

Original Paper

A Review of the Research on Linguistic Prosody in Conversation Entrainment

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: November 7, 2023 Accepted: November 22, 2023 Online Published: December 3, 2023

doi:10.22158/jpbr.v5n4p80

URL: <http://dx.doi.org/10.22158/jpbr.v5n4p80>

Abstract

Having conversation is a joint action in which interacting individuals coordinate their behavior and adapt their linguistic choices to each other in order to make smooth and successful communication. This coordination or adaptation is called entrainment. In daily conversation, interlocutors entrain at linguistic or nonlinguistic levels. The present research focuses on linguistic prosody in entrainment, and reviews the theoretical and empirical studies of linguistic prosody in conversation entrainment in order to supply references for future studies in the similar fields.

Keywords

prosody, conversation, entrainment

1. Introduction

Social interaction involves participants' mutual coordination or adaptation. This coordination is called entrainment. Entrainment in speech means that speakers adapt their communicative behavior to their conversational partners. The joint nature of language processing in communication requires the interpersonal coordination in minds and actions (Brennan *et al.*, 2010). There is much evidence that entrainment is critical to humans' assessment of dialogue success, overall quality, and their evaluation of conversational partners (Goleman, 2006).

Often this entrainment or accommodation produces convergence in conception, syntactic forms, lexicon choices, prosody, postures and other behavior of interlocutors (Condon & Sander, 1974; Meltzoff & Moor, 1977; Maurer & Tindall, 1983; Bavelas *et al.*, 1986; Bernieri & Rosenthal, 1991; Bernieri *et al.*, 1994; Richardson *et al.*, 2007; Shockley *et al.*, 2007; Shockley *et al.*, 2009; Hess & Blairy, 2011). This paper focuses on the speech prosody and entrainment in conversation, and aims to

review the theories and empirical research on the interrelation between prosody and entrainment in order to supply references for future studies of prosodic entrainment.

2. Review of the Theoretical Base

Entrainment or adaptation in speech is ubiquitous in interaction as it facilitates, through the alignment of cognitive representations, comprehension and understanding between speakers. The theoretical research on entrainment mainly includes Communication Accommodation Theory, The Chameleon Effect, Perception-Behavior Link, and Interactive Alignment Model. All these theories provide influential and convincing explorations for alignment in conversation from the perspectives of sociolinguistics or psycholinguistics.

2.1 Communication Accommodation Theory (CAT)

In the theories of entrainment, Communication Accommodation Theory (CAT) is the most influential. CAT founded by Giles (1973) states that speakers adjust their speech, vocal patterns and gestures to accommodate to their partners in conversation (Turner & West, 2010). Various reasons for the facts that individuals emphasize or minimize the social differences between themselves and their interlocutors through verbal and nonverbal communication are supplied in this theory (Gallois *et al.*, 2005).

CAT supplies a wide-ranging framework to predict and explain interlocutors' adjustments in creation, maintenance, or decrease of social distance in interaction (Giles & Ogay, 2007, p. 325). This theory explores the different ways in which people accommodate communication, motivations for doing adjustments, and the consequences.

CAT first appeared as Speech Accommodation Theory (SAT) in Giles (1973), which is a study of accent mobility. As a socio-psychological model exploring accent and bilingual shifts in interactions (Giles, 1973; Giles, Taylor, & Bourhis, 1973), CAT becomes an interdisciplinary model of relational and identity processes in communicative interaction (Coupland & Jaworski, 1997).

CAT is established on social psychology, especially from four main socio-psychology theories: similarity-attraction, social exchange, causal distribution and intergroup distinctiveness. The similarity-attraction theory points out that the more similar our attitudes and beliefs are to those of others, the more likely it is for them to be attracted to us; the social exchange process theory states that people tend to choose the action which bring greater rewards and less costs; the causal attribution process theory points out that people interpret others' behavior, and evaluate the individuals in terms of the motivations and intentions that people attribute as the cause of their behavior; the process of intergroup distinctiveness theory suggests that when people from different groups are in contact, they compare themselves on dimensions which are important to them, such as personal attributes, abilities, material possession and so on (Giles & Smith, 1979). These theories supply the foundation to explain the convergence or divergence in interlocutors' speaking.

According to CAT, the benefits of adapting communication to others are to increase communication efficiency, accommodate the differences in language, ability, culture, etc., and gain social approval or desired level of social distance.

Two main accommodation processes described by this theory: convergence and divergence (Giles *et al.*, 1991).

Convergence refers to the strategies through which individuals adapt to each other's communicative behaviors, in order to reduce these social differences. It is a process through which an individual shifts his or her speech patterns in interaction so that they more closely resemble the speech patterns of his interlocutor(s) (Giles & Smith, 1979). People can converge through many features of communication such as their use of language, their "pronunciation, pause and utterance lengths, vocal intensities, nonverbal behaviors, and intimacy of self-disclosures" (Giles & Smith, 1979, p. 46).

Divergence refers to the instances in which individuals accentuate the speech and non-verbal differences between themselves and their interlocutors. This helps to sustain a positive image of one's in-group and hence to strengthen one's social identity. Divergence can thus be a way for members of different groups to maintain their cultural identity, a mean to contrast self-images when the other person is considered a member of an undesirable group, and a way to indicate power or status differences (Turner & West, 2010).

