

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Journal of Phonetics

journal homepage: [www.elsevier.com/locate/Phonetics](http://www.elsevier.com/locate/Phonetics)

Special Issue: Vocal Accommodation in Speech Communication. eds. Pardo, Pellegrino, Dellwo &amp; Möbius

## Special issue: Vocal accommodation in speech communication

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

## ARTICLE INFO

## Article history:

Received 15 July 2022

Received in revised form 15 October 2022

Accepted 18 October 2022

## Keywords:

Vocal accommodation  
Phonetic convergence  
Speech production

## ABSTRACT

This introductory article for the Special Issue on Vocal Accommodation in Speech Communication provides an overview of prevailing theories of vocal accommodation and summarizes the ten papers in the collection. Communication Accommodation Theory focusses on social factors evoking accent convergence or divergence, while the Interactive Alignment Model proposes cognitive integration of perception and production as an automatic priming mechanism driving convergence language production. Recent research including most of the papers in this Special Issue indicates that a hybrid or interactive synergy model provides a more comprehensive account of observed patterns of phonetic convergence than purely automatic mechanisms. Some of the fundamental questions that this special collection aimed to cover concerned (1) the nature of vocal accommodation in terms of underlying mechanisms and social functions in human–human and human–computer interaction; (2) the effect of task-specific and talker-specific characteristics (gender, age, personality, linguistic and cultural background, role in interaction) on degree and direction of convergence towards human and computer interlocutors; (3) integration of articulatory, perceptual, neurocognitive, and/or multimodal data to the analysis of acoustic accommodation in interactive and non-interactive speech tasks; and (4) 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.

© 2022 Elsevier Ltd. All rights reserved.

## 1. Introduction: vocal accommodation in speech communication

Communication accommodation is a ubiquitous and multifaceted phenomenon observed in human communication, animal communication, and in human–computer interactions, encompassing numerous linguistic, para-linguistic, and extra-linguistic features (see [Gasiorek, Giles & Soliz, 2015](#); [Shepard, Giles, & Le Poire, 2001](#)). Vocal accommodation in particular is a tendency for talkers to adapt acoustic–phonetic attributes to those of another talker. Diverse approaches to vocal accommodation have involved multiple terms—*alignment* (e.g., [Pickering & Garrod, 2004, 2013](#)), *convergence* (e.g., [Babel, 2010](#); [Pardo, 2006](#)), *entrainment* (e.g., [Borrie & Liss, 2014](#); [Levitan & Hirschberg, 2011](#)), *imitation* (e.g., [Babel, 2012](#); [Goldinger, 1998](#)), *synchrony* (e.g., [Edlund, Heldner, & Hirschberg, 2009](#); [Levitan & Hirschberg, 2011](#)), and *mimicry* (e.g., [Chartrand & Bargh, 1999](#); [Pentland, 2008](#);

see [Wynn & Borrie 2022](#) for a review of types of entrainment, classification criteria, and terminology). Regardless of the terminology, all of these approaches assess increasing or decreasing similarity in speech and propose some form of internal/cognitive and/or external/social mechanisms driving these changes. Vocal accommodation is a phenomenon that is seemingly deeply rooted in communication. Convergence can also be found in a variety of animal species ([Ruch, Zürcher, & Burkart, 2018](#)), suggesting that the tendency to signal similarity in communication predates the language age in humans.

In the years since Giles first coined the term *Speech Accommodation* ([Giles, 1973](#)), research on vocal accommodation in interactive and non-interactive tasks has shown that speakers converge acoustically in both suprasegmental and segmental features, including *speech rate* (e.g. [Manson, Bryant, Gervais, & Kline, 2013](#); [Putman & Street, 1984](#); [Schweitzer & Lewandowski, 2013](#); [Staum Casasanto, Jasmin, & Casasanto, 2010](#)), *pause rates* (e.g., [Bilous & Krauss, 1988](#); [Pardo, Cajori Jay, Hoshino, Hasbun,](#)

\* Corresponding author.

E-mail address: [pardoj@mail.montclair.edu](mailto:pardoj@mail.montclair.edu) (J.S. Pardo).

Sowemimo-Coker, & Krauss, 2013), *fundamental frequency* (F0; e.g., Gregory & Webster, 1996; Levitan & Hirschberg, 2011), *long-term average spectra* (LTAS; e.g., Gregory & Webster, 1996), *mel-frequency cepstral coefficients* (e.g., Bailly & Martin, 2014), *voice quality* (e.g., Levitan & Hirschberg, 2011), *voice onset time* (e.g., Nielsen, 2011; Sanchez, Miller, & Rosenblum, 2010; Shockley, Sabadini, & Fowler, 2004; Solanki, 2017), *vowel formants* (e.g., Babel, 2012; Delvaux & Soquet, 2007; Pardo, Urmanche, Wilman, & Wiener, 2017), *clicks* (e.g., Gold, French, & Harrison, 2013), *utterance duration* (e.g., Matarazzo, Weitman, Saslow, & Wiens, 1963; Cappella & Planalp, 1981), and *amplitude envelope signals* (Lewandowski, 2012).

