World Journal of Social Science Research
ISSN 2375-9747 (Print) ISSN 2332-5534 (Online)
Vol. 11, No. 1, 2024
www.scholink.org/ojs/index.php/wjssr

Original Paper

A Study of Prosodic Entrainment and Social Factors in Mandarin Conversations

Zhihua Xia[1*]

[1] The School of Foreign Studies, Jiangsu Normal University, Xuzhou, Jiangsu Province, China

[*] Zhihua Xia, The School of Foreign Studies, Jiangsu Normal University, Xuzhou, Jiangsu Province, China

Received: December 1, 2023 Accepted: December 14, 2023 Online Published: December 29, 2023

doi:10.22158/wjssr.v11n1p42 URL: http://dx.doi.org/10.22158/wjssr.v11n1p42

Abstract

In conversations, interlocutors usually adopt prosody to that of their partner, and they become similar in prosodic production for successful communication. This phenomenon of prosodic entrainment is related to complex factors. This study aims to explore the relationship between prosodic entrainment and social factors. Two analyses are accomplished: the analysis of prosodic entrainment and gender, and the analysis of prosodic entrainment and role. In terms of prosodic entrainment and gender, it is found that the most prosodic features are entrained in female-male conversations, and the least in male-male conversations. In terms of prosodic entrainment and roles, it is found that different roles have influence on the entrainment degree, and information givers entrain more to followers in conversation.

Keywords

prosodic entrainment, Mandarin conversation, social factors, gender, role

1. Introduction

The importance of prosody in interaction and the ubiquity of entrainment in conversation call for intensive and extensive research on prosodic entrainment.

Prosodic entrainment means speakers adapt prosody to that of their interlocutors and become similar in conversation for better, smooth and successful communication (Levitan & Hirschberg, 2011; Levitan et al., 2012; Xia *et al.*, 2004, 2023; Xia, 2013; Xia & Ma, 2016a, 2016b, 2019).

Prosodic entrainment is important in social interaction. It assists the smooth expression and comprehension, and it reveals the alignment of cognitive, expressive, and comprehensive layers in interaction, by which communication is fulfilled accurately and effectively (Boylan, 2004; Parrill &

Kimbara, 2006; Pickering & Garrod, 2004, 2006). In addition, the accommodation in prosody improves interaction by establishing rapport (harmonious relation and mutual attention) and affiliation (Sheperd *et al*., 2001; Lakin & Chartrand, 2003; Pickering & Garrod, 2006; Tickle-Degnen & Rosenthal, 2007; Miles *et al.*, 2009; Kopp, 2010; Chi-Chun *et al*., 2010).

Prosodic accommodation varies according to the speakers' social status. For example, the participants having lower social status adapt more to the interlocutors having higher status (Giles *et al*., 1991). According to Speech Accommodation Theory (CAT; Giles *et al*., 1991), the benefits of adapting communication to accommodate others are to increase communication efficiency and gain social approval or shorten desired level of social distance. The interlocutors adapting to their partners have been evaluated more positively than those who do not, in terms of power, attractiveness and intelligence (Gallois & Callan, 1988; Giles *et al.*, 1991). Pardo (2006) has examined the degree to which interlocutors increased similarity in phonetic performances during conversational interaction, and find that both the role of a participant in the task and the sex of the pair of the talkers affected the degree of convergence. Levitan *et al*. (2012) have accomplished research on the relationship between prosodic entrainment and social behavior, and find that mixed gender pairs entrain more than same gender pairs, and entrainment is more important to the perception of mixed gender pairs than it is for the same gender pairs.

How do social factors affect prosodic entrainment in Mandarin conversation? Social factors cover various aspects. It is difficult to control simultaneously several social factors in one experiment. This research focuses on two of them: gender and role. Therefore, two kinds of analyses are made in this study: the analyses of entrainment and gender, and the analyses of entrainment and role.

Therefore, this study aims to explore the relationship between prosodic entrainment and social factors, specifically the gender and role in Mandarin conversations. The analyses of entrainment and gender include two parts: proximity by gender groups, and entrainment degree by gender groups. The analyses of entrainment and role include two parts: Role Influence Test and Role Direction Test.

2. The Functions of Prosody in Entrainment

Main functions of prosody in communication are involved in entrainment. Prosody expresses text/discourse meaning, organizing meaning, and interpersonal meaning in entrainment.

Entrainment involves the discourse/textual meaning of prosody. The prosody in entrainment is realized at discourse level which is larger than a single sentence. Levitan and Hirschberg (2011) test conversation-level proximity and find that speakers are significantly more similar to their partners than to speakers with whom they are not paired in any conversation, and assume that the coordination in the conversation improves over the course of the conversation. De Looze *et al.* (2014) find that the prosodic accommodation changes dynamically over the course of a conversation and across

conversations, and that the dynamics inform the naturalness of the conversation flow, the speakers' degree of involvement, and their affinity in the conversation.

Entrainment involves the organizing function in interaction. Prosodic entrainment is found to be realized at turn level, and it supplies cues for turn holding or yielding in the collaborative speaking context. Levitan and Hirschberg (2011) find that at the turn level, all the prosodic parameters examined in their research exhibit proximity and convergence, which indicates that entrainment is a dynamic process of continuous matching at turn exchanges, even in the prosodic feature that does not display conversation-level proximity. Beňuš, Gravano and Hirschberg (2011) have done research on the temporal alignment of turn-initial single words including mm, hm, um, okay, uh, or yeah, in English task-oriented dialogues, and find that the alignment in the prosodic forms is linked to the achievement of pragmatic goals by interlocutors, and that the degree to which temporal and metrical characteristics of interlocutors' speech become similar, and the directionality of this phenomenon is one aspect of turn-taking behavior that signals asymmetrical dominance relationships between the interlocutors.

The prosody in backchannels supplies good evidence for cooperation in interaction. Levitan, Granvano and Hirschberg (2011) have done research on the entrainment in speech preceding backchannels. They examined the pitch, intensity, duration, and voice quality and find that the similar prosodic patterns are used in the partners' backchannels, and that this similarity increases over the course of a dialogue, and that this similarity is associated with measures of dialogue coordination and task success. Backchannels or overlappings demonstrate coordination and adaptation of interlocutors in conversation and show that "conversation is not simply a start-stop interaction, and social interaction is a dynamic and joint activity" (De Looze, *et al*., 2014, p. 12). Prosody in backchannels supplies the evidence of continuous engagement by the interlocutors.