Some motives are supplied for convergence. An important one is the desire to gain approval from the other. A psychological reason is similarity attraction: The more similar we are to our conversational partner, the more he or she will like or respect us, and the more social rewards we can expect (Byrne, 1971). Another motive for convergence is that converging into a common state can improve the effectiveness of communication, increase the predictability of the speaking partner, and hence make lower the uncertainty, interpersonal anxiety (Gudykunst, 1995).

Some motives are supplied for divergence, too. One is the desire to emphasize distinctiveness from one's interlocutor. Social Identity Theory (e.g., Tajfel & Turner, 1986) supplies the explanation: when speakers communicate with each other, an inter-individual interaction is entirely on the basis of their individual differences in temperament and personality and where their ethnicity, gender, age etc.

Although the motives for convergence and divergence exist, speakers' converging are generally viewed more favorably than diverging, and are perceived as more efficient in their communication as well as more cooperative (Giles & Ogay, 2007). From the perspective of perception, convergence is generally considered as the positive evaluation by receivers (Bourhis *et al.*, 1975). For instance, increasing similarity in communicative behavior, such as the Speaking-rate increases in both sides of conversation is perceived attractiveness as well as their ability to gain addressees' compliance (Buller *et al.*, 1992).

Therefore, accommodation most often takes the form of convergence, when a speaker chooses a language variety that seems to fit the style of the other speaker. Less frequently, accommodation may take the form of divergence, when a speaker signals social distance or disapproval by using a language

variety that differs from the style of the other speaker.

2.2 The Chameleon Effect and Perception-Behavior Link

Different from CAT, which proposes speakers intentional and goal-oriented adjustment to converge, maintain, or diverge in interaction, Chartrand and Bargh (1999) put forward the Chameleon Effect in human's communication, that is, people communicate passively and unconsciously reflecting their interlocutors' social behavior like chameleons.

The Chameleon Effect refers to “non-conscious mimicry of the postures, mannerisms, facial expressions, and other behaviors of one's interaction partners, such that one's behavior passively and unintentionally changes to match that of others in one's current social environment” (Chartrand & Bargh, 1999, p. 893).

It is easy to discover in our daily life that one has taken on the accent, speech patterns, and even behavioral mannerisms of one's interaction partners. This process is always considered to be natural and unconscious. It has been proved early that a bi-directional relationship between non-conscious mimicry and liking, rapport, and affiliation of the speakers (Bandura, 1977; Galef, 1988; Heyes, 1993; Chen & Bargh, 1997; Chartrand & Bargh, 1999). That is to say, affiliation can be conveyed by unconscious mimicry, and non-conscious mimicry creates affiliation. This kind of mimicry has been proved to play an important role in human evolution: initially, mimicry has survival value in helping people in communication; then mimicry has evolved to serve a social function; non-conscious behavioral mimicry produces affiliation, and serves to foster favorable relationship with others (Lakin *et al.*, 2003).

Human beings automatically imitate different aspects of interaction patterns, including the speech patterns, facial expressions, emotions, moods, postures, gestures, mannerisms, or even idiosyncratic movements, etc. Studies in these fields are briefly listed below.

People mimic non-consciously the accent (Giles & Powesland, 1975; Pardo, 2006), rates of speech (Webb, 1969, 1972; Levitan *et al.*, 2012), and speech rhythms (Cappella & Panalp, 1981). We also imitate the facial expressions of other people. Imitation of others' facial expressions can result in actually adopting the emotions and moods of others. For instance, when we see or hear others laugh, we are easy to laugh more than ourselves (Young & Frye, 1966), or when we listen to happy or sad persons, we are easy to imitate their tone and take on their mood state (Neumann & Strack, 2000).

Studies also prove the interpersonal coordination, which means “degree to which the behaviors in an interaction are nonrandom, patterned, or synchronized in both timing and form” (Bernieri & Rosenthal, 1991, p. 403). Interpersonal coordination can be further divided into interactional synchrony and behavior matching. Interactional synchrony means the rhythm of an interaction is smooth and synchronic (Bernieri *et al.*, 1994); behavior matching refers to the tendency to mimic or mirror the behaviors of interactional partners (Bavelas *et al.*, 1986).

Perception-behavior link is assumed to be the underlying mechanism for The Chameleon effect. Perception-behavior link means “the unintentional, non-conscious effects of social perception on social behavior” and more illustration about Perception-behavior link comes below (Chartrand *et al.*, 2005, p. 334).

Perception and behavior are inextricably intertwined such that people automatically behave as they perceive. The perception of observables may activate specific behavioral representations, trait constructs, or stereotypes. Mimicry is a manifestation of the perception-behavior link at its most fundamental level. It is no more than copying another’s observables and requires only the ability to perceive the behavior in the other person and the ability to form the behavior oneself.

Dijksterhuis and Bargh (2001) provide the neurological evidence for perception-behavior link. They propose that in the process of thinking, neurological data have also accumulated for a link between the perception of others’ behavior and one’s own behavior; particularly, and they prove that perceiving someone else engaging in a behavior is neurologically similar to performing that behavior.

