Absolute and Relative Entrainment in Mandarin Conversations

Qiuwu Ma[1], Zhihua Xia[12], Ting Wang[1]

[1] Tongji University, Shanghai, China
[2]Jiangsu Normal University, Jiangsu, China
tjnkmqw@126.com; xzhlf@163.com; sweetywangting@gmail.com

ABSTRACT

Based on Tongji Games Corpus, this study analyzes acoustic-prosodic entrainment in Mandarin conversations. Analyses have been accomplished at the levels of conversation, turn, and tone unit. Absolute entrainment in prosody is found at the levels of conversation and turn, and relative entrainment is found over tones. Therefore, this study identifies evidence for the existence of two kinds of entrainment in Mandarin conversations.

Keywords: prosodic entrainment, tones, Mandarin conversations

1. INTRODUCTION

Having conversation is a joint action in which interacting individuals coordinate their behaviour and adapt their linguistic choices to each other. Often this entrainment or accommodation produces convergence in conception, syntactic forms, lexicon choices, prosody, pronunciations, postures and other behaviour of interlocutors. This research focuses on prosody in entrainment.

The languages involved in studies of prosodic entrainment mainly are English, Dutch, Swedish, Japanese, Arabic, etc. (Natale 1975; Gregory & Hoyt 1982; Levelt & Kelter 1982; Gregory *et al*. 1993; Edlund *et al*. 2009; Levitan & Hirschberg 2011; De Looze *et al*. 2011; Levitan *et al.* 2012; Levitan 2013; De Looze *et al.* 2014).

Different from the languages mentioned, Chinese is a tone language. Tones are the use of pitch in language to distinguish lexical or grammatical meaning. Besides, the features of F0 in Chinese play their roles in forming intonation and expressing pragmatic meaning in communication. It is quite interesting to ask: How is pitch used to identify lexical meaning and at same time to show pragmatic meanings in interaction? How does Chinese prosody work in entrainment?

In English conversations, evidence of an overall coordination between interlocutors is found at conversation level, and evidence of turn-by-turn coordination, which shows speakers' closely match over turns, is also found at the turn level in Columbia Games Corpus (Levitan & Hirschberg 2011; Levitan *et al*. 2012).

In this paper, prosodic entrainment in Mandarin conversations has been studied at the both levels and over tone units. Our goal is to identify how Mandarin speakers exhibit prosodic entrainment in conversation.

In Section 2, we describe the corpus and annotation of this study. In Section 3, we make analyses of entrainment at conversation level. In Section 4, we make analyses at the turn level. In Section 5, we make analyses over tones. In Section 6, we discuss our results and describe future research.

2. CORPUS AND ANNOTATION

The analyses of this research are based on Tongji Games Corpus, which contains approximately 12 hours of spontaneous, task-oriented Mandarin conversations. 115 conversations are elicited by two games (Picture Ordering game and Picture Classifying game). Average duration of a conversation is 6 minutes.

In order to make the analyses over tone units valid, target tone units are designed in carrier sentences, "下一个图标是 xx" (the next picture is xx), in which "xx" carries the target tone unit. All the tone units in this study are carried by bi-syllabic words, and both syllables carry tone 1. The reason why bi-syllabic words are used is that according to the formation of the Chinese words, about 80% are bi-syllabic (Liu & Liang 1990). The reason why both syllables in each word carry tone 1 is to simplify the analyses of tones combinations in the present research. The matching within tone units with different tones combinations or tone sandhis could be studied in the future.

IPUs are adopted as the minimal units in analyses. The threshold for IPUs of this study is 80ms. IPUs are automatically labelled by SPPAS (Bigi & Hirst 2012), and their boundaries are checked manually in Praat. Seven parameters from 3 main aspects are set in this study including the feature of duration (Speaking-rate), the features of F0 (F0 min, F0 mean, F0 max), and the features of intensity (Intensity min, Intensity mean and Intensity max).