Prosody in entrainment is by nature related to the stance and interpersonal meaning of interaction. According to the definition of prosodic entrainment, prosodic entrainment means speakers adapt prosody to that of their interlocutors in conversation, and become similar in speaking for better smooth and successful communication (Levitan & Hirschberg, 2011; Levitan *et al*., 2012). Therefore, the prosody in entrainment shows cooperation in conversation, and it illustrates a collaborative stance and attitude.

Prosody in entrainment plays crucial roles not only in expressing text/discourse meaning, organizing meaning, and interpersonal meaning but also in psychological and cognitive coordination of conversation. The prosodic entrainment is explained in a psychological model --Interactive Alignment Model. Studies show that prosodic adaptation is a subconscious method of obtaining social approval and acceptance (Matarazzo & Wiens, 1967; Giles *et al.*, 1991; Chartrand & Bargh, 1999). Chartrand and Bargh (1999, p. 901) point out that "the chameleon effect operates in a passive, non-goal dependent manner to create greater liking and ease of interaction". The social desirability test is adopted by Natale

(1975) to study the speakers' prosodic accommodation, and it is found that those who have higher scores in the social desirability are more likely to accommodate their intensity level and timing patterns to those of their partners comparing with those with a low social desirability score.

In the research on human--computer interaction, prosodic entrainment has been proved to be an important aspect of interaction. The users' consistency to a machine increases when it adapts to their prosody (Suzuki & Katagiri, 2007). The telephone system adapting its speaking-rate with the users' is evaluated more favorably (Ward & Nakagawa, 2002). The prosodic adaptation is found to be such a crucial constituent of social interaction, usually considered as unconscious process, and accomplished automatically in the research of human computer interaction (Oviatt, 1996; Coulston *et al.*, 2002; Breazeal, 2002; Bell *et al.*, 2003). In spoken dialogue systems, various language technologies are used to facilitate and enhance human communication, but automatic systems which are able to identify and recreate social cues and behaviors is still a challenge and ongoing process. Therefore, the need of improving the naturalness in human-machine interaction motivates more work in this area. Before machines produce human-like communication, it is necessary to know comprehensively how human beings interact in real life. The present research is a study of this kind.

3. Prosodic Realization in Entrainment

Since prosody plays such crucial roles in entrainment, how is it realized?

Prosody mainly covers 4 aspects: fundamental frequency, duration, intensity, and voice quality (Crystal, 1969; Bolinger, 1989; Ladd, 1996; Cruttenden, 1997). The studies of prosody in entrainment have been accomplished mainly from these 4 aspects.

Adaptation in pitch and intonation contours is found in interlocutors' speaking (Putman & Street, 1984; Giles *et al.*, 1991; Zebrowitz *et al.*, 1992; Gregory *et al.*, 1993; Stanford & Webster, 1996; Gregory & Dangan, 1997; Shepard *et al.*, 2001; De Looze *et al.*, 2011). The global pitch level (mean F0) adaptation is found in interviews of English in the research of Collins (1998). Pitch level and pitch range (median F0, and sd F0) accommodation is found in task-based conversations (Vaughan, 2011; De Looze *et al.*, 2011; De Looze & Rauzy, 2011). Levitan and Hirschberg (2011) find the global level (conversation level) proximity by the measurement of pitch mean and pitch max, which means speakers are significantly more similar to their partners than to speakers with whom they are not paired in any conversation, and they also find local level (turn level) proximity, which means interlocutors exhibit similar pitch features at turn exchanges.

Adaptation in durational feature is found in interlocutors' speaking (Matarazzo & Wiens, 1967; Webb, 1972; Welkowitz & Kuc, 1973; Street *et al.*, 1983; Woodall & Burgoon, 1983; Giles *et al.*, 1991; Jaffe, 2001; McGarva & Warner, 2003; Kousidis *et al.*, 2008; Edlund *et al.*, 2009; De Looze *et al.*, 2011). Pause duration of interviewee is found to be directly influenced by those of the interviewer in

Matarazzo and Wiens's research (1967). The similar results are found in natural conversations of English and Swedish (Edlund *et al.*, 2009; De Looze *et al.*, 2011; De Looze & Rauzy, 2011). Rhythm accommodation in conversational dyads is found by McGarva and Warner (2003). Levitan and Hirschberg (2011) find the accommodation in Speaking-rate at conversation level and turn level.

Adaptation in features of intensity is found in interlocutors' speaking (Black, 1949; Natale, 1975; Gregory & Hoyt, 1982; Coulston *et al.*, 2002; De Looze *et al.*, 2011). Mean intensity adaptation is found in English interviews in Natale's research (1975). According to Levitan and Hirschberge (2011), the features of intensity are the most prominent in entrainment at both conversation and turn levels. Other studies also confirm the adaptation in intensity during interaction (Kousidis *et al*., 2009; Heldner *et al*., 2010; Vaughan, 2011; De Looze *et al*., 2011; De Looze & Rauzy, 2011).

Adaptation in voice quality is also found in interlocutors' speaking. Levitan and her group (Levitan & Hirschberg, 2011; Levitan *et al*., 2012) examined entrainment on voice quality: Jitter, Shimmer, Noise-to-harmonics Ratio (NHR). In these researches, jitter describes varying pitch in the voice, and is perceived as a rough sound; shimmer describes fluctuation of loudness in the voice; noise-to-harmonics ratio is associated with perceived hoarseness. The results of the research show that shimmer and NHR exhibit significantly coordination over the course of the conversation.

From what are stated above, we can find that the studies of prosody in entrainment have been accomplished mainly from the 4 aspects of prosody. However, the research on adaptation in the features of fundamental frequency, duration, and intensity is much more than that in the features of voice quality. The present research examines the prosodic features from the former 3 aspects for the sake of relatively sufficient references from the previous studies.

4. Methods

4.1 Corpus

Tongji Games Corpus with approximately 12 hours of spontaneous, task-oriented conversations between pairs of subjects is used for this study. Two games elicited these spontaneous conversations. One is Picture Ordering Game, and the other is Picturing Classifying Game. The common feature of these two games is that there is cooperation between interlocutors in conversation. These cooperative games elicited the collaborative conversations, because in the present research it is assumed that entrainment in collaborative conversations is prominent.