Because of the perception-behavior link, people entrain in a variety of ways, including speech, facial expressions, physical mannerisms, moods, emotions, etc. The present research focuses on the prosodic alignment in dialogues.

According to the theories of The Chameleon Effect and Perception-Behavior Link, in the real communication, if it is not an absolute duplication or perfect imitation between speakers, a linked perceptual-productive system might produce convergence. A central phenomenon produced by such approaches is an increase in similarity among linguistic components, which is termed convergence, accumulating common ground, or alignment (Pardo, 2006).

2.3 Interactive Alignment Model

From psycho-linguistic aspect, Interactive Alignment Model lays the important foundation for the research on entrainment.

Interactive Alignment Model provides an account of psycho-linguistic mechanisms that interlocutors employ in dialogues. Pickering and Garrod (2004) propose this systematic account of dialogue, and assumes that “in dialogue, the linguistic representations employed by the interlocutors become aligned at many levels, as a result of a largely automatic process (p. 169)”. This theory helps to generalize the process of production and comprehension in dialogue, one of most natural and basic forms in communication.

The interactive alignment includes mainly six points (Pickering & Garrod, 2004, p. 172):

- 1) Alignment of situation model forms the basis of successful dialogue.
- 2) The way that alignment of situation models is achieved is by a primitive and resource-free priming mechanism.
- 3) The same priming mechanism produces alignment at other levels of representation, such as the lexical and syntactic.

- 4) Interconnections between the levels mean that alignment at one level leads to alignment at other levels.
- 5) Another primitive mechanism allows interlocutors to repair misaligned representations interactively.
- 6) More sophisticated and potentially costly strategies that depend on modeling the interlocutor's mental state are only required when the primitive mechanisms fail to produce alignment.

Interactive Alignment Model explains how different levels of representation interact in a dialogue.

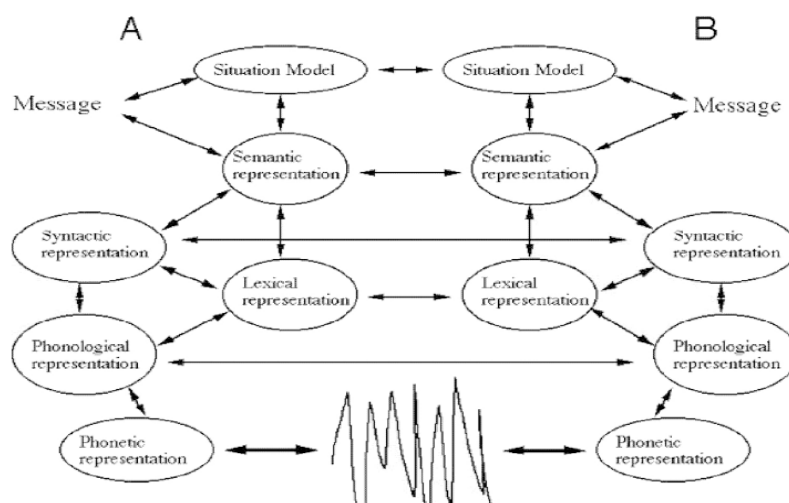


Figure 1. Shows how this Model Works (Pickering & Garrod, 2004, p. 176)

In Figure 1, A and B represent two interlocutors in a dialogue. This schematic representation of the stages of comprehension and production processes is used to illustrate the overall architecture in dialogue. A situation model is located at the top, and other levels of linguistic representation are semantic, syntactic, lexical, phonological, and phonetic forms.

According to Pickering and Garrod (2004), alignment of situation model is the basis of successful dialogue. A situation model is a multi-dimensional representation of the situation under discussion (Johnson-Laird, 1983). The key dimensions encoded in situation models are space, time, causality, intentionality, and reference to main individuals under discussion (Zwaan & Radvansky, 1998). Based on some evidences for alignment of situation models (Garrod & Anderson, 1987; Tanenhaus *et al.*, 1995; Chambers *et al.*, 2002), Pickering and Garrod (2004) assume that successful dialogue involves approximate alignment at the level of the situation model at least, and interlocutors develop aligned situation models.

In addition to the situation model, linguistic elements and structures indicate alignment at other levels. The successful dialogues are produced through the alignment at different levels of language

representations.

Several studies supplied evidence for the lexical alignment in dialogue. Garrod and Anderson (1987) demonstrate the alignment of lexical processing in conversation. Studies show that interlocutors tend to develop the similar set of referring expressions to particular object (Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992; Brennan & Clark, 1996).

A series of research provide evidence for the syntactic alignment in dialogue. Levelt and Kelter (1982) find that subjects tend to use consistent syntactic pattern in conversation. Bock (1986) finds that speakers tended to repeat syntactic form. Branigan *et al.* (2000) find that syntactic structure in dialogue is influenced mutually between two interlocutors. The further study of this group finds that syntactic alignment is not restricted between speaker-addressee (Branigan *et al.*, 2007). A study of bilingual monologue and dialogue find that the integration of languages in sentence production relied on process of syntactic co-activation between languages and on the processes of alignment between dialogue partners (Kootstra *et al.*, 2010).

