AN INVESTIGATION OF ENGLISH PROSODY PRODUCED 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: This study is concerned with the prosodic properties of English speech produced by Chinese students. The database used in this study is a reasonably large-scale, multilingual speech corpus designed for prosody research. The speech data contains three recordings: 1) 40 English passages read by 10 native English speakers; 2) the same English passages read by 10 Chinese native speakers; 3) 40 Chinese passages read by the same Chinese speakers. Momel algorithm was employed to calculate the melody metrics on this database, and 14 prosodic parameters were compared among the three types of speech. The discrimination of Chinese English from native English and native Chinese regardless of gender reached approximately 76.5%. The results demonstrated that the pitch movements in Mandarin Chinese were greater and faster than in English, and the English produced by Chinese speakers showed greater and faster pitch movements than native English speech, but smaller and slower pitch movements than Mandarin Chinese speech. However, there was a slight difference between Chinese male and female speakers, but larger differences between English male and female speakers. Chinese female speakers even displayed smaller magnitudes of pitch movements on the basis of Momel targets than English female speakers in their English speech. Further investigations should be carried out to find out the reasons of such pitch movement patterns.

1 Introduction

It is well known that Pike [9] and Abercrombie [1] proposed that the languages of the world can be classified into two types of rhythm patterns: a) stress-timed rhythm, and b) syllable-timed rhythm. Ramus [10] showed that stress-timed languages have a higher standard deviation of consonantal intervals (ΔC) and a relatively lower proportion of the vocalic intervals ($\%V$); while syllable-timed languages have a lower ΔC and a higher $\%V$. A similar evaluation approach was adopted by Grabe and Low [5]. They based their pairwise comparison of successive vocalic and intervocalic intervals and calculated speech rhythm with the *Pairwise Variability Index (PVI)*, which computes the sum of the durational differences between adjacent vocalic or consonantal intervals in an utterance. They found that stress-timed languages have a higher variation in vowel durations, whereas syllable-timed languages (including Mandarin) show a lower variation in vowel length. Following the studies of Ramus et al. [10] and Grabe and Low [5], we measured $\%V$, ΔC, normalised variation of the pairs of two adjacent vowel intervals (*nPVI-V*), and raw variation of the pairs of two adjacent consonant intervals (*rPVI-C*) of German speech produced by Chinese speakers, and described that German L2 was influenced by L1 of Chinese in terms of ΔC and $\%V$, *nPVI-V* and *nPVI-V* in our previous investigations [4, 3].

However, the analysis of rhythm in speech and language should be extended to more variables than just the duration of segments or syllables, as suggested by Kohler [8]. The recurring phonetic patterns in the production of speech forms the characteristic of rhythm. The phonetic variables include not only patterns of syllabic timing, but also patterns of fundamental frequency and energy. Pitch movement with the change of fundamental frequency plays an important role in the perception of different types of speech melody. This study aims to investigate the prosodic differences in terms of pitch movements.

In order to find a melodic typology based on the recurrent pitch patterns commonly found in languages, Hirst proposed the Momel algorithm, and measured mean, standard deviation and coefficient of variation for pitch intervals and slopes of rises and falls in continuous passages in French, English and Mandarin Chinese. These metrics were calculated from consecutive target points derived automatically from the recordings using the Momel algorithm. The identification of language regardless of gender attained 76% correct discrimination from these metrics, which were based solely on the distribution of the pitch target-points, without any consideration of the relationship of the target points to phonological, lexical or syntactic constituents [6].

Adopting the same technique proposed by Hirst [6], we aim to discover the prosodic deviance in the English speech produced by native Chinese speakers, and endeavour to illustrate whether it is the result of negative transfer from the native language.

10 Chinese students, including 5 females and 5 males, were recruited as subjects. 40 five-sentence continuous passages taken from Eurom1 corpus were selected as reading material. Theses 10 subjects were asked to read both English and Chinese texts. The prosodic features of the English produced by the Chinese speakers will be compared with those of the Chinese produced by themselves and the English taken from the OMProDat database [7]. Momel is employed to model the pitch contours as target points, and from the scaled target points the mean and standard deviation of each passage was calculated for 1) value of target points; 2) absolute difference from previous target; 3) absolute difference from previous target divided by distance in seconds [6]. With the results of the analysis we can demonstrate whether the prosodic characteristics of English produced by Chinese speakers deviate from native speakers, whether a gender difference exists, and how about the dynamic variations of pitch movements.