In Picture Ordering Games, one subject played as the information giver, and the other subject played as the information follower. The information giver's screen displayed 18 cards in the correct order; the information follower's screen displayed the same 18 cards but in the incorrect order. The information giver instructed the information follower to put the 18 cards in correct order. 18 pictures were labeled by the numbers. The Picture Ordering Game started from the picture numbered as 1, continued to the

picture numbered as 2, 3, ... and ended by the picture numbered as 18. When the task was completed, the same pair switched roles and accomplished the picture ordering game again.

In Picture Classifying Game, what is presented on the screens of computers for the two interlocutors are the same (the same 18 pictures in the same order). Each pair accomplished attributive classification over 18 pictures. Interlocutors had discussion, reached agreement on the common attributes they found for the pictures, and grouped the pictures with the common attributes together. Every picture was assigned to a category by the common attributes accepted by both of the interlocutors. The classification also started from the picture numbered as 1, continued to the picture numbered as 2, 3, ... and ended by the picture numbered as 18.

4.2 Gender Groups Control

Three gender groups were set in this study: female and female groups, male and male groups, female and male groups. All the subjects formed 70 pairs, of which there are 23 female and female pairs, 17 male and male pairs, and 30 female and male pairs. A series of analyses over prosodic entrainment and gender in Chapter 7 are based on the conversations produced by these gender groups.

4.3 Roles Control

Two types of roles are controlled in this research: unequal roles (information giver and information follower in Picture Ordering Games), and equal roles (the equal roles in the Picture Classifying Games). The effect of unequal roles on prosodic entrainment was analyzed by the comparisons with the effect of equal roles in the Role Effect Test. In addition, for the unequal roles, the entrainment direction was also analyzed to find out whether information givers entrain more to followers, or followers entrain more to givers.

4.4 Variables

The present research focuses on prosodic entrainment in Mandarin conversations. Seven variables are set in the analyses of Tongji Games Corpus in this study. These parameters come from 3 main aspect of prosody, 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).

Speaking-rate equals the number of syllables per second in this study. Syllable counts are made automatically by Praat according to the orthographic transcriptions.

Fundamental frequency of a voice is used to measure how often the sound wave repeats itself. The present research adopts min, mean and max values of fundamental frequency, which are obtained automatically by Praat.

The features of intensity are used to describe the degree of energy in a sound wave. It is perceived as the volume of a sound. The present research adopts min, mean and max values of intensity, which are also obtained automatically by Praat.

5. Results and Discussions

5.1 Entrainment and Gender

The analyses of entrainment and gender include two parts: proximity and the entrainment degree of pairs with different gender combination. In the present research, there are three kinds of gender combination in conversation (female-female groups, female-male groups, male-male groups).

5.1.1 Paired T-tests over different gender groups

Paired T-tests are accomplished between the partner distances and non-partner distances in this analysis. For each speaker in a conversation, a partner distance and non-partner distance are calculated. 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-partners are restricted to those of the same gender and conversational role (the information giver or follower). It is hypothesized that partner distance is smaller than the non-partner distance in this analysis. That is, if there is prosodic proximity, the partner distance should be smaller than the non-partner distance, which can supply the evidence for entrainment at conversation level.

The paired T-tests are accomplished over the pairs with three kinds of gender combination (female-female groups, female-male groups, male-male groups). 39 female subjects participated in the female-female conversations; 20 male subjects participated in the male-male conversations; 40 subjects (20 female subjects and 20 male subjects) participated in the mixed-gender conversations.

For female-female groups, 39 pairs of partner distances and non-partner distances are calculated for one variable (one prosodic feature), and are put in a paired T-test. There are 7 variables in the present research, so 7 series of paired T-tests are accomplished over all the prosodic features examined.

For male-male groups, 20 pairs of partner distances and non-partner distances are calculated for one variable (one prosodic feature), and are put in a paired T-test. There are 7 variables in the present research, so 7 series of paired T-tests are accomplished over all the prosodic features examined.

For female-male groups, 40 pairs of partner distances and non-partner distances are calculated for one variable (one prosodic feature), and are put in a paired T-test. There are 7 variables in the present research, so 7 series of paired T-tests are accomplished over all the prosodic features examined.

Since the paired T-tests are accomplished over 3 gender groups respectively, the results of analyses are listed separately.

For female-female group, the results of paired T-tests are listed in Table 1.

Table 1. Paired T-tests of Female-female Groups

Feature	*t*	*df*	*p-value*	*Sig.*
Speaking-rate	-5.792	38	0.0	*
F0 min	-1.283	38	0.207	/
F0 mean	-0.523	38	0.604	/
F0 max	-0.241	38	0.811	/
Intensity min	-1.726	38	0.092	/
Intensity mean	-4.765	38	0.0	*
Intensity max	-5.075	38	0.0	*

In Table 1, an asterisk * indicates the significant difference, and the symbol / indicates no significant difference. This table shows that in female-female conversations, speakers show significant entrainment over 3 prosodic features: Speaking-rate (p= 0.0 < 0.05), Intensity min (p= 0.0 < 0.05), and Intensity max (p= 0.0 < 0.05).

For male-male group, the results of paired T-tests are listed in Table 2.

Table 2. Paired T-tests of Male-male Group

Feature	*t*	*df*	*p-value*	*Sig.*
Speaking-rate	-6.804	19	0.0	*
Pitch min	0.453	19	0.656	/
Pitch mean	0.228	19	0.822	/
Pitch max	0.595	19	0.559	/
Intensity min	-1.869	19	0.077	/
Intensity mean	0.176	19	0.862	/
Intensity max	0.099	19	0.922	/

In Table 2, an asterisk * indicates the significant difference, and the symbol / indicates no significant difference. Table 24 shows that in male-male conversations, speakers show significant entrainment over 1 prosodic feature: Speaking-rate (p=0.0 < 0.05).

For female-male group, the results of paired T-tests are listed in Table 3.

Table 3. Paired T-test of Female-male Group

Feature	*t*	*df*	*p-value*	*Sig.*
Speaking-rate	-3.024	39	0.004	*
Pitch min	1.032	39	0.309	/
Pitch mean	0.972	39	0.337	/
Pitch max	-4.281	39	0.000	*
Intensity min	0.684	39	0.498	/
Intensity mean	-3.379	39	0.002	*
Intensity max	-3.225	39	0.003	*

In Table 3, an asterisk * indicates the significant difference, and the symbol / indicates no significant difference. Table 26 shows that in female-male conversations, speakers show significant entrainment over 4 prosodic features: Speaking-rate ($p=0.004 < 0.05$), F0 max ($p=0.0 < 0.05$), Intensity mean ($p=0.002< 0.05$)), and Intensity max ($p=0.003 < 0.05$).