Alignment is also found at the level of articulation. Fowler and Housum (1987) find that when speakers repeat expressions, articulation becomes reduced (the shortened expressions are the examples). Other studies also find the alignment in phonetic elements and prosodic expressions (Cummins, 2009; Levitan & Hirschberge, 2011; Levitan *et al.*, 2012).

Alignment happens not only at a single level. There is interaction among the alignment from different levels, that is, alignment at one level leads to alignment at other levels. For instance, syntactic alignment is enhanced by lexical overlap, that is, syntactic alignment is considerably enhanced if the verb is repeated (Brannigan *et al.*, 2000); syntactic alignment is enhanced by semantic overlap (Cleland *et al.*, 2002). Cleland and Pickering (2003) find that the exact repetition at one level is not required, but if the relationship at one level is closer, the tendency to align at the other level is stronger. According to Pickering and Garrod (2004), speakers intend to align expressions at different levels simultaneously; when all the levels are entrained, speakers will repeat each other's' expressions in the same way, for example, with the same intonation.

Interactive Alignment Model supplies an account for the alignment in dialogues. This model assumes that the linguistic representations used by speakers become aligned at levels, that the alignment at one level contributes to the alignment at other levels, and that the alignment at all the linguistic levels makes the state of speakers of conversation similar, which produces smooth and successful communication.

For the underlying mechanism of alignment, based on Interactive Alignment Model, Pickering & Garrod (2004) propose that alignment is unconscious, that is, interlocutors may be almost entirely unaware that alignment has taken place.

3. Review of the Empirical Research

The previous empirical researches are reviewed mainly from two aspects: the studies of entrainment over various language components, and pragmatic functions of entrainment.

3.1 Studies of Entrainment over Language Components

Motivated by the theories claiming entrainment's realization at various language representations, some studies have explored entrainment over different language components.

Lexical entrainment has been studied. There is extensive evidence for alignment in word choice. Garrod and Anderson (1987) find that interlocutors use the same referring expressions, and they tend to interpret words in the same way when subjects did a cooperative maze game. Brennan and Clark (1996) find that directors and matchers prefer to use subordinate-level terms (eg., pennyloafer) rather than basic-level terms (eg., shoe) in card games, in which directors describe a set of cards depicting common objects to matchers so that they could reconstruct the directors' array. An explanation for lexical entrainment in this research is that interlocutors negotiate a given conceptualization or achieve conceptual pacts in referring an object. In the experiment following up, Metzger and Brennan (2003) show that when partners are changed and conceptual pact is broken by new referring expressions to refer to a previously discussed object, subjects experienced difficulty in locating the objects. Similarly, Horton and Gerrig (2005) find such partner-specific effects, and suggest interlocutors represent as a conjoined cue information about linguistic expressions and the person with whom the expressions is used, so that the presence of that particular interlocutor activates that expression.

Entrainment has been studied at syntactic level. Levelt and Kelter (1982) find alignment in syntactic structure between interlocutors in the study over telephone conversations. For instance, interlocutors tend to say *At 5 o'clock* as the answer to *At what time does your shop close?* And tend to say *5 o'clock* as the answer to *What time does your shop close?* Garrod and Anderson (1987) find coordination in syntactic structure between partners in conversation in the study of the cooperative maze games. They find that participants tend to align on the same sentence structure, although there are various methods to describe their positions to each other. For example, one subject said: "I am two along, four up". and the partner tended to say: "I am one along, five up". When this player changed to say: "I am at B4", the partner tended to say: "I am at A5". The subjects tended to coordinate the sentence structure rather than specific words. Branigan *et al.* (2000) find that syntactic structure of a subject's utterance is influenced by that of a confederate's previous utterance in the study of card games. Reitter *et al.* (2014) find that interlocutors are more likely to use a given syntactic rule soon after their partners have used that rule in the study of the task-oriented conversations. Alignments in the syntactic representations are also found in bilinguals' conversation (Hartsuiker *et al.*, 2004). Strong evidence for alignments in syntactic structure is found in other studies (Cleland & Pickering 2003; Haywood *et al.*, 2005). Entrainment has been studied at acoustic-prosodic level.

The earlier studies of prosodic entrainment are concerned with one or two aspects of prosody separately. Matarazzo and Wiens (1967) focus on adaptation (increasing or decreasing) in silence duration of an interviewer according to an interviewee's response time latency. Natale (1975) finds that subjects entrained to new intensity level when they were engaged in open-ended conversation from within-conversation intensity levels, and this entrainment increased over the course of a conversation. Street (1984) finds that interviewees converged towards their interviewers on response latency and speech-rate. Similarly, Giles *et al.* (1991) find that interlocutors tend to align on speech rate. Gregory *et al.* (1993) find by the measurement of pitch and intensity that similarity is greater in true conversations than in conversations simulated by splicing together utterances from speakers who do not actually interact. Ward and Litman (2007) find by the measurement of loudness and pitch through regression lines that students in tutorial dialogues converged to their tutor on max and mean amplitude and diverged on min pitch.

