



## Forms, factors and functions of phonetic convergence: Editorial

## ARTICLE INFO

## Keywords:

Phonetic convergence  
Speech production  
Speech perception  
Computer modeling

## ABSTRACT

This introductory article for the Special Issue on Forms, Factors and Functions of Phonetic Convergence offers a comprehensive overview of the dominant theoretical paradigms, elicitation methods, and computational approaches pertaining to phonetic convergence, and discusses the role of established factors shaping interspeakers' acoustic adjustments. The nine papers in this collection offer new insights into the fundamental mechanisms, factors and functions behind accommodation in production and perception, and in the perception of accommodation. By integrating acoustic, articulatory and perceptual evaluations of convergence, and combining traditional experimental phonetic analysis with computational modeling, the nine papers (1) emphasize the roles of cognitive adaptability and phonetic variability as triggers for convergence, (2) reveal fundamental similarities between the mechanisms of convergence perception and speaker identification, and (3) shed light on the evolutionary link between adaptation in human and animal vocalizations.

## 1. Introduction

Speech communication permeates our daily experiences, often appearing deceptively simple. The process is, instead, remarkably complex, as demonstrated by the multitude of components incorporated into theoretical models to best illustrate the communication flow (see review in [Elleström, 2018](#)). Whether viewed as a linear ([Jakobson, 1960](#); [Shannon, 1948](#)), interactive ([Schramm, 1955](#)), or transactional process ([Barnlund, 1970](#)), speech communication and its success hinge on individuals' remarkable ability to adapt in perception and production to the specific sender-receiver combinations as well as to the contextual, environmental and channel-related characteristics of the interaction ([Burchfield et al., 2023](#); [Choi and Perrachione, 2019](#); [Lavan et al., 2019](#)).

Beyond adaptations driven by interlocutor, context, channel and environment, a subtler form of speech adjustments with yet a strong influence on communicative success involves modifying one's communicative behavior toward or away from that of dialogue partners or model talkers, across a broad spectrum of linguistic, paralinguistic, and extralinguistic features (cf. par. 1.3. Forms). Depending on the theoretical framework (cf. par. 1.1 Theoretical paradigms), the communicative dimensions under examination (cf. par. 1.3. Forms) and the computational approach to quantifying these adjustments in

communication (cf. par. 1.4. Computational Approaches), the phenomenon is variously referred to as communicative accommodation (e.g., [Giles et al., 1991](#)), alignment (e.g., [Pickering and Garrod, 2004](#)), entrainment (e.g., [Levitan and Hirschberg, 2011](#)), and synchrony or chameleon effect (e.g., [Chartrand and Bargh, 1999](#)). The ability of vocal modulations is also observed in animals, thus implying that accommodation may operate as an evolutionary bridge from non-human primate vocal communication to human speech (e.g., [Ruch et al., 2018](#); [Zürcher et al., 2021](#)). Beyond the spoken/vocal modality, accommodation is also observed in written communication, as discussed by [Adams et al. \(2018\)](#), and in sign-language communication, as explored by [Stamp et al. \(2016\)](#).

When it comes to interspeaker adjustments in speech acoustics (henceforth phonetic convergence), it is notable that the last 50 years have seen a proliferation of cross-disciplinary research on the nature of phonetic convergence, the functions it serves, the forms that it can take, and the factors affecting the degree and direction of acoustic adjustments. This is showcased by the sheer volume of publications, encompassing journal articles, special issues, books, book sections, and conference proceedings, devoted to the multifaceted aspects of phonetic convergence. According to Pubmed, for example, from 1973 (the year when the Accommodation Theory was first proposed by [Giles et al.](#),

<https://doi.org/10.1016/j.specom.2024.103142>

Received 3 July 2024; Accepted 13 September 2024

Available online 30 September 2024

0167-6393/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1973) to 2024, 103 journal papers have been published related to the topic of speech accommodation.<sup>1</sup>

Mendeley retrieved a total of 367 publications from the last 10 years (2014–2023) having “speech accommodation”, “phonetic convergence”, or “acoustic-prosodic entrainment” as keywords.<sup>2</sup>

This specific introductory special issue article firstly aims to offer an overview of the current scientific state-of-the-art in phonetic accommodation and convergence, in terms of theoretical frameworks, and methods for eliciting and measuring convergence, along with an explanation of factors modulating the degree and direction of inter-speaker acoustic adjustments. Secondly, we will introduce the current special issue and its relationship with past and present collections on the topic of phonetic accommodation. Then, we will present the nine papers featured in this collection, and their contributions to the field in terms of investigated factors, computational approaches, and measures of convergence. Finally, drawing on their findings and insights, we will outline new avenues to be explored in future studies.

### 1.1. Theoretical paradigms

Accommodation has been conceptualized in several ways, with different theoretical paradigms placing emphasis on cognitive, social, linguistic, and conversational factors, or a combination thereof, depending on the specific framework. There seems to be common consent across several approaches and theories that convergence in both linguistic and non-linguistic aspects of communication occur subconsciously, driven by an automatic priming mechanism. Within the realm of language, these theories suggest that perceiving a linguistic unit automatically triggers the production of that unit due to the inherent link between perception and production. In particular, The Interactive Alignment Model by Pickering and Garrod (2004) propose that in dialogues, this priming mechanism operates across multiple linguistic levels, including phonological, syntactic, and semantic levels. This allows language users to synchronize their representations and make predictions about forthcoming utterances, thereby enhancing mutual understanding. The view of linguistic and paralinguistic alignment as an automatic process finds support in a series of studies demonstrating convergence in non-interactive contexts (e.g., shadowing recorded speech) where participants are not instructed to imitate the model speaker or explicitly told to refrain from imitation (Dufour and Nguyen, 2013; Goldinger, 1998; Shockey et al., 2004; Walker and Campbell-Kibler, 2015). A comparable unconscious link between perception and production is thought to regulate alignment also in non-linguistic behaviors such as body posture, mannerisms, and facial expressions (cf. the chameleon effect described by Chartrand and Barg, 1999). The authors suggest that “merely perceiving another’s behavior automatically increases the likelihood of engaging in that behavior oneself” (Chartrand and Barg, 1999:893). Nevertheless, given the assumption of automaticity, this paradigm does not adequately address cases of asymmetric convergence across speakers (where one speaker adjusts more than the other), unimodal convergence (where only certain features on certain linguistic levels are adjusted) or cases of divergence

<sup>1</sup> For replication purposes, the search query was made with the following keywords and logical function (((Phonetic Convergence[Title/Abstract]) OR (Speech Convergence[Title/Abstract]) OR (Speech Accommodation [Title/Abstract]) OR (Phonetic Accommodation [Title/Abstract]) OR (Vocal Accommodation[Title/Abstract]) OR (Acoustic Convergence [Title/Abstract]) OR (Speech imitation Title/Abstract]) OR (Acoustic entrainment [Title/Abstract]) OR (Acoustic prosodic Entrainment [Title/Abstract])).

<sup>2</sup> Detailed statistics: 164 publications on the topic of phonetic convergence (138 journals; 17 book sections; 8 conference proceedings; 8 generics, 1 book); 174 publications on the topic speech accommodation (138 journals; 17 book sections; 8 conference proceedings, 8 generics, 1 book); 29 publications on the topic of acoustic-prosodic entrainment (8 journal papers; 21 conference proceedings.

(where the similarity to the interlocutor decreases).

An alternative theoretical approach to the Interactive Alignment Model is offered by the Communication Accommodation Theory (e.g., Giles et al., 1991; Giles and Ogay, 2007). According to this framework, accommodation serves two main functions, regulating social distance, affiliation to social groups and managing comprehension. The directions of inter-speaker fluctuations are convergence, divergence, and maintenance. Convergence and divergence refer to adjusting one’s communicative behaviors to be more similar or dissimilar to one another, respectively. Maintenance refers to “sustaining one’s default level of communicating, without adjusting for others” (Dragojevic et al., 2016:37). Among the three directions of accommodation, the default accommodative move is convergence, which participants use to indicate cooperation in interaction (e.g., Pellegrino and Dellwo, 2023), and which holds significant implications for accent changes and dialect formation (e.g., Nguyen and Delvaux, 2015; Pinget, 2022). Non-accommodative strategies such as divergence and maintenance can be deliberately employed in situations where individuals seek to reinforce their personal or social identity, intentionally create communication difficulties, or counteract their interlocutor’s extreme speech patterns (e.g., Giles et al., 1991). Non-deliberate limitations to cooperative accommodation are also observed among interlocutors with differing language backgrounds (e.g., native versus nonnative speakers; Costa et al., 2008; Olmstead et al., 2021; Tobin, 2022), and with atypical neurological conditions (e.g., dysarthria; Borrie et al., 2015) or neuropsychological predispositions (Yu et al., 2013).