The results of paired T-tests over the pairs with 3 kinds of gender combination show that in female-female conversations, speakers show significant entrainment over 3 prosodic features: Speaking-rate, Intensity min, and Intensity max; in male-male conversations, speakers show significant entrainment over 1 prosodic feature: Speaking-rate; in female-male conversations, speakers show significant entrainment over 4 prosodic features: Speaking-rate, F0 max, Intensity mean, and Intensity max.

Then it is found that more features are entrained in female-male conversations. The number of prosodic features entrained in mix gender group's conversations is the most, and the number of the prosodic features in male-male group's conversations is the least.

In conversations of the pairs with the same gender (female-female and male-male conversations), the features of duration and intensity are entrained; in mixed gender group's conversations, including the feature of duration, features of F0 are also entrained.

5.1.2 Entrainment degree of the pairs with different gender combination

In addition to the number of features speakers entrained over, the present research also analyzes the degree of entrainment in the pairs with different gender combination.

From the results above in section 5.1.1, over Speaking-rate, Intensity mean and Intensity max, pairs in female-female, female-male and male-male conversations exhibit entrainment. It is necessary to explore further what is the difference of entrainment degrees over three gender combinations for anyone feature of the three.

ANOVA tests are adopted in the analyses of entrainment degree, in which adaptation degrees are considered as dependent variables, and 3 gender combinations are independent variables.

In this analysis, *sim* represents the adaptation degree of the prosodic feature f. *disp* represents the partner distance as in Formula 1, and *disnp* represents the non-partner distance as in Formula 2. More details are shown in Formula 3.

For each conversation, the present research defines disp as the partner distance between two partners (speaker A, speaker B) on the prosodic feature f:

$$disp = \left| A_f - B_f \right| \quad \text{(Formula 1)}$$

In Formula 1, disp represents the partner distance, Af and Bf are weighted average (as what mentioned in section 3.2.6) for the feature f over the whole conversation of the two partnered speakers A and B.

The present research define disnp as the non-partner distance on the feature f:

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

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 restriction to the speakers with the same gender and role as the speakers' partner is to decrease the influence of gender and role in the results. Af and Xif are also the weighted mean (mentioned in section 3.2.6) for the feature f over the whole conversation of the two non-paired speakers A and X. | Af - Xif | represents the distance between non partners

The second step is to define a parameter--- *sim,* in this analysis, which is calculated by Formula 3.

$$sim = 1 - \frac{disp}{disnp} \quad \text{(Formula 3)}$$

In Formula 3, *sim* represents the adaptation degree of one prosodic feature. *disp* represents the *partner* distance as in Formula 1, and *disnp* represents the *non-partner* distance as in Formula 2.

In this analysis, non-partner distance is considered as the baseline in the distance of the two interlocutors, the distance without the mutual adaptation. Thus, disp/disnp, disp normalized by disnp, represents the ratio of the remained distance after adaption between interlocutors to the primitive distance. Then, sim, 1- disp/disnonp, represents a similarity ratio, the percentage of becoming similar. This similarity ratio or normalization in Formula 3 has an advantage to control for speaker' differences.

For Speaking-rate, Intensity means, and Intensity max, over which all the pairs with 3 gender combinations exhibit entrainment, ANOVA tests are conducted over female-female, female-male, and male-male groups.

39 female subjects participated in the female-female conversations; 20 male subjects participated in the male-male conversations; 40 subjects (20 female, 20 male) participated in the mixed-gender conversations. Sim is calculated by Formula 3 over the values of disp and disnp from these conversations.

39 *sim* come from 39 female-female conversations; 20 sim come from 20 male-male conversations; 40 sim come from female-male conversations. ANOVA tests are accomplished over 3 prosodic features, taking sim as the dependent variable and 3 gender groups the independent variables.

For Speaking-rate, Intensity mean, and Intensity max, over which all pairs with 3 gender combinations exhibit entrainment, ANOVA tests are conducted over the female-female, female-male, and male-male groups, and the results of these tests are listed respectively below. The results of ANOVA test in terms of Speaking-rate over 3 gender combinations are listed in Table 4.

Table 4. An ANOVA Test in Terms of Speaking-rate over 3 Gender Combinations

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.056	2	1.028	4.508	.013
Within Groups	21.886	96	.228		
Total	23.942	98			

Table 4 shows that over Speaking-rate, *sim* has significant difference in the pairs with 3 gender combinations ($p=0.013 < 0.05$, $F=4.508$). That is, in terms of Speaking-rate, entrainment degree is significantly different over the pairs with three gender combinations. It is necessary to test further the difference between every two gender combinations. Therefore, the multiple comparisons are made below.The results of post-doc comparison are listed in Table 5.

Table 5. Post hoc Comparison of ANOVA Test in Terms of Speaking-rate

Multiple Comparison						
LSD						
					95% Confidence Interval	
(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
ff	fm	.23622*	.10745	.030	.0229	.4495
	mm	-.12486	.13132	.344	-.3855	.1358
fm	ff	-.23622*	.10745	.030	-.4495	-.0229

	mm	-.36108*	.13076	.007	-.6206	-.1015
mm	ff	.12486	.13132	.344	-.1358	.3855
	fm	.36108*	.13076	.007	.1015	.6206

*. The mean difference is significant at the 0.05 level.

Table 5 shows that over Speaking-rate, *sim* is the smallest in female-male conversations within 3 gender combinations, because *sim* in the female-male group is significantly smaller than that in the female-female group ($p=0.030 < 0.05$) and *sim* in the female-male group is significant smaller than that in the male-male group ($p=0.007 < 0.05$). *Sim* is not significantly different between the female-female group and the male-male group. That is to say, in terms of Speaking-rate, the entrainment degree over the female-male group is the smallest, and there is no significant difference between female-male group and the male-male group.

The results of ANOVA test in terms of Intensity Mean in the pairs with 3 gender combinations are listed in Table 6.