For instance, speakers are proved to converge in Speaking-rate (Giles *et al.*, 1991), subvocal frequency or amplitude contour (Gregory, 1990), vocal intensity (Natale, 1975), etc. over longer speaking. Different from these studies, the present research has accomplished relatively comprehensive analyses of convergence in prosody, which involve main features of duration, fundamental frequency and intensity.

The recent researches involve relatively comprehensive and thorough aspects of prosody. Levitan and Hirschberg (2011) investigate four acoustic and prosodic dimensions (including 8 features from the duration, fundamental frequency, intensity, and voice quality) in the research on entrainment at multiple levels. In this research, they find the difference in speakers' coordination with each other in these dimensions over the conversation as a whole as well as on a turn-by-turn basis, and they find the coordination between interlocutors improve in the course of the conversation over certain prosodic features. De Looze *et al.* (2014) measure three prosodic parameters including pitch range, voice intensity and articulation rate. In this research, they propose an automatic system for the capture of dynamic manifestation in prosodic entrainment, and they find that prosodic accommodation changes dynamically over the course of a conversation and across conversations, and that these dynamics inform about the naturalness of the conversation flow, the speakers' degree of involvement and their affinity in the conversation. Reichel *et al.* (2018) examined prosodic entrainment in cooperative game dialogs for new feature sets describing register, pitch accent shape, and rhythmic aspects of utterances, and found that features sets undergo entrainment in different quantitative and qualitative ways.

Other language components are also involved in the studies of entrainment. Speech preceding backchannels are evidence of continued engagement by one's dialogue partner. Levitan *et al.* (2011) examine the backchannels in the Columbia Games Corpus and find that speaking partners tend to use similar sets of backchannels, and that the similarity increases over the course of a dialogue. Filled pauses such as *uh* or *um* seem ideal candidates for studying performances of language in

communication because they are frequently used in spontaneous conversations, they are low information load, and they are subject to accommodation to the interlocutors (Beňuš, 2009). Beňuš *et al.* (2012) investigate immediately adjacent filled pauses produced by justices and lawyers during oral arguments of the Supreme Court. They find that lawyers' filled pauses are more similar to those of the justices who voted for them and to those of the justices who voted against them. Xia *et al.* (2023) set a hierarchical framework by three layers of conversations, turns and tone units in Mandarin, investigated prosodic entrainment at each level and compared the three, and found that the global and local entrainment exist independently, and local entrainment is more evident than global.

3.2 Studies of Entrainment and its Pragmatic Functions

Some studies have explored the relationship between entrainment and its pragmatic functions.

Some studies have explored the relation between entrainment and task success. Reitter and Moor (2007) find that task success can be predicted by lexical and syntactic entrainment in just five minutes of an interaction in the Map Task Corpus. In their recent research, Reitter and Moor (2014) has accomplished deeper study of entrainment and task success. They develop two predictions from The Interactive Alignment Model, and use two methods to quantify the known structural priming effects in the full inventory of syntactic choices. In this research, they find a positive correlation of long-term adaptation and a quantifiable task success measure, and propose that lexical and syntactic repetition are reliable and computationally exploitable predictors of task success. Nenkova *et al.* (2008) find that there is significant correlation between the degree of entrainment on high-frequency words and affirmative cues words and game score in the Columbia Games Corpus, and that such entrainment is significantly correlated to dialogue naturalness and smooth interaction flow. Friedberg *et al.* (2012) find that the students who exhibited high level of alignment in the use of project-related words became more similar after the experiments, while those who exhibited low level tended to become less familiar. Levitan *et al.* (2012) have examined acoustic/prosodic entrainment in Columbia Games Corpus according to four objective measures of dialogue success: the mean latency between turns, the percentage of turns that are interruptions, the percentage of turns that are overlaps and the number of turns in a task. The results that latency is negatively correlated with entrainment and overlaps and the number of turns is positively correlated prove that entrainment is highly related to task success. Thomason *et al.* (2013) show that the students who exhibited higher degree of entrainment on pitch features gained higher scores in tests of a tutoring dialogue system.

Some studies are concerned with entrainment and interaction outcome. Niederhoffer and Pennebaker (2002) consider the relationship between Linguistic Style Matching (LSM), the degree to which participants in conversation match each other in the use of word categories representing a variety of psychological and linguistic dimensions, and interaction outcome. They compare speakers' scores over 18 linguistic dimensions of language and find that in all cases, interlocutors exhibit significant LSM on both conversational level and a turn-by-turn level. Similarly, Taylor and Thomas (2008) have

accomplished the research on Linguistic Style Matching (LSM) and Negotiation outcome. Correlational analyses over 18 linguistic categories in the negotiations between police negotiators and hostage takers show that successful negotiations are associated with higher aggregate levels of LSM than unsuccessful negotiations. Ireland *et al.* (2011) have investigated whether similarity in interlocutors' use of function words (LSM) predicts outcomes for romantic relationships. On the basis of the analyses in two experiments, they find that LSM appears to reflect implicit interpersonal processes central to romantic relationships. Levitan *et al.* (2012) has examined the relation between entrainment and social behaviors. Four parameters (including *gives encouragement*, *conversation awkward*, *trying to be liked*, and *like more*) are set as variables for social behaviors. Through the correlation analyses, they find that entrainment is correlated with the most social variables for female-male pairs, and these correlations are also the strongest. Lee *et al.* (2014) has done research on prosodic entrainment in affective spontaneous spoken interactions of married couples. They quantify prosodic entrainment (one of aspects of interaction synchrony), especially over pitch and energy, in married couples' problem-solving interactions. A statistical testing in this research shows that higher values in entrainment over the features indicate highly positive affect, and these features are proposed to be effective predictors of session-level annotation on positive or negative affect.