Data extraction is accomplished over the smallest analysis units---IPUs. If some analyses cover a unit

which contain more than one IPU, weighed averages of all the IPUs within these units are used in analyses. Methods of Caspers (2003) are used for identification of turns in Mandarin conversations.

3. ENTRAINMENT AT CONVERSATION LEVEL

The analyses of entrainment proximity are accomplished at conversation level. The aim of these analyses is to test whether there is overall similarity of prosodic features from two interlocutors over a whole conversation.

Paired T-tests are accomplished over two sets of distances (Levitan & Hirschberg 2011; Levitan *et al*. 2012): *partner* distances and *non-partner* distances. The *partner* distance is the distance of a prosodic feature between the speaker and his partner; *non-partner* distance is the mean of the distances of a prosodic feature between the speaker and other speakers, with whom he is not partnered in any conversations. The non-partner is restricted to the speaker with the same gender and conversational role as the partner in a dialogue. Thus, hypothesis of these analyses is that the *partner* distance should be smaller than the *non-partner* distance, which can supply the evidence for entrainment at conversation level.

For each conversation, this study defines *disp* in Formula 1 as the *partner* distance between two partners (speaker A, speaker B) on the prosodic feature f:

$$disp = \left| A_f - B_f \right| \quad (1)$$

$$disnp = \frac{\sum_i \left| A_f - Xi_f \right|}{\left| X \right|} \quad (2)$$

In Formula 2, *disnp* represents the *non-partner* distance, X(i) are the set of speakers, which are selected randomly in the Tongji Games Corpus. These speakers have the same gender and role as the speaker's partner, and are not paired with the speaker in any conversations. The results are shown in Table 1.

According to Table 1, for all the 99 speakers, the distances of paired speakers are significantly smaller than those of non-paired speakers in terms of 4 prosodic features: Speaking-rate (p= 0.0 < 0.05), F0 max (p= 0.001 < 0.05), Intensity mean (p= 0.0 < 0.05), and Intensity max (p= 0.0 < 0.05).

Table 1: Paired T-test between partner and non-partner distances

Feature	t	df	p-value	Sig.
Speaking-rate	-7.994	98	0.0	*
F0 min	0.454	98	0.650	/
F0 mean	0.665	98	0.507	/
F0 max	-3.442	98	0.001	*
Intensity min	-1.156	98	0.251	/
Intensity mean	-5.054	98	0.0	*
Intensity max	-5.128	98	0.0	*

The results above indicate that speakers exhibit proximity at conversation level over the prosodic features of Speaking-rate, F0 max, Intensity mean, and Intensity max. It is found that in Mandarin conversations, proximity in prosodic entrainment at conversation level is realized in the 3 main aspects of prosody: duration, F0 and intensity.

4. ENTRAINMENT AT TURN LEVEL

The analyses of entrainment proximity are also accomplished at the turn level. The aim of these analyses is to test whether there is overall similarity of prosodic features from two interlocutors over turns.

Paired T-tests are conducted between the distance of *adjacent* IPUs at turn exchanges and the average distance of ten *non-adjacent* IPUs (Levitan & Hirschberg's 2011). The *adjacent* distance is the distance of a prosodic feature between the final IPU in turn *i-1*, uttered by speaker A, and the initial IPU in turn *i*, uttered by B (A's partner). The *non-adjacent* distance is the distance of a prosodic feature between the turn *(i-1)*'s final IPU spoken by A and the other ten randomly chosen IPUs uttered by B, which is not adjacent to the final IPU of turn *i-1*. If for a feature, adjacent IPUs are more similar than non-adjacent IPUs, it indicates that speakers entrain at turn level over the feature.

This method is explained by following formulas.

$$disa = \left| \mathrm{IPU_t} - \mathrm{IPU_p} \right| \quad (3)$$

$$disna = \frac{\sum_{i=1}^{10} \left| \mathrm{IPU_t} - \mathrm{IPU_i} \right|}{10} \quad (4)$$

The results is listed in Table 2.

