Phonetic Accommodation During Conversational Interactions: An Overview

Acomodación fonética durante las interacciones conversacionales: una visión general

Leonardo Barón-Birchenall[i, ii] (iD) (✉)

[i.] *Corporación Universitaria Minuto de Dios - UNIMINUTO; Bogotá; Colombia.*
[ii.] *Laboratoire Parole et Langage; Aix-en-Provence; France.*

Abstract

During conversational interactions such as tutoring, instruction-giving tasks, verbal negotiations, or just talking with friends, interlocutors' behaviors experience a series of changes due to their counterpart's characteristics and the interaction itself. These changes are pervasively present in every social interaction; most of them occur in the sounds and rhythms of our speech, which is known as acoustic-prosodic accommodation, or simply phonetic accommodation. The consequences, linguistic and social constraints, and underlying cognitive mechanisms of phonetic accommodation have been studied for at least 50 years, due to the importance of the phenomenon to several disciplines such as linguistics, psychology, and sociology. Based on the analysis and synthesis of the existing empirical research literature, in this paper, we present a structured and comprehensive narrative review of the qualities, functions, onto- and phylogenetic development, and modalities of phonetic accommodation.

Keywords: phonetic accommodation; speech; conversation; convergence; similarity; entrainment; synchronization; phonetics; social interaction.

Resumen

Durante las interacciones conversacionales, como dar una tutoría, brindar instrucciones, las negociaciones verbales o simplemente hablar con amigos, los comportamientos de las personas experimentan una serie de cambios, debido a las características de su interlocutor y a la interacción en sí. Estos cambios se encuentran presentes en cada interacción social y la mayoría de ellos ocurre en los sonidos y ritmos del habla, lo cual se conoce como acomodación acústico-prosódica o simplemente, acomodación fonética. Las consecuencias, las limitaciones lingüísticas y sociales, y los mecanismos cognitivos subyacentes a la acomodación fonética se han estudiado durante al menos cincuenta años, en virtud de la importancia del fenómeno para varias disciplinas como la lingüística, la psicología, y la sociología. A partir del análisis y de la síntesis de la literatura de investigación empírica existente, en este artículo se presenta una revisión narrativa estructurada y exhaustiva de las cualidades, las funciones, el desarrollo onto- y filogenético, y las modalidades de la acomodación fonética.

Palabras clave: acomodación fonética; habla; conversación; convergencia; similitud; arrastre; sincronización; fonética; interacción social.

Introduction

During conversational interactions between humans (or even between humans and machines), interlocutors' behaviors experience a series of changes due to their counterpart's characteristics and the interaction itself. This phenomenon, known as behavioral accommodation, is pervasively present in every social interaction we have and is an important subject for several areas of linguistics, psychology, sociology, and applied areas related to these fields.

The term phonetic accommodation, particularly, describes the tendency of humans to adapt their speech acoustics to each other during conversational interactions. This process of adaptation has important implications for communicative success, an increase of empathy and positive evaluations towards interlocutors, accent change and dialects formation, and even acquisition of the phonology and phonetics of a second language.

Some of the questions that we address in this paper include: (a) To which extent can we find the capacities that enable phonetic accommodation in the early stages of life or even in other species; (b) Which purposes related to linguistic development, as well as social behavior, serves phonetic accommodation; (c) Are phonetic accommodation processes influenced by the gender, the role in the conversation, or social characteristics of the interlocutors; (d) To which extent is phonetic accommodation an automatic process or a process that requires a certain degree of awareness; (e) Is phonetic accommodation a consistent phenomenon across different modalities (speaking rate, speech rhythm, etc.) and different types of conversational interactions.

Trying to answer these questions, we conducted an analysis and synthesis of the existing empirical research literature from the last 50 years. As far as we know, this is the most detailed and comprehensive review of phonetic accommodation up to date. The contents of the paper are organized as follows: In the second section, "Generalities About Linguistic Accommodation", we address the generalities about linguistic accommodation. In Sections 3.1 to 3.3, functions, automaticity, degree of awareness, and task difficulty matters related to phonetic accommodation are discussed. In Section 3.4 we consider the impact of the conversational role, social biases, and gender of the interlocutors during phonetic accommodation. Sections 3.5 and 3.6 are dedicated to the development of the behavioral and psychological mechanisms responsible for phonetic accommodation during the first months of life, as well as the possible existence of such kind of mechanisms in species other than humans. In Section 3.7 we discuss the empirical evidence regarding the different modalities of phonetic accommodation. Finally, we present the conclusions of the paper.

Given that in the literature about linguistic accommodation the terminology tends to get mixed up, which hinders the compilation and comparison of data, for this review we will use, when possible, the following homogenized definitions (based mostly on Louwerse *et al.*, 2012):

Accommodation: a phenomenon in which talkers alter diverse linguistic and paralinguistic features in response to specific characteristics of received stimuli.

Imitation: a form matching process that can occur immediately or after a determined period, both consciously and subconsciously. Unlike convergence or proximity, the increase or maintenance of the degree of similarity between the original and the imitated behavior is not relevant in this case.

Convergence: a symmetric or asymmetric increase of similarity of diverse linguistic and paralinguistic features of two or more individuals during an interaction (or one individual towards

a computerized interlocutor). This process is likely to occur unintentionally, but it can occur intentionally at times as well (see Section 3.2 for further details).

Proximity: the maintenance of a certain degree of similarity during a conversational interaction.

Generalities About Linguistic Accommodation

Accommodation between interlocutors has been studied concerning several linguistic features, including phonetic characteristics (detailed in this paper), segmental duration (e.g., Edlund *et al.*, 2009; Lee *et al.*, 2018; Pardo *et al.*, 2013), linguistic style (e.g., Manson *et al.*, 2013; Thomson *et al.*, 2001), syntactic complexity (Xu & Reitter, 2016), and lexical choices (e.g., Reverdy *et al.*, 2020; Ward & Litman, 2007). Thorough reviews may be found in Barón-Birchenall (2018), Bonin *et al.* (2013), Delaherche *et al.* (2012), De Looze *et al.* (2014), Louwerse *et al.* (2012), and Pardo (2006).

Linguistic accommodation may be influenced by social, cultural, and personal aspects, such as perceived social status, social biases, language background, perception of attractiveness, and gender (e.g., Babel, 2011; Louwerse *et al.*, 2012). It can occur during natural conversations (Pardo, 2006), or in response to both natural and manipulated recorded stimuli (Goldinger, 1998; Nielsen, 2011), even if the listener is not instructed to listen to them (e.g., Delvaux & Soquet, 2007). Additionally, speech accommodation may result in modifications of the phonetic repertoire of one or more speakers (Heath, 2014).

Typically, convergence has been treated as the default form of accommodation, but not converging with an interlocutor is also common, and may be due to different reasons, such as an infrequent behavior that might not provide enough exposure to allow the increase of similarity (Louwerse *et al.*, 2012). Such lack of convergence may be seen as a sign of creativity in linguistic choices: an attractive quality that would lead to a positive impression of the speaker (Schoot *et al.*, 2016).

Divergent communicative behaviors may as well serve to convey and reinforce social roles (which is known as *speech complementarity*). This would be especially true in contexts such as organizational hierarchies, which often present high expectations about appropriate behavior at different levels (Muir *et al.*, 2017).

Regarding the underlying architecture of the systems responsible for linguistic accommodation, it has been proposed the existence of several interconnected and multiple-functionally processes working at different levels, such as lexical, syntactic, and phonological (Pickering & Garrod, 2004). The degree to which these processes are linked and structured is not clear (Weise & Levitan, 2018). From another point of view, a cognitive central system would control accommodation processes at various levels, modulating different behavioral channels depending on the interactive context (a complete discussion in Louwerse *et al.*, 2012).

The architecture, functionalities, and interactions of the cognitive-behavioral system, or systems, driving linguistic accommodation are the subject matter of the two predominant theoretical models regarding the topic: the *communication accommodation theory* –CAT– (Giles *et al.*, 1991) and the *interactive alignment model* –IAM– (Pickering & Garrod, 2004; an overview of these models can be found in Barón-Birchenall [2018] and Ruch *et al.* [2017]; see also the *interpersonal synergies* concept [Fusaroli *et al.*, 2014] as a possible rationale for linguistic accommodation).

Although both the CAT and the IAM rely on extensive empirical support, they are quite different. Even if some authors have proposed that a combination of intentional-social factors, emphasized by the CAT, and automatic-unintentional conditions, emphasized

by the IAM, may explain the vocal accommodation phenomena (e.g., Babel [2012] and Pardo [2006]). The CAT belongs to the social psychology and sociolinguistics tradition and emphasizes the adaptive benefits of accommodation for survival and reproduction. It also entails the idea of a link between the perceived behavioral similarity of a person and the ascription of positive attributes to that person (Ruch *et al.*, 2017). In this sense, speakers may promote social approval and efficient communication by adapting to their interlocutors' communicative behavior (Levitan *et al.*, 2012).