2 Method

For this study the English recording taken from the OMProDat database was used as the reference of target language. OMProDat contains recordings of 40 five-sentence passages, each was read by 5 male and 5 female speakers of each language, currently the database includes recordings for English, French, Chinese and Korean [7]. Though we provided the Chinese recordings in the database [2], to facilitate the comparison of source language and interlanguage of the speakers in this study, we made extra recordings of Chinese and English produced by the same Chinese speakers. The reading passages were kept the same as in the OMProDat.

This study employed the same method which was described by Hirst [6] to investigate metrical features. The target points were scaled using the OMe (Octave-Median) scale with the formula to reduce the inter-subject variability:

$$\text{ome} = \log_2(\text{Hz/Median}) \tag{1}$$

where *median* is the median value of f0 for the whole five-sentence passage.

From the scaled target points the mean and standard deviation for each passage was calculated for:

1. octave value of target points on the OMe scale

2. interval absolute difference from previous target
3. rise difference from previous target
4. fall difference from previous target
5. slope absolute difference from previous target divided by distance in seconds
6. rise slope from previous target
7. fall slope from previous target

Therefore 14 values (7 means and 7 standard deviations) were calculated for each passage. Since these values have been offset to the speaker's median f0 by formula (1), they can be compared between English and Chinese subjects.

2.1 Subjects

We recruited 10 native Chinese speakers, including 5 men and 5 women. These 10 undergraduates whose majors are English at Tongji University in Shanghai come from Shanghai. Normally students originating from Shanghai have the advantage of learning English earlier and enjoying more opportunities to practice their English, they have better performance in the pronunciation of English. They have less difficulty in the accuracy in pronouncing vowels and consonants, and they can speak in a more fluent way. However, deviation can still be perceived in their English speeches from the native ones. It is interesting to find out whether the above mentioned parameters can distinguish the interlanguage from the target language and source language.

2.2 Data

The recordings consist of three parts.

1. The first part represents the target language taken from OMProDat database, which contains recordings of 10 native English speakers (5 women and 5 men) reading 40 passages.
2. The second part represents the interlanguage, which contains recordings of 10 native Chinese speakers (5 women and 5 men) reading the same 40 passages.
3. The third part represents the source language, which contains recordings of the same 10 native Chinese speakers reading 40 passages. These passages are the Chinese version of the English text [2].

The first part was taken from the open database. The second and third parts were recorded in the studio at Tongji University in Shanghai. The speakers had no difficulty in reading either the English or Chinese passages. Each part contains 400 passages and more than 3 hours speech, and all of them were recorded with 16 bit and 44.1k Hz.

3 Results

The mean and standard deviation of each of the seven parameters described in the beginning of this section were calculated for each 5-sentence passage, therefore 14 parameters were collected for each passage. Every recording contains 10 subjects and every subject read 40 passages, which resulted in 400 sets of values for each part of speech. We have altogether gathered 16,800 values (3 types x 10 subjects x 40 passages x 14 parameters). Few outliers were found in the

preliminary statistics, a further investigation revealed that in some cases two successive target points were extremely close, and in other cases pitch values were doubled, both led to very large values of slope. In order to remove some extreme values, any passage with an absolute value of rising or falling slope greater than 5 octaves per second was eliminated from the analysis. Thus the following results were based on 400 passages for native English, 395 passages for Chinese English, and 396 passages for Mandarin Chinese.

3.1 Linear discriminant analysis

First, it is essential to test whether these parameters can distinguish English from Mandarin Chinese produced by native speakers.

Table 1 - Confusion matrix for discriminant analysis on English and Mandarin Chinese

	Predicted	
	English	Mandarin Chinese
English	**380(95%)**	20(5%)
Chinese	24(6.1%)	**372(93.9%)**

It can be observed in Table 1 that the discrimination of English from Chinese with these 14 parameters is approximately 94.5% (=(380+372)/796).

Table 2 - Confusion matrix for discriminant analysis on English produced by native speakers (native English) and English produced by Chinese speakers (Chinese English)

	Predicted	
	native English	Chinese English
native English	**294(73.5%)**	106(26.5%)
Chinese English	86(21.8%)	**309(78.2%)**

It can be observed in Table 2 that the discrimination of English produced by native speakers (native English) from English produced by Chinese speakers (Chinese English) with these 14 parameters is over 75.8% (=(294+309)/795), which is much higher than the chance rate of 50%. Now Mandarin Chinese produced by native speakers is included, the discrimination rate can be observed in Table 3.