The complexity of explanations of accommodation does not seem to suggest a binary choice between automatic mechanisms and social factors. Numerous hybrid approaches have therefore been proposed. According to this perspective, social, linguistic, cognitive, and personality factors function as modulators of accommodation, affecting the strength or weakness of internal cognitive connections between perception and production (e.g., Babel, 2012; Pardo, 2012; Pardo et al., 2017; Ross et al., 2021; Wagner et al., 2021). Conversational dynamics are given particular emphasis in the hybrid model of Interpersonal Synergy (Fusaroli et al., 2014; Fusaroli and Tylén, 2016). In their view of dialogue as interpersonal synergy, interlocutors couple and constrain their interactional patterns in multiple and variable manners, where reciprocal imitation is just one of the possible strategies.

Reciprocal imitation is thus reconceptualized “as part of a complex process in which interactional patterns are jointly curbed and shaped by situational and task constraints” (Fusaroli et al., 2014:174).

### 1.2. Elicitation methods

Phonetic convergence has been elicited and noticed in diverse communicative situations along the interactive-non-interactive continuum. Examples of employed interactive laboratory tasks aimed at studying short-term accommodation are the map task (e.g., Pardo et al., 2010; Ulbrich and Canzi, 2023; Weise et al., 2020), diapix task (e.g., Ruch, 2021), maze navigation task (e.g., Lee et al., 2018; Lee et al., 2021), picture or word matching games (e.g., Biro et al., 2022; Kalmanovitch, 2016), rhyme game (Lelong and Bailly, 2011), dominoes game (Pellegrino and Dellwo, 2023), building Lego constructions (Abel and Babel, 2017), mock police interviews (Earnshaw, 2021), semi-spontaneous interactions (e.g., Kachel et al., 2024; Ludusan and Wagner, 2022), and mother-infant free-play (Van Puyvelde et al., 2015). Forms of medium- or long-term convergence have been examined in real-life interactive scenarios including married couples’ problem-solving interactions (Lee et al., 2010) and TV series (Sonderegger et al., 2017). Furthermore, evidence of convergence is observed in non-interactive scenarios such as shadowing and imitation tasks (e.g., Pardo et al., 2017, 2018) and synchronous reading tasks (Cerdeña-Oñate et al., 2021).

### 1.3. Forms of convergence

Accommodation is multidimensional. Evidence of convergence has been found in various linguistic and extra-linguistic features [for syntax, e.g., Branigan et al. (2000); for lexicon (e.g., Bell, 2002); laughter (e.g., Ludusan and Wagner, 2022), facial expressions (e.g., Lakin, 2013), and body movements (e.g., Dijksterhuis and Bargh, 2001)]. In terms of phonetic convergence, it is hardly possible to identify a feature that has not been the object of investigation. A non-exhaustive inventory of segmental and suprasegmental features examined in accommodation research includes: vowel formants, voice onset time, voicing contrasts, clicks, mel-frequency cepstral coefficients, speech rate, long-term average spectra, pause rate, utterance duration, fundamental frequency, rhythm, and voice quality. References to studies on individual features are provided in nearly all papers in this collection. For a summary of research on speaking rate, fundamental frequency, vocal intensity, voice onset time, vowel spectra and rhythm, cf. Barón-Birchenall (2023). Apart from analyzing speech acoustics, investigations into accommodation during real-time interactions have expanded to include articulatory data (Lee et al., 2018; Lee et al., 2021) and EEG data for pinpointing neural oscillatory indicators of phonetic convergence (Mukherjee et al., 2019).

### 1.4. Quantitative approaches

Accommodation has been quantified in numerous ways, with one of the most commonly used methods being the *Difference in Distance* paradigm. This involves comparing the distance in a specific feature between speakers before exposure to an interaction partner or model speakers (baseline distance) and during or after the exposure phase. A decrease in post-exposure distance indicates convergence, while an increase suggests divergence, and no change implies maintenance in vocal behavior. This paradigm, however, has known limitations, particularly regarding the representativeness of baseline production for speakers' general behavior and the influence of initial distance on quantification of convergence. Large baseline distances often result in evidence of convergence, while smaller distances may lead to maintenance or divergence being incorrectly inferred. In addition, accommodation is a highly multi-dimensional phenomenon. Thus, it is not necessarily clear, whether the dimension(s) that are being monitored, are representative for the observed direction of accommodation or whether the actual accommodating takes place in a complex interaction of different dimension. An alternative approach, proposed by Cohen Priva and Sanker (2019), is the linear combination approach, which models speaker behavior by considering subject and interlocutor performance outside of the conversation, along with additional variability factors. Other approaches Levitan and Hirschberg, 2011) distinguish between proximity (similarity of a feature over the entire conversation), convergence (degree of increased similarity over the conversation), and synchrony (turn-by-turn coordination between partners). These measures can be calculated globally by comparing interspeaker behavior between the dialogue's initial and final parts, or locally by analyzing conversational turns (Levitan and Hirschberg, 2011). A comprehensive framework has been proposed by Wynn and Borrie (2022), who categorize the major entrainment classes (proximity and similarity) along global/local and dynamic/static axes, providing a unified framework to study accommodation phenomena. Evaluations of phonetic convergence are also conducted in perception, holistically through AXB or XAB tasks, where naïve raters are asked to judge the similarities between the model speaker (X), and the flanking items (A and B) consisting in baseline utterances and utterances produced during or after the conversation (e.g., Dias et al., 2021). In much research a combination of acoustic and perceptual measures of convergence is used (e.g., Pardo, 2013; Ross et al., 2021; Wagner et al., 2021). Another difficulty related to all approaches discussed in this section is the concept of the baseline. Given that speakers constantly interact with others, it is unclear whether the

speech obtained prior to an interaction during an experiment is a baseline representing a speaker's non-accommodated speech productions or whether it represents speech accommodated to a previous interactant (e.g. an experimenter) or a social situation (e.g. a formal experimental setup).

### 1.5. Factors

Many variables affecting speech communication have been examined also regarding their ability to prompt or hinder phonetic convergence. In this paragraph, we provide a critical summary of speaker characteristics, conversational or task-related elements, and linguistic aspects that have been shown to modulate convergence. In terms of speaker-specific factors, research has focused on the speakers' sex (Namy et al., 2002; Weise et al., 2020), age (Nielsen, 2014; Paquette-Smith et al., 2022; Van Puyvelde et al., 2015), visual attractiveness and likeability (Michalsky and Schoormann, 2017, 2021), as well as physical and mental states (for healthy vs Parkinsonian disease, cf. Borrie and Liss (2014), Späth et al. (2016); for personality traits cf. Yu et al., 2013). As for the effect of the speakers' sex, the usual narrative is that female speakers tend to converge more than male speakers (Namy et al., 2002), but gender and role in interaction (information giver vs information receivers) have shown to interact in a complex fashion (Pardo et al., 2017). Considering the influence of age, multiple lines of evidence show that younger individuals are more inclined to converge compared to adults (Nielsen, 2014; Smith et al., 2007; Tagliamonte and Molfenter, 2007). This inclination may stem from their greater plasticity and stronger integrative motivations, making them more receptive to the pronunciation norms prevalent in their primary linguistic environments. However, when the effect of age was examined in conjunction with other variables, such as the accent of model speakers, the age differences in accommodation diminishes (Paquette-Smith et al. 2022). The presence of pitch synchrony and pitch imitation in mother-infant dialogues (van Puyvelde et al. 2015) introduces additional complexity and reinforces the view that accommodation cannot be examined without considering individuals' communicative intentions. Regarding the effect of perceived attractiveness and likeability, studies have shown that the degree of phonetic convergence depends on these evaluations but, again, the effect may vary from feature to feature, elicitation method (shadowing task vs conversations) and operationalization of convergence (global vs local, synchrony vs convergence) (Michalsky and Schoormann, 2017, 2021). Concerning the effect of physical states on convergence, research has shown that conversations disrupted by dysarthria exhibit noticeably lower levels of acoustic-prosodic entrainment compared to dialogues between healthy controls (Borrie and Liss, 2014; Späth et al., 2016). This suggests that the presence of pathological speech, marked by rhythmic disturbances that impede entrainment, contributes to a deficit in entrainment during face-to-face spoken interaction. In studies examining personality traits within the framework of the Big Five personality model, it has been observed that individuals with neurotic tendencies tend to exhibit a preference for convergence (Yu et al., 2013).