Table 6. An ANOVA Test in Terms of Intensity Mean over 3 Gender Combinations

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.204	2	1.102	3.131	.048
Within Groups	33.791	96	.352		
Total	35.996	98			

Table 6 shows that over Intensity mean, *sim* is significantly different in the pairs with 3 gender combinations ($p=0.048 < 0.05$, $F=3.131$). That is, in terms of Intensity min, entrainment degree is significantly different over the pairs with three gender combinations. It is necessary to test further the difference between every two gender combinations. Therefore, the multiple comparisons are made below. The results of post-hoc comparison are listed in Table 7.

Table 7. Post hoc Comparison of ANOVA Test in Terms of Intensity Mean

Multiple Comparison						
LSD						
					95% Confidence Interval	
(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
ff	fm	.12049	.13351	.369	-.1445	.3855
	mm	.40764*	.16317	.014	.0837	.7315
fm	ff	-.12049	.13351	.369	-.3855	.1445
	mm	.28715	.16248	.080	-.0354	.6097
mm	ff	-.40764*	.16317	.014	-.7315	-.0837
	fm	-.28715	.16248	.080	-.6097	.0354

*. The mean difference is significant at the 0.05 level.

Table 7 shows that over Intensity mean, *sim* is the smallest in male-male conversations within 3 gender groups, because *sim* in the male-male group is significantly smaller than in the female-female group ($p= 0.014 < 0.05$), and *sim* in the male-male group tends to be smaller than in the female-male group ($p= 0.080 < 0.01$). *Sim* in the male-male group is not significantly different from that in female-female groups or female-male groups. That is to say, in terms of Intensity mean, the entrainment degree over male-male group tends to be the smallest, and there is no significant difference between the female-female group and the female-male group. The results of ANOVA test in terms of Intensity max in the pairs with 3 gender combinations are listed in Table 8.

Table 8. An ANOVA test in terms of Intensity max over 3 gender combinations

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.446	2	1.223	3.729	.028
Within Groups	31.484	96	.328		
Total	33.930	98			

Table 8 shows that over Intensity max, *sim* is significantly different in the pairs with 3 gender combinations ($p=0.028 < 0.05$, $F=3.729$). That is, in terms of Intensity max, entrainment degree is significantly different over the pairs with three gender combinations. It is necessary to test further the

difference between every two gender combinations. Therefore, the multiple comparisons are made below.The results of post-hoc comparisons are listed in Table 9.

Table 9. Post hoc Comparison of ANOVA Test in Terms of Intensity Max

Multiple Comparison						
LSD						
					95% Confidence Interval	
(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
ff	fm	.18379	.12887	.157	-.0720	.4396
	mm	.42689*	.15750	.008	.1143	.7395
fm	ff	-.18379	.12887	.157	-.4396	.0720
	mm	.24310	.15683	.124	-.0682	.5544
mm	ff	-.42689*	.15750	.008	-.7395	-.1143
	fm	-.24310	.15683	.124	-.5544	.0682
*. The mean difference is significant at the 0.05 level.						

Table 9 shows that over Intensity max, *sim* tends to be the smallest in male-male conversations among the 3 gender groups, because *sim* in the male-male group is significantly smaller than that in the female-female group ($p= 0.008 <0.05$), and *sim* in the male-male group is smaller than that in the female-male group, although the result is not significant ($p= 0.124 > 0.05$). *Sim* in the male-male group is not significantly different from that in the female-female group or the female-male group. That is to say, in terms of Intensity max, the entrainment degree in the male-male group tends to be the smallest, and there is no significant difference between the female-female group and female-male group.

5.1.3 Summary of the analyses of gender and prosodic entrainment

According to the results of a series of ANOVA tests over Speaking-rate, Intensity mean, and Intensity max, it is found that in terms of Speaking-rate, the entrainment degree over female-male group is the smallest, and there is no significant difference over the female-male group and the male-male group; in terms of Intensity mean, the entrainment degree of the male-male group tends to be the smallest, and there is no significant difference between the female-female group and the female-male group; in terms of Intensity max, the entrainment degree of male-male group tends to be the smallest, and there is no significant difference between the female-female group and the female-male group.

A summary of the research on entrainment and gender was made.

It is found that mixed gender pairs entrain over the greatest number of features, and male pairs on the least in both languages. It is also found that 3 gender groups show significant difference in entrainment degree, and male-male pairs tend to entrain least. Entrainment of mixed gender pairs is most prevalent but not necessarily strongest. Therefore, it can be concluded that in both number of features entrained and degree of entrainment: males entrain the least.

5.2 Entrainment and Role

The analyses of entrainment and role are accomplished by two steps. The first step is the Role Influence Test, which aims to test whether the unequal roles affect the degree of entrainment or not. If the unequal roles prove to affect the degree of entrainment in the test of first step, it is necessary to test where the differences in the degree of entrainment come from. Then the second step is accomplished, which is called Role Direction Test. This test aims to test the direction of entrainment between two roles in conversations. In detail, in the present research, this test aims to find out whether givers entrain more to followers or followers entrain more to givers?

5.2.1 Role influence test

In order to find out whether the unequal roles affect the degree of entrainment or not, the present research compares the entrainment degree of the pairs with unequal roles and that of the pairs with equal roles, which are controlled in Picture Ordering Games and Picture Classifying Games. So, research question in this section is converted to find whether the entrainment degree is different between Picture Ordering Games and Picture Classifying Games or not.

The entrainment degree in Picture Ordering Games, *odis*, refers to the distance of prosodic features over the whole conversation between two speakers in Picture Ordering Games. The entrainment degree in Picture Classifying Games, *cdis*, refers to the distance of prosodic features over the whole conversation between two speakers in Picture Classifying Games. Paired T-tests are accomplished between *odis* and *cdis*.

In terms of *odis*, the distance of prosodic features over the whole conversation between two speakers in the Picture Ordering Games conversations, the average distance from two conversations is calculated as its value, because every pair of subjects plays the Picture Ordering Game twice by switching the roles from information giver to information follower in Picture Ordering Games. Thus, 14 *odis* are computed for one prosodic parameter over 28 conversations in Picture Ordering Games by getting their averages.

In terms of *cdis*, the distance of prosodic features over the whole conversation between two speakers in Picture Classifying Games conversations, one value of *cdis* comes from one conversation, because every pair plays classifying game once. 14 pairs participated in 14 conversations, so 14 *cdis* are computed over 14 conversations for one prosodic parameter.