Table 3 - Confusion matrix for discriminant analysis on English produced by native speakers, English produced by Chinese speakers and Chinese produced by native speakers

	Predicted		
	native English	Chinese English	native Chinese
native English	**270(67.5%)**	109(27.3%)	21(5.3%)
Chinese English	85(21.5%)	**293(74.2%)**	17(4.3%)
native Chinese	17(4.3%)	31(7.8%)	**348(87.9%)**

The discrimination of English produced by Chinese speakers (Chinese English) from English (native English) and Chinese (native Chinese) produced by native speakers with these 14 parameters reaches approximately 76.5% (=(270+293+348)/1191), which is also much higher than the chance rate of 50%.

3.2 Significance levels for parameters

The significance level of analysis of variance for each of the 14 parameters analysed for language and gender as well as for the interaction language*gender is shown in Table 4. It is clear all these parameters show significant or moderate correlations with the language types. Most of these parameters display no correlation with gender.

Table 4 - Significance levels of ANOVA for each parameter. [-] : n.s., [*] = $p < 0.05$, [**] = $p < 0.01$, [***] $p < 0.001$

	Mean			Standard deviation		
	Language	Gender	Language*Gender	Language	Gender	Language*Gender
Octave	***	-	***	***	*	***
Interval	***	-	*	***	-	***
Rise	***	-	***	***	**	***
Fall	***	-	***	***	***	***
Slope	***	-	-	***	-	**
Rise-slope	***	-	***	***	-	**
Fall-slope	***	**	***	**	-	-

3.3 Rising and falling means

The mean values of falling and rising interval and slope for Chinese English are greater than those for native English, but smaller than those for Mandarin Chinese. However, the values are different for males and females.

The mean values of rising interval can be ranked in the following ascending sequence: English male (EN-m: 0.241) < Chinese English female (CNEN-f: 0.276) < Chinese English male (CNEN-m: 0.296) < English female (EN-f: 0.299) < Chinese female (CN-f: 0.361) < Chinese male (CN-m: 0.391), which can be approximately observed in Figure 1.

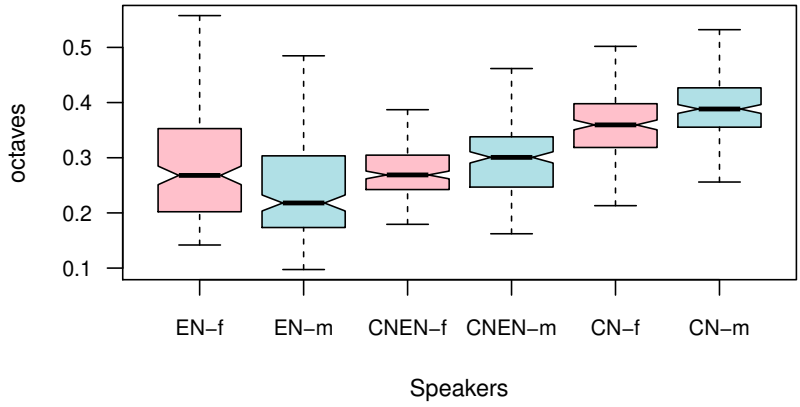

Figure 1 - Boxplot of rising intervals using the Octave-Median scale

The mean values of falling interval keep the same order as the rising interval: English male (EN-m: 0.233) < Chinese English female (CNEN-f: 0.279) < Chinese English male (CNEN-m: 0.296) < English female (EN-f: 0.299) < Chinese female (CN-f: 0.387) < Chinese male (CN-m: 0.401), which can be generally observed in Figure 2.

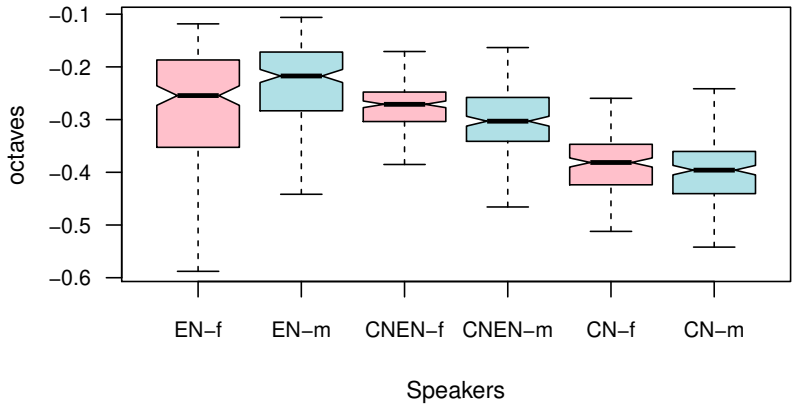

Figure 2 - Boxplot of falling intervals using the Octave-Median scale

4 Discussion

Comparing the results in the investigation with those in the previous investigations, we try to give some explanations:

- Hirst reported that discrimination of English and French from Chinese was 93% [6]. In this investigation, new Chinese database was employed, and the discrimination rate of English from Chinese reached 94.5%. It is clear that Chinese can be separated successfully from English with these 14 parameters.

