Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

Prosodic Transfer: A Comparison Study of F0 Patterns in L2 English by Chinese Speakers

Hongwei Ding[1,2], *Rüdiger Hoffmann*[2], *Daniel Hirst*[3]

[1]School of Foreign Languages, Shanghai Jiao Tong University, China
[2]Institute of Acoustics and Speech Communication, TU Dresden, Germany
[3]Laboratoire Parole et Langage, CNRS & Aix-Marseille University, Aix-en-Provence, France
hwding@sjtu.edu.cn, ruediger.hoffmann@tu-dresden.de, daniel.hirst@lpl-aix.fr

Abstract

A comparison was made among the fundamental frequency (F0) patterns of continuous speech in English, Mandarin Chinese and L2 English produced by Chinese speakers. Ten adult native Chinese speakers were asked to read narrative text written in both English and Chinese. The comparative analysis of 300 sentences was performed in the following aspects: F0 mean, pitch range, pitch change rate and pitch change amount. It is found that in terms of both pitch range on the phoneme level and pitch change amount on the utterance level, L2 English speech by Chinese subjects displayed a significantly larger value than the English speech by native speakers. Moreover, the same Chinese subjects demonstrated still a larger value in these two pitch-related variables in their Chinese speech. The dynamic characteristic of L2 English can be attributed to the negative transfer of L1 Chinese. The findings can shed some light on the understanding of the difference in F0 patterns between a tone language and a non-tone language, and can also provide some implications for L2 speech learning.

Index Terms: prosodic transfer, L2 English, L1 Mandarin Chinese

1. Introduction

It is clear that speakers of tone languages such as Chinese display systematically different fundamental frequency (F0) patterns from speakers of non-tone languages such as English [1]. Furthermore, it is well known that the learned patterns of articulatory and prosodic behaviour can be transferred from the native language (L1) to the foreign language (L2) [2, 3, 4]. This study aims to find out whether F0 patterns of L2 English produced by Chinese speakers are different from those of English, and whether the prosodic deviation is transferred from Chinese.

1.1. Negative transfer in L2 English

Mandarin Chinese and English are two cases of tone vs. non-tone languages: Chinese is a typical tone language, whereas English is a typical non-tone language. In Mandarin Chinese the tones occur on all syllables in a sentence. The pitch contour is an interaction of syllable tones and the sentence intonation, which was compared to "small ripples on large waves" by Chao [5]. While the F0 pattern of English speech is determined by the placement of primary stress on a few of the syllables in a sentence. It is suggested that the pitch patterns of an adult's L2 can be characterized by the acquired pattern of F0 in L1 [6]. With the help of ProZed, which is a tool designed by Hirst [7], the difference of F0 patterns between English, L2 English and Mandarin Chinese can be visualized in Figure 1.

(a) (b) (c)

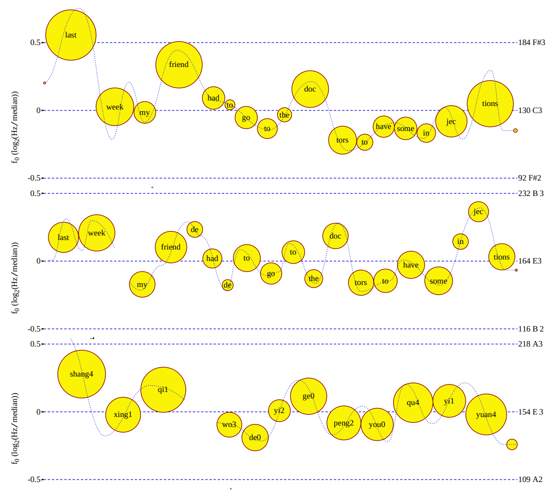

Figure 1: Prosody display of English (a), Chinese English (b) and Mandarin Chinese (c). Note: Though the same English sentence in (a) was read by the Chinese speaker in (b), two epentheses occurred, which resulted in two more syllables (*de*).