Table 2: Paired t-tests between adjacent and non-adjacent distances of IPUs

disa vs. *disna*	t	df	p	Sig.
Speaking-rate	-3.613	69	0.001	*
F0 min	-0.17	69	0.986	/
F0 mean	1.280	69	0.205	/
F0 max	0.813	69	0.419	/
Intensity min	0.747	69	0.458	/
Intensity mean	-3.716	69	0.0	*
Intensity max	-4.163	69	0.0	*

Table 2 shows that the distance of the adjacent IPUs is significantly smaller than the distance of the non-adjacent IPUs over 3 variables: Speaking-rate ($p=0.001 < 0.05$), Intensity mean ($p=0.0 < 0.05$), and Intensity max ($p=0.0 < 0.05$), and there is not the significant difference between the two kinds of distances over other 4 variables.

The results show that these 3 prosodic features (Speaking-rate, Intensity mean, Intensity max) exhibit significant prosodic proximity at turn level. Comparing the results of prosodic proximity at conversation level in Section 3, we can find that these 3 features show prosodic proximity at both conversation level and turn level. F0 max only shows prosodic proximity at conversation level, not at turn level.

5. ENTRAINMENT OVER TONES

This section aims to find out whether there is entrainment over tones. In female and male pairs' production of tones, because of difference in pitch range, absolute convergence of tonal pitch is not perceived to be collaborative. It is hypothesized that there is convergence in relative pitch registers between mixed gender pair. Pearson's correlation analyses are adopted to find out their relationship.
The analysis is accomplished in three steps.

The first step is to find the corresponding tone units produced by female and male subjects in the corpus of the present research. In order to reduce the computation burden, 10 conversations are randomly chosen from all the 30 conversations produced by mixed gender pairs in the corpus. As mentioned above, the two-tone1 combinations are carried by 18 bi-syllabic words in the carrier sentences in each conversation. There should be 180 (18 x 10=180) two-tone1 combinations in the chosen conversations for the female or male subjects respectively. Because some of the carrier sentences are missed in the spontaneous conversations, 161 corresponding two-tone1 units are found out in the 10 chosen conversations respectively for female or male subjects.

The second step is to calculate the values representing the relative pitch registers of all the two-tone1 units. Because all the tones in this analysis are the two-tone1 combinations, the mean of pitch values over the whole tone units are used to represent the positions where the pitch registers occupy within the pitch range of the subject. Thus, the absolute mean of pitch value is normalized by the pitch range of the speaker.

The method of Five Point Scale (FPS Chao 1930) is adopted in the normalization. In this method, pitch mean is converted to the five-point-scale value by the formula: $T = [(lgx-lgmin) / (lgmax-lgmin)] \times 5$ (Shi 1986), in which x is the mean of pitch values of a two-tone1 unit, min is the minimal pitch value within the pitch range of the speaker, and max is the maximal pitch value.

Thus, the comparisons of the female and male pith registers are accomplished in the comparisons of their corresponding five-point-scale values.
The third step is to do Pearson's correlation analyses. The results of Pearson's correlation analyses are listed in Table 3.

Table 3: Pearson's correlation analyses between female and male pitch registers

Correlations

		dmale	dfemale
dmale	Pearson Correlation	1	.501**
	Sig. (2-tailed)		.000
	N	161	161
dfemale	Pearson Correlation	.501**	1
	Sig. (2-tailed)	.000	
	N	161	161

**. Correlation is significant at the 0.01 level (2-tailed).

In Table 3, dmale refers to five-point-scale values of the pitch mean over the tone unit of a male speaker,

and dfemale to the similar value of the female partner. The table 20 shows that there is significantly positive correlation between dmale and dfemale (r= 0.501, p= 0.000).

These results indicate that there is significantly positive correlation between the relative pitch registers of tone units produced by female and male pairs, although they are in different pitch range in speaking. Therefore, the hypothesis that there is relative entrainment between the relative pitch registers of tones in the conversations of mixed gender pairs is proved.