Some studies are concerned with entrainment and social factors. The studies of entrainment and gender in dialogue are in relatively bigger number. Several theories of entrainment predict that females will entrain to a greater degree than males. The male dominance hypothesis (Bilous & Krauss, 1988) asserts that differences in speech between males and females can be attributed to women's subordinate social status. Speech Accommodation Theory (Giles *et al.*, 1987) claims that when a power imbalance exists between interlocutors, the less dominant or powerful speaker will converge more. Chartrand and Bargh (1999) posits that the perception-behavior link is the mechanism behind entrainment, and thus, women should entrain more than men, regardless of the gender of their conversational partner, because women are known to have greater perceptual sensitivity to vocal characteristics. Namy, Nygaard, and Sauerteig (2002) explain that female speakers are perceived to accommodate more in a shadowing task than male speakers in this way. Pardo (2006), on the other hand, finds that female pairs are less similar to each other than male pairs, and that functions outside the domain of perception appear to influence the degree of phonetic convergence. In English conversations, Levitan and her group finds (Levitan *et al.*, 2012) that female-male pairs entrain on every prosodic feature examined; in addition, the degree of entrainment on Intensity mean and max is greatest for female-male pairs; male pairs show the least evidence of entrainment, entraining only on Intensity mean, Intensity max, and syllables per second; all these results support the hypothesis that entrainment is less prevalent among males; their degree of entrainment on these features is also lower than that displayed by female or mixed gender pairs; female pairs entrain on all features except pitch mean, pitch max, and jitter. Xia and Ma (2016a, 2016b, 2019) found that in Mandarin conversation, mixed gender groups entrain on the greatest number of prosodic

features, and males entrain on the least. Xia *et al.* (2014) made a cross-linguistic comparison between Mandarin Chinese and English in prosodic entrainment and found the striking similarities over the number of prosodic features and the degree of prosodic entrainment.

In the study of entrainment and social dominance in dialogue, Worgan and Moore (2011) proposes that social dominance can be understood as an interaction affordance, revealing action potentials for each signaling participant, and can be detected as a feature of rapport not of the individual. This study finds that speakers adjust their speech to match the current dominant individual.

Some studies investigate the relation between entrainment and role in dialogue. Motivated by the hypothesis of *audience design* (Bell, 1984), which means speakers might choose to use expressions that they believe will enhance communicative success, entrainment has been explored between speaker and addressee. Galati and Brenna (2010) have explored to what extent speakers adapt to their addressee. They find lexically identical expressions by the same speaker are more intelligible to another group of listeners when the expressions have been addressed to new addressees than when they had been addressed to old addressees, and they conclude that speakers' entrainment in spontaneous discourse is driven as least in part by addressees. Relative studies also involve the conversations with more than two participants. Branigan *et al.* (2007) have investigated whether the linguistic behavior of participants in a dialogue is affected by their role within that interaction by three experiments, and this study finds that syntactic alignment is not limited to speaker addressee dyads, and that the prior participant role of the current speaker affected alignment, that is, prior addressees aligned more than prior side-participants.

4. Conclusion and Discussion

This study supplies a comprehensive review of research on linguistic prosody and conversation entrainment. It is found that the theories of Communication Accommodation Theory (CAT), The Chameleon Effect and Perception-Behavior Link, Interactive Alignment Model set main theoretical bases for the research of prosody and entrainment, and the empirical studies provide references in research design, methodologies, data analyses, etc. to the further studies in the similar fields. Therefore, this review is expected to supply references to the future researches on linguistic prosody and conversation entrainment.

Is alignment intentional or unconscious in conversation? The theories reviewed seem to stand in two sides for this question. CAT proposes goal-oriented mechanism, while on the other side, The Chameleon Effect, Perception-Behavior Link, and Interactive Alignment Model support unmediated mechanism. However, some other scholars propose the coexistence of both, and interlocutors make balance between these two mechanisms in different communicative contexts. Branigan *et al.* (2010) point out that alignment in a particular context may have both unmediated and mediated components. Haywood *et al.* (2005) propose that unmediated alignment might occur alongside mediated alignment

designed to enhance communicative effectiveness.

Although there are differences in the view of underlying mechanism for alignment in conversation, all these theories mentioned above have made influential contributions for the explanations of interlocutors' alignment, among which two points form solid foundation for the future research.