In Figure 1 the pitch contour is demonstrated in a continuous dotted line with each circle corresponding to one syllable. The vertical level of the circle represents the pitch and the diameter represents the duration of the syllable. The unit of pitch has already been normalized to the logarithmic scale $\log_2$(Hz/median). The scale is OMe (Octave Median), which was proposed by De Looze and Hirst [8]. A clear declarative intonation contour with a few stressed syllables (circles with larger diameters) in English can be observed in Figure 1 (a); while more fluctuations in the pitch contour and comparable diameters of the circles in Chinese can be found in Figure 1 (c); finally, the Chinese characteristics of a pitch contour with many small ups and downs and circles of comparable diameters are also reflected in Figure 1 (b). How to describe the differences of pitch change pattern between languages and the transfer of F0 pattern of L1 in that of L2 speech has been a fascinating task for many prosody experts.

1.2. Comparison researches in pitch changes

However, few researches have been devoted to compare the pitch patterns between tone language and non-tone languages because of the complexities of pitch calculation. In the early

10.21437/SpeechProsody.2016-155

period Eady [1] claimed that F0 patterns of Chinese have a greater amount of dynamic movements than those of English. But the comparison was based on F0 values calculated in Hz rather than normalized unit (such as semitone), and the results may not be convincing. In recent times Keating and Kuo [9] compared the speaking fundamental frequency in English and Mandarin, but no comparison between native English and L2 English was made. Hirst and Ding [10] employed 18 metrics for the comparison, and found that Chinese English was intermediate between English and Mandarin Chinese. But these metrics were only extracted from the acoustic signal without reference to phonetic information. As we have observed in Figure 1, the pitch changes based on syllables may provide more information for rhythmic differences in perception. Therefore, this study aims to compare the pitch changes based on phonetic information among English, Chinese and L2 English and to determine whether F0 patterns in the interlanguage (L2 English) are indeed systematically different from those of the target language (English), and whether they resemble the characteristics of the source language (Mandarin Chinese).

2. Method

2.1. Corpus

For this study the English recording was taken from the OMProDat database, and it is regarded as the reference of the target language (English). L2 English and Mandarin Chinese speech data were specially recorded for the study, and they are treated as the interlanguage (Chinese English) and the source language (Chinese) respectively. For each language, OMProDat contains recordings from 5 male and 5 female speakers, and each speaker read the same 40 five-sentence passages in a specific language [11].

2.1.1. Stimuli

Though we provided the Chinese recordings in the database [12], to facilitate the comparison of the source language and the interlanguage of the speakers in the study, we made extra recordings of Chinese and English produced by the same Chinese speakers. The reading passages were kept the same as those in the OMProDat. Regarding the large amount of manual work in phoneme annotation and pitch mark correction, only two passages on different themes were selected from the OMProDat for this study. Moreover, two passages which evoke less emotional responses were used. In this way, the emotional effects on F0 patterns can be minimized and attention can be focused on the language-specific factors.

2.1.2. Subjects

We recruited 10 native Chinese speakers, including 5 men and 5 women. All of them were born and raised in Shanghai, and studying at Tongji University in Shanghai at the time of data collection. Normally, students originating from Shanghai have the advantage of learning English earlier and enjoying more opportunities to practice their oral English; therefore, they are more accurate in pronouncing vowels and consonants and can speak in a more fluent way. However, the deviation of F0 patterns away from those of the native English can still be perceived in their English speech. The subjects in this study, who had few segmental problems but more suprasegmental problems, could serve as good subjects for prosodic investigation in L2 English speech learning.

2.1.3. Data

The recording was conducted in a recording room at Tongji University. Before the recording started, the subjects were asked to get familiarized with the texts and given some practice in reading them at a normal rate and with a natural intonation. During recording, the subjects were asked to repeat the whole passage when they mispronounced some words. The speech was recorded at a sampling rate of 44.1 kHz with 16-bit resolution.

The data employed in this study consist of three types:

1. The English data contain 10 native speakers reading 2 passages, which include 92 and 93 syllables respectively.
2. The L2 English data contain 10 Chinese speakers reading the same 2 English passages as those of the English.
3. The Mandarin Chinese data contain the same 10 Chinese speakers reading the corresponding 2 Chinese passages, which include 96 and 112 syllables respectively.

