Perception and Production of Prominence Distribution Patterns of Chinese EFL Learners

WANG Xia[1,2], LI Aijun[1], JI Xiaoli[1,3]

[1] Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China;
[2] Nokia Research Beijing; [3] Zhejiang University, Hangzhou, China

xia.s.wang@nokia.com, liaj@cass.org.cn,jixiaoli2005@126.com

Abstract

This paper explores the relationship between the perception and production of prominence distribution patterns through a perceptual experiment on both Chinese EFL (English as foreign language) learners and native English speakers. Seven American English native speakers and twelve Chinese EFL learners contributed to the production database, in which each speaker was required to read 32 utterances with different sentence types and focus types (narrow or broad). Six Americans and nine Chinese EFL learners were recruited for the perceptual experiment on the judgment of prominence distributions. The similarities of the distribution patterns of prominences were calculated and MDS analysis was conducted from the speakers' and listeners' perspectives. The results exhibited a close correlation between perception and production of prominence distribution patterns and learners' oral English proficiency levels.

1. Introduction

The relationship between speech production and speech perception has usually been assumed to be straightforward. However, Neufeld [1] cautioned that a distinction must be made between acquisition phenomena and performance phenomena because "there is frequent asymmetry in the adult's receptive and production performance in L2 at the phonological level". Chun [2] pointed out that more cross-linguistic research of this nature is needed: first, on the perception of intonation patterns in L2; second, on whether correct perception must always precede correct production (or whether accurate perception and production can be achieved at about the same time or in either order); and third, whether perceptual training improves perception and/or production.

Intonation has long been noted as an important linguistic phenomenon in verbal communication, where it commonly serves the function of helping to convey an intended meaning. Understanding prosodic structures plays an important role not only to native listeners, but also to second language learners' perception and production of speech. In the phonological study of intonation, Pirrehumbert [3] specifies three types of tonal events for the tonal inventory of English intonation, namely, seven pitch accents, two boundary tones and two phrase accents. Research has proved that intonation boundary location influences the interpretation of globally ambiguous particle constructions [4]. Also, pitch accent placement and type have been shown to affect sentence comprehension by conveying the focus and information structure of an utterance.

Metrically, English is a stress language, while Chinese is a tone language. Among the three factors of pitch, length and loudness, pitch proves to be the most efficacious and important cue to the perception of English stress [5]. In Chinese, pitch also plays an essential role in the production and perception of contrastive stress, weak stress and normal stress [6]. Xu [7] 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 [8] 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 [9], 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.

Thus, the present study focuses on the relationship between the perception and production of prominence distribution patterns of Chinese EFL learners. In general, the question is whether one needs an 'English' ear to learn English? Does one need to perceive correctly in order to produce correctly? Concretely there are two research aspects that this study attempts to address. Firstly, is there any relation between Chinese EFL learners' oral English proficiency and their production of prominence distribution patterns? If so, the Chinese EFL learners with high proficiency level should have similar prominence distribution patterns as the American speakers. This assumption can be verified by comparing the perceptual results of prominence distribution pattern produced by native speakers with those of the learners from the American native listeners. Secondly, based on the first question, this study also intends to see if the learners' perception of the prominence distribution patterns is closely related to their English proficiency levels. This can be tested by comparing the perceptual results given by native listeners with those given by Chinese listeners on the sub-corpora produced by American native speakers between native listeners and Chinese listeners. Hopefully, the previous two assumptions can provide important evidence on the relationship between the Chinese EFL learners' speech production and speech perception.

2. Perceptual experiment

2.1. The speech database

The speech materials consist of 8 sentence pairs as listed in Table 1 from the CELSCOM corpus [10]. Each sentence was uttered as two sentence types (statement vs. *yes-no* question) and two focus types (narrow vs. broad).

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. This production database contains 608 (16*2*19) utterances in total.

The sentences were digitized at 16 kHz sampling rate and 16 bit precision. The sentences with narrow focus were embedded in dialogue to make the narrow focus self-explanatory in the recording prompt. Therefore what the speaker read was the complete dialogue. During the perceptual experiment, the context was removed and only the sentences with narrow focus were presented to the listeners to eliminate the effect of context on perception.

It is worth to clarify that the reason why we use different groups of people as the speakers and listeners: the speakers have to know the focus of all the sentences in reading, so they must not attend the perceptual experiment.