In collaborative conversations, for the mixed gender pairs, in the production of tones, convergence of their absolute pitch sounds weird, because two speakers with gender differences stay within different pitch ranges in speaking.

How do mixed gender pairs exhibit collaboration in the production of tones? The analysis in this section shows that there is relative entrainment between relative pitch registers of tones within mixed gender pairs' pitch ranges.

6. DISCUSSION AND FUTHER RESEARCH

The major finding of this research is that two kinds of entrainment are found in Mandarin conversations, in which relative entrainment is found over tones and absolute entrainment is found at global units (conversation and turn).

Convergence in the relative pitch registers of tones is found between mixed gender pairs in Mandarin conversations.

Different from relative entrainment in tones, entrainment in previous research on prosody in English conversations (Levitan & Hirschberg 2011; Levitan *et al*. 2012) is considered to be absolute, in which the distance of the pairs in conversation is used to measure entrainment. If the distance becomes smaller, entrainment is proved. Thus, reduced distance of one pair in conversation means the absolute approach of two speakers, and it indicates absolute entrainment.

This kind of absolute entrainment is also found in Mandarin conversations at levels of conversation and turn.

Two main directions are suggested in future research. The entrainment should be analyzed further over four tones' combinations in Mandarin. Relative entrainment should be investigated in other languages.

Acknowledgements

This work was supported by 2009BYY005 (汉语语调音系学研究). Great thanks should be given to Professor Julia Hirschberg, Professor Daniel Hirst, and Doctor Rivka Levitan for their instruction and help.

7. REFERENCES

[1] Bigi, B. & D. Hirst. 2012. Speech phonetization alignment and syllabification (sppas): a tool for the automatic analysis of speech prosody. In *Proceedings of Speech Prosody 2012* (pp.19–22). Shanghai: Tongji University.

[2] Caspers, J. 2003. Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics*, 31: 251-276.

[3] Chao, Y. R. 1930. A system of tone letters. *Le Maĭter Phonétuque*, 45: 24-27.

[4] De Looze, C., C. Oertel, S. Rauzy & N. Campbell. 2011. Measuring dynamics of mimicry by means of prosodic cues in conversational speech. *Proceedings of ICPhS* (pp.1294-1297). Springer.

[5] De Looze, C., S. Scherer, B. Vaughan & N. Campbell. 2014. Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, 58:11-34.

[6] Edlund, J., M. Heldner & J. Hirschberg. 2009. Pause and gap length in face-to-face interaction. In *Proceedings of 10th Annual Conference of the International Speech Communication Association* (pp.2779-2782).

[7] Gregory, S. W. & B. R. Hoyt. 1982. Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psychological Research*, 1: 35-46.

[8] Gregory, S. W., S. Webster & G. Huang. 1993. Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication*, 13:195-217.

[9] Levelt, W. J. M. & S. Kelter. 1982. Surface form and memory in question answering. *Cognitive Psychology*, 14: 78–106.

[10] Levitan R. 2013. Entrainment in spoken dialogue systems: adopting, perdicting, and influencing user behavior. In *Proceedings of the NAACL HLT 2013 Student Research Workshop* (pp. 84–90), Atlanta, Georgia.

[11] Levitan R. & J. Hirschberg. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of Interspeech* (pp.3081–3084), Florence, Italy.

[12] Levitan R., A. Gravano, L. Willson, S. Benus, J. Hirschberg & N. Nenkova 2012. Acoustic-prosodic entrainment and social behavior. In *Proceedings of Conference of the North American*, Chapter of the Association for Computational Linguistics: Human Language Technologies (pp.11-19). Montréal, Canada.

[13] Natale, M. 1975. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32: 790–804.

[14] 刘源，梁南元（Liu Y. & N. Y. Liang），1990，现代汉语统计词典，宇航出版社，北京。

[15]石锋（Shi F.），1986，天津方言双字组声调分析，《语言研究》（1）：66-83。