For each type of data, 5 females and 5 males were recorded. The English data were taken from the OMProDat database. The L2 English and Mandarin Chinese were recorded in Shanghai. Since each passage contains 5 sentences, the whole database contains 300 sentences: 3 categories x 10 speakers x 2 passages x 5 sentences = 300 sentences.

2.2. Analysis procedure

The 300 sentences were first automatically annotated by an automatic aligner SPPAS [13]. Then the annotation were corrected manually with Praat program [14] by two phonetics experts, and checked by the first author. Since the automatic extraction of F0 values from acoustic natural speech was still unsatisfactory, a manual correction of F0 was carried out. The waveform, spectrogram, pitch marking and annotation were displayed simultaneously for the correction with a Praat script ProsodyPro [15].

The pitch markings were manually corrected to ensure utmost accuracy of F0. Some speakers (including both English and Chinese, male and female) demonstrated much glottalization [16] in their speech, and pitch mark mistakes usually took place when the glottalization occurred, which required much manual adjustment. We tried our best to correct the voiced parts with little glottalization, the voiced parts with strong glottalization were specially annotated, and were not included in the calculation of F0 values. Fortunately, the glottalized parts which were not included occupied only a small part at the beginning or at the end of the voiced speech segments. In this way, some extreme F0 values were avoided. F0-related variables were then also extracted with ProsodyPro [15].

It has been well acknowledged that what matters in characterizing pitch and intonation patterns is *pitch changes*, i.e. pitch rises and pitch falls. Measurements of F0 values in hertz were converted to semitones to normalize across English and Chinese, male and female speakers according to the equation with a reference of 100 Hz. The following conversion equation proposed by Fant et al. [17] was used in this study.

$$f(\text{st}) = 12\log_2\left(\frac{f(\text{Hz})}{100(\text{Hz})}\right) \quad (1)$$

With this normalisation in frequency level, it is possible to display a qualitative derivation of the essentials of an intonation contour and facilitates inter-speaker comparisons and specifications of group average data. Except for F0 means, other F0 variables discussed in this study are presented in semitone (st).

The edited output was then analyzed to determine the averages of the following variables for each subject over the two entire passages of speech:

1. *F0 mean*: mean F0 value for voiced speech (in *Hz*);
2. *Speech rate*: rate of speech calculated on the basis of speech signal (in *syllables/second*);
3. *Pitch range on sentence level*: absolute difference between maximum F0 and minimum F0 of voiced speech within a sentence (in *semitone*);
4. *Pitch range on phoneme level*: absolute difference between maximum F0 and minimum F0 of voiced speech within a phoneme (in *semitone*). Because there can be more than one peak or valley in some Chinese syllables, pitch range on phoneme level is calculated instead of that on syllable level to capture possible fluctuations.
5. *Pitch change rate*: absolute pitch change amount of all voiced speech parts in every 10-msec interval (in *semitone every 10 milliseconds*);
6. *Pitch change amount*: sum of the absolute pitch change amount of every voiced speech part in the whole passage (in *semitone*).

The group mean is based on the averages of each subject, which are further the averages of two passages.

2.3. Results

The comparison of the above-listed variables were made among three groups of data, which represent English, L2 English, and Mandarin Chinese, respectively. Suppose that a difference may exist between males and females. All comparisons were performed within males or females across groups, and between males and females within the same group. Independent t-tests were employed to find whether there was a significant difference between any two groups or between males and females. The value of each speaker was the average of two passages. In the following tables, *EN-EN*, *CN-EN*, and *CN-CN* represent the three groups: English, L2 English (Chinese English), and Mandarin Chinese, respectively.

Apart from F0 means, no other values demonstrate significant differences between males and females within any group.

2.3.1. F0 Mean

F0 means of the English males (124Hz) and females (205Hz) were lower than those of the Chinese males and females respectively. The Chinese speakers showed similar F0 means when they speak Chinese or English. The values of male speakers are 144Hz (L2 English) and 141Hz (Chinese), while those of the female speakers are 230Hz (L2 English) and 229Hz (Chinese). The overview of F0 mean values (*Mean*) and standard deviations (*SD*) are described in Table 1.