Among conversational/task factors, for instance, the dynamics of accommodation have been tested in relation to the pair composition in terms of gender (same/mixed gender pair), conversational role

(information giver vs information receiver), the progression of the conversation (initial vs final part of the conversation), and task instruction and difficulty. Information receivers tend to be more resistant to convergence than information givers (Pardo, 2006; Pardo et al., 2010), with female sex pairs maintaining their individual acoustic behavior to a higher degree than male sex pairs (Pardo, 2006; Pardo et al., 2010; for alternative results, Pardo et al., 2018). It has been shown that the degree of accommodation varies as the conversation unfolds (De Looze et al., 2014; Edlund et al., 2009; Tobin, 2022). Instruction to attend to the conversational partner's speech, and more engaging and less cognitively demanding tasks are also triggers of convergence (Abel

and Babel, 2017; Biro et al., 2022; Tausczik and Pennebaker, 2013).

In terms of language-related characteristics, convergence is shaped by the known effects of previous exposure to lexical items and their frequency characteristics. Previously heard lexical items and more frequent linguistic units are more susceptible to convergence-induced variations (Goldinger, 1998; Goldinger and Azuma, 2004). In cross-dialectal and cross-linguistics settings, larger geographical distribution of acoustic properties (Ross et al., 2021), dialect prestige (Ross et al. 2021; Ruch, 2021), and positive attitude toward the variety of the interlocutors (Babel, 2010) evoke convergence. The role of acoustic distance in speech acoustics between interlocutors or between shadowers and model talkers is more controversial, with findings pointing to both conducive and detrimental effects of large acoustic distance on convergence (e.g., Babel and Bulatov, 2012; Kim et al., 2011; Lin et al., 2021; Pellegrino, 2024; Ross et al., 2021).

## 2. The present collection

The present special issue, along with the one published in the Journal of Phonetics (Pardo et al. 2022), was motivated by the "Workshop on Accommodation in Speech Communication" held in Zurich, Switzerland, in December 2018 (<http://tiny.uzh.ch/RC>). The two special issues feature papers stemming from, but not restricted to, the workshop presentations. In line with the workshop's broad scope of topics, both collections adopted a heterogeneous approach to explore the forms, factors, and functions of phonetic convergence. Questions that both special issues aimed to address concerned:

- The underlying mechanisms and social functions of vocal accommodation.
- The influence exerted by sources of between-speaker variability (e.g., age, gender, cultural and language background, dominance in conversation) and of speech-extrinsic variability (i.e., channel variability, background conditions) on the degree and direction of phonetic convergence.
- The combination of articulatory, perceptual, neurocognitive, and/or multimodal data to the analysis of speech accommodation in interactive and non-interactive tasks.

Topics specific to the present special issue related to

- The effect of audio-visual information on acoustic, articulatory, and perceived convergence.
- Methodological issues for measuring and analyzing phonetic convergence.

(See Table 1 for a comparison of topics between the collection published in Journal of Phonetics and the current special issue). While

**Table 1**  
Common and issue-specific topics for the special issue in Journal of Phonetics (Pardo et al., 2022) and the current collection.

JPHON (Pardo et al., 2022)	Current special issue
Mechanisms and social functions leading to accommodation in speech perception and production	
The effect of task-specific and talker-specific characteristics on degree and direction of convergence in human-human and human-computer interactions	
Forms which convergence can take (acoustics, articulatory kinematics, perceptual assessments)	
The contribution of short/long-term accommodation in human-human and human-computer interactions to the diffusion of linguistic innovation and, ultimately, language variation and Change	Effect of audio-visual information on acoustic, articulatory, and perceived convergence
	Methodological issues for measuring and analyzing phonetic convergence.

the papers in the previous special issue (Pardo et al., 2022) had a narrower focus on phonetic and phonological features, the papers in the current collection adopt a comprehensive view of a wide variety of elements in speech communication, encompassing elements such as laughter, and incorporating data from various domains, such as acoustics, articulation and perception. Moreover, some of the papers combine results from traditional acoustic analysis alongside those derived from computational modeling.

Both collections, however, differ from previous collections on communicative accommodation (e.g., Coupland and Giles, 1988; Giles, 1984; Giles et al., 2015, 2023; Pardo et al., 2022; Yaeger, 1992) in that they primarily focus on phonetic convergence, referencing multiple theoretical paradigms to explain vocal convergence. Furthermore, the contributions featured in this collection have applied exclusively quantitative methods to measure accommodation in both interactive and non-interactive speech settings. Conversely, most papers in past special issues have evolved from the framework of the Communication Accommodation Theory and use this paradigm as a lens to investigate convergence and divergence in both oral and written communication, blending qualitative and quantitative analytical approaches.

The collection of papers was published as a 'Virtual Special Issue' (VSI), which means that papers belonging to the special issue are published immediately after their acceptance. Upon completion, the special issue is subsequently available as an online collection. This avoids waiting times for authors completing the review process faster. For this reason, the articles presented in the following sections appeared between 2021 and 24.

### 2.1. Presentation of individual papers

The nine original research articles compiled in the present collection mainly addressed the first three questions outlined in the special issue. In particular, some of them concentrate on mechanisms and conditions that enhance individuals' capacity to accommodate in speech acoustics and articulation (Lee et al., 2021) and in perception (Burchfield et al., 2023), and to perceive accommodation in speech (Dias et al., 2021). Two papers focus on task-related characteristics such as the co-presence of a live interlocutor (Cerdeña-Oñate et al., 2021) and the degree of engagement elicited by the experimental setting (Biro et al., 2022). In terms of speaker-specific factors affecting convergence. Kachel et al. (2024) delve into the effect of participants' sexual orientation and experimenters' sex on convergence.

Most of the papers in this collection investigated convergence in monolingual settings [e.g. Chilean Spanish (Cerdeña-Oñate et al., 2021); varieties of American English (Dias et al., 2021; Lee et al., 2021), German (Kachel et al., 2024)], but two have a cross-linguistic orientation, including interaction between speakers of two dialectal varieties of Swiss German (e.g., Zurich and Grison German dialects

- Ruch, 2021), of L1 and L2 varieties of Belfast English and Northern Standard German (Ulbrich and Canzi, 2023), or Chinese-English bilinguals (Burchfield et al., 2023). In the paper by Ludusan and Wagner (2022) the entrainment in laughter is examined cross-linguistically and cross-culturally in speakers of French, German, and Mandarin Chinese. Of the range of elicitation methods employed in accommodation research, this collection combines research applying traditional methods, like map task (Ulbrich and Canzi, 2023), diapix task (Ruch, 2021), maze navigation task (Lee et al., 2021), or spontaneous conversations (Kachel et al., 2024), with studies implementing synchronous reading tasks (Cerdeña-Oñate et al., 2021) or word matching tasks in a virtual game environment (Biro et al., 2022).

In terms of evaluation of convergence, the perception of convergent speech is examined using AXB similarity tasks (Dias et al., 2021), and accommodation in perception is studied using a lexical decision task (Burchfield et al., 2023). In production, a variety of segmental and suprasegmental measures have been extracted along with articulatory measures. Computationally, accommodation is quantified through

established methods (e.g., difference in distance, proximity/synchrony/convergence) or methods (e.g. coefficient of variation, Lee et al., 2021) tailored to test specific research hypotheses. Two papers combined traditional and computational modeling techniques [e.g., attunement model (Lee et al., 2021); agent-based modeling (Ruch, 2021)], hence permitting to touch upon the sixth topic of the special issues, i.e., methodological considerations for measuring and analyzing phonetic convergence (cf. par. 3. Future lines of research).