- These 14 parameters can also discriminate Chinese English from English with an accuracy rate of 75.8%, which is much higher above the chance rate of 50%. However, Chinese English is more like English than Chinese, 21.5% of the Chinese English was confused as English, but only 4.3% as Chinese.

- In this investigation most parameters have significant correlations with the language types, which include target language (English), interlanguage (Chinese English), and source language (Chinese). Few parameters have correlation with gender. Similar results were found by Hirst in [6].

- Hirst found that a gender difference observed in both English and French was not observed in Chinese [6]. This evidence is supported by the new database in this investigation. We found that for English there is a significant gender difference, but for Chinese English or Chinese, the difference is not significant. Moreover, female speakers showed greater amplitudes and larger slopes of pitch movements for English, while male speakers demonstrated slightly larger and quicker pitch movements in Chinese and Chinese English, but the differences are not significant. It should also be investigated why there are no significant differences between Chinese male and female speakers.

- However, most values of the Chinese English produced by male speakers revealed that Chinese English displayed greater and faster pitch movements than English and much smaller and slower pitch movements than Chinese, which can be explained by the negative transfer of native language. However, it is not the case for Chinese female speakers. The reason may be that ample pitch movements by Chinese speakers usually take place on syllable level, for targets larger than syllable level (e.g. Momel targets), the variation of pitch movements by Chinese speakers may be less than native speakers of intonation language, such as English. However, further investigation should be carried out to explain this phenomenon.

5 Conclusion

In this investigation it is proved that the parameters calculated on the basis of Momel targets can distinguish Chinese English from English produced by Chinese males, but not so well for Chinese females. Further investigation should be conducted to find more relevant differences.

6 Acknowledgements

The first author is sponsored by the National Social Science Foundation of China (13BYY009) and the Interdisciplinary Program of Shanghai Jiao Tong University (14JCZ03) for this research work. We are very thankful to Xinping Xu from Tongji University for her support in the collection of the data.

References

[1] ABERCROMBIE, D.: *Elements of general phonetics*. Aldine: Chicago, 1967.

[2] DING, H. and D. HIRST: *A Preliminary Investigation of the Third Tone Sandhi in Standard Chinese with a Prosodic Corpus*. In *Proc. ISCSLP*, pp. 436–439, 2012.

[3] DING, H. and R. HOFFMANN: *A Durational Study of German Speech Rhythm by Chinese Learners*. In *Speech Prosody 2014*, pp. 295–299, 2014.

[4] DING, H., R. JÄCKEL and R. HOFFMANN: *A Preliminary Investigation of German Rhythm by Chinese Learners*. In WAGNER, P. (ed.): *Tagungsband der 24. Konferenz Elektronische Sprachsignalverarbeitung*, vol. 65, pp. 79–85. TUDpress, 2013.

[5] GRABE, E. and E. L. LOW: *Durational variability in speech and the rhythm class hypothesis*. In GUSSENHOVEN, C. and N. WARNER (eds.): *Laboratory Phonology 7*, pp. 515–546. Berlin: Mouton, 2002.

[6] HIRST, D.: *Melody Metrics for Prosodic Typology: Comparing English, French and Chinese*. In *Interspeech 2013*, pp. 572–576, 2013.

[7] HIRST, D., B. BIGI, H. CHO, H. DING, S. HERMENT and T. WANG: *Building OMPro-Dat: an open multilingual prosodic database*. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody*, pp. 11–14, 2013.

[8] KOHLER, K. J.: *Rhythm in Speech and Language: A New Research Paradigm*. Phonetica, 66:29–45, 2009.

[9] PIKE, K. L.: *The intonation of American English*. University Press: Michigan, 1945.

[10] RAMUS, F., M. NESPOR and J. MEHLER: *Correlates of linguistic rhythm in the speech signal*. Cognition, 73:265–292, 1999.