Table 1: *Comparison of F0 means (Hz).*

	$EN-EN$		$CN-EN$		$CN-CN$	
$Gender$	$Mean$	SD	$Mean$	SD	$Mean$	SD
$Male$	124	11.7	144	19.1	141	16.3
$Female$	205	21.5	230	28.8	229	20.2

Though Chinese speakers showed a higher pitch than the English speakers, the differences was not significant. However, significant differences between males and females were found within each group.

2.3.2. Speech rate

The speech rate was calculated by dividing the number of syllables in the passage by the duration (in seconds) of this passage without pauses, which is also called articulation rate. Without consideration of the pauses, articulation rates of English, L2 English and Chinese are listed in Table 2.

Table 2: *Comparison of articulation rate (syl/sec).*

	$EN-EN$		$CN-EN$		$CN-CN$	
$Gender$	$Mean$	SD	$Mean$	SD	$Mean$	SD
$Male$	5.62	0.37	3.75	0.24	4.19	0.26
$Female$	6.04	0.34	4.06	0.42	4.54	0.48

Both English male and female speakers displayed a higher articulation rate on the syllable level than the Chinese males and females respectively. Significant differences were found between English and L2 English, English and Chinese for both males and females. No significant differences were shown between L2 English and Chinese for neither males nor females, though they showed a higher rate when they read Chinese.

2.3.3. Pitch range on sentence level

Each passage consists of 5 sentences. The value was first averaged over 5 sentences, and then over 2 passages for each speaker. The group averages are shown in Table 3.

Table 3: *Comparison of pitch range on sentence level (st).*

	$EN-EN$		$CN-EN$		$CN-CN$	
$Gender$	$Mean$	SD	$Mean$	SD	$Mean$	SD
$Male$	14.05	3.59	13.02	2.84	12.10	1.24
$Female$	12.73	1.44	10.67	1.12	11.85	1.84

A significant difference was shown between English and L2 English for female speakers ($t(8)=2.51$, $p=0.036$), but not for male speakers. No other significant differences can be found between other groups.

2.3.4. Pitch range on phoneme level

Pitch ranges on each voiced phoneme were first averaged over each passage, and then over two passages for each speaker. The group averages are illustrated in Table 4.

Table 4: *Comparison of pitch range on phoneme level (st).*

	$EN-EN$		$CN-EN$		$CN-CN$	
$Gender$	$Mean$	SD	$Mean$	SD	$Mean$	SD
$Male$	1.62	0.19	2.15	0.32	2.92	0.22
$Female$	1.62	0.13	1.92	0.15	3.13	0.23

For female speakers, there was a significant difference between any two groups: between Chinese and English with $t(8)=12.95$ $p<0.005$; between Chinese and L2 English with $t(8)=3.39$, $p=0.05$; and between L2 English and English with $t(8)=9.76$, $p<0.005$. The same applied to male speakers, and significant differences were found: between Chinese and English with t(8)=10.07, p<0.005; between Chinese and L2 English with t(8)=4.78, p=0.001; and between L2 English and English with t(8)=2.58, p<0.033.

2.3.5. Pitch change rate

The pitch change amount in every 10-msec interval of each voiced speech was first averaged over the whole passage, and then over the two passages for each speaker. The group values are presented in Table 5.

Table 5: *Comparison of pitch change rate (st/msec).*

	$EN-EN$		$CN-EN$		$CN-CN$	
$Gender$	$Mean$	SD	$Mean$	SD	$Mean$	SD
$Male$	0.258	0.034	0.235	0.054	0.294	0.033
$Female$	0.308	0.042	0.249	0.036	0.327	0.015

For female speakers, no significant difference was found between Chinese and English, but between English and L2 English with t(8)=2.42, p=0.042, and between L2 English and Chinese with t(8)=4.51, p=0.002. For male speakers, no significant difference was shown between any of these three groups.

2.3.6. Pitch change amount on utterance level

The pitch change amount on the utterance level was the average of two passages for each speaker, and the group averages are presented in Table 6.