One paired T-test is accomplished between 14 *odis* and 14 *cdis* for one prosodic parameter. 7 paired T-tests are accomplished for 7 prosodic parameters. The results in paired T-tests of entrainment degree

for the same interlocutors between two games are listed in Table 10.

Table 10. Paired T-tests in Entrainment Degree between Two Games

Feature	*t*	*df*	*p-value*	*Sig.*	*meano*	*meanc*
Speaking-rate	2.225	13	0.044	*	0.7782	0.5194
F0 min	2.887	13	0.013	*	21.4646	17.7399
F0 mean	2.119	13	0.054	/	25.8304	23.4078
F0 max	2.506	13	0.026	*	38.2213	33.9959
Intensity min	-1.391	13	0.187	/	4.5517	5.4831
Intensity mean	0.746	13	0.469	/	4.7587	4.3888
Intensity max	0.875	13	0.397	/	5.0007	4.5094

In Table 10, an asterisk * indicates the significant difference, and the symbol / indicates no significant difference. According to this table, the entrainment degree between the Picture Ordering Games and Picture Classifying Games are significantly different over 3 prosodic parameters: Speaking-rate ($p= 0.044 < 0.05$), F0 min ($p= 0.013 < 0.05$) and F0 max ($p= 0.026 < 0.05$). The entrainment degree between two games is significantly different over F0 mean ($p= 0.054 < 0.1$) at confidence level 0.1.

These results indicate that the degree of entrainment is significantly different between Picture Ordering Games and Picture Classifying Games over three prosodic parameters: Speaking-rate, F0min, and F0max, and the feature of F0 mean tends to be different between the two games.

These indicate that the entrainment degree in the picture ordering game is larger than the entrainment degree in the picture classifying game over almost every prosodic feature tested.

The results above prove that the unequal roles (information giver or follower) affect the entrainment degree in conversation over the prosodic features of duration and F0, and the entrainment in the picture ordering games is stronger than that in the picture classifying games. That is, there is more entrainment in the conversations with unequal roles than in those with equal roles.

5.2.2 Role direction test

It is proved in the Section 5.2.1 that the unequal roles (information giver or follower) affect the entrainment degree in conversation over the prosodic features of duration and F0. Questions can be asked further: what is the entrainment direction? Do information givers entrain more to followers, or do followers entrain more to givers?

As mentioned at the beginning of this section, the Role Direction Test is the second step in the analyses of entrainment and role. In the first step, the Role Influence Test, it is found that the degree of

entrainment is significantly different between Picture Ordering Games and Picture Classifying Games over the features of duration and F0. Then it is necessary to analyze where the differences in the degree of entrainment come from.

As mentioned above, some subjects are chosen to participate in both Picture Ordering Games and Picture Classifying Games. In Picture Ordering Games, two interlocutors (A and B) have different roles as information giver and information follower. When they finished the first game, they switched the roles in another ordering game. That is, two interlocutors participated in two ordering games, in which they switched the roles between information giver and information follower. In picture classifying, two speakers have the same roles in discussion. Thus, every subject has 3 speaking states.

For instance, a subject in a pair (A & B) has 3 speaking states.

Ag: A plays as an information giver in the first picture ordering game with B.

Af: A plays as an information follower in the second picture ordering game with B.

A0: A plays in an equal role with B in a picture classifying game.

Bg: B plays as an information follower in the first picture ordering game with A.

Bf: B plays as an information giver in the second picture ordering game with A.

B0: B plays in an equal role with A in a picture classifying game.

The same pair plays two kinds of games: the Picture Classifying Games and Picture Ordering Games. Taking one side of the interlocutors as an example, A0 is A's speaking state with equal roles; Ag is A's speaking state when A is an information giver; Af is A's speaking state when A is an information follower. The difference between Ag and Af is the change in speaking state caused by the change in A's roles from the information giver to the information follower, while the partner is still B. What is changed in B in Picture Ordering Games is also the state of speaking from the information follower to the information giver. Therefore, as the other side of the pair, the partner of A, B also has 3 similar speaking states.

The differences in the entrainment degree between two kinds of games found in the Section 5.1 are assumed to be caused by the difference in the roles. In the present research, among the 3 speaking states of A, A0 is considered as a reference state, because this speaking state is produced under the equal role with the partner. The deviation between Ag and A0 is caused only by A's role of a giver, and the deviation between Af and A0 is caused only by A's role of a follower. Therefore, $|Af-A0|$ is the deviation caused by the role of giver; $|Af-A0|$ is the deviation caused by the role of follower. The key point to make such comparisons valid is that in this analysis the partner of A is not changed, and what is changed is only the role of A. And all of this attributes to the experiment design of the present research that several pairs of subjects participated in both games.

A paired T-test between $|Ag-A0|$ and $|Af-A0|$ is adopted in this section in order to find out which

role causes bigger deviation. If ｜Ag-A0｜is bigger than｜Af-A0｜, the deviation caused by the role of giver is bigger than that caused by the role of follower. That is, the giver makes larger deviation in entrainment than the follower does in conversation, so the giver entrains more to the follower. If｜Ag-A0｜is smaller than /Af-A0/, the deviation caused by the role of giver is smaller than that caused by the role of follower. That is, the follower makes a larger derivation in entrainment than the giver does in conversation, so the follower entrains more to the giver.

One point should be mentioned in the use of this method. The post-conversation influence is not considered in this analysis. Fourteen pairs participated in both the Picture Ordering Games and Picture Classifying Games. Every two interlocutors of them participated in two ordering games, in which they switched the roles between information giver and information follower, and a picture classifying game, in which two speakers in the pair have the same roles in discussion. One pair of these subjects actually produced three conversations. Thus, a question is raised: is there post-conversation influence among these conversations? That is, is the entrainment state in the followed conversations influenced by the previous ones?

In the present research, the post-conversation influence of these three conversations produced by every pair of the 14 is not considered, because the requirements of these games are different. After the first picture ordering game, the roles of two interlocutors are changed in the second picture ordering game, and the content of picture classifying game is quite different from the previous two. Thus, the requirements of the games are always different. Therefore, the post-conversation influence is excluded in these analyses.

According to the corpus description in Chapter 3, 14 pairs (28) subjects participated in both the Picture Ordering Games and Picture Classifying Games. Then 28 subjects made 28 conversations. One subject (A) has 3 speaking states, Ag, Af and A0 from the Picture Ordering Games and Picture Classifying Games. For 28 subjects, there are 28 groups of Ag, Af and A0.