Table 1. *List of production materials. The words in bold were the focus when the sentences were read with narrow focus.*

S	Y-N questions vs. statements
1	Could you be **mistaken**? / You could be **mistaken**.
2	Can I **count** on that? / I can **count** on that.
3	Does it **matter** all that much? / It does **matter** all that much.
4	Have you got enough **money** in hand? /You have got enough **money** in hand.
5	Would you say it's a practical **proposition**? / You would say it's a practical **proposition**.
6	Can you find out whether John will **help**? /You can find out whether John will **help**.
7	Can I see him if I come back **later**? /I can see him if I come back **later**.
8	Shall we be able to finish the job on **time**? / We shall be able to finish the job on **time**.

2.2. Oral English proficiency evaluation

The oral English proficiency levels of Chinese speakers and listeners were evaluated by some American teachers from the United States, working or visiting China. The criteria for evaluation were mainly the learners' intonation. The American teachers were asked to rank the Chinese speakers and listeners separately according to their oral English proficiency levels (see Table 2). In the following tables and figures, 'A' represents American and 'C' symbolizes Chinese; 'F' represents female and 'M' represents male; 'LA' indicates American listeners and 'LC' indicates Chinese listeners. For example, CF01 means the first Chinese female speaker, LC01 means the first Chinese listeners. 'H' stands for high grade, 'M' for middle grade and 'L' for low grade.

Table 2. *Ranking of Chinese EFL learners' oral English proficiency levels*.

Spks.	grade	rank	listeners	grade	rank
CF07	H	1	LC02	H	1
CF04	H	2	LC07	H	2
CF05	M	3	LC08	H	3
CM10	M	4	LC06	H	4
CF01	M	5	LC01	M	5
CM08	M	6	LC09	M	6
CF06	L	7	LC04	M	7
CF02	L	8	LC05	L	8
CF03	L	9	LC03	L	9
CM09	L	10			
CM12	L	11			
CM11	L	12			

2.3. The setup of perceptual experiment

Six Americans and nine Chinese EFL learners were invited to participate in the perceptual experiment. All the sentences were randomized before being presented to the listeners. The listeners were instructed to mark the prominent words while listening to these sentences. There was no limitation on the number of prominences for each sentence.

The word marked as prominent by the listener was coded as '1' while the opposite were coded as '0'. The prominence distribution patterns can be represented by calculating the prominence ratio of each word to the total number of words. The 4 sub graphs in Figure 1 illustrate the prominence distribution patterns of the sentence 'Have you got enough money in hand?' and its statement counterpart perceived by the 6 American and 9 Chinese listeners respectively. From the figure, we can see that for the question "Have you got enough money in hand?", five Americans marked prominence on the word 'money' while five Chinese listeners marked the word 'hand' as prominent; four Chinese listeners marked the word 'got' as prominent as well. For the statement 'You have got enough money in hand', all the listeners (six Americans and nine Chinese were included) marked the word 'money' as prominent.

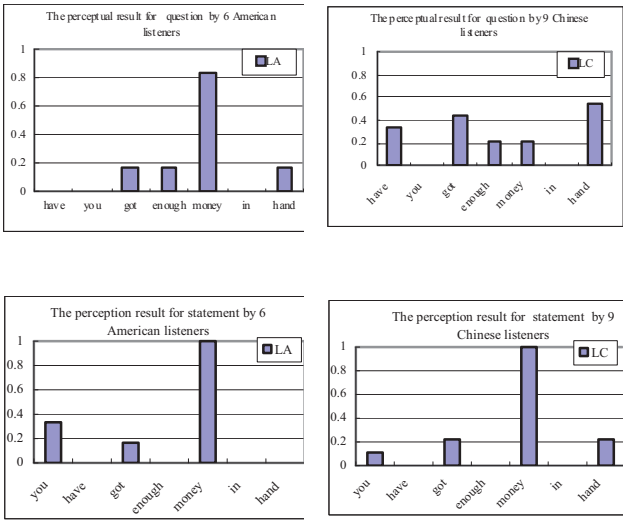

Figure 1: *The prominence distribution patterns of the sentence "Have you got enough money in hand?" and its statement counterpart spoken by AM05 perceived by six American listeners and nine Chinese listeners.*

3. Result Analysis

3.1. Similarity and MDS analysis

In order to estimate the similarity of perceived prominence distribution patterns among the listeners, we used the following e-index distance [11] to calculate the similarity of the patterns of prominence distribution for each sentence,

$$S(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{\|x_i\|\|x_j\|}}$$

where x_i and x_j represent two vectors of the listeners' perceptual results of the same sentence. The higher the value of $S(x_i, x_j)$, the more similar the two distribution patterns are. For the above example, the e-index distance between Americans and Chinese listeners for the question and statement respectively is 0.38 and 0.91.