Table 6: *Comparison of pitch change amount on utterance level (st).*

	$EN-EN$		$CN-EN$		$CN-CN$	
$Gender$	$Mean$	SD	$Mean$	SD	$Mean$	SD
$Male$	2.31	0.26	3.34	0.61	3.59	0.29
$Female$	2.48	0.20	3.61	0.32	4.27	0.27

For female speakers, significant differences were found between Chinese and L2 English with t(8)=12.0, p<0.001; between L2 English and English with t(8)=6.61, p<0.001; and between Chinese and L2 English with t(8)=3.55, p=0.007. For male speakers, significant difference were demonstrated between Chinese and English with t(8)=7.35, p<0.001; between L2 English and English with t(8)=3.49, p=0.008, but the differences between Chinese and L2 English was not significant.

3. Discussions

Among all the variables, only pitch range on the phoneme level and pitch change amount on the utterance level are consistently significantly different between English and Chinese, and between English and L2 English for both males and females.

1. *F0 means* of females are significantly higher than those of males, but not between English and Chinese speakers, whose body sizes can be different. These results are consistent with previous findings that the correlation of F0 and body size is very weak within sex [18, 19].
2. *Speech rate* of the English speakers were higher compared with that reported in the literature [20], while the Chinese subjects read at a normal rate as required. One fact which has been proved repeatedly is that L2 speakers spoke slower than the native speakers [21].
3. *Pitch range on the sentence level* is not significantly different across different groups, and similar results were found in our previous study in learning L2 German by Chinese speakers [21]. The larger value of the English female speakers might due to their relatively emotional expressions, while the Chinese females were not able to speak in such an emotional way.
4. *Pitch range on the phoneme level* is higher for Chinese speakers. Because of the negative transfer of L1, Chinese speakers tend to attach tones to syllables even when they speak English. Rising and falling tones increase the fluctuation amount in the pitch contour. Though the pitch range on few stressed syllables in English can be larger, the average values are still smaller compared with Chinese or Chinese English, which is also consistent with the previous investigation in L2 German speech [21].
5. *Pitch change rate* is surprisingly insignificantly different between Chinese and English speakers, which is not consistent with the findings reported by Eady [1]. One reason is that pitch changes were calculated in hertz in his investigation [1], and the higher F0 means of Chinese speakers can raise the values of F0 changes in hertz. Another reason is that the high articulation rate of the English speakers in this study also increases the pitch change rate. Finally, there is also evidence showing that pitch change by speakers of a lexical tone language like Chinese is not significantly faster than that produced by speakers of languages with no lexical tone by Xu [22].
6. *Pitch change amount* is larger for Chinese speakers. This can be attributed to two reasons: 1) the English speakers reduced many unstressed vowels, so many pitch changes disappeared; 2) the English speakers spoke much faster, many pitch changes may be compressed.

4. Conclusion

This study conducted a systematic investigation on F0 patterns of English, Chinese and L2 English by Chinese speakers with the comparison of several F0-related variables, and has shown that F0 patterns of L2 English produced by Chinese speakers are systematically different from those of native English speakers, which can be transferred from their native tone language. The Chinese accent in English may not be related to F0 means, pitch change rate, or pitch range on the sentence level. The results claim that larger pitch range on the phoneme level and greater pitch change amount on the utterance level can better represent the dynamic characteristics of Chinese accent of English. We will employ various speech data to find more robust variables which can better capture the prominent features of the differences in F0 patterns between English and L2 English by Chinese speakers in the future.

5. Acknowledgements

The first author is sponsored by the National Social Science Foundations of China (13BYY009, 10CYY009, 13&ZD189) and the Interdisciplinary Program of Shanghai Jiao Tong University (14JCZ03) for this research work. We are very thankful to Xinping Xu at Tongji University and Rainer Jäckel at TU Dresden for their support in the collection and annotation of the data.

6. References

[1] S. J. Eady, "Differences in the f0 patterns of speech: Tone language versus stress language," *Language and Speech*, vol. 25, pp. 29–42, 1982.

[2] J. E. Flege and R. D. Davidian, "Transfer and developmental process in adult foreign language and speech production," *Applied Psycholinguistics*, vol. 5, pp. 323–347, 1985.

[3] T. Visceglia, C.-Y. Tseng, Z.-Y. Su, and C.-F. Huang, "Interaction of lexical and sentence prosody in Taiwan L2 English," in *SLaTE Workshop*, 2010.