On the other hand, the IAM belongs to the cognitive psychology and psycholinguistics tradition and emphasizes the causal mechanistic cognitive processes that result in accommodation (Ruch *et al.*, 2017). According to this model, mutual understanding in dialogue relies on a variety of interconnected adaptation processes that occur at multiple levels of linguistic representation, such as lexical, syntactic, and semantic. Alignment at these levels leads to the alignment of the speakers' mental representations of the things being mentioned, which, in turn, is the ultimate goal of a successful conversation (Xu & Reitter, 2016). Here, *alignment* is understood as "a state in which two or more dialogue partners have an identical (or at least highly similar) representation at a particular linguistic level" (Oben & Brône, 2015, p. 550).

Phonetic Accommodation

Phonetic accommodation may occur when a *model speaker* is presented via audio, as well as visually, during a lip-reading task (e.g., Gentilucci & Bernardis, 2007; Sanchez *et al.*, 2010; a model speaker is someone, usually a human, from whom natural or modified speech sounds are taken as a reference to imitate, or to interact with, during experimental tasks). In addition, speech modifications due to phonetic convergence can be abstracted from particular interactions and generalized across the speaker's linguistic system (Babel, 2011).

Nevertheless, complete phonetic convergence between talkers is impossible to reach, because even for a single speaker, two productions of the same speech segment are different in terms of phonetic detail. Thus, phonetic convergence between individuals tends to be graded, and its effects under experimental conditions are typically subtle (Nguyen & Delvaux, 2015; Pardo, 2006). Moreover, although instructions to imitate tend to lead to greater convergence in general terms (Clopper & Dossey, 2020), even experiments explicitly demanding impersonation do not attain a complete degree of phonetic convergence (Wretling & Eriksson, 1998).

Phonetic accommodation can occur with respect to several acoustic features of speech, such as speaking rate (or speech rate): the velocity of our speech, it can be expressed in different ways, including words per minute and syllables per second; fundamental frequency (or *f0*): the frequency at which vocal cords vibrate when we make certain speech sounds, it is perceived by the ear as *pitch*; and vocal intensity: the amplitude of the vibrations of the vocal cords when we speak, it is perceived by the ear as *loudness* and it is referred to as *volume*.

In addition, phonetic accommodation can occur with respect to minimal speech features, such as the patterns of sounds that constitute vowels (vowel spectra), or the time gap between the release of certain speech sounds and the beginning of the vocal cords' vibrations (voice onset time), as well as with respect to wider features, such as the speech rhythm, or rhythms (this is a tough one, but generally speaking speech rhythm is related to the arrangement and emphasis of the sounds of our voice; it is not clear how

it works, but we try to shed some light in "Rhythmic Accommodation"; see also Turk & Shattuck-Hufnagel [2013]).

All the speech features just mentioned, and more, each contributes to the process of accommodation and can be individually affected during conversations. However, Babel and Bulatov (2011) suggest that there is not a single acoustic feature in the speech that serves as the only, or primary, imitable feature.

Lastly, phonetic accommodation can occur with respect to different accents (distinctive modes of pronunciation of a specific language) or different dialects (particular forms of a specific language, characteristic of determined regions or social groups). Even if the distinction between these two is not totally clear, phonetic accommodation between speakers with different accents has been more studied than accommodation between speakers with different dialects.

Functions

Generally speaking, imitative behaviors may improve interpersonal exchanges by increasing affiliation and empathy between interactional partners and supporting vicarious learning. Imitating other people's actions may also facilitate anticipation and understanding of such actions, particularly when their conveyed meaning is unclear (Adank *et al.*, 2010).

Regarding social functions, overall linguistic convergence between speakers contributes to communication success by facilitating sense-making, common goal attainment, and exchange of information, as well as preventing misunderstandings and establishing rapport and intimacy (Bonin *et al.*, 2013; Borrie *et al.*, 2015; De Looze *et al.*, 2014; Levitan *et al.*, 2011; Reitter & Moore, 2007). In addition, converging during conversations may help interlocutors to define their identity by categorizing others and themselves into groups, and to establish a mutual comprehension by decreasing social distance (Lelong & Bailly, 2011).

Phonetic convergence, in particular, correlates with a positive evaluation towards the conversational partner, enables communication efficiency, and helps to establish common ground during interactions, reinforcing thus social affiliation (Kousidis *et al.*, 2008, 2009; Lee *et al.*, 2010; Louwerse *et al.*, 2012). From a listener's standpoint, the degree of *pulse* convergence between two interlocutors (but not *meter* convergence) has been associated with the degree of friendliness and social bonding between them (Polyanskaya *et al.*, 2019; understanding pulse as a regularly recurring acoustic event, and meter as a structured pattern of accentuation of groups of pulses).

Furthermore, positive correlations have been found between the amount of involvement in an interaction and the degree of synchrony between interlocutors' *f0* and voice intensity level (De Looze *et al.*, 2011; De Looze & Rauzy, 2011), between the degree of speaking rate convergence and the amount of cooperation during conversations (Manson *et al.*, 2013), and between the amount of convergence of *f0* and the degree of learning gain during interactions between students and computer tutors (Thomason *et al.*, 2013).

On the other hand, several linguistic functions have been associated with convergence between speakers. For instance, it is believed that phonetic convergence plays an important role in the acquisition of the phonology and phonetics of a second language. This process partly relies on the ability to reproduce foreign speech sounds, so individuals' different capacities to imitate speech may result in foreign accent differences in late L2 learners (Nguyen & Delvaux, 2015; Sancier & Fowler, 1997). Imitating the pronunciation of

sentences being listened may as well improve unfamiliar accent comprehension (Adank *et al.*, 2010).

Phonetic accommodation is also considered as one of the mechanisms responsible for channeling linguistic variation towards dialect formation, and eventually into language change (Nguyen & Delvaux, 2015). Particularly, accommodation during small-scale conversational interactions would influence population-level linguistic variations through the elimination of unpredictable grammatical alterations. Speakers who make variable use of a linguistic constituent would accommodate speakers who use the same constituent, particularly in specific grammatical contexts, rather than vice versa because for the latter "accommodate" means to violate grammatical rules (Fehér *et al.*, 2019).

Nevertheless, physiological or learned restrictions on articulatory movements may hinder convergence along incompatible phonetic features such as voice onset time (VOT) or stop closure duration. According to Heath (2015), in such cases, it should be expected divergence in at least one of the measured features. Ultimately, speech modifications due to divergent accommodation do not tend to persist beyond the interaction in which they are realized, so it is unlikely that divergent behaviors can generate stable language variations (Heath, 2014, 2015).

Converging during conversations that include task coordination may also function as a recovery device, by marking a point in time to which interlocutors can return if a communication breakdown occurs (Louwerse *et al.*, 2012). Besides, convergence may relieve the speaker's cognitive system of some of the burden of constantly computing the next behavior of her or his interlocutor during an interaction (Louwerse *et al.*, 2012).

From a different perspective on the specific functions of behavioral accommodation, "interpersonal coordination is not beholden to any single functional explanation, but can strategically adapt to diverse conversational demands" (Duran & Fusaroli, 2017, p. 1). Moreover, even if it is a common belief that convergence plays an important role in social interactions, understanding the phenomenon as a linear causation, the other way around is also a possibility. Cooperation and closer relationships, for instance, could increase the attention paid to the behavior of others, improving the representation of their motor behavior, and thus facilitating convergence (Koban *et al.*, 2017).

Automaticity and Degree of Awareness

In the opinion of Louwerse *et al.* (2012), in general terms, behavioral convergence is immediate and involuntary, rather than strictly intentional. It comprises different features in different channels, such as postural sway, eyebrow movements, and speech rate, which would be very difficult to control intentionally during a conversation. However, relations between speech perception and production are constrained by situational aspects that also influence the direction and magnitude of accommodation during a conversational interaction (Pardo, 2006, referring specifically to phonetic convergence). Consequently, an ongoing debate takes place over whether or not speech perception produces linguistically significant parameters (gestural, lexical, syntactic, semantic, and/or phonological), which lead automatically to imitation (Garrod & Pickering, 2004; Pardo, 2006; Pickering & Garrod, 2004, 2021).

In this respect, Goldinger (1998) proposed that the automaticity of imitation might rely on the structural and functional characteristics of episodic memory systems. "In such systems, frequency and repetition effects are an expected outcome, and the data patterns from shadowing imitation closely followed the predicted impact of frequency and repetition" (Pardo, 2013a, p. 2).

Furthermore, Koban *et al.* (2017) suggest that interpersonal spontaneous motor synchronization is a consequence of individual brains' interaction with each other, operating under a general optimization principle of neural computation. Motor behavior convergence between individuals would be thus computationally more efficient and energetically less costly (than the lack of it): therefore, it would arise automatically. As a result, greater optimization would improve coordination, and greater coordination would in turn promote optimization.

Empirically speaking, the automaticity of phonetic imitation has been observed in several studies, such as the one of Delvaux and Soquet (2007), in which the simple exposure of participants to a different regional dialect, without specifically asking them to imitate or even to listen to it, was enough to trigger imitation.