Aligned with the range of themes targeted by this special issue, the overview of papers begins with those that delve into questions pertaining mechanisms and social functions leading to accommodation in speech production and perception. Lee et al. (2021) offers an original approach to understanding accommodation that attributes real-time adaptive capability in production to three primary cognitive abilities: the variability in generating phonetic units, the direct correspondence between sensory and motor functions in the human facial and oral mechanisms, and the social pressure to align behavior with others. Authors employ an ‘attunement model’ to illustrate how real-time accommodation can arise from the individual ability of flexible adaptation. This model predicts patterns of convergence among interacting speakers, notably (a) that the speaker who naturally exhibits more variability will be the one to accommodate with their partner by demonstrating a marked reduction in their variability, and (b) that there will be a notable difference in their variability when speaking alone versus interacting with a partner. These hypotheses were also evaluated through real-time interactions between three pairs of speakers engaged in a maze navigation task and a corresponding solo speech task. Specifically, the authors evaluated whether the variability in acoustic and kinematic speech properties can indicate which speaker in a pair will align with the other. Acoustically, they measured the duration of sentences and the fundamental frequency at the end of utterances; in articulation, the time-to-peak velocity (TPV) of the tongue tip movement. Methods for quantifying convergence included assessing the coefficient of variation in these acoustic and articulatory measures to observe changes in variability from solo speaking to interaction for each maze task. The findings from both model simulation and empirical tests reinforce the assumptions that structured variability enhance speakers’ capacity for adaptation, which is crucial for convergence in speech accommodation. Specifically, the speaker within a dyad who inherently displayed more variability is identified as the one more likely to adapt to their partner. This speaker not only shows a significant decrease in variability during dialogue but also a pronounced difference in variability when comparing solo to interactive speaking scenarios.

According to Burchfield et al. (2023), variability, operationalized in terms of conversational experience, is the key ingredient to explain accommodation in perception (e.g., perceptual learning). In their exploration of factors contributing to listeners’ capacity to perceptually adjust phoneme categories when exposed to speech from unfamiliar talkers, the authors hypothesized that neither the listener’s native language (L1) nor specific task characteristics, but the size of the interlocutor set, play a crucial role in facilitating accommodation in perception. To investigate this hypothesis, the researchers tested two groups of bilingual speakers: (a) heritage language users, whose family language was Mandarin and who acquired English as their environmental language, and (b) international students, who had Mandarin as their first language (L1) and later acquired English as their second language (L2). Through a self-reported questionnaire assessing the frequency of language use (ranging from always to never) in various situational contexts (interactions with partners, relatives, friends, work settings, church, and shops), the authors found that the two groups differed in their language usage patterns. One of the two languages was reported to be used extensively in diverse social settings (English for the heritage speakers and Chinese for the international students), while the other was reported to be used in a more restricted range of social situations. Following a training phase where participants were taught to perceive an ambiguous sound as either [f] or [s], they underwent a post-test consisting of a

lexical decision task in both English and Mandarin. In this task, participants were required to determine whether the presented stimulus was a word or a nonword. Critical items included 40 disyllabic words containing the sounds /f/ or /s/ in the medial position, while the ambiguous sound was taken from an acoustic continuum reaching from a canonical /f/ to a canonical /s/. The findings revealed that heritage speakers, when exposed to an ambiguous English sound within lexically disambiguating contexts, exhibited the anticipated adjustment of phonemic boundaries during categorization. Conversely, for Mandarin, the opposite trend was observed: international students, but not heritage speakers, showed evidence of adaptation. In instances where learning did not manifest, participants reported a less frequent use of the respective language with fewer interlocutors. These results suggested that successful retuning in any language is contingent upon regular conversational interactions with novel talkers.

While the paper by Burchfield et al. (2023) investigated factors enhancing accommodation in perception, Dias et al. (2021) explored methods to improve the detection of convergence within speech. Based on insights from research indicating that (1) recognizing a speaker entails learning speaker-specific phonetic patterns, (2) the familiarity with speaker-specific phonetic attributes aids in processing speech from those particular speakers, (3) learning speaker-specific phonetic details can extend to previously unheard segments, the authors investigate to what extent training the raters to learn shared speaker-specific dimensions resulting from convergence has comparable effect on the detection of convergence. In two separate experiments, raters were trained to identify an utterance as a shadow of a previously heard model within the AXB task including feedback. The testing phase of Experiment 1 involved raters listening to the two trained models alongside four new shadowers. Conversely, in the testing phase of Experiment 2, raters evaluated the similarity of new shadowed utterances spoken by novel shadowers of new models. Results from Experiment 1 indicated that listeners could indeed enhance their detection of phonetic convergence with feedback, and this learning extended to novel words and shadowers. Given that this learning did not transfer when new models, together with new shadowers and new words, were introduced (Experiment 2), the overall findings of the two experiments were taken as evidence that during the training the raters were improving their ability to perceive the shared talker-specific phonetic dimensions of the model, rather than simply becoming more adept at detecting convergence irrespective of the talkers involved. The fact that this learning extended to new shadowers of the models and novel words suggested that flexibility in perception is essential to identify shared attributes across individuals with varying vocal tracts and dialects.

The papers described so far have tackled general questions related to production and perception of accommodation and accommodation in perception. With the following two papers, the attention is shifted to the effects of talker-specific information and task-specific factors on convergence. The study conducted by Kachel et al. (2024) expanded on a debated topic in accommodation research, viz. the effect of sex. They focused on the interaction between the sex of the experimenter and the sexual orientation of the study participants, self-assessed as primarily or exclusively lesbian or straight through a Kinsey-like scale, as a modulator of convergence. The specific form under study was fundamental frequency (F0) mean. Recordings of both spontaneous speech (path description and picture description) and read speech (sentences) were collected twice from lesbian participants and straight women in the presence of the female experimenter first and then with a male experimenter, or vice-versa. In the analysis of the read speech, it was evident that participants adapted their mean fundamental frequency to match that of the first experimenter, and this adjustment remained consistent even in the presence of the second experimenter. Specifically, women who initially interacted with a male experimenter read with a lower pitch compared to those who first interacted with a female experimenter. This discrepancy in fundamental frequency persisted even after subsequent interactions with the opposite experimenter. Regarding

spontaneous speech, the evidence suggesting that women who initially encountered the male experimenter spoke with a lower pitch than those who initially encountered the female experimenter was replicated among participants who identified as exclusively lesbian or straight, rather than those identified as primarily lesbian or straight. Whether the observed differences in accommodation patterns have exclusively to do with the degree of straightness/lesbianness still remains to be understood, thus opening new directions of research on the complex interaction between interlocutors' sex and the sexual orientation and many other factors affecting convergence in interaction.

Not only the sex of the interlocutor as studied by Kachel et al. (2024) or the interlocutor set size as explored by Burchfield et al. (2023), but also the presence itself of a live co-speaker is a key factor which may influence accommodation. The study of Cerda-Oñate et al. (2021) explored how the presence or absence of a co-speaker and the metrical aspects of a text (weak or strong meter) influence speech rhythm synchronization during chorus reading. Speech rhythm was conceptualized as a two-level phenomenon. At the lower level, rhythm was operationalized in terms of the regularity in the distribution of vowel onsets and in the timing properties of consonantal and vocalic intervals. At the higher level (meters), syllable grouping into metrical structures based on lexical stress and phrasal prominence were considered in the operationalization of rhythm. To evaluate the influence of co-speaker presence/absence and meter on speech synchronization, the authors gathered data from 30 speakers of Chilean Spanish. Each participant engaged in synchronous reading tasks alongside a model speaker of the same language. Different synchronous reading conditions and texts were employed:

1. Synchrony-live condition: Participants read aloud in synchrony with the live model speaker.
2. Synchrony with recording from live condition: Experimental participants synchronized their reading with a recording of the model speaker from the live condition.
3. Synchrony with recording from non-live condition: Experimental participants synchronized their reading with a recording of the model speaker reading the texts alone.

The researchers utilized both a poetic text with a strong meter and a narrative text with a weak meter as reading materials. The findings indicated that the presence of a co-speaker significantly affected interspeaker synchronization, with meter offering additional advantages for coordinated speech in more demanding situations. Participants adjusted their speech to enhance synchronization, particularly during simultaneous reading activities. They did it by regularizing the durations of vowels, consonantal intervals, and syllables, thus making the distribution of vowel and syllable onsets – p-centers or beats – more equally-timed. The effect of strong meter on rhythmic convergence was obtained when the reading task was performed in synchrony with a recording of the model speaker reading solo. Although synchronization is possible in the absence of a live co-speaker, this ability is enhanced by bidirectional rhythmic adaptations that make speech patterns more predictable and thus easy to align.