[4] A. Rosenberg and J. Hirschberg, "Production of English prominence by native Mandarin Chinese speakers," in *Workshop on Prosodic Prominence: Perceptual, Automatic Identification*, 2010.

[5] Y. R. Chao, *A Grammar of Spoken Chinese*. Berkeley: University of California Press, 1968.

[6] U. Gut, J. Trouvain, and W. J. Barry, "Bridging research on phonetic descriptions with knowledge from teaching practice – the case of prosody in non-native speech," in *Non-Native Prosody. Phonetic Description and Teaching Practice*. Mouton de Gruyter, 2007.

[7] D. J. Hirst, "Prozed: A speech prosody analysis-by-synthesis tool for linguists," in *Proceedings of the 6th International Conference on Speech Prosody 2012*, Q. Ma, H. Ding, and D. J. Hirst, Eds., vol. 1. Tongji University Press, 2012, pp. 15–18.

[8] C. De Looze and D. J. Hirst, "The OMe (octave-median) scale: a natural scale for speech prosody," in *Proceedings of the 7th International Conference on Speech Prosody (SP7)*, N. Campbell, D. Gibbon, and D. J. Hirst, Eds. Trinity College, Dublin, Ireland, May 2014.

[9] P. Keating and G. Kuo, "Comparison of speaking fundamental frequency in English and Mandarin," *J. Acoust. Soc. Am.*, vol. 132, no. 2, pp. 1050–1060, 2012.

[10] D. J. Hirst and H. Ding, "Using melody metrics to compare English speech read by native speakers and by L2 Chinese speakers from Shanghai." in *Interspeech*, 2015, pp. 1942–1946.

[11] D. J. Hirst, B. Bigi, H. Cho, H. Ding, S. Herment, and T. Wang, "Building OMProDat: an open multilingual prosodic database," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody*, 2013, pp. 11–14.

[12] H. Ding and D. J. Hirst, "A preliminary investigation of the third tone sandhi in standard Chinese with a prosodic corpus," in *Proc. ISCSLP*, 2012, pp. 436–439.

[13] B. Bigi and D. J. Hirst, "Speech phonetization alignment and syllabification (SPPAS): a tool for the automatic analysis of speech prosody," in *Proceedings of Speech Prosody 2012*, 2012.

[14] P. Boersma and D. Weenink. Praat: doing phonetics by computer [computer program]. Version 6.0.05, retrieved 01 January 2016. [Online]. Available: http://www.praat.org

[15] Y. Xu, "Prosodypro — a tool for large-scale systematic prosody analysis," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 2013, pp. 7–10.

[16] H. Ding, O. Jokisch, and R. Hoffmann, "Glottalization in inventory construction: A cross-language study," in *Proc of 4th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Hongkong, 2004, pp. 37–40.

[17] G. Fant, A. Kruckenberg, K. Gustafson, and J. Liljencrants, "A new approach to intonation analysis and synthesis of Swedish," in *Proceedings of Speech Prosody 2002*, Aixen-Provence, France, 2002, pp. 283–286.

[18] J. Gonzalez, "Formant frequencies and body size of speaker: a weak relationship in adult humans," *Journal of Phonetics*, vol. 32, p. 277–287, 2004.

[19] C. G. Henton, "Fact and fiction in the description of female and male pitch," *Language & Communcation*, vol. 9, no. 4, pp. 299–311, 1989.

[20] J. Trouvain, "Tempo variation in speech production: Implications for speech synthesis," Ph.D. dissertation, Saarland University, 2003.

[21] H. Ding, O. Jokisch, and R. Hoffmann, "A phonetic investigation of intonational foreign accent in Mandarin Chinese learners of German," in *Proc. 6th International Conference on Speech Prosody*, Q. Ma, H. Ding, and D. J. Hirst, Eds., vol. 1, Shanghai, China, 2012, pp. 374–377.

[22] Y. Xu and X. Sun, "Maximum speed of pitch change and how it may relate to speech," *Journal of the Acoustical Society of America*, vol. 111, pp. 1399–1413, 2002.