From another point of view, also assuming automaticity of accommodation, "research has suggested that prosodic adaptation [accommodation] is a subconscious method of achieving social approval and acceptance and is utilized to identify with a particular social group" (De Looze *et al.*, 2014, p. 13; see also Giles *et al.*, 1991; Chartrand & Bargh, 1999). As believed by Lakin and Chartrand (quoted by De Looze *et al.*, 2014, p. 14): "Accommodation would have become automatic over the course of human evolution, playing an important role as a necessary pre-requisite for communicating and for maintaining harmonious relationships within a group."

On the other hand, authors such as Heath (2014) and Koban *et al.* (2017) consider that behavioral convergence occurs both consciously and subconsciously. In this scenario, and under normal circumstances, people do not need to be aware of spontaneous convergence to occur. Still, the consequences of converging are consciously accessible, allowing a person, for example, to easily note that they are clapping along with the rest of the audience (Koban *et al.*, 2017).

Task Difficulty and Timing

It has been proposed that during conversations, overall behavioral coordination may take a few seconds to occur, and its effects may persist after the end of the interaction, perhaps to be carried to the next interaction (Louwerse *et al.*, 2012; Pardo, 2006). In particular, it has been found that phonetic accommodation may start during the first minutes of an interaction (Goldinger, 1998; Pardo *et al.*, 2010), and its effects may persist even up to a week after the initial exposure (Goldinger & Azuma, 2004).

In Kousidis *et al.*'s (2008) study, for instance, speakers were found to converge early during the interaction regarding voice intensity and speech rate. In the study conducted by Delvaux and Soquet (2007), only a couple of trials were necessary to obtain the imitation effect of a different regional dialect. Such effect was observable in the speakers' realizations up to 10 minutes after the last exposure to the stimuli.

Other than that, although multiple examples of phonetic convergence are presented in this paper, the amount of similarity between speakers related to certain linguistic features may remain stable rather than increase during the interaction. This was the case in Lelong and Bailly's (2011) study, in which no sign of an increase of resemblance was found regarding renditions of vowels.

Furthermore, considering that interlocutors pass through different phases during conversations (i.e., reflecting, arguing, giving feedback), the amount of phonetic resemblance between them may also fluctuate depending on their mental state and degree of involvement (De Looze *et al.*, 2014; Edlund *et al.*, 2009; Kousidis *et al.*, 2009). For

example, analyzing prosodic accommodation in Japanese dyadic telephone conversations, De Looze *et al.* (2014) found that the resemblance between interlocutors did not continuously increase or decrease over time. On the contrary, it varied several times during the conversations. Similar data indicating fluctuation during phonetic accommodation rather than linear increasing or linear decreasing have been reported for English speakers by De Looze and Rauzy (2011), and Vaughan (2011).

Role, Gender, and Social Biases

According to Louwerse *et al.* (2012), during instruction-giving tasks, asymmetry in roles tends to cause asymmetry in accommodation, because the instruction follower is more likely to do what the instruction giver has just done than vice versa. Additionally, role assignation may produce a social asymmetry because instruction givers know what the next subgoal of the assignment is, and they are likely to initiate subtasks and determine strategies.

Additionally, individuals in a low-power role are thought to be motivated to seek social approval from individuals in a high-power role. This would be obtained through behavioral modifications depending on specific situations, for example, a job interview or a courtroom situation (in line with this idea, see Muir *et al.*'s [2017] research on linguistic style accommodation during face-to-face interactions).

Accordingly, it is believed that the speech of individuals of lower social status tends to converge towards their interlocutor's speech if such interlocutor is considered to be of higher social status (Giles *et al.*, 1991). Partially in line with this idea, Gregory and Webster (1996) observed a higher level of *f0* proximity between a television host and higher-status guests versus lower-status ones during dyadic interviews. In addition, Hay *et al.* (1999) found that the ethnicity of Oprah Winfrey's guests in her television show influenced the phonetic implementation of /ay/ in different words uttered by Ms. Winfrey.

On the other hand, several studies have yielded mixed results regarding the influence of role and gender (sometimes both at the same time) on the process of phonetic accommodation. For instance, Namy *et al.* (2002) reported that during a *shadowing* (close repetition) task of isolated words, women converged with their male and female interlocutors more than men did, with respect to similarity judged by external listeners. In addition, women converged more with men than with other women, whereas men exhibited a similar degree of convergence with both sexes.

The results of Namy *et al.* (2002) just mentioned contrast with the ones of Pardo's (2006) study, in which pairs of male talkers converged more than pairs of females with respect to phonetic features during a map task (similar results in Pardo *et al.*, 2010). Likewise, Thomason *et al.* (2013) found a greater degree of loudness minimum and maximum convergence in verbal interactions of male vs. female students with a computerized tutor voice. Nonetheless, in Pardo (2006), instruction givers within female pairs converged towards receivers, but receivers did not converge towards givers. On the contrary, in male pairs, instruction receivers converged towards givers more than vice versa (Pardo, 2013a).

It is worth noting that roles assumed during asymmetrical tasks and conversations are not necessarily constant nor cause an asymmetrical convergence pattern. Support for this idea comes from a related linguistic field, particularly from Xu and Reitter (2016), who investigated the syntactic complexity of utterances during dyadic conversations. Results showed that the syntactic complexity of the topic leader's utterances decreased whereas the syntactic complexity of the topic follower's utterances increased.

Further information about the role of gender on phonetic accommodation has been provided by Lelong and Bailly (2011), who found a stronger effect of phonetic proximity between dyads of the same sex, particularly female-female, as opposed to mixed-gender dyads, regarding renditions of French peripheral oral vowels. Contrarily, Levitan *et al.* (2012) reported that female-male pairs converge the most, whereas male-male pairs converge the least, with respect to phonetic features in the context of a cooperative computer game.

Conversely, Kawasaki *et al.* (2013) did not find differences between men and women with respect to speech rhythm accommodation. Additionally, research from a related linguistic field also found no differences between men and women related to the preferential use of linguistic style during e-mail interactions (i.e., the proportion of adjectives, opinions, apologies, insults, and personal information, which predicts the gender of the person writing the message; Thomson *et al.*, 2001).

As for the influence of social biases on the process of phonetic accommodation, research has mostly focused on characteristics such as desirability, attractiveness, and social status. For instance, Natale (1975) found that persons who obtained high scores on a social desirability test were more likely to converge towards their interlocutors in terms of voice intensity level, compared to persons with lower scores on the same test.

Babel (2012), for her part, reports that the more attractive female participants rated a "White model talker", the more likely they were to imitate his vowels' renditions. On the contrary, the more attractive male participants rated the same model talker, the less likely they were to imitate his vowels. However, no significant relation between attractiveness and vowel imitation was found for the "Black model talker." Additionally, it was observed a higher amount of vowel convergence in the condition involving a visual image of the model talker as opposed to the condition with no image.

Babel (2010) also found that the more positive the implicit social biases toward a person's place of origin are, the more that person is imitated. In this study, the convergence of New Zealand participants towards an Australian talker in terms of vowel formant frequencies was positively affected by the implicit bias of the New Zealanders towards Australia.

Concerning attitudes and accommodation, Yu *et al.* (2013) reported that a positive attitude towards a male narrator, along with the personality trait of *openness*, correlated positively with the degree in which speakers imitated the narrator's extended VOT. Accordingly, Lewandowski and Jilka (2019) found that the personality traits *openness* and *neuroticism* have a positive impact on the degree of phonetic convergence between German speakers and English speakers during a task-oriented interaction.

Phylogenetic Development

Both the phylogenetic (related to a species) and ontogenetic (related to an individual organism) approaches to the development of the behavioral and psychological mechanisms responsible for phonetic (and overall linguistic) accommodation are crucial to understand the subject, even if they are easily neglected in the literature of the field.

The ability to coordinate movements or vocalizations, or both, with a shared, repeating interval of time, would have evolved from specific primate behaviors, such as the so-called *carnival display* (i.e., groups of chimpanzees engaged in a chaotic voice and movement exhibition; stomping, running, and slapping trees, without any explicit indication of inter-individual coordination; Merker *et al.*, 2009). In this scenario, the

human ability to coordinate in pairs, or groups, with a steady beat source of a sound, is seen as a refinement of an ancient connection between calls and movements already present among our hominoid ancestors. This ability may have evolved for purposes of mate attraction, by enabling the voice coordination needed for enhancing the signal directed to a distant partner.

In terms of empirical research, the existence of vocal accommodation between non-human species has hardly been investigated (Duranton & Gaunet, 2016). There are a few empirical studies of birds, monkeys, and bats. Several species of non-human primates, for instance, modify the structure of their calls in response to conspecifics' vocalizations indicating a degree of predation threat, environmental events, and spatial relations within the group, among others (Barón-Birchenall, 2016). Furthermore, long-term vocal accommodation in non-human primates has been observed during pair and group formation, apparently aimed to reinforce dyadic bonds and group identity (Ruch *et al.*, 2017).

Additional instances of vocal accommodation in the animal kingdom include the first vocalizations of certain bat pups, which gradually converge, in terms of the resemblance of acoustic features, toward adult-like calls during the first months of life (Prat *et al.*, 2015). Also, birds' subsongs (generic variations of a future song, similar to infants' babbling and mice ultrasonic vocalizations), gradually converge toward the song of the bird's tutor (an adult bird) in terms of acoustic features (Doupe & Kuhl, 1999; see also Arriaga *et al.*, 2012).