Biro et al. (2022) authored a paper in the collection that delves into the influence of task-related characteristics on convergence. Their aim was to investigate the influence of task engagement on phonetic convergence in primary and secondary phonetic features of voicing. Drawing on Cheyne et al. (2009)'s definition, they conceptualized engagement as the level to which task characteristics sustain participants' attention toward task-related goals. In their experiment, varying levels of engagement were induced by having participants complete word-matching puzzles either in a highly immersive video game environment (navigating a virtual world in *Minecraft* TM) or a less engaging task (clicking on words from a list). Using monosyllabic voicing minimal pairs as stimuli, the study measured changes in VOT, vowel length, first formant (F1) onset, and F0 onset over the duration of the experiment.

Convergence was assessed by comparing the mean acoustic feature values across normalized trial numbers when participants were acting as directors for their partners. Within dyads, individuals displayed variations in voicing characteristics like VOT duration, vowel length (VL), F1 height and F0 onset relative to their conversational partner. Based on the average values of the first trials, each dyad member was designated as having long vs short VOT, long vs short vowel length, high vs low initially F1 height, high vs low F0. Results showed that task engagement positively affected convergence but only for secondary voicing features (F1 onset and vowel length). In high-engagement task, participants with higher initial F1 onsets and longer initial VL converged toward participants with lower initial F1 onsets for voiceless tokens and with shorter initial VLs, respectively. Participants with lower initial F1 onsets converged toward participants with higher initial F1 onsets for voiced tokens. Inconsistent patterns of convergence were documented for VOT, whereas for F0 onset, the changes observed for speakers with higher initial F0 onset values in the high engagement task were not in a direction indicating convergence. Overall, these findings suggested that the engagement of participants with their task environment can influence the extent of phonetic convergence. However, engagement interacts in a complex fashion with the salience of certain cues for conveying linguistic contrasts. The results were interpreted within the framework of interpersonal synergies by Fusaroli et al. (2014) suggesting that reciprocal imitation is only one of the strategies that interlocutors use to coordinate in dialogue situations.

The remaining three papers examined various forms of convergence in a cross-linguistic context. Ruch (2021), for example, focused on vowel accommodation across different dialects, specifically the varieties of Swiss German spoken in the Cantons of Zurich and the Grisons. These dialects exhibit significant differences in their speech patterns, both in terms of individual sounds and prosodic characteristics, thus providing opportunities for cross-dialectal accommodation. Despite these differences, however, they are mutually intelligible and hold equal status, which could potentially hinder the process of accommodation. Like the study conducted by Lee et al., 2021, Ruch employed a dual approach involving computer simulations (i.e., agent-based modeling techniques) and investigations of how pairs of Zurich and Grison German speakers varied their vowel pronunciation after interacting with speakers from the other dialect as opposed to before the interaction. The acoustic data from these initial recordings served as input for the computational modeling, which simulated how the vowel systems of the dialects evolved when speakers from different dialects interacted without any social influence. The simulation results, based on interactions between computer-generated agents, suggested a clear trend towards convergence, driven primarily by the initial variability of vowels in the acoustic space prior to interaction. The analysis of data from real interactions relied on the paradigm of the difference in distance, in that it compared the acoustic distance between vowel formants before and after dialogues. The results of real interaction data, which includes pre- and post-dialogue lexical items, revealed a different pattern from computational modeling pointing to vowel maintenance. The discrepancy between these real interaction findings and the simulation results was attributed to unique sociolinguistic dynamics present in the German-speaking region of Switzerland. Factors such as the equal prestige of dialects, mutual intelligibility, and frequent interactions between speakers of different dialects in daily situations triggered dialectal maintenance. In contrast, in simulated dialogues between computer agents, linguistic factors such as acoustic variability play a more dominant role in driving convergence.

The study by Ulbrich and Canzi (2023) centered on L1-L2 intonational alignment. The authors aimed to explore whether L1-L2 contrastive phonological features undergo accommodation and whether the phenomenon extends also to gradient acoustic realizations. The L1 and L2 combination under study were L1 Belfast English vs L2 Belfast English (BfE) with Northern Standard German (NSG) as the subjects' first language, and vice versa. The phonological features under

investigation were pitch accent (a categorical feature), pitch in utterance-final position, and peak alignment in nuclear pitch accents (gradient phonetic realizations) during a collaborative map task. BfE and NSG exhibit distinct intonation patterns for nuclear pitch accents in declarative utterances, with NSG having a falling pitch accent pattern and BfE using a rising intonation contour. Two subject groups participated in a collaborative map task, wherein the BfE speakers and the NSG speakers were led through the map by a NSG speaker and a BfE speaker, respectively. Accommodation was assessed by comparing the realization of rising and falling accents, pitch range, and peak alignment for high-tone falling (HL) and low-tone rising (LH) nuclear pitch accents among NSG and BfE speakers. This comparison was conducted during early and late parts of the recordings, which corresponded to the initial and final 35 % of the recording duration for each individual collaborative map task session. Evidence of accommodation was observed for all three prosodic phenomena, with asymmetries between the speakers' groups, modulated by factors such as linguistic function, perceptual salience, and familiarity. Regarding the realization of nuclear pitch accents, L1 BfE learners of German exhibited a greater convergence towards the falling nuclear pitch accent. This was attributed to the limited functions in information structure associated with falling intonation in their L1. Conversely, the variable functions served by rising intonation in NSG were accounted for by the comparatively limited accommodation observed in NSG L1 learners of BfE. In the context of peak alignment, the resistance observed towards alignment in default patterns, coupled with the variable alignment seen in non-default patterns, was attributed to the entrenchment of pitch accent patterns in episodic memory. As proposed by Goldinger (1998), more familiar linguistic units were interpreted as being more deeply entrenched compared to less familiar ones, leading to a scarcity of memory traces to draw upon, hence to higher vulnerability and susceptibility to change.

The last paper in this collection is by Ludusan and Wagner (2022). The paper is the only paper of the collection that does not focus on speech as a carrier of language but on laughter, a paralinguistic phenomenon common across all languages and cultures in spontaneous interactions. They examined convergence of laughter in French, German, and Mandarin Chinese speech from the DUEL corpus, a multilingual dataset annotated for laughter. They examined temporal and form-related entrainment at two organizational levels: the laughter-token level and the turn level. The goals were to determine whether (1) entrainment at one level implies entrainment at another organizational level; (2) language- or culture-specific trends can be observed, and (3) the familiarity of conversational partners plays a role in the entrainment process. Along the lines of the computational approach introduced by Levitan and Hirschberg (2011) the researchers used five measures of convergence. Two temporal measures at the token level, synchrony and convergence, represented the dynamic coordination between partners during conversation and the degree to which partners become more similar over the interaction, respectively. At the turn level, two temporal measures, alignment and congruence, captured the degree of alignment between laughter and turn boundaries and the increased use of this marking over the conversation, respectively. Form-related entrainment analysis examined degree of voicing, i.e., the similarity of consecutive laughter from a signal intensity perspective. The findings provided evidence for temporal entrainment at the laughter-token level, revealing that speakers distributed their laughter events throughout a conversation. At the turn level, speakers across all three languages displayed entrainment by aligning their laughter more with the beginning and end of their turns, with this tendency becoming more pronounced in the second half of the recordings compared to the first half. Moreover, the study supported form-related entrainment, indicating that conversational partners tended to employ more similar intensity levels for consecutive laughter events compared to non-consecutive ones. Importantly, these entrainment patterns were shown to be independent of the degree of familiarity between the speakers.

### 3. Future lines of research

The contributions in this special issue introduce novel inquiries about the driving forces and capabilities behind accommodation in production and perception, highlighting the roles of cognitive adaptability and phonetic diversity. The experiments and findings by Lee et al. (2021) illustrate that speakers who naturally exhibit more variability in real and computer-simulated interaction were more likely to adapt their speech acoustics and articulation to match their conversational partners. Similarly, the simulation results in Ruch (2021) about asymmetric accommodation between Zurich and Grison German dialect, revealed that the most significant adjustments appeared in dialects and vowels that initially show greater phonetic diversity. Taken together, these findings suggest that speakers who start with a wider range of phonetic realization are more adaptable, as they can produce phonetic alternatives that fit within their current range of acoustic and articulatory patterns. In the realm of accommodation in perception, the research by Burchfield et al. (2023) demonstrated that variability, conceptualized as size of conversational experience, enhances perceptual learning. Essentially, by frequent interaction with diverse talkers, speakers fine-tune their language processing abilities. Future research could explore the impact of a broader spectrum of within-speaker variation in perception and production on accommodation. Such studies might, for instance, compare groups of speakers or imitators and model speakers, matched according to a gradient of acoustic variability within the speaker, prompted by different speaking styles that range in expressiveness (from reading aloud to baby talk). It would thus be interesting to examine if speakers with greater variability show more significant convergence. Additionally, investigating individuals with acting backgrounds or varying personalities (extroverts vs introverts) might shed light on whether speakers accustomed to assuming varied roles or those more socially active (and thus having a broader circle of contacts) are more inclined to adapt in speech production and perception than speakers lacking acting experience, for example, or speakers with more introverted personalities (who typically have fewer social interactions).