One point is the existence of entrainment in communication. Although the explanations come from different perspectives, agreement can be made on the existence of entrainment in communication among the theories mentioned above (CAT, The Chameleon Effect, Perception-Behavior Link, and Interactive Alignment Model). Based on this agreement, the present research aims to find out how prosodic entrainment is made in Mandarin conversation. The findings of this research are expected to confirm the existence of convergence in interaction from aspect of prosody, and to help people make deeper understanding of entrainment in communication.

The other is the multilevel exhibition of entrainment. The main claim Interactive Alignment Model has made is that interlocutors' alignment in conversation is realized at various levels of linguistic representations. According to this model, as one of linguistic representations, prosody is bound to exhibit alignment. Therefore, it is valid to do research on alignment at prosodic level to explore interlocutors' alignment in prosody, and aims to find out how people entrain by prosody in communication.

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

- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bavelas, J. B., Black, A., Lemery, C. R., MacInnis, S., & Mullet, J. (1986). Experimental methods for studying "elementary motor mimicry". *Journal of Nonverbal Behavior*, 10, 102-119.
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145-204.
- Beňuš, Š. (2009). Variability and stability in collaborative dialogues: Turn-taking and filled pauses. *Proceedings of the 10th INTERSPEECH* (pp. 709-799).
- Beňuš, Š., Levitan, R., & Hirschberg, J. (2012). Entrainment in spontaneous speech: The case of filled pauses in Supreme Court hearings. *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications* (pp. 793-797). Kosice, Slovakia.

- Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman, & B. Rimé (Eds.), *Fundamentals of Nonverbal Behavior* (pp. 401-432). New York: Cambridge University Press.
- Bernieri, F. J., Davis, J. M., Rosenthal, R., & Knee, C. R. (1994). Interactional synchrony and rapport: Measuring synchrony in displays devoid of sound and facial affect. *Personality and Social Psychology Bulletin*, 20, 303-311.
- Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviors of same- and mixed-gender dyads. *Language & Communication*, 8, 183-194.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355-387.
- Bourhis, R. Y., Giles, H., & Lambert, W. E. (1975). Social consequences of accommodating one's style of speech: A cross-national investigation. *International Journal of the Sociology of Language*, 6, 55-72.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition*, 75, B13-B25.
- Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, 104(2), 163-197.
- Branigan, H. P., Pickering, M. J., Pearson, J., & Mclean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, 42, 2355-2368.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482-1493.
- Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two Minds, One Dialog: coordinating Speaking and Understanding. In H. R. Brian (Ed.), *The Psychology of Learning and Motivation* (Vol. 53, pp. 301-344). Academic Press.
- Buller, D. B., LePoire, B. A., Aune, R. K., & Eloy, S. V. (1992). Social perceptions as mediators of the effect of Speaking-rate similarity on compliance. *Human Communication Research*, 19, 286-311.
- Byrne, D. (1971). *The Attraction Paradigm*. New York: Academic Press.
- Cappella, J. N., & Panalp, S. (1981). Talk and silence sequences in informal conversations: III. Inter-speaker influence. *Human Communication Research*, 7, 117-132.
- Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains in real-time sentence comprehension. *Journal of Memory and Language*, 47, 30-49.
- 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.
- Chartrand, T. L., Maddux, W.W., & Lakin, J. L. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The New Unconscious* (pp. 334-361). Oxford: Oxford University

Press.

- Chen, M., & Bargh, J. A. (1997). Non-conscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology, 33*, 541-560.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.
- Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun phrase structure. *Journal of Memory and Language, 49*, 214-230.
- Condon, W. S., & Sander, L. W. (1974). Synchrony demonstrated between movements of the neonate and adult speech. *Child Development, 54*, 456-462.
- Coupland, N., & Jaworski, A. (1997). Relevance, accommodation, and conversation: Modeling the social dimension of communication. *Multilingua, 16*, 235-258.
- Cummins, F. (2009). Rhythm as entrainment: the case of synchronous speech. *Journal of Phonetics, 37*, 16-28.
- 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.
- Dijksterhuis, A., & Bargh, J. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology, 33*, 1-40.
- Fowler, C. A., & Housum, J. (1987). Talkers' signaling "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489-504.
- Friedberg, H., Litman, D., & Susannah, B. S., & Paletz, B. F. (2012). Lexical entrainment and success in student engineering groups. In *Proceedings of 2012 IEEE, Spoken Language Technology Workshop (SLT)* (pp. 404-409). Miami, FL.
- Galef, B. G. (1988). Imitation in animals: History, definition and interpretation of data from the psychological laboratory. In T. Zentall, & B. G. Galef (Eds.), *Comparative Social Learning* (pp. 3-28). Hillsdale, NJ: Erlbaum.
- Galati, A., & Brenna, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35-51.
- Gallois, C., Ogay, T., & Giles, H. (2005). Communication Accommodation Theory: A look back and a look ahead. In W. B. Gudykunst (Ed.), *Theorizing About Intercultural Communication* (pp. 121-148). SAGE Publications.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study inconceptual and semantic co-ordination. *Cognition, 27*, 181-218.
- Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics, 15*(2), 87-105.
- Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation*. London: Academic press.