In terms of functionality, vocal coordination has an adaptive value for clusters of animals, increasing the effectiveness of protection against predators (Duranton & Gaunet, 2016). Rapidly matching acoustic signals allow the vocalizer to address individual conspecifics in a context where a signal can be directed at a multitude of listeners. Interestingly, in the animal kingdom, the timing of a response is a key factor in vocal matching. Whereas a prolonged interval between emissions may not be perceived as a response to the first signal, a hasty reply may be perceived as a sign of aggression. However, overlapping of the signal is not a common occurrence, and sometimes serves an affiliative purpose in birds' duet signing (King *et al.*, 2014).

Furthermore, it has been suggested that the function of vocal accommodation consisting in signaling social closeness or distance to a partner or a group, along with some level of vocal control, evolved before the emergence of the human language rather than being the result of it (Ruch *et al.*, 2017). From this standpoint, vocal accommodation is seen as a pre-adaptation that would have paved the way for language evolution. Crucially, the capacity for behavioral interactional synchrony could even be shared between humans, chimpanzees, bonobos, and macaques (see Yu *et al.*, 2018, and references therein).

Conversely, an advanced ability to imitate may have represented a major precursor of the evolution of the human language, as well as one of the main steps in the evolution beyond the great ape (MacNeilage, 1998). In the words of Fitch (2010, p. 163): "the capacity of human infants and children ... to imitate motor actions (as well as vocalizations) remains unparalleled in its richness, despite clear homologs in ape behavior."

Ontogenetic Development

Behavioral coordination between infants and their caregivers allows them to create and maintain a strongly attached relationship that is essential for the development of the child (Duranton & Gaunet, 2016). This process seems to rely on brain mechanisms operating by means of coupling coordinated rhythmic oscillators, such as the biological clock and heart rhythms (Feldman *et al.*, 2005; Trevarthen, 1998).

Chronologically speaking, a series of accommodation-related capacities, including imitation, coordination, and rhythmicity, develop early during infancy. Initially, a seemingly universal proto-language is created between mother and infant during their early interactions. Rhythmic and intonational modulations play a key role during these interactions, to the extent that even deaf mothers vocalize at some point to their deaf infants, although neither of them can hear the sound (MacNeilage, 1998).

Already within the first hours of life, newborns can imitate tongue protrusion two or three minutes after seeing the model, and also the protrusion of the lips, mouth opening, smiles, and an expression of surprise (Beebe *et al.*, 2003). Later, during the first days of life, neonates can detect the rhythm of adult speech and synchronize their movements with it (Condon & Sander, 1974), and their cries exhibit tonal contours similar to those of their mother tongue (Mampe *et al.*, 2009).

At about three months of age, infants begin to open and close their mouths and move their tongues while paying attention to the adult's face and voice during episodes of interaction that involve eye-to-eye contact and sometimes also voicing (Bloom, 1998). During these interactions, they learn to take turns in vocal exchanges and match their partner's gaze directions and facial expressions (Feldman *et al.*, 2005). Additionally, during the third month of life, infants alter the *f0* ratios of their utterances in such a way that the dyadic vocal exchanges with their mothers become tonally synchronized (Van Puyvelde *et al.*, 2015).

At roughly seven months of age, infants begin to babble, rhythmically opening and closing their mouths. From that point on, utterances will typically have a fixed rhythm (MacNeilage, 1998). Approaching 14 months of age, infants show more exploratory behavior (including gaze direction) and smile more toward adults who imitate them than toward adults who do not (Beebe *et al.*, 2003).

However, despite all of the above, the ability to synchronize with an external acoustic isochronous signal does not develop until late in infancy and becomes steady just until puberty (see Merker *et al.*, 2009, and references therein).

Modalities of Phonetic Accommodation

Accent/Dialect

Research on accent/dialect accommodation usually finds convergence between speakers, towards a model speaker, or towards a typical style of pronunciation. In a study by Evans and Iverson (2007), for example, the speech of young adults from Northern England was evaluated before and after moving to Southern England. Acoustic analyses showed that most participants altered their typically northern pronunciation of vowels towards a more southern accent. Perceptual analyses revealed an increasing resemblance towards the southern accent over time (interestingly, this is one of the few studies in which perceptual and acoustic evaluations yielded equivalent results; see also Clopper & Dossey, 2020).

Similar effects were found in French female talkers who were exposed to a different regional dialect via loudspeakers. After the exposition, they produced vowels significantly different from their typical realizations and significantly closer to the model speaker's realizations (Delvaux & Soquet, 2007).

It has also been reported that American English native speakers converged with both native English-speaking and Spanish-accented English-speaking model talkers regarding

f0, duration, and vowel spectra (Lewandowski & Nygaard, 2018). In this case, however, although participants did not exhibit differential convergence patterns regarding acoustic measures, perceptual assessments revealed that native English speakers converged more toward the non-native English models. In this respect, Clopper and Dossey (2020) argue that, while convergence on stereotyped variants tends to be avoided, converging towards a talker with a "non-prestigious" variety is not.

In contrast to the aforementioned studies, Aubanel and Nguyen (2010) carried out automatic recognition analyses of conversations between pairs of one Northern- and one Southern-French speakers, finding no compelling evidence of phonetic convergence between them. In this regard, it is worth noting that interlocutors who speak the same language and the same dialect may converge more than interlocutors who speak different languages or different dialects (Kim *et al.*, 2011).

Speaking Rate

A seminal study by Street (1984) on this subject found that persons being interviewed for 20 to 30 minutes converged with their interviewers in terms of speech rate. However, these results contrast with another series of studies from the same author in which no consistent patterns of convergence in speech rate were found (Putman & Street, 1984).

Moreover, speech rate convergence has been found between American English speakers spontaneously conversing during a cooperative game (Levitan & Hirschberg, 2011), between unacquainted English speakers during telephonic conversations (Cohen *et al.*, 2017), and between North American English speakers and *confederates* (someone who collaborate with the experimenters) during the reading of scripted dialogues (in this case, the confederate's rate influenced the participant more than the other way around; Schultz *et al.*, 2015). Additional evidence of speech rate convergence during English conversations can be found in Kousidis *et al.* (2008), Levitan *et al.* (2012), and Manson *et al.* (2013).

It has also been found that healthy individuals synchronize their speech rate with recorded stimuli of speech with abnormal rhythmic production parameters, increasing it in response to productions from individuals with hypokinetic dysarthria (fast speech rate), and decreasing it in response to productions from individuals with ataxic dysarthria (slow speech rate; Borrie & Liss, 2014).

In contrast, mixed patterns of speaking rate accommodation have been found between pairs of English-speaking students working together on a mathematical problem (face-to-face or via shared workspaces; Lubold & Pon-Barry, 2014), between American English adult speakers (Wynn & Borrie, 2020), and between adult speakers of Hebrew and a confederate (Freud *et al.*, 2018).

Further studies have also found mixed evidence in terms of speech rate accommodation (Wynn *et al.*, 2018, 2019). In these studies, the speech rate of typically developed adults converged towards the manipulated speech rate (slow/fast) of a female model speaker during a laboratory task. However, no evidence of convergence was observed in autistic adults and children, and typically developed children, with respect to the model speaker.

With respect to the perceptual assessment of speaking rate accommodation, Pardo and colleagues (Pardo, 2013b; Pardo *et al.*, 2010) examined the degree of *articulation rate* convergence between speakers using acoustic measures and AXB tasks. Note that for establishing the articulation rate of speech, the rate of syllable onsets within an utterance is measured while eliminating pauses and silences, which, in turn, removes important temporal information that contributes to speakers' speech rate and prosody (Schultz

et al., 2015). In both studies (Pardo, 2013b; Pardo *et al.*, 2010), convergence between speakers was detected by listeners during the perceptual task, but the acoustic analysis did not show talkers to converge.

As we have seen until this point, there is no conclusive evidence of a tendency of talkers to synchronize their speech rate. In fact, there is even evidence of divergence effects in articulation rate in spontaneous dialogues between unacquainted German speakers (Schweitzer & Lewandowski, 2013).

Nonetheless, speech rate convergence does not necessarily happen over contiguous sequences, but can rather have an extended influence over time (Duran & Fusaroli, 2017). For instance, an interlocutor can echo an increase of the speech rate made by the other speaker earlier during a conversation. Additionally, variations in speaking rate may be due to variations in the number of pauses in speech and their mean duration rather than to variations in the actual articulation rate (De Looze *et al.*, 2014). According to Bonin *et al.* (2013, p. 542), "while speakers' articulation rate is rather constant in nature, one may rather accommodate their speech in terms of pause duration." In consequence, the method used to establish the existence of speaking rate convergence, as well as the conceptualization of the phenomenon, may determine the actual degree of convergence that can be recognized (see Wynn & Borrie [2020] for a discussion about methodological issues related to convergence analyses).