Afterwards, MDS (multi-dimensional scaling) analysis was made to get the 3-D spacial relationship of the perceived results, as shown in Figures 3 and 4.

3.2. The speakers' oral English proficiency levels and their prominence distribution patterns

Before further analysis of the speakers' prominence distribution pattern, we first checked whether the six native listeners were consistent in their perception of prominence. According to the perceptual results on the 7 American speakers' speech materials, there was significant difference among the six American listeners ($F(5,204) = 5.226$, $P<0.001$). Turkey HSD post-hoc test indicated that, only the listener LA03 was different from the other five native listeners. One-way ANOVA analysis showed that there was no significant difference among the other five American listeners ($F(4, 170) = 1.393$, $P = 0.239 > 0.05$), which meant that the perception of the prominence distribution by native listeners was a valid method for analyzing the speakers' prominence distribution patterns. Therefore, similarities of speakers' prominence distribution patterns through the perception were calculated based on the results of these five American listeners.

Results of one-way ANOVAs demonstrated that there was a significant difference among the nineteen speakers' prominence distribution patterns (F (18, 737) = 4.269, P<0.001). Thus MDS was adopted to estimate the distance among them. Perceptually similar stimuli showed up in nearby positions in the space. In Figure 2, we can see that the prominence distribution patterns produced by Chinese EFL learners CF04, CF07, CF01 are closer to those of the American speakers.

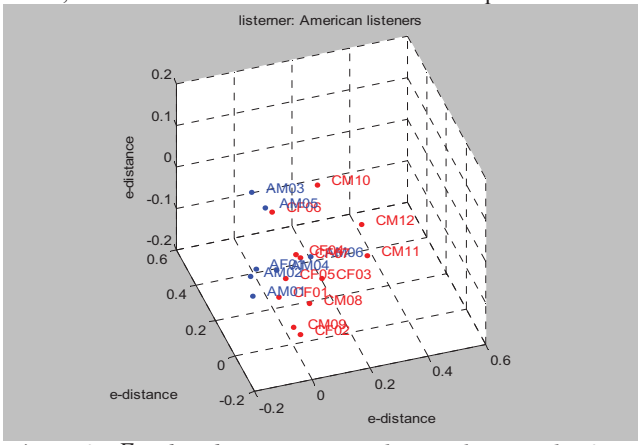

Figure 2: *E-index distance among the speakers in the 3-D space.*

Table 3. *Comparison between the speakers' oral English proficiency levels and their similarities on prominence distribution patterns. 'e-Distance' stands for the average e-index distance for a speaker, 'cluster' is the K-Means cluster based on e-index distances, 'Grade' is the rank of the speakers' oral English proficiency.*

Speakers	e-Distance	Cluster	Grade
AM03	0.618	1	-
AM04	0.615	1	-
AM01	0.613	1	-
AF01	0.613	1	-
CF04	0.610	1	H
AM02	0.610	1	-
CF07	0.604	1	H
AM05	0.599	2	-
CF01	0.595	2	M
CF06	0.594	2	L
CF02	0.593	2	L
CM08	0.591	2	M
CM09	0.591	2	L
CF03	0.590	2	L
CF05	0.590	2	M
AM06	0.565	2	-
CM10	0.546	3	M
CM12	0.536	3	L
CM11	0.524	3	L

Comparing the result of MDS with the ranking according to the speakers' oral English proficiency by American English teachers, we can see that there was a close relation (see Table 3) between the Chinese EFL learners' oral English proficiency and their prominence distribution patterns (liner regression analysis: r=0.829). For most of them, the learners with higher oral proficiency level had prominence distribution patterns more similar to the American speakers. This provided the answer to our first question: speakers with high oral English proficiency produce prominence distribution patterns more similar to the American speakers. One thing to note is that though most of the American speakers come from western America, the speaker AM06 is from California. So there is a different between AM06 and the other five American speakers.