The role influence test in Section 5.1 is the precondition of the Role Direction Test in Section 5.2. That is, those features which show role influence could be tested further for the role direction. According to the results of role influence test, different entrainment degree is found between unequal roles (the role of information giver and follower) and equal roles (the role of picture classifier) over the features of duration and F0. So as the further test of Role Influence Test, in the Role Direction Test, paired T-tests are accomplished over the features of duration and F0 including Speaking-rate, F0min, F0mean, and F0max.

For the features of duration and F0, 28 groups of｜Ag-A0｜and｜Af-A0｜are listed. The paired T-tests are accomplished between these groups.

The results of paired T-tests between｜Ag-A0｜and｜Af-A0｜are listed in Table 11.

Table 11. Paired T-tests between｜Ag-A0｜and /Af-A0/

Feature	*t*	*df*	*p-value*	*Sig.*	｜Ag-A0｜	｜Af-A0｜
Speaking-rate	2.481	27	0.02	*	0.5472	0.3448
F0 min	-2.437	27	0.8	/	3.7020	5.9698
F0 mean	1.912	27	0.067	/	6.2957	4.1347
F0 max	2.790	27	0.01	*	12.6457	6.4339

In Table 11, an asterisk * indicates the significant difference, and the symbol / indicates no significant difference. According to this table, the｜Ag-A0｜is significantly and consistently bigger than /Af-A0/ over speaking rate and F0 max ($p=0.02, 0.01 < 0.05$). In addition, ｜Ag-A0｜is significantly bigger than｜Af-A0｜over F0 max ($p=0.067 < 0.1$) at confidence level 0.1. These show that Ag makes the significantly bigger deviation than Af. That is, when A plays as the information giver, deviation in entrainment is bigger than that when A plays as the information follower.

The results indicate that givers make larger adjustments towards their followers, and followers make smaller adjustments towards their givers. When a giver makes larger adjustment and a follower makes smaller adjustment, the giver entrains more to the follower.

Based on the results above, it is found that there is more entrainment when speakers are givers than when they are followers. Information givers entrain more than followers in conversation.

In conversation, givers entrain more to followers. And the entrainment from givers to followers is prominent over Speaking-rate and F0 max which are the main features that show global entrainment as well.

5.2.3 Summary of the analyses of role and prosodic entrainment

A summary was made for the analyses of entrainment and role.

This section includes two tests: Role Influence Test and Role Direction Test.

Through the Role Influence Test, it is found that the unequal roles (information giver or follower) affect the entrainment degree in conversation over the prosodic features of duration and F0, and there is more entrainment in the conversations with unequal roles than in those with equal roles.

Through the Role Direction Test, it is found that there is more entrainment when speakers are givers than when they are followers. Information givers entrain more than followers in conversation.

In conversations, givers entrain more to followers. And the entrainment from givers to followers is prominent in Speaking-rate and F0 max.

6. Conclusions

This study tested the relationship between prosodic entrainment and two social factors (gender and role) in conversations.

The analyses of entrainment and gender include two parts: proximity by gender groups, and entrainment degree by gender groups. In the analysis of proximity by gender groups, it is found the most prosodic features are entrained in female-male conversations, and the least in male-male conversations; in the same gender pairs' conversations (female-female and male-male conversations), the features of duration and intensity are entrained; in mixed gender pairs' conversations (female-male), besides the same features entrained above, features of F0 are also entrained. In the analysis of entrainment degree by gender groups, it is found that male-male pairs tend to entrain least in MC conversations.

The analyses of entrainment and role include two parts: Role Influence Test and Role Direction Test. In the analysis of Role Influence Test, it is found that the entrainment degree in Picture Ordering Games is significantly bigger than that in the Picture Classifying Games. Two interlocutors play unequal roles in Picture Ordering Games (information giver and information follower), and pairs play equal roles in Picture Classifying Games. This result of Role Influence Test shows that different roles have influence on the entrainment degree. In the analysis of Role Direction Test, it is found that information givers entrain more to followers in conversation.

Based on all the findings above, a conclusion can be made that in Mandarin Chinese conversations, prosodic entrainment has close relationship with the interlocutors' gender and role.

Acknowledgments

This work was supported by the National Social Science Fund of China (NSSF Grant No. 20BYY099), the Fund for English Teaching from Jiangsu Higher Education Association (2022WJZD006), the Fund for Courses Construction in Ideological and Political Education from Jiangsu Normal University (22KCSZkc10), and the Fund for English Lesson Construction from the School of Foreign Studies in Jiangsu Normal University (2022YYKC3, 2022YYKC6).

References

Bell, L., Gustafson, J., & Heldner, M. (2003). Prosodic adaptation in human-computer interaction. *In Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS), vol.3*, 833-836. *Citeseer.*

Beňuš, Š., Gravano, A., & Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, 43, 3001-3027.

Black, J. W. (1949). The intensity of oral responses to stimulus words. *Journal of Speech and Hearing Disorders*, *14*(1), 16-22.

Bolinger, D. (1989). *Intonation and Its Use*: *Melody in Grammar and Discourse*. London: Edward Arnold.

Boylan, P. (2004). *Accommodation theory revisited*. Retrieved September 10, 2012, from http://www.docin.com/p-688557528.html

Breazeal, C. (2002). Regulation and entrainment in human-robot interaction. *The International Journal of Robotics Research*, *21*(10-11), 883-902.

Chartrand, T. L., & Bargh, J. A. (1999). The Chameleon Effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893-910.

Chi-Chun, L., Matthew, B., Athanasios, K., Adam, L., Brian, B., Andrew, C., Panayiotis G., & Shrikanth, N. (2010). Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. In *Proceedings of Eleventh Annual Conference of the International Speech Communication Association* (pp.793-796). Makuhari, Japan.

Collins, B. (1998). Convergence of fundamental frequencies in conversation: If it happens, does it matter? In *Proceedings of Fifth International Conference on Spoken Language Processing (ICSLP)*.

Coulston, R., Oviatt, S., & Darves, C. (2002). Amplitude convergence in children's conversational speech with animated personas. In *Proceedings of the Seventh International Conference on Spoken Language Processing*, vol. 4. (pp. 2689-2692) Citeseer.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

De Looze, C., & Rauzy, S. (2011). Measuring speakers' similarity in speech by means of prosodic cues: Methods and potential. *Proceedings of Interspeech 2011* (pp. 1393-1396) (publishing).