- Giles, H., & Smith, P. (1979). Accommodation Theory: Optimal Levels of Convergence. In H. Giles, St. Clair, & N. Robert (Eds.), *Language and Social Psychology*. Baltimore: Basil Blackwell.
- Giles, H., & Ogay, T. (2007). Communication Accommodation Theory. In B. B. Whaley, & W. Samter (Eds.), *Explaining Communication: Contemporary Theories and Exemplars* (pp. 325-344). Lawrence Erlbaum Associates, Inc.
- Giles, H., Taylor, D. M., & Bourhis, R. Y. (1973). Towards a theory of interpersonal accommodation through speech: Some Canadian data. *Language in Society*, 2, 177-192.
- 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.
- Goleman, D. (2006). *Social Intelligence: The New Science of Human Relationships*. Bantam.
- Gregory, S. W. (1990). Analysis of fundamental frequency reveals co-variation in interview partners' speech. *Journal of Nonverbal Behavior*, 14, 237-251.
- 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.
- Gudykunst, W. B. (1995). Anxiety, uncertainty management theory: Current status. In R. L. Wiseman (Ed.), *Intercultural Communication Theory* (pp. 8-58). SAGE Publications INC.
- Hartsuiker, R. J., Pickering, M. J., & Elena, V. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish/English bilinguals. *Psychological Science*, 15(6), 409-414.
- Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science*, 16(5), 362-366.
- Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, 40(2), 129-141.
- Heyes, C. M. (1993). Imitation, culture and cognition. *Animal Behavior*, 46, 999-1010.
- Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, 96(2), 127-142.
- Johnson-Laird, P. N. (1983). *Mental Models: Toward A Cognitive Science of Language, Inference and Consciousness*. Harvard University Press.
- Kooststra, G. J., Hell, J. G. V., & Dijkstra, T. (2010). Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language*, 63, 210-231.
- Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Journal of Psychological Science*, 14(4), 334-339.

- Lee, C. C., Katsamanis, A., Black, M. P., Baucom, B. R., Christensen, A., Georgiou, P. G., & Narayanan, S. S. (2014). Computing vocal entrainment: A signal-derived PCA-based quantification scheme with application to affect analysis in married couple interactions. *Computer Speech & Language*, 28(2), 518-539.
- Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, 14, 78-106.
- 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.
- Maurer, R. E., & Tindall, J. H. (1983). Effect of postural congruence on client's perception of counselor empathy. *Journal of Counseling Psychology*, 30(2), 158.
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects in the comprehension of referring expressions. *Journal of Memory and Language*, 49(2), 201-213.
- Meltzoff, A. N., & Moor, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198(4312), 75-78.
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, 22(1), 39-44.
- Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Personality and Social Psychology*, 21(4), 422-432.
- 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.
- Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies* (pp. 169-172). Ohio, USA.
- Neumann, R., & Strack, F. (2000). "Mood contagion": The automatic transfer of mood between persons. *Journal of Personality and Social Psychology*, 79, 211-223.

- Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21*, 337-360.
- Pardo J. S. (2006). On phonetic convergence during conversational interaction. *Acoustical Society of America, 119*(4), 2382-2393.
- Pickering, M. J., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169-226.
- Reichel, U. D., Beňuš, S., & Mady, K. (2018). Entrainment profiles: Comparison by gender, role, and feature set. *Speech Communication, 100*, 46-57.
- Reitter, D., & Moore, J. D. (2007). Predicting success in dialogue. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (pp. 808-815). Prague, Czech Republic.
- Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language, (76)*, 29-46.
- Richardson, M. J., Marsh., K. L., Isenhower, R. W., Goodman, J. R. L., & Schmdt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science, 26*(6), 867-891.
- Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance, 33*(1), 201.
- Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science, 1*(2), 305-319.
- Street, Jr., Richard, L., James, N., & Van Kleek, A. (1983). Speech convergence among talkative and reticent three-year-olds. *Language Sciences, 5*(1), 79-96.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 632-634.
- Taylor, J. P., & Thomas, S. (2008). Linguistic Style Matching and Negotiation outcome. *Negotiation and conflict management research, 1*(3), 263-281.
- Thomason, J., Nguyen, V. H., & Litman, D. (2013). Prosodic entrainment and tutoring dialogue success. *International Journal of Artificial Intelligence in Education, 750-753*.
- Turner L. H., & West, R. L. (2010). *Introducing Communication Theory: Analysis and Application* (4th ed.). New York, NY: McGraw-Hill.
- 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. (1969). Subject Speaking-rates as a function of interviewer behavior. *Language & Speech, 12*, 54-67.

- 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.
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31, 183-194.
- Worgan, S. F., & Moor, R. K. (2011). Towards the detection of social dominance in dialogue. *Speech Communication*, 53, 1104-1114.
- 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.
- Xia, Z. H., & Ma, Q. W. (2019). *Prosodic Entrainment in Mandarin Chinese Conversations: An Experimental Study*. Tongji University Press, Shanghai, China.
- 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* (pp. 65-69). Dublin, Ireland.
- Young, R. D., & Frye, M. (1966). Some are laughing; some are not—Why? *Psychological Reports*, 18, 747-752.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.