Another promising avenue of inquiry introduced by Dias et al. (2021) focuses on exploring how the process of recognizing convergent speech aligns with the mechanisms involved in speaker identification and the impact of talker familiarity on speech processing. The underlying hypothesis suggests that these phenomena depend on the individual's capacity to identify unique acoustic features of speakers and the ability to apply this knowledge flexibly to new speakers and speech items after learning them. This suggests that acoustic convergence, where one speaker adopts the vocal characteristics of their conversational partner, may have negative impact on the recognition of a speaker based on their voice. Existing findings from naturally occurring forms of convergence (Pellegrino and Dellwo, 2023) or computer-generated exaggerated forms of convergence (Farrús et al., 2010; Kinnunen and Li, 2010) support this idea, showing that acoustic convergence can increase the vocal similarity between conversational partners, thereby diminishing the uniqueness of individual voices and affecting their identifiability by both humans and voice recognition systems. With these premises, it would be worth investigating whether gaining familiarity with the phonetic variability induced by convergence benefits the processing of speaker identities, likewise the recognition of speech. Another related field of study with potential insights about the impact of convergent speech on vocal individuality is synchronous speech. The findings reported by Cerda-Onate et al. (2021) indicate that when speakers align their speech rhythm with another speaker in real time, their rhythmic properties tend to synchronize, facilitating alignment. Further research should therefore clarify the relationships between synchronous speech, vocal accommodation, and voice recognition.

The findings about synchronous reading are intriguing also for understanding the evolutionary force driving convergence. It has been shown that capacity for accommodation predates the evolution of articulated speech, hence chorus reading and other cooperative

communication forms, manifested through vocal convergence, may serve the same purpose as the need to signal group cohesion and cooperation in animal communication (Cerda-Oñate et al., 2021). An intriguing research direction would thus be to delve deeper into the evolutionary connection between changes in animal vocalizations and vocal accommodation in humans, aiming to illuminate the social/cooperative aspect of accommodation.

In this collection, two studies (Lee et al., 2021 and Ruch, 2021) explored interspeaker accommodation through both acoustic analysis and computational modeling, yielding results that were consistent across methods in one instance (Lee et al., 2021) but divergent in the other (Ruch, 2021). Specifically, Lee et al. (2021) found agreement between computer simulations and real-time interactions, supporting the hypothesis that higher within-speaker acoustic-phonetic variability promotes accommodation. Conversely, Ruch (2021) observed vowel convergence in agent-based modeling simulations, while analyses of pre- and post-dialogue recordings indicated maintenance. The authors noted methodological challenges when comparing traditional analyses to computational modeling: the outcomes of simulations may not be directly applicable to real-life scenarios. For example, the attunement model could only predict convergence, while real interactions are characterized by the co-occurrence of accommodative and non-accommodative moves. In the case of the agent-based modeling, the interaction happening between the agents and between humans were not comparable in terms of the number of interlocutors from the other dialect (one to one in real time interaction; many to many in the computational models). Moreover, the frequency of lexical items may play a different role in the short-term linguistic experience and memories of the agents as compared to the life-long one of the participants. Despite these challenges, comparing computational modeling with real-life interactions remains a valuable approach. Further research is needed to support the notion that, in the absence of social, attitudinal, or ideological factors, linguistic factors primarily drive convergence as a default communicative strategy. As also envisaged by Ruch (2021), future investigations combining these methodologies could illuminate the interplay between mechanical, phonetic and social factors in accommodation. Such research may provide insights into behaviors influencing convergence when the dynamics of these influences in human interaction are difficult to test.

## Funding

This project was supported by the URPP Language and Space of the University of Zurich and the NCCR Evolving Language, Swiss National Science Foundation Agreement #51NF40\_180888.

## CRediT authorship contribution statement

**Elisa Pellegrino:** Conceptualization, Investigation, Methodology, Project administration, Resources, Software, Validation, Writing – original draft, Writing – review & editing. **Volker Dellwo:** Conceptualization, Funding acquisition, Supervision, Validation, Writing – review & editing. **Jennifer S. Pardo:** Conceptualization, Writing – review & editing. **Bernd Möbius:** Conceptualization, Writing – review & editing.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Elisa Pellegrino reports financial support, administrative support, article publishing charges, and travel were provided by Swiss National Science Foundation. Elisa Pellegrino reports financial support, administrative support, article publishing charges, and travel were provided by University of Zurich URPP Language and Space. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## References

- Abel, J., Babel, M., 2017. Cognitive load reduces perceived linguistic convergence between dyads. *Lang. Speech*. 60 (3), 479–502. <https://doi.org/10.1177/0023830916665652>.
- Adams, A., Miles, J., Dunbar, N.E., Giles, H., 2018. Communication accommodation in text messages: exploring liking, power, and sex as predictors of textisms. *J. Soc. Psychol.* 158 (4), 474–490. <https://doi.org/10.1080/00224545.2017.1421895>.
- 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., 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J. Phon* 40 (1). <https://doi.org/10.1016/j.wocn.2011.09.001>.
- Babel, M., Bulatov, D., 2012. The role of fundamental frequency in phonetic accommodation. *Lang. Speech*. 55 (2), 231–248. <https://doi.org/10.1177/0023830911417695>.
- Barnlund, D.C., 1970. A transactional model of communication. *Language Behavior*. De Gruyter, pp. 43–61. <https://doi.org/10.1515/9783110878752.43>.
- Barón-Birchenall, L., 2023. Phonetic accommodation during conversational interactions: an overview. *Revista Guillermo de Ockham*, 21 (2). <https://doi.org/10.21500/22563202.6150> press.
- Bell, A., 2002. Back in style: reworking audience design. *Style and Sociolinguistic Variation*. Cambridge University Press, pp. 139–169. <https://doi.org/10.1017/CBO9780511613258.010>.
- Biro, T., Toscano, J.C., Viswanathan, N., 2022. The influence of task engagement on phonetic convergence. *Speech Commun* 138, 50–66. <https://doi.org/10.1016/j.specom.2022.02.002>.
- Borrie, S.A., Liss, J.M., 2014. Rhythm as a coordinating device: entrainment with disordered speech. *J. Speech Lang. Hear. Res.* 57 (3), 815–824. [https://doi.org/10.1044/2014\\_JSLHR-S-13-0149](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. *Front Psychol* 6. <https://doi.org/10.3389/fpsyg.2015.01187>.
- Branigan, H.P., Pickering, M.J., Cleland, A.A., 2000. Syntactic co-ordination in dialogue. *Cognition* 75 (2), B13–B25. [https://doi.org/10.1016/S0010-0277\(99\)00081-5](https://doi.org/10.1016/S0010-0277(99)00081-5).
- Burchfield, L.A., Antoniou, M., Cutler, A., 2023. The dependence of accommodation processes on conversational experience. *Speech Commun* 153. <https://doi.org/10.1016/j.specom.2023.102963>.
- Cerda-Oñate, K., Vega, G.T., Ordín, M., 2021. Speech rhythm convergence in a dyadic reading task. *Speech Commun* 131, 1–12. <https://doi.org/10.1016/j.specom.2021.04.003>.
- Chartrand, T.L., Bargh, J.A., 1999. The chameleon effect: the perception–behavior link and social interaction. *J. Pers. Soc. Psychol.* 76 (6), 893–910. <https://doi.org/10.1037/0022-3514.76.6.893>.
- Cheyne, J.A., Solman, G.J.F., Carriere, J.S.A., Smilek, D., 2009. Anatomy of an error: a bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition* 111 (1). <https://doi.org/10.1016/j.cognition.2008.12.009>.
- Choi, J.Y., Perrachione, T.K., 2019. Time and information in perceptual adaptation to speech. *Cognition* 192, 103982. <https://doi.org/10.1016/j.cognition.2019.05.019>.
- Cohen Priva, U., Sanker, C., 2019. Limitations of difference-in-difference for measuring convergence. *Laboratory Phonol.: J. Associat. Laboratory Phonol.* 10 (1), 15. <https://doi.org/10.5334/labphon.200>.
- Costa, A., Pickering, M.J., Sorace, A., 2008. Alignment in second language dialogue. *Lang. Cogn. Process* 23 (4), 528–556. <https://doi.org/10.1080/01690960801920545>.
- Coupland, N., Giles, H., 1988. Communicative accommodation: recent developments. *Lang. Commun.* 175–327.
- De Looze, C., Scherer, S., Vaughan, B., Campbell, N., 2014. Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Commun*. 58, 11–34. <https://doi.org/10.1016/j.specom.2013.10.002>.
- Dias, J.W., Vazquez, T.C., Rosenblum, L.D., 2021. Perceptual learning of phonetic convergence. *Speech Commun*. 133, 1–8. <https://doi.org/10.1016/j.specom.2021.07.004>.
- Dijksterhuis, A.J., Bargh, J.A., 2001. The perception–behavior expressway: automatic effects of social perception on social behavior. *Adv. Exp. Soc. Psychol.* 33, 1–40.
- Dragojevic, M., Gasiorek, J., Giles, H., 2016. Accommodative strategies as core of the theory. *Communication Accommodation Theory: Negotiating Personal Relationships and Social Identities across Contexts* 36–59. <https://doi.org/10.1017/CBO9781316226537.003>.
- Dufour, S., Nguyen, N., 2013. How much imitation is there in a shadowing task? *Front Psychol* 4. <https://doi.org/10.3389/fpsyg.2013.00346>.
- Earnshaw, K., 2021. Examining the implications of speech accommodation for forensic speaker comparison casework: a case study of the West Yorkshire face vowel. *J Phon* 87, 101062. <https://doi.org/10.1016/j.wocn.2021.101062>.