3.3. The listeners' oral English proficiency levels and their perception of prominence distribution patterns

In the previous session, results of one-way ANOVA showed that the five American listeners (LA03 excluded) were consistent in their perception of prominence. Yet, there was significant difference among the Chinese listeners (F (8, 369) =26.590, p< 0.001). Figure 3 shows the 3D MDS analysis on the perceived results for all American speakers by American and Chinese listeners. From the figure, we obtained that the American listeners were closer than Chinese listeners, and the Chinese listeners LC06, LC08 and LC01 were closer to American listeners, which means that they had more similar prominence distribution patterns with the American listeners.

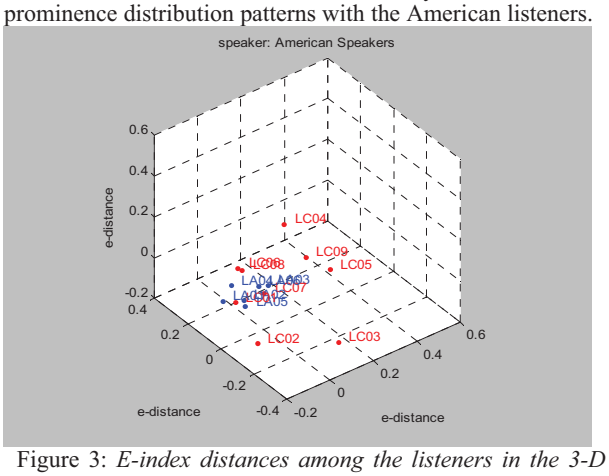

Figure 3: *E-index distances among the listeners in the 3-D space.*

Table 4. *Comparison between the listeners' oral English proficiency levels and their similarities on perception of prominence distribution patterns.*

Listeners	e-Distance	Cluster	Grade
LA04	0.650	1	-
LA05	0.636	1	-
LA02	0.635	1	-
LA01	0.622	1	-
LC06	0.621	1	H
LC08	0.606	2	H
LA06	0.606	1	-
LC02	0.606	1	H
LC01	0.573	2	M
LA03	0.563	2	-
LC07	0.545	2	H
LC09	0.535	2	M
LC04	0.505	3	M
LC05	0.495	3	L
LC03	0.464	3	L

As shown in Table 3, the American speaker AM06 produced different prominence distribution patterns from the other six speakers. One-way ANOVAs showed that there was no significant difference among the other six American speakers (F (5, 210) =0.233, P= 0.948>0.05). So we used the listeners' perception result from the six American speakers (AM06 excluded) to calculate the e-index distances among the listeners.

From Table 4, we can see that there is a close relationship between the listeners' oral English proficiency and their perception on prominence distribution patterns (liner regression analysis: r=0.854). The listeners with higher oral English proficiency have patterns more similar to the American listeners. In other words, listeners with higher oral English proficiency have a more similar perception of prominence distribution patterns to the American listeners, which answers our second question. The analysis tells us that there is a close

relationship between the perception and production of prominence distribution patterns by Chinese EFL learners.

4. Discussion and Conclusion

This research is a part of a long-term research project on Chinese learner' intonation patterns. In our previous paper [12], we performed acoustic analysis on the intonation patterns of yes-no questions on the basis of the same speech database. We found that Chinese EFL learners' nuclear pitch patterns of yes-no questions are quite different from those of American speakers. American speakers apply a low rising tone (*L*H*) on nuclear words in yes-no questions regardless of whether the nuclear accent is on the medial or final part of a sentence. When the nuclear accent is not on the final syllable of the sentences, Chinese EFL learners always apply a high-level tone (*H**) or a falling tone (*H*L*) on the nuclear accent. A low rising tone (*L*H*) is one variety that can be found only when a nuclear accent falls on the final syllable of a sentence.

Figure 4 shows the pitch contours of sentences in Figure 1 spoken by American and Chinese speakers. On the nuclear word 'MONEY', both the American speakers and Chinese EFL learners apply a falling tone (H*L) in statement, and the American speakers apply a low rising tone (*L*H*) while Chinese EFL learners adopt a falling tone (*H*L*) for question.