In this regard, Edlund *et al.* (2009) analyzed the length of pauses (within-speaker silences) and gaps (between-speaker silences) during Swedish spontaneous dialogues. Results were inconclusive regarding the convergence of pause duration. Likewise, De Looze *et al.* (2011) did not find compelling evidence of convergence of pause duration during spontaneous conversations of English speakers. These results, however, contrast with the ones presented by Gregory and Hoyt (1982), indicating that participants in dyadic interviews tend to converge with respect to the duration of pauses in speech.

Regarding human-computer interactions and interactions in virtual environments, it has been found that users of spoken *dialogue systems* (automatized tools of information) adapt their speech rate to that of the system, maintaining a rate suitable for automatic speech recognition. Users show a preference towards a spoken dialogue system that adapts its speech rate to the user's speech rate as opposed to a non-adaptive system (Kousidis *et al.*, 2009).

In line with these findings, Casasanto *et al.* (2010) reported that Dutch speakers modify their speech rate towards the speech rate of a pre-recorded human model speaker within an immersive virtual reality environment. Likewise, using also a fast-slow speech approach, Bell *et al.* (2003) observed that Swedish speakers modify their speech rate towards the speech rate of an animated character in a simulated spoken dialogue system (note that Bell *et al.*'s experiment was a *Wizard of Oz study*, a type of study in which participants interact with a computer system under the impression that it is autonomous, but it is operated totally or partially by a concealed person).

Fundamental Frequency (f0)

There is some evidence of speakers synchronizing their *f0* during communicative interactions. For instance, convergence between male dyads' *f0* during American English and Egyptian Arabic interviews (Gregory *et al.*, 1993), and convergence of the average *f0* of English speakers during unrestricted conversations (Collins, 1998). In addition, it has been observed that, during spontaneous conversations, speakers synchronize the *f0* of their *backchannels* (short vocalizations or gestures indicating that a speaker is following

and understanding a conversation) with the *f0* of their interlocutors' preceding utterance (Heldner *et al.*, 2010; Levitan *et al.*, 2011).

However, most studies have shown inconsistent patterns of *f0* accommodation, mostly during informal and cooperative English conversations, including Kousidis *et al.* (2008), Levitan and Hirschberg (2011), Levitan *et al.* (2012), Lubold and Pon-Barry (2014), Manson *et al.* (2013), Vaughan (2011), and Ward and Litman (2007).

Vocal Intensity

Several studies have found evidence of convergence of vocal intensity in different contexts, including non-directive interviews of English speakers (Gregory & Hoyt, 1982; Natale, 1975), informal English conversations (Kousidis *et al.*, 2008), interactions between children and a virtual animated tutor (Coulston *et al.*, 2002), and human-human tutoring dialogs (Ward & Litman, 2007).

Conversely, as in the case of *f0*, mixed patterns of vocal intensity accommodation have also been reported. For instance, De Looze and Rauzy (2011) found that, during informal conversations, English speakers exhibited patterns of proximity between each other rather than convergence. Likewise, Kousidis *et al.* (2009) did not find consistent patterns of vocal intensity convergence between English speakers during cooperative dialogues. Additional studies presenting non-conclusive results regarding vocal intensity convergence include Levitan *et al.* (2012), Levitan and Hirschberg (2011), and Vaughan (2011).

Voice Onset Time (VOT) and Vowel Spectra

Some studies indicate that voice onset time (VOT) tends to synchronize between interlocutors, under certain conditions. Sancier and Fowler (1997), for instance, analyzed the VOT of voiceless stops in a bilingual speaker of English and Portuguese. After staying a couple of months in Brazil, the participant's VOT was shorter in both languages. Correspondingly, it was longer after staying a comparable amount of time in the USA. Likewise, Shockley *et al.* (2004) found evidence of convergence between American English speakers and a model speaker regarding lengthened VOT of word-initial voiceless stops. Sanchez *et al.* (2010), for their part, reported convergence between the VOT of persons shadowing visual speech tokens of a face articulating /pa/ syllables at two different rates and the VOT of the model.

Mixed evidence regarding VOT convergence has also been presented: Nielsen (2011), for example, found that participants imitated the VOT of a model speaker when the VOT was artificially extended but not when it was shortened. Furthermore, Yu *et al.* (2013) reported no overall effect of convergence between female and male participants' and a male narrator's VOT, and Heath (2014) found that only some of the participants in his experiment converged towards a model talker with artificially extended VOT.

Regarding vowel spectra, Babel (2012) reported convergence of vowel formants between male and female talkers and two male speakers during a word-shadowing task. These results contrast with evidence of a lack of convergence (Pardo *et al.*, 2010), and inconsistent patterns of convergence (Pardo *et al.*, 2012), of vowel spectra during conversational interactions. In the same vein, Clopper and Dossey (2020) found convergence on vowel fronting, but not on a particular reduction of a diphthong to a long vowel, between English speakers and a model talker with a different accent.

Moreover, in a series of experiments, Pardo *et al.* (2013) analyzed phonetic accommodation in shadowed monosyllabic words using a perceptual AXB task and acoustic

measures of vowel duration, *f0*, and vowel formants. None of the three analyzed parameters yielded a significant degree of convergence in the acoustic measurement, whereas convergence was detected in all three of them in the perceptual task.

Rhythmic Accommodation

Several factors must be considered regarding rhythmic accommodation, including the definition of rhythm, the role of rhythm in turn-taking accommodation, and the level of analysis of the phenomenon (i.e., syllables, sentences, paragraphs).

Taking a sentence-level approach, Späth *et al.* (2016) conducted an experiment on rhythmic accommodation with healthy persons and patients suffering from Parkinson's disease. The authors found that speech rhythm resemblance between participants and a model speaker is greater in sentences with a metrically regular structure (an entirely uniform alternation of stressed and unstressed syllables) as opposed to sentences with an irregular structure (a less regular succession of stressed and unstressed syllables), especially in individuals with Parkinson's.

Following Späth *et al.*'s (2016) approach to speech rhythm analysis, Barón-Birchenall (2018, 2022) observed that regular rhythmic sentences, arranged in accentual groups, generate a greater amount of resemblance between pairs of Spanish speakers in terms of rhythm and *f0* range, as opposed to irregular rhythmic sentences and sentences arranged in accentual feet. The author's findings indicate that during conversational interactions, both rhythmic regularity and phonological phrasing have an influence not only on the degree of resemblance between speakers but also on the average and the range of the interlocutors' *f0*.

Other studies have approached the phenomenon of rhythm accommodation by examining the speech rate. For instance, based on studies of the phonetics of imitation, Wretling and Eriksson (1998) suggest that speakers may be able to vary their speech rate to produce phrases or other major segments within a given period, but articulatory timing patterns at a more local level are more rigid and very difficult to change. Under this assumption, imitation of someone's speech, or convergence with another interlocutor in terms of metrical patterns, may occur mostly with respect to the overall rate of speech, whereas the relative durations of words or other minor segments would remain almost invariant for a given speech rate.

Taking a different approach, Kawasaki *et al.* (2013) had two human subjects, or a human subject and a machine, alternately pronounce letters of the alphabet. They measured both the duration of the pronounced sounds and the intervals between them. The analyses revealed a higher degree of convergence during the human-human tasks compared to the human-machine tasks (in which the machine pronounced the letters at a fixed interval). In a similar vein, Lelong and Bailly (2011) found a significant degree of speech rhythm convergence (based on syllabic durations) between renditions of French vowels uttered by two interlocutors involved in a turn-by-turn game.

For their part, McGarva and Warner (2003) approach the subject by defining *vocal activity rhythms* as a periodic fluctuation observed in on-off vocal activity during conversations. Following this approach, the authors analyzed vocal rhythm accommodation between dyads of English female speakers during informal conversations. Convergence between dyads was only found in some of the interactions, and it did not occur at the start of conversations, but rather gradually.

Rhythmic accommodation has also been analyzed by means of spectral measures, which include disfluencies and pauses, and "provide an account of syllable prominence, stressed and unstressed syllable variation and their distribution" (Rao & Smiljanic, 2011, p. 1662). Following this approach, Rao *et al.* (2013) conducted a study involving female dyads and male dyads reading English syllables and a short-written paragraph before and after performing an interactive map task. Regarding speech rhythm, male participants were more likely to converge with each other than female participants.

Regarding the role of the speech rhythm in turn transitions during conversations, Ten Bosch *et al.* (2005) conducted a study involving face-to-face and telephonic dyadic interactions. Results showed a significant correlation between the average duration of the between-turn pauses in the telephonic condition, which was interpreted as a form of accommodation between the members of the dyad. However, no further signs of accommodation were observed for the duration of pauses between utterances within turns, or during face-to-face interactions. In a subsequent study, Himberg *et al.* (2015) found that when two persons were creating stories together, in turns, one word at a time, their word rhythms were strongly *entrained* (understanding word-rhythm entrainment as a phase-locking of the temporal sequences of the interlocutors' words onset times).

Lastly, based on the analysis of twelve spontaneous dyadic English conversations, Mooney and Sullivan (2015) propose that rhythmically coordinated speech is more likely to occur during a transition of the speaker floor from one interlocutor to the other than during other parts of the conversation.