- Eldlund, J., Heldner, M., Hirschberg, J., 2009. Pause and gap length in face-to-face interaction. In: *Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech*, pp. 2779–2782.
- Elleström, L., 2018. A medium-centered model of communication. *Semiotica* 188 (224), 269–293. <https://doi.org/10.1515/sem-2016-0024>.
- Farrús, M., Wagner, M., Erro, D., Hernando, J., 2010. Automatic speaker recognition as a measurement of voice imitation and conversion. *Int. J. Speech Lang. the Law* 17 (1). <https://doi.org/10.1558/jisll.v17i1.119>.
- Fusaroli, R., Rączaszek-Leonardi, J., Tylén, K., 2014. Dialog as interpersonal synergy. *New Ideas Psychol.* 32, 147–157. <https://doi.org/10.1016/j.newideapsych.2013.03.005>.
- Fusaroli, R., Tylén, K., 2016. Investigating conversational dynamics: interactive alignment, interpersonal synergy, and collective task performance. *Cogn. Sci.* 40 (1). <https://doi.org/10.1111/cogs.12251>.
- Giles, H., 1984. The dynamics of speech accommodation. *Special issue. Int J Soc Lang.*
- Giles, H., Coupland, N., Coupland, J., 1991. Accommodation theory: communication, context, and consequence. *Contexts of Accommodation*. Cambridge University Press, pp. 1–68. <https://doi.org/10.1017/CBO9780511663673.001>.
- Giles, H., Edwards, A.L., Walther, J.B., 2023. Communication accommodation theory: past accomplishments, current trends, and future prospects. *Language Sciences* 99, 101571. <https://doi.org/10.1016/j.langsci.2023.101571>.
- Giles, H., Gasiorek, J., Soliz, J., 2015. Recent developments in communication accommodation theory: innovative contexts and applications. *Lang. Commun.* 41.
- Giles, H., Ogay, T., 2007. *Communication Accommodation Theory. Explaining communication: Contemporary theories and Exemplars*. Lawrence Erlbaum Associates Publishers, pp. 293–310.
- Giles, H., Taylor, D.M., Bourhis, R.Y., 1973. Towards a theory of interpersonal accommodation through speech: some Canadian data. *Lang. Soc* 2 (2), 177–192. <https://doi.org/10.1017/S0047404500000701>.
- Goldinger, S.D., 1998. Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 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. *Psychon. Bull. Rev.* 11, 716–722. <https://doi.org/10.3758/bf03196625>.
- Jakobson, R., 1960. Closing statement: linguistics and poetics. *Style in language* 350.
- Kachel, S., Simpson, A.P., Steffens, M.C., 2024. Speakers' vocal expression of sexual orientation depends on experimenter gender. *Speech Commun.* 156, 103023. <https://doi.org/10.1016/j.specom.2023.103023>.
- Kalmanovitch, Y., 2016. *Speech in interaction – the zurich tangram corpus germany. tagungsband Der 12. Tagung Phonetik Und Phonologie Im Deutschsprachigen Raum* 79–81.
- Kim, M., Horton, W.S., Bradlow, A.R., 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Lab. Phonol.* 2 (1), 125–156.
- Kinnunen, T., Li, H., 2010. An overview of text-independent speaker recognition: from features to supervectors. *Speech Commun* 52 (1), 12–40. <https://doi.org/10.1016/j.specom.2009.08.009>.
- Lakin, J.L., 2013. 18 Behavioral mimicry and interpersonal synchrony. *Nonverbal Communication. De Gruyter*, pp. 539–576. <https://doi.org/10.1515/9783110238150.539>.
- Lavan, N., Burton, A., Scott, S.K., et al., 2019. Flexible voices: Identity perception from variable vocal signals. *Psychon. Bull. Rev.* 26, 90–102. <https://doi.org/10.3758/s13423-018-1497-7>.
- 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. *Interspeech 2010*, 793–796. <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), 1–26. <https://doi.org/10.1371/journal.pone.0201444>.
- Lee, Y., Goldstein, L., Parrell, B., Byrd, D., 2021a. Who converges? Variation reveals individual speaker adaptability. *Speech Commun.* 131, 23–34. <https://doi.org/10.1016/j.specom.2021.05.001>.
- Lelong, A., Bailly, G., 2011. Study of the Phenomenon of Phonetic Convergence Thanks to Speech Dominoes. In: *Esposito, A., Vinciarelli, A., Vicsi, K., Pelachaud, C., Nijholt, A. (Eds.), Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues. Lecture Notes in Computer Science*, 6800. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-25775-9\\_26](https://doi.org/10.1007/978-3-642-25775-9_26).
- Levitan, R., Hirschberg, J., 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In: *Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech*, pp. 3081–3084, 10.1.1.296.8560.
- Lin, Y., Yao, Y., Luo, J., 2021. Phonetic accommodation of tone: reversing a tone merger-in-progress via imitation. *J Phon* 87, 101060. <https://doi.org/10.1016/j.wocn.2021.101060>.
- Ludusan, B., Wagner, P., 2022. Laughter entrainment in dyadic interactions: temporal distribution and form. *Speech Commun.* 136, 42–52. <https://doi.org/10.1016/j.specom.2021.11.001>.
- Michalsky, J., Schoormann, H., 2017. Pitch convergence as an effect of perceived attractiveness and likability. In: *Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech*, pp. 2253–2256. <https://doi.org/10.21437/Interspeech.2017-1520>.
- Michalsky Jan and Schoormann, H. (2021). Birds of a feather flock together but opposites Attract! On the Interaction of F0 Entrainment, Perceived Attractiveness, and Conversational Quality in Dating Conversations. In J. and B.-D. M. and O. J. J. Weiss Benjamin and Trouvain (Eds.), *Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speakers* (pp. 215–242). Springer Singapore. [https://doi.org/10.1007/978-981-15-6627-1\\_12](https://doi.org/10.1007/978-981-15-6627-1_12).
- Mukherjee, S., Badino, L., Hilt, P.M., Tomassini, A., Inuggi, A., Fadiga, L., Nguyen, N., D'Ausilio, A., 2019. The neural oscillatory markers of phonetic convergence during verbal interaction. *Hum. Brain Mapp.* 40 (1), 187–201. <https://doi.org/10.1002/hbm.24364>.
- Namy, L.L., Nygaard, L.C., Sauerteig, D., 2002. Gender differences in vocal accommodation. *J. Lang. Soc. Psychol.* 21 (4), 422–432. <https://doi.org/10.1177/026192702237958>.
- Nguyen, N., Delvaux, V., 2015. Role of imitation in the emergence of phonological systems. *J. Phon.* 53, 46–54. <https://doi.org/10.1016/j.wocn.2015.08.004>.
- Nielsen, K., 2014. Phonetic imitation by young children and its developmental changes. *J. Speech Lang. Hear. Res.* 57 (6), 2065–2075. <https://doi.org/10.1044/2014.JSLHR-S-13-0093>.
- Olmstead, A.J., Viswanathan, N., Cowan, T., Yang, K., 2021. Phonetic adaptation in interlocutors with mismatched language backgrounds: a case for a phonetic synergy account. *J. Phon.* 87, 101054. <https://doi.org/10.1016/j.wocn.2021.101054>.
- Paquette-Smith, M., Schertz, J., Johnson, E.K., 2022. Comparing phonetic convergence in children and adults. *Lang. Speech.* 65 (1), 240–260. <https://doi.org/10.1177/00238309211013864>.
- Pardo, J.S., 2006. On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am.* 119 (4), 2382–2393. <https://doi.org/10.1121/1.2178720>.
- Pardo, J.S., 2012. Reflections on phonetic convergence: speech perception does not mirror speech production. *Linguistics Lang. Compass.* 6 (12), 753–767. <https://doi.org/10.1002/lnc3.367>.
- Pardo, J.S., 2013. Measuring phonetic convergence in speech production. *Front Psychol* 4 (AUG), 1–5. <https://doi.org/10.3389/fpsyg.2013.00559>.
- Pardo, J.S., Jay, I.C., Krauss, R.M., 2010. Conversational role influences speech imitation. *Attention. Percept. Psychophys.* 72 (8), 2254–2264. <https://doi.org/10.3758/APP.72.8.2254>.
- Pardo, J.S., Pellegrino, E., Dellwo, V., Möbius, B., 2022. Special issue: vocal accommodation in speech communication. *J. Phon.* 95, 101196. <https://doi.org/10.1016/j.wocn.2022.101196>.
- Pardo, J.S., Urmanche, A., Wilman, S., Wiener, J., 2017. Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics* 79 (2), 637–659. <https://doi.org/10.3758/s13414-016-1226-0>.
- Pardo, J.S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., Ward, M., 2018. A comparison of phonetic convergence in conversational interaction and speech shadowing. *J Phon* 69 (April), 1–11. <https://doi.org/10.1016/j.wocn.2018.04.001>.
- Pellegrino, E., 2024. After self-imitation prosodic training L2 learners converge prosodically to the native speakers. *Languages* 9 (1), 33. <https://doi.org/10.3390/languages9010033>.
- Pellegrino, E., Dellwo, V., 2023. Speakers are more cooperative and less individual when interacting in larger group sizes. *Front Psychol* 14. <https://doi.org/10.3389/fpsyg.2023.1145572>.
- Pickering, M.J., Garrod, S., 2004. Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27 (2), 169–190. <https://doi.org/10.1017/S0140525X04000056>.
- Pinget, A.-F., 2022. Individual differences in phonetic imitation and their role in sound change. *Phonetica* 79 (5), 425–457. <https://doi.org/10.1515/phon-2022-2026>.
- Ross, J.P., Lilley, K.D., Clopper, C.G., Pardo, J.S., Levi, S.V., 2021. Effects of dialect-specific features and familiarity on cross-dialect phonetic convergence. *J Phon* 86, 101041. <https://doi.org/10.1016/j.wocn.2021.101041>.
- Ruch, H., 2021. Dialect contact in real interactions and in an agent-based model. *Speech Commun* 134, 55–70. <https://doi.org/10.1016/j.specom.2021.09.003>.
- Ruch, H., Zürcher, Y., Burkart, J.M., 2018. The function and mechanism of vocal accommodation in humans and other primates. *Biol. Rev.* 93 (2), 996–1013. <https://doi.org/10.1111/brv.12382>.
- Schramm, W., 1955. Information Theory and Mass Communication. *Journalism Quarterly* 32 (2), 131–146. <https://doi.org/10.1177/107769905503200201>.
- Shannon, C.E., 1948. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 27 (3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- Shockley, K., Sabadini, L., Fowler, C.A., 2004. Imitation in shadowing words. *Percept. Psychophys.* 66 (3), 422–429. <https://doi.org/10.3758/BF03194890>.
- Smith, J., Durham, M., Fortune, L., 2007. Mam, my trousers is fa'in doon!": community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Lang Var Change* 19 (01). <https://doi.org/10.1017/S09544394507070044>.
- Sonderegger, M., Bane, M., Graff, P., 2017. The medium-term dynamics of accents on reality television. *Language (Baltim)* 93 (3). <https://doi.org/10.1353/lan.2017.0038>.
- Späth, M., Aichert, I., Ceballos-Baumann, A.O., Wagner-Sonntag, E., Miller, N., Ziegler, W., 2016. Entraining with another person's speech rhythm: evidence from healthy speakers and individuals with Parkinson's disease. *Clin. Linguist. Phon.* 30 (1), 68–85. <https://doi.org/10.3109/02699206.2015.1115129>.
- Stamp, R., Schembri, A., Evans, B.G., Cormier, K., 2016. Regional Sign Language Varieties in Contact: investigating Patterns of Accommodation. *J. Deaf Stud. Deaf Educ.* 21 (1), 70–82. <https://doi.org/10.1093/deafed/env043>.
- Tagliamonte, S.A., Molfenter, S., 2007. How'd you get that accent?: acquiring a second dialect of the same language. *Language in Society*, 36 (5). <https://doi.org/10.1017/S0047404507070911>.
- Tausczik, Y.R., Pennebaker, J.W., 2013. Improving teamwork using real-time language feedback. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 459–468. <https://doi.org/10.1145/2470654.2470720>.
- Tobin, S.J., 2022. Effects of native language and habituation in phonetic accommodation. *J Phon* 93, 101148. <https://doi.org/10.1016/j.wocn.2022.101148>.