Convergence is also apparent when measured holistically through perceptual similarity tasks. Studies using lexical shadowing tasks have demonstrated that utterances produced after exposure to a model talker are perceptually more similar to the model talker's utterances relative to the shadower's baseline utterances (e.g., Babel, McAuliffe, & Haber, 2013; Babel, McGuire, Walters, & Nicholls, 2014; Dias & Rosenblum, 2016; Goldinger, 1998; Miller, Sanchez, & Rosenblum, 2010, 2013; Namy, Nygaard, & Sauerteig, 2002; Nielsen, 2011; Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013; Pardo et al. 2017; Sanchez et al., 2010; Shockley et al., 2004; Walker & Campbell-Kibler, 2015). Similar effects of mutual exposure on perceived similarity between talkers' pronunciation has also been documented during conversational interactions (e.g., Aguilar, Downey, Krauss, Pardo, Lane, & Bolger, 2016; Dias & Rosenblum, 2011; Kim, Horton, & Bradlow, 2011; Pardo, Cajori Jay, & Krauss, 2010; Pardo, Cajori Jay, et al., 2013). In this line of research, it has been shown that perceived similarity also increases as the conversation unfolds and can persist after the conversation (e.g., Pardo, 2006).

Despite an abundance of acoustic and perceptual evidence demonstrating that convergence occurs in many characteristics, patterns of phonetic convergence are extremely variable across measures and speakers, irrespective of the degree of interactivity of communicative tasks (i.e., shadowing task or conversational interaction), frequency of occurrence of lexical items, and talkers' sex or role in conversation (information giver vs receiver) (see Pardo et al., 2017; Pardo, Urmanche, Wilman, Weiner, Mason, Francis, & Ward, 2018). Studies have shown that talkers do not imitate all acoustic phonetic attributes in the same manner—talkers can converge on some acoustic dimensions while diverging on others (e.g., Babel 2012; Bilous & Krauss, 1988; Levitan & Hirschberg, 2011; Pardo et al., 2010; Pardo Gibbons, Suppes, & Krauss, 2012; Pardo, Cajori Jay, et al., 2013; Pardo, Jordan, et al., 2013; Pardo et al. 2017; Walker & Campbell-Kibler, 2015), or even converge on one set of attributes for one set of speakers/model talkers and another set of attributes for another set of speakers (Pardo et al., 2017). Patterns of convergence in one measure within a pair or within a speaker, indeed, do not represent pairs and speakers' overall convergence patterns in other measures except for closely related features (Sazker, 2015; Cohen Priva & Sanker, 2019). Mismatches between patterns of convergence in perceptual and acoustic measures have also been observed and imputed to the fact that listeners use multiple acoustic-phonetic dimensions

simultaneously when asked to judge the global similarity of speech excerpts (e.g., Babel & Bulatov, 2012; Babel, McAuliffe, & Haber, 2013; Clopper & Dossey, 2020; Pardo, 2010; Pardo et al., 2010, 2012, 2013, 2017; Phillips & Clopper, 2011).

This extreme variability in research findings, attributable to the use of different speakers, different data collection and analysis methods, and/or different phonetic features across studies, makes it difficult to determine why speakers make these adjustments in their acoustic-phonetic behaviour. Two major theoretical frameworks address accommodation—The automatic mechanistic account of the *Interactive Alignment Model* (IAM) proposed by Pickering & Garrod (2004, 2013, 2021) and the social approach of Giles' *Communication Accommodation Theory* (CAT) (e.g., Giles, 2016; Giles et al. 1991; Shepard et al., 2001). Detailed descriptions of the two models are offered in nearly all contributions of this collection. According to IAM, convergence in conversation is evoked by an internal talker-based priming mechanism resulting from an automatic link between perception and production. According to CAT, the two major directions of accommodation, convergence and divergence, reflect or express social closeness to or distance from an interlocutor (Soliz & Giles, 2014).

Both accounts of accommodative behavior suggest that convergence is the unmarked pattern. In contrast, divergence occurs when speakers strive to increase social distance or to counteract an exaggerated speech pattern of their interlocutor, possibly trying to entice them to converge (e.g., slowing down a very fast-talking speaker). Divergence after an insult to a talker's language was noted early on in work by Bourhis and Giles (1977), but has rarely been systematically examined in studies of phonetic accommodation. In such cases, the natural tendency to converge may be superseded by a strong motivation to diverge (Lewandowski, 2012). Schweitzer et al. (2017) observed that a decrease in likability between interlocutors can trigger either convergence or divergence on pitch accent realization, suggesting individual differences in accommodative behavior.

Some of the papers in this special issue contribute to an understanding of divergence. For example, Earnshaw (2021) finds that participants diverged from an interlocutor when playing the mock police interview task, which is characterized by a high degree of formality and asymmetric power dynamics, as compared to when interacting in informal, symmetric paired casual conversations. Weise, Silber-Varod, Lerner, Hirschberg, & Levitan (2020) suggest that speakers of Hebrew may have a stronger tendency to diverge than speakers of Indo-European languages. Gessinger, Möbius, Le Maguer, Raveh, and Steiner (2021) find very few cases of divergence that can perhaps even be argued to be outliers in the light of overwhelