrase.

Here we also look at the locally normalized duration of intermediate phrase (noted as DurST3) and intonational phrase (noted as DurST4). These parameters reflect the lengthening effect within a prosodic phrase context. Figure 8 shows that the duration of accentuated vowels in both intermediate phrase and intonational phrase has clear lengthening effect, e.g., about 40% - 60% more than that of other vowels. It's of course more obvious in intonational phrases. Again, the behavior of Chinese EFL learners is clearly different from native speakers.

Therefore, native speakers make good use of vowel lengthening in connected speech. They put more emphasis on accentuation, lengthening the accentuated vowels and shortening the non-accentuated vowels. Chinese EFL learners tended to lengthen the accentuated vowel as well, but with much less durational dynamics.

### 3.5. Pitch range enlargement effect of accentuated vowels

We have summarized our findings on the durational parameters. Here we will look at pitch or $F_0$ related parameters.

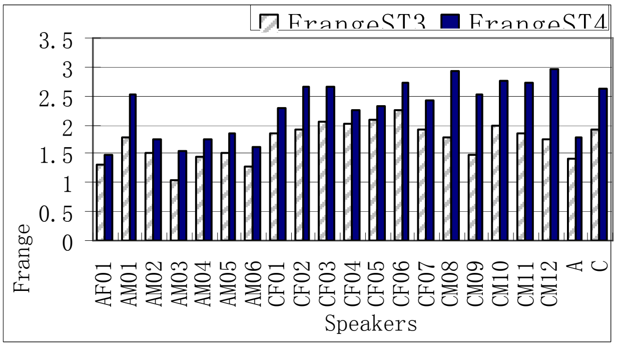

Figure 9. Local normalized $F_0$ range of ST3 and ST4

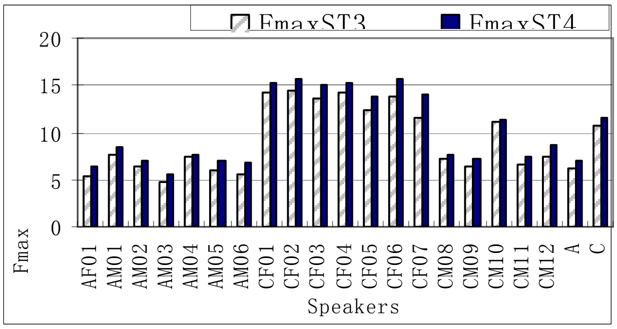

Figure 10. Local normalized $F_0$ maximum of ST3 and ST4

For Chinese language, generally speaking, the higher the level of accentuation is, the higher the $F_0$ goes, the larger the pitch range is on the accentuated syllables. While English is stress timed language, there is no concept of tones on each syllable. Therefore, intonation will impact on the $F_0$ movements as well. To simplify, in this section, we will put sentence intonation aside and focus on the pitch movements on accentuation.

For each vowel, we will calculate the maximum value of its $F_0$, noted as Fmax, and the minimum value of its $F_0$, noted as Fmin. The pitch range of this vowel is defined as Frange, and Frange = Fmax – Fmin.

Similarly to the local normalization process, we also calculated the normalized maximum pitch value and normalized pitch range parameters based on the local context within intermediate phrases, noted as FmaxST3 and

FrangeST3, respectively. Same process is applied to intonational phrases, and we get Fmax ST4 and FrangeST4.

Figure 4 shows the local normalized $F_0$ range on accentuated vowels of intermediate phrase (ST3) and intonational phrase (ST4), in the scale of semi tones. Figure 5 shows the local normalized $F_0$ maximum value for accentuated vowels in ST3 and ST4.

From Figure 9 and Figure 10, the higher the accentuation level is, the higher the $F_0$ range is, the higher the $F_0$ maximum value is. Some American speakers such as AF01, AM03, etc. does not show enlarging effect on $F_0$ rising or pitch range on accentuation. The reason is that, for Native American speakers, lowering pitch is also one mechanism for the mark of accentuation, which is seldom used in Chinese EFL learners.

The similarity matrix of FmaxST3, FmaxST4, FrangeST3, and FrangeST4 among different speakers shows that: American speakers are close to each other in the space; Chinese female speakers are close to each other as well, but distant to American group; Chinese male speakers are scattered in the space. Chinese EFL learners tend to overuse pitch movement, raising $F_0$, enlarging $F_0$ range, to implement accentuation, which is in line with the accentuation behavior in Chinese. This is a negative transfer of their mother tongue, Chinese.

### 3.6. Prosodic parameters vs. English proficiency level

We nail down the research question to different functional sentences, more specifically to statements and yes/no questions. The observation is in line with the findings described before on duration, $F_0$ range and $F_0$ maximum value. Chinese ELF learners exhibit less lengthening effect on durational of accents than Native American English speakers; while they show bigger pitch movements than Native Speakers due to negative transfer of Chinese, a tonal language as mother tongue.

This is especially obvious in average $F_0$ value on accentuation in yes/no questions. In [9], we found out that Native American speakers applied low-rising tone (L*H) on nuclear accent, while Chinese EFL learners show difficulty to realize nuclear accent in that way. If the nuclear accent is not on the final syllable of the sentence, they tended to use high level tone or falling tone on nuclear words. In F0 parameter analysis, we get the same results on yes/no questions. The average value of $F_0$ on accentuation is much lower in American Speakers than in Chinese EFL learners. This is because of the negative transfer of Chinese, which makes Chinese EFL learners tend to raise $F_0$ on accents.

## 4. CONCLUSIONS

Based on the analysis of large parallel corpus with accurate annotations, we draw the following conclusions:

1. The speech rate of Chinese EFL learners is much slower than that of Native American speakers. The lower the Chinese EFL learner’s English proficiency is, the more the pauses are in their speech;
2. Length wise, there are generally less syllables in Chinese EFL learners’ prosodic phrases, including both intermediate phrases and intonational phrases. In comparison with native American, the higher the EFL Learner’s English proficiency level is, the more similar the prosodic boundary patterns are;
3. Similarly, comparing to Native American speakers, the higher the EFL learner’s English proficiency level is, the more similar the accentuation positions are;Acoustic parameter analysis shows that both Chinese EFL Learners and Native American speakers lengthen the vowel duration on accents, and enlarge the $F_0$ range. Native English speakers lower the $F_0$ to implement accentuation, while Chinese EFL learners typically raise the $F_0$ to implement accentuation. Chinese ELF learners tend to use pitch movement, not durational lengthening to implement accentuation.
4. Statistical analysis on $F_0$ supports the previous findings on the difficulty of L*H pitch accent pattern [9].

## 5. REFERENCES

[1] A. Cruttenden, *Intonation*, 2 ed., Beijing, Peking University Press, 1997.

[2] S. Duanmu, The *phonology of Standard Chinese*, Oxford, Oxford University Press, 2000.

[3] F. LIU, and Y. XU, "Question intonation as affected by word stress and focus in English," *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 2007.

[4] Y. JIA, et al., "The Effect of Focal Accents upon Sentential Pitch in Standard Chinese," *Report of Phonetic Research, vol. 2006*, pp. 61-68, 2006.

[5] J. CHEN, "Contrastive study on prosodic aspects for standard and Shanghai-accented Chinese," *Master thesis*, Zhejiang University, Hangzhou, 2004.

[6] X. Wang, “Phonetic Research on English Prosody Acquisition of Chinese Learners based on a Large Comparative Speech Corpus”, *Ph.D Dissertation*, Chinese Academy of Social Sciences, 2010.

[7] X. Wang, A.J. Li, Z.Y. Xiong, Z.G. Yin, "Multi-accent and multi-lingual speech corpus," *Proceedings of O-COCOSDA 2008*, Japan, 2008.

[8] J. Dang, et al., “Comparison of Emotion Perception among Different Cultures”, *APSIPA proceedings*, Sapporo, Japan, 2009

[9] X. Wang, A.J. Li, X.L. Ji, “Intonation Patterns of Yes-No Questions for Chinese EFL learners”, *Proceedings of O-COCOSDA 2009,* China, 2009

[This paper was published in O-COCOSDA 2011]