- Ulbrich, C., Canzi, M., 2023. Intonational alignment in second language acquisition. *Speech Commun* 146, 70–81. <https://doi.org/10.1016/j.specom.2022.11.007>.
- 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>.
- Wagner, M.A., Broersma, M., McQueen, J.M., Dhaene, S., Lemhöfer, K., 2021. Phonetic convergence to non-native speech: acoustic and perceptual evidence. *J Phon* 88, 101076. <https://doi.org/10.1016/j.wocn.2021.101076>.
- Walker, A., Campbell-Kibler, K., 2015. Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Front Psychol* 6 (MAY). <https://doi.org/10.3389/fpsyg.2015.00546>.
- Weise, A., Silber-Varod, V., Lerner, A., Hirschberg, J., Levitan, R., 2020. Entrainment in spoken Hebrew dialogues. *J. Phon.* 83, 101005. <https://doi.org/10.1016/j.wocn.2020.101005>.
- Wynn, C.J., Borrie, S.A., 2022. Classifying conversational entrainment of speech behavior: an expanded framework and review. *J. Phon.* 94, 101173. <https://doi.org/10.1016/j.wocn.2022.101173>.
- Yaeger, D.M. (1992). Communicative Accommodation: a New Perspective on “Hypercorrect” Speech. *Language and Communication*.
- 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>.
- Zürcher, Y., Willems, E.P., Burkart, J.M., 2021. Trade-offs between vocal accommodation and individual recognisability in common marmoset vocalizations. *Sci. Rep.* 11 (1), 15683. <https://doi.org/10.1038/s41598-021-95101-8>.

### Further reading

Pickering, M.J., 2004. Toward a mechanistic psychology of dialogue 1–58.

Elisa Pellegrino<sup>a,\*</sup>, Volker Dellwo<sup>a</sup>, Jennifer S. Pardo<sup>b</sup>, Bernd Möbius<sup>c</sup>  
<sup>a</sup> University of Zurich, Switzerland  
<sup>b</sup> Montclair State University, United States  
<sup>c</sup> Saarland University, Germany

\* Corresponding author at: University of Zurich, Switzerland.  
 E-mail address: [elisa.pellegrino@uzh.ch](mailto:elisa.pellegrino@uzh.ch) (E. Pellegrino).