Further Types of Phonetic Accommodation

A non-exhaustive list of additional kinds of phonetic accommodation that have been investigated includes sentence duration, backchannels, pre-voicing, and pitch accent realization. For instance, Lee *et al.* (2018) tested four English-speaking dyads before, during, and after a cooperative maze navigation task. Speakers in three out of four dyads synchronized the duration of their sentences, whereas one dyad showed significant divergence. Levitan *et al.* (2011), for their part, found that interlocutors tend to use similar sets of backchannel-preceding cues increasingly over time during spontaneous dyadic conversations. Mitterer and Ernestus (2008) found that native speakers of Dutch converged towards a female model speaker with respect to the presence of pre-voicing in initial voiced stops of nonwords during a shadowing task. Lastly, Gessinger *et al.* (2021) reported that German speakers converge towards a natural model talker in terms of pitch accent realization also during a shadowing task.

Conclusions

For at least 50 years, linguistic modifications exhibited by interlocutors during conversational interactions have been examined. This phenomenon, known as accommodation (among other names), relies on a set of capacities developed during the earliest stages of life, which includes imitation, coordination, and rhythmicity. Analogously, the ability of two or more organisms to engage in group behavioral coordination with the help of an external signal can be found to some extent in animal species not necessarily closely related to humans.

One of the most studied types of accommodation has been the phonetic type, with respect to which mixed, and sometimes opposite, and ambiguous results arise. Such inconsistency is especially marked in acoustic and perceptual analysis comparisons, which

may be because the acoustic analysis can focus on a single feature, while perceptual evaluation tends to take a holistic approach to the signal.

Furthermore, attaining phonetic convergence may be hindered by speech features being markedly different from one interlocutor to another. Conversely, speech features that are already present in the behavioral repertoire of both interlocutors may ease and speed up convergence between them. In any case, the particular details of the methodological approach, as well as the study design used to empirically test phonetic accommodation, play an important role in the findings. Such findings, in turn, may differ depending on several contextual factors such as the participants' accents or dialects.

Phonetic accommodation serves diverse purposes and facilitates diverse processes related to linguistic development as well as social behavior, including improvement of social exchanges; increasing of cohesion, empathy, and positive evaluation towards interlocutors; communicative success; goal attaining; accent change; dialects formation; and acquisition of the phonology and phonetics of a second language.

Besides, even though convergence between speakers is believed to be the default form of accommodation, an increase in similarity between interlocutors is only one of the possible outcomes of an interaction. An increasing difference between speakers' behaviors, as well as the maintenance of a certain degree of similarity, among further outcomes, may also occur. Crucially, accommodation seems to be ephemeral in response to short-term interactions, hence, the adoption of new behaviors should not be expected from one or both interlocutors in such cases.

Apart from that, research results regarding gender differences in phonetic accommodation (including mixed versus same-sex dyads) are diverse, and in some cases, convergence must not be expected due to the great difference usually found between sexes and the consequent effort that would take to attain a middle ground (e.g., $f0$ mean).

Concerning the influence of social roles on phonetic accommodation, is somehow expected that, during dyadic interactions, the "less prominent" individual will become more similar to the "prominent" one than the contrary. However, such roles may vary during interactions, and depend also, to some degree, on how one interlocutor feels about the other. On top of that, nominal roles assigned or expected may not necessarily correspond to the actual assumed role.

Regarding the amount of automaticity and degree of awareness responsible for phonetic accommodation, some studies indicate that to some extent a strong link between sensory and motor processes allows non-conscious behavioral changes during conversational interactions. However, the influence of high cognitive functions and social factors on accommodation has also been observed. At present, this is one of the main distinctions between the two predominant theoretical explanations of linguistic accommodation.

In this respect, future research could benefit from a theoretical paradigm able to unify the interactive alignment model and the communication accommodation theory, explaining why accommodation may happen automatically and at the same time be influenced by contextual factors such as the ones analyzed here. Likewise, explaining the discrepancies between the results of acoustic and perceptual accommodation analyses would be a major advance in the field.

Acknowledgements

References

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, *21*(12), 1903-1909. https://doi.org/10.1177/0956797610389192

Arriaga, G., Zhou, E. P., & Jarvis, E. D. (2012). Of mice, birds, and men: The mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLoS ONE*, *7*(10), e46610. https://doi.org/10.1371/journal.pone.0046610

Aubanel, V., & Nguyen, N. (2010). Automatic recognition of regional phonological variation in conversational interaction. *Speech Communication*, *52*(6), 577-586. https://doi.org/10.1016/j.specom.2010.02.008

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437-456. https://doi.org/10.1017/s0047404510000400

Babel, M. (2011). Imitation in speech. *Acoustics Today*, *7*(4), 16-22. https://doi.org/10.1121/1.3684224

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177-189. https://doi.org/10.1016/j.wocn.2011.09.001

Babel, M., & Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, *55*(2), 231-248. https://doi.org/10.1177/0023830911417695

Barón-Birchenall, L. (2016). Animal communication and human language: An overview. *International Journal of Comparative Psychology*, *29*, 1-27. https://doi.org/10.46867/ijcp.2016.29.00.07

Barón-Birchenall, L. (2018). *Influence of sentence-level rhythmic regularity and phonological phrasing on linguistic accommodation during conversational interactions: The case of Spanish-speaking dyads* [Doctoral dissertation, Aix Marseille Université]. HAL. https://hal.archives-ouvertes.fr/tel-01993542

Barón-Birchenall, L. (2022). *Sentence-level rhythmic regularity and accentual-group phrasing increase linguistic resemblance between Spanish-speaking interlocutors*. Manuscript submitted for publication.

Beebe, B., Sorter, D., Rustin, J., & Knoblauch, S. (2003). A comparison of Meltzoff, Trevarthen, and Stern. *Psychoanalytic Dialogues*, *13*(6), 777-804. https://doi.org/10.1080/10481881309348768

Bell, L., Gustafson, J., & Heldner, M. (2003). *Prosodic adaptation in human-computer interaction* [Conference presentation]. 15th International Congress of Phonetic Sciences, Barcelona, Spain. http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_2453.pdf

Bloom, K. (1998). The missing link's missing link: Syllabic vocalizations at 3 months of age. *Behavioral and Brain Sciences*, *21*(4), 514-515. https://doi.org/10.1017/s0140525x98251260

Bonin, F., De Looze, C., Ghosh, S., Gilmartin, E., Vogel, C., Polychroniou, A., Salamin, H., Vinciarelli, A., & Campbell, N. (2013). Investigating fine temporal dynamics of prosodic and lexical accommodation. In *Proceedings of Interspeech 2013* (pp. 539-543). ISCA. https://eprints.gla.ac.uk/93668/1/93668.pdf

Borrie, S. A., & Liss, J. M. (2014). Rhythm as a coordinating device: Entrainment with disordered speech. *Journal of Speech, Language, and Hearing Research*, *57*(3), 815-824. https://doi.org/10.1044/2014_jslhr-s-13-0149

Borrie, S. A., Lubold, N., & Pon-Barry, H. (2015). Disordered speech disrupts conversational entrainment: A study of acoustic-prosodic entrainment and communicative success in populations with communication challenges. *Frontiers in Psychology*, 6, 1187. https://doi.org/10.3389/fpsyg.2015.01187

Casasanto, L. S., Jasmin, K., & Casasanto, D. (2010). *Virtually accommodating: Speech rate accommodation to a virtual interlocutor* [Conference presentation]. 32nd Annual Meeting of the Cognitive Science Society, Austin, United States. https://pure.mpg.de/rest/items/item_458220/component/file_529234/content

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. https://doi.org/10.1037/0022-3514.76.6.893

Clopper, C. G., & Dossey, E. (2020). Phonetic convergence to Southern American English: Acoustics and perception. *The Journal of the Acoustical Society of America*, 147(1), 671-683. https://doi.org/10.1121/10.0000555

Cohen, U., Edelist, L., & Gleason, E. (2017). Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *The Journal of the Acoustical Society of America*, 141(5), 2989-2996. https://doi.org/10.1121/1.4982199

Collins, B. (1998). *Convergence of fundamental frequencies in conversation: If it happens, does it matter?* [Conference presentation]. 5th International Conference on Spoken Language Processing, Sydney, Australia. https://doi.org/10.21437/icslp.1998-111

Condon, W. S., & Sander, L. W. (1974). Synchrony demonstrated between movements of the neonate and adult speech. *Child Development*, 45(2), 456-462. https://doi.org/10.2307/1127968

Coulston, R., Oviatt, S., & Darves, C. (2002). *Amplitude convergence in children's conversational speech with animated personas* [Conference presentation]. 7th International Conference on Spoken Language Processing, Denver, United States. https://doi.org/10.21437/icslp.2002-671

De Looze, C., & Rauzy, S. (2011). Measuring speakers' similarity in speech by means of prosodic cues: Methods and potential. In *Proceedings of Interspeech 2011* (pp. 1393-1396). ISCA. https://doi.org/10.21437/interspeech.2011-457

De Looze, C., Oertel, C., Rauzy, S., & Campbell, N. (2011). *Measuring dynamics of mimicry by means of prosodic cues in conversational speech* [Conference presentation]. 17th International Congress of Phonetic Sciences, Hong Kong. https://hal.archives-ouvertes.fr/hal-01705525/document

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. https://doi.org/10.1016/j.specom.2013.10.002

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012). Interpersonal synchrony: A survey of evaluation methods across disciplines. *EEE Transactions on Affective Computing*, 3(3), 349-365. https://doi.org/10.1109/T-AFFC.2012.12

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64(2-3), 145-173. https://doi.org/10.1159/000107914

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, 22, 567-631. https://doi.org/10.1146/annurev.neuro.22.1.567

Duran, N. D., & Fusaroli, R. (2017). Conversing with a devil's advocate: Interpersonal coordination in deception and disagreement. *PLoS ONE, 12*(6), e0178140. https://doi.org/10.1371/journal.pone.0178140

Duranton, C., & Gaunet, F. (2016). Behavioural synchronization from an ethological perspective: Overview of its adaptive value. *Adaptive Behavior*, *24*(3), 181-191. https://doi.org/10.1177/1059712316644966

Edlund, J., Heldner, M., & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. In *Proceedings of Interspeech 2009* (pp. 2779-2782). ISCA. https://doi.org/10.21437/interspeech.2009-710

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, *121*(6), 3814-3826. https://doi.org/10.1121/1.2722209

Fehér, O., Ritt, N., & Smith, K. (2019). Asymmetric accommodation during interaction leads to the regularisation of linguistic variants. *Journal of Memory and Language*, *109*, 104036. https://doi.org/10.1016/j.jml.2019.104036

Feldman, R., Mayes, L. C., & Swain, J. E. (2005). Interaction synchrony and neural circuits contribute to shared intentionality. *Behavioral and Brain Sciences*, *28*(5), 697-698. https://doi.org/10.1017/s0140525x0529012x

Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511817779

Freud, D., Ezrati-Vinacour, R., & Amir, O. (2018). Speech rate adjustment of adults during conversation. *Journal of Fluency Disorders*, *57*, 1-10. https://doi.org/10.1016/j.jfludis.2018.06.002

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147-157. https://doi.org/10.1016/j.newideapsych.2013.03.005

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8-11. https://doi.org/10.1016/j.tics.2003.10.016

Gentilucci, M., & Bernardis, P. (2007). Imitation during phoneme production. *Neuropsychologia*, *45*(3), 608-615. https://doi.org/10.1016/j.neuropsychologia.2006.04.004

Gessinger, I., Schweitzer, A., Andreeva, B., Raveh, E., Möbius, B., & Steiner, I. (2018). Convergence of pitch accents in a shadowing task. In *9th International Conference on Speech Prosody 2018* (pp. 225-229). ISCA. https://doi.org/10.21437/speechprosody.2018-46

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1-68). Cambridge University Press. https://doi.org/10.1017/cbo9780511663673.001

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251-279. https://doi.org/10.1037/0033-295x.105.2.251

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review*, *11*(4), 716-722. https://doi.org/10.3758/bf03196625

Gregory, S. W., & Hoyt, B. R. (1982). Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psycholinguistic Research*, *11*(1), 35-46. https://doi.org/10.1007/bf01067500

Gregory, S. 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. https://doi.org/10.1037/0022-3514.70.6.1231

Gregory, S. W., Webster, S., & Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language & Communication*, *13*(3), 195-217. https://doi.org/10.1016/0271-5309(93)90026-j

Hay, J., Jannedy, S., & Mendoza-Denton, N. (1999). *Oprah and /ay/: Lexical frequency, referee design and style* [Conference presentation]. 14th International Congress of Phonetic Sciences, Ber-

keley, United States. http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_1389.pdf

Heath, J. (2014). Accommodation can lead to innovated variation. *University of California, Berkeley Phonology Lab Annual Report*, *10*, 119-145. https://doi.org/10.5070/p78f86j4k3

Heath, J. (2015). Convergence through divergence: Compensatory changes in phonetic accommodation. *LSA 2015 Annual Meeting Extended Abstracts*, *6*, 1-4. http://journals.linguisticsociety.org/proceedings/index.php/ExtendedAbs/article/view/3002

Heldner, M., Edlund, J., & Hirschberg, J. (2010). Pitch similarity in the vicinity of backchannels. In *Proceedings of Interspeech 2010* (pp. 3054-3057). ISCA. https://doi.org/10.21437/interspeech.2010-58

Himberg, T., Hirvenkari, L., Mandel, A., & Hari, R. (2015). Word-by-word entrainment of speech rhythm during joint story building. *Frontiers in Psychology*, *6*, 797. https://doi.org/10.3389/fpsyg.2015.00797

Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., & Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Scientific Reports*, *3*, 1692. https://doi.org/10.1038/srep01692

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125-156. https://doi.org/10.1515/labphon.2011.004

King, S. L., Harley, H. E., & Janik, V. M. (2014). The role of signature whistle matching in bottlenose dolphins, *Tursiops truncatus*. *Animal Behaviour*, *96*, 79-86. https://doi.org/10.1016/j.anbehav.2014.07.019

Koban, L., Ramamoorthy, A., & Konvalinka, I. (2017). Why do we fall into sync with others? Interpersonal synchronization and the brain's optimization principle. *Social Neuroscience*, *14*, 1-9. https://doi.org/10.1080/17470919.2017.1400463

Kousidis, S., Dorran, D., McDonnell, C., & Coyle, E. (2009). *Convergence in human dialogues: Times series analysis of acoustic feature* [Conference presentation]. SPECOM 2009, St. Petersburg, Russia. https://arrow.tudublin.ie/dmccon/2/

Kousidis, S., Dorran, D., Wang, Y., Vaughan, B., Cullen, C., Campbell, D., McDonnell, C., & Coyle, E. (2008). Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues. In *Proceedings of Interspeech 2008* (pp. 22-26). ISCA. https://doi.org/10.21437/interspeech.2008-369

Lee, C. C., Black, M., Katsamanis, A., Lammert, A. C., Baucom, B. R., Christensen, A., Georgiou, P. G., & Narayanan, S. S. (2010). Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. In *Proceedings of Interspeech 2010* (pp. 793-796). ISCA. https://doi.org/10.21437/Interspeech.2010-287

Lee, Y., Danner, S. G., Parrell, B., Lee, S., Goldstein, L., & Byrd, D. (2018). Articulatory, acoustic, and prosodic accommodation in a cooperative maze navigation task. *PLoS ONE*, *13*(8), e0201444. https://doi.org/10.1371/journal.pone.0201444

Lelong, A., & Bailly, G. (2011). Study of the phenomenon of phonetic convergence thanks to speech dominoes. In A. Esposito, A. Vinciarelli, K. Vicsi, C. Pelachaud & A. Nihjolt (Eds.), *Analysis of verbal and nonverbal communication and enactment: The processing issue* (pp. 273-285). Springer. https://doi.org/10.1007/978-3-642-25775-9_26

Levitan, R., Gravano, A., & Hirschberg, J. (2011). Entrainment in speech preceding backchannels. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics. https://academiccommons.columbia.edu/doi/10.7916/D8DN4DFM/download

Levitan, R., Gravano, A., Wilson, L., Beňuš, Š., Hirschberg, J., & Nenkova, A. (2012). *Acoustic-prosodic entrainment and social behavior* [Conference presentation]. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Montréal, Canada. https://aclanthology.org/N12-1002.pdf

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of Interspeech 2011* (pp. 3081-3084). ISCA. https://www.isca- speech.org/archive/interspeech_2011/levitan11_interspeech.html

Lewandowski, E. M., & Nygaard, L. C. (2018). Vocal alignment to native and non-native speakers of English. *The Journal of the Acoustical Society of America*, *144*(2), 620-633. https://doi.org/10.1121/1.5038567

Lewandowski, N., & Jilka, M. (2019). Phonetic convergence, language talent, personality and attention. *Frontiers in Communication*, *4*, 1-19. https://doi.org/10.3389/fcomm.2019.00018

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multi-modal communication is synchronized. *Cognitive Science*, *36*(8), 1404-1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x

Lubold, N., & Pon-Barry, H. (2014). A comparison of acoustic-prosodic entrainment in face-to-face and remote collaborative learning dialogues. In *2014 IEEE Spoken Language Technology Workshop (SLT)* (pp. 288-293). IEEE. https://ieeexplore.ieee.org/document/7078589

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, *21*(4), 499-511. https://doi.org/10.1017/s0140525x98001265

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, *19*(23), 1994-1997. https://doi.org/10.1016/j.cub.2009.09.064

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, *34*(6), 419-426. https://doi.org/10.1016/j.evolhumbehav.2013.08.001

McGarva, A. R., & Warner, R. M. (2003). Attraction and social coordination: Mutual entrainment of vocal activity rhythms. *Journal of Psycholinguistic Research*, *32*(3), 335-354. https://doi.org/10.1023/A:1023547703110

Merker, B. J., Madison, G. S., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, *45*(1), 4-17. https://doi.org/10.1016/j.cortex.2008.06.011

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168-173. https://doi.org/10.1016/j.cognition.2008.08.002

Mooney, S., & Sullivan, G. C. (2015). Investigating an acoustic measure of perceived isochrony in conversation: Preliminary notes on the role of rhythm in turn transitions. *University of Pennsylvania Working Papers in Linguistics*, *21*(2), 128-135. https://repository.upenn.edu/pwpl/vol21/iss2/15