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

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

Edlund, J., Heldner, M., & Hirschberg, J. (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).

Gallois, C., & Callan, V. J. (1988). Communication accommodation and the prototypical speaker: Predicting evaluations of status and solidarity. *Language and Communication*, *8*(3), 271-283.

Giles, H., Coupland, J., & Coupland, N. (1991). Accommodation Theory: Communication, Context, and Consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of Accommodation*. Cambridge: Cambridge University Press.

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

Gregory, S. W., & Dagan, K. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior*, *21*(1), 23-43.

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

Heldner, J., Edlund, M., & Hirschberg, J. (2010). Pitch similarity in the vicinity of backchannels. In *Proceedings of Interspeech* (pp. 1-4).

Jaffe, J., & Feldstain, S. (1970). *Rhythms of Dialogue*. New York: Academic Press.

Kousidis, S., Dorran, D., Mcdonnell, C., & Coyle, E. (2009). Convergence in human dialogues time series analysis of acoustic feature. In *Proceedings of the International Workshop SPEECH and COMPUTER (SPECOM)*. St. Petersburg, Russian Federation. Retrieved June 10, 2013, from http://arrow.dit.ie/cgi/viewcontent.cgi?article=1002&context=dmccon

Kopp, S. (2010). Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication*, *52*(6), 587-597.

Kousidis, S., Dorran, D., McDonnell, C., & Coyle, E. (2008). Times series analysis of acoustic feature convergence in human dialogues. In *Proceedings of Interspeech*.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge University Press.

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Journal of Psychological Science*, *14*(4), 334-339.

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

Levitan, R., Gravano, A., & Hirschberg, J. (2011). Entrainment in speech-proceeding backchannels. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (pp. 113-117). Portland, Oregon.

Levitan, R., Gravano, A., Willson, L., Benus, S., Hirschberg, J., & Nenkova, N. (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.

Matarazzo, J. D., & Wiens, A. N. (1967). Interviewer influence on durations of interviewee silence. *Journal of Experimental Research in Personality*, *2*(1), 59-69.

McGarva, A. R., & Warner, R. M. (2003). Attraction and social coordination: Mutual entrainment of vocal activity rhythms. *Journal of Psychology*, *32*(3), 335-354.

Miles, L. K., Nind, L. K., & Macrae, C. N. (2009). The rhythm of rapport: interpersonal synchrony and social perception. *Journal of Experimental Social Psychology*, *45*(3), 585-589.

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.

Oviatt, S. (1996). User-centered modeling for spoken language and multimodal interfaces. *IEEE Multimedia*, *3*(4), 26-35.

Pardo J. S. (2006). On phonetic convergence during conversational interaction. *Acoustical Society of America*, *119*(4), 2382-2393.

Parrill, F., & Kimbara, I. (2006). Seeing and hearing double: The influence of mimicry in speech and gesture on observers. *Journal of Nonverbal Behavior*, *30*(4), 157-166.

Pickering M. J., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169-226.

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*, 203-228.

Putman, W. B., & Street, R. L. (1984). The conception and perception of non-content speech performance: Implications for speech-accommodation theory. *International Journal of the Sociology of Language*, *46*, 97-114.

Shepard, C. A., Giles, H., & Le Poire, B. A. (2001). Communication accommodation theory. *The New Handbook of Language and Social Psychology*, vol. 1.2 (pp. 33-56). John Wiley & Sons Incorporated.

Stanford, G. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, *70*(6), 1231-1240.

Street, Jr., Richard, L., James, N., & Van Kleek, A. (1983). Speech convergence among talktive and reticent three-year-olds. *Language Sciences*, *5*(1), 79-96.

Suzuki, N., & Katagiri, Y. (2007). Prosodic alignment in human-computer interaction. *Connection Science*, *19*(2), 131-141.

Tickle-Degnen, L., & Rosenthal, R. (2007). The nature of rapport and its nonverbal correlates. *Psychological inquiry*, *1*(4), 285-293.

Vaughan, B. (2011). Prosodic synchrony in cooperative task-based dialogues: A measure of agreement and disagreement. In *Proceedings of Interspeech* (pp. 1865-1868).

Ward, N., & Nakagawa, S. (2002). Automatic user-adaptive speaking rate selection for information delivery. In *7th International Conference on Spoken Language Processing* (ICSLP

2002-INTERSPEECH 2002).

Webb, J. T. (1972). Interview synchrony: An investigation of two Speaking-rate measures in an automated standardized interview. In B. Pope, & A. W. Siegman (Eds.), *Studies in Dyadic Communication* (pp. 115-133). New York: Pergamon.

Welkowitz, J., & Kuc, M. (1973). Interrelationships among warmth, genuineness, empathy, and temporal speech patterns in interpersonal interaction. *Journal of Consulting and Clinical Psychology*, *41*(3), 472-473.

Woodall, G. W., & Burgoon, J. K. (1983). Talking fast and changing attitudes: A critique and clarification. *Journal of Nonverbal Behavior*, *8*(2), 126-142.

Xia, Z. H., Hirschberg, J., & Levitan, R. (2023). Investigating prosodic entrainment from global conversations to local turns and tones in Mandarin conversations. *Speech Communication*, *153*.

Xia, Z. H., Levitan, R., & Hirschberg, J. (2014). *Prosodic entrainment in Mandarin and English: A cross-linguistic comparison*. In Proceedings of Speech Prosody, Dublin, Ireland, 65-69.

Xia, Z. H., & Ma, Q. W. (2019). *Prosodic Entrainment in Mandarin Chinese Conversations: An Experimental Study*. Tongji University Press, Shanghai, China.

Xia, Z. H., & Ma, Q. W. (2016a). The influence of gender on prosodic entrainment in Mandarin conversations. *Journal of Phonetics in China*, 118-128.

Xia, Z. H., & Ma, Q. W. (2016b). *Gender and prosodic entrainment in Mandarin conversations*. In Proceedings of 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP 2016), Tianjin, China.

Zebrowitz, L. A., Brownlow, S., & Olson, K. (1992). Baby talk to the baby-faced. *Journal of Nonverbal Behavior*, *16*(3), 143-158.