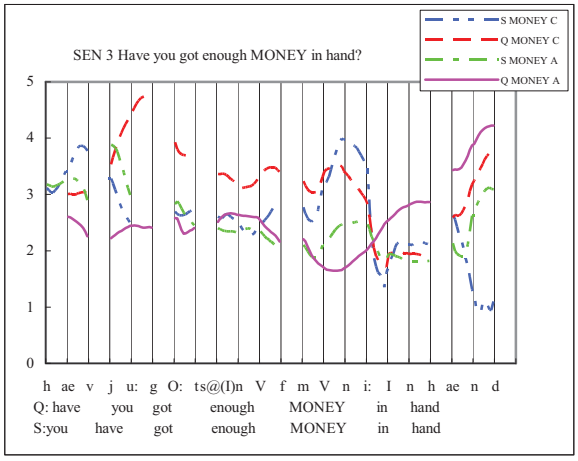

Figure 4: *The Time-normalized F0 contours of sentence 'Have you got enough MONEY in hand?' and its statement counterpart by American and Chinese speakers.*

Figure 1 illustrates the perceived prominence distribution patterns of the question and its statement counterpart produced by the American speaker AM05 (the pitch contours are in legends of 'Q MONEY A' and 'S MONEY A' in Fig.4). Figure 1 indicates that the American listeners considered the word "money" as the nuclear accent either in the question and statement, while Chinese listeners perceived the word "hand" as the nuclear accent in the question and "money" as the nuclear accent in the respective statement.

With respect to perception, we can see that the Chinese EFL learners tended to perceive words with *H* target tones as prominent ones, i.e. 'hand' in the question and 'money' in the statement, while for American nuclear tones of 'L*H', Chinese ELF learners had difficulties to make correct judgments.

With respect to production, we can see that the Chinese EFL learners applied a falling tones (H*L) on the focus word 'MONEY' both in the question and statement (the pitch patterns are shown in Figure 4 as 'Q MONEY C' and 'S MONEY C'), they always had problems to produce L*H tones, indicating a kind of close relationship between the EFL learners' perception and production of the prominence.

To summarize, this paper focuses on the relationship between the perception and production of prominence distribution patterns by Chinese EFL learners; it leaves the study of linguistic or phonological meaning of the prominence patterns for future studies. The present study demonstrated that there is a close relationship between the Chinese EFL learners' oral English proficiency and their production and perception of prominence distribution patterns. Together with our previous research, it is implied that the EFL learners lack 'a good English ear' to produce the English nuclear tones in L*H or 'L*'correctly. Such kind of negative transfer of intonation patterns needs to be compared between Chinese and American intonation patterns phonetically and phonologically with more sentence types and materials. In fact, the perceptual experiments on more utterances with pausing and boundary perception [13] have been done. We believe that together with the current research, further research will help greatly the Chinese EFL learners and CALL systems on prosodic aspects.

Acknowledgement: The project is funded by Nokia Research Beijing and 'Speech and Discourse Processing' Key Lab of Chinese Academy of Social Sciences.

5. Reference

[1] Neufield, G. G., "On the adult's ability to acquire phonology," *TESOL Quanterly,* vol. 14, pp. 285-298, 1980.

[2] Chun, D. M., *Discourse intonation in L2: From theory and research to practice*. Philadephia: John Benjamins Publishing Company, 2002.

[3] Pierrehumbert, J. B., "The phonology and phonetics of English intonation," Docter degree, MIT, 1980.

[4] Kang, S. and S. R. Speer, "Prosodic disambiguation of participle construction in English," presented at the Speech prosody 2004 proceedings, Nara, Japan, 2004.

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

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

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

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

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

[10] WANG, X., *et al.*, "Multi-accent and multi-lingual speech corpus," presented at the Proceedings of O-COCOSDA 2008, Japan, 2008.

[11] DANG, J., *et al.*, "Comparison of emotion perception among difference cultures," presented at the APSIPA, sapporo, Japan, 2009.

[12] JI, X., *et al.*, "Intonation patterns of yes-no questions for Chinese EFL learners," presented at the O-COCOSDA 2009, Beijing, 2009.

[13] JIA, Y. and A. LI, "An Introduction to IViE English Labeling System," *Report of Phonetic Research,* vol. 1, pp. 76-80, 2005.

[This paper was published in Speech Prosody 2010]