Muir, K., Joinson, A., Cotterill, R., & Dewdney, N. (2017). Linguistic style accommodation shapes impression formation and rapport in computer-mediated communication. *Journal of Language and Social Psychology*, *36*(5), 525-548. https://doi.org/10.1177/0261927x17701327

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, *21*(4), 422-432. https://doi.org/10.1177/026192702237958

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*(5), 790-804. https://doi.org/10.1037/0022-3514.32.5.790

Nguyen, N., & Delvaux, V. (2015). Role of imitation in the emergence of phonological systems. *Journal of Phonetics*, *53*, 46-54. https://doi.org/10.1016/j.wocn.2015.08.004

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132-142. https://doi.org/10.1016/j.wocn.2010.12.007

Oben, B., & Brône, G. (2015). What you see is what you do: On the relationship between gaze and gesture in multimodal alignment. *Language and Cognition*, *7*(4), 546-562. https://doi.org/10.1017/langcog.2015.22

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382-2393. https://doi.org/10.1121/1.2178720

Pardo, J. S. (2013a). Reconciling diverse findings in studies of phonetic convergence. *Proceedings of Meetings on Acoustics*, *19*(1), 060140. https://doi.org/10.1121/1.4798479

Pardo, J. S. (2013b). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, *4*, 559. https://doi.org/10.3389/fpsyg.2013.00559

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*(1), 190-197. https://doi.org/10.1016/j.wocn.2011.10.001

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, *72*(8), 2254-2264. https://doi.org/10.3758/app.72.8.2254

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, *69*(3), 183-195. https://doi.org/10.1016/j.jml.2013.06.002

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169-190. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge University Press. https://doi.org/10.1017/9781108610728

Polyanskaya, L., Samuel, A. G., & Ordin, M. (2019). Speech rhythm convergence as a social coalition signal. *Evolutionary Psychology*, *17*(3), 1-11. https://doi.org/10.1177/1474704919879335

Prat, Y., Taub, M., & Yovel, Y. (2015). Vocal learning in a social mammal: Demonstrated by isolation and playback experiments in bats. *Science Advances*, *1*(2), e1500019. https://doi.org/10.1126/sciadv.1500019

Putman, W. B., & Street, R. L. (1984). The conception and perception of noncontent speech performance: Implications for speech-accommodation theory. *International Journal of the Sociology of Language*, *46*, 97-114. https://doi.org/10.1515/ijsl.1984.46.97

Rao, G., & Smiljanic, R. (2011). *Effects of language, speaking style and age on prosodic rhythm* [Conference presentation]. 17th International Congress of Phonetic Sciences, Hong Kong. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Rao/Rao.pdf

Rao, G., Smiljanic, R., & Diehl, R. (2013). Individual variability in phonetic convergence of vowels and rhythm. *Proceedings of Meetings on Acoustics*, *19*(1), 060084. https://doi.org/10.1121/1.4800750

Reitter, D., & Moore, J. (2007). *Predicting success in dialogue* [Conference presentation]. 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic. https://era.ed.ac.uk/bitstream/handle/1842/4165/MooreJ_Predicting%20Success%20in%20Dialogue.pdf?sequence=1&isAllowed=y

Reverdy, J., Koutsombogera, M., & Vogel, C. (2020). Linguistic repetition in three-party conversations. In A. Esposito, M. Faundez-Zanuy, F. Morabito & E. Pasero (Eds.), *Neural approaches*

to dynamics of signal exchanges (vol. 151, pp. 359-370). Springer. https://doi.org/10.1007/978-981-13-8950-4_32

Ruch, H., Zürcher, Y., & Burkart, J. M. (2017). The function and mechanism of vocal accommodation in humans and other primates. *Biological Reviews*, *93*(2), 996-1013. https://doi.org/10.1111/brv.12382

Sanchez, K., Miller, R. M., & Rosenblum, L. D. (2010). Visual influences on alignment to voice onset time. *Journal of Speech, Language, and Hearing Research*, *53*(2), 262-272. https://doi.org/10.1044/1092-4388(2009/08-0247)

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421-436. https://doi.org/10.1006/jpho.1997.0051

Schoot, L., Heyselaar, E., Hagoort, P., & Segaert, K. (2016). Does syntactic alignment effectively influence how speakers are perceived by their conversation partner? *PLoS ONE*, *11*(4), e0153521. https://doi.org/10.1371/journal.pone.0153521

Schultz, B. G., O'Brien, I., Phillips, N., McFarland, D. H., Titone, D., & Palmer, C. (2015). Speech rates converge in scripted turn-taking conversations. *Applied Psycholinguistics*, *37*(5), 1201-1220. https://doi.org/10.1017/s0142716415000545

Schweitzer, A., & Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. In *Proceedings of Interspeech 2013* (pp. 525-529). ISCA. https://www.isca-speech.org/archive/interspeech_2013/schweitzer13_interspeech.html

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*(3), 422-429. https://doi.org/10.3758/bf03194890

Späth, M., Aichert, I., Ceballos-Baumann, A. O., Wagner-Sonntag, E. W., Miller, N., & Ziegler, W. (2016). Entraining with another person's speech rhythm: Evidence from healthy speakers and individuals with Parkinson's disease. *Clinical Linguistics and Phonetics*, *30*(1), 68-85. https://doi.org/10.3109/02699206.2015.1115129

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, *11*(2), 139-169. https://doi.org/10.1111/j.1468-2958.1984.tb00043.x

Ten Bosch, L., Oostdijk, N., & Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Communication*, *47*(1-2), 80-86. https://doi.org/10.1016/j.specom.2005.05.009

Thomason, J., Nguyen, H. V., & Litman, D. (2013). Prosodic entrainment and tutoring dialogue success. In H. C. Lane, K. Yacef, J. Mostow & P. Pavlik (Eds.), *Artificial intelligence in education* (pp. 750-753). Springer. https://doi.org/10.1007/978-3-642-39112-5_104

Thomson, R., Murachver, T., & Green, J. (2001). Where is the gender in gendered language? *Psychological Science*, *12*(2), 171-175. https://doi.org/10.1111/1467-9280.00329

Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. In S. Braten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 15-46). Cambridge University Press.

Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology*, *4*(1), 93-118. https://doi.org/10.1515/lp-2013-0005

Van Puyvelde, M., Loots, G., Gillisjans, L., Pattyn, N., & Quintana, C. (2015). A cross-cultural comparison of tonal synchrony and pitch imitation in the vocal dialogs of Belgian Flemish-speaking and Mexican Spanish-speaking mother–infant dyads. *Infant Behavior and Development*, *40*, 41-53. https://doi.org/10.1016/j.infbeh.2015.03.001

Vaughan, B. (2011). Prosodic synchrony in co-operative task-based dialogues: A measure of agreement and disagreement. In *Proceedings of Interspeech 2011* (pp. 1865-1868). ISCA. https://doi.org/10.21437/interspeech.2011-507

Ward, A., & Litman, D. (2007). Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education 2007* (pp. 57-60). http://d-scholarship.pitt.edu/23210/1/cpSlate.pdf

Weise, A., & Levitan, R. (2018). Looking for structure in lexical and acoustic-prosodic entrainment behaviors. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (vol. 2, pp. 297-302). Association for Computational Linguistics. https://doi.org/10.18653/v1/n18-2048

Wretling, P., & Eriksson, A. (1998). Is articulatory timing speaker specific? – Evidence from imitated voices. In P. Branderud & H. Traunmüller, *Proceedings of Fonetik 98* (pp. 48-51). Stockholm University. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.39.6122&rep=rep1&type=pdf

Wynn, C. J., & Borrie, S. A. (2020). Methodology matters: The impact of research design on conversational entrainment outcomes. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1352-1360. https://doi.org/10.1044/2020_jslhr-19-00243

Wynn, C. J., Borrie, S. A., & Pope, K. A. (2019). Going with the flow: An examination of entrainment in typically developing children. *Journal of Speech, Language, and Hearing Research*, *62*(10), 3706-3713. https://doi.org/10.1044/2019_jslhr-s-19-0116

Wynn, C. J., Borrie, S. A., & Sellers, T. P. (2018). Speech rate entrainment in children and adults with and without autism spectrum disorder. *American Journal of Speech-Language Pathology*, *27*(3), 1-10. https://doi.org/10.1044/2018_ajslp-17-0134

Xu, Y., & Reitter, D. (2016). Convergence of syntactic complexity in conversation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (vol. 2, pp. 443-448). ACL. https://doi.org/10.18653/v1/p16-2072

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PLoS ONE*, *8*(9), e74746. https://doi.org/10.1371/journal.pone.0074746

Yu, L., Hattori, Y., Yamamoto, S., & Tomonaga, M. (2018). Understanding empathy from interactional synchrony in humans and non-human primates. In L. D. Di Paolo, F. Di Vincenzo & F. De Petrillo (Eds.), *Evolution of primate social cognition* (vol. 5, pp. 47-58). Springer. https://doi.org/10.1007/978-3-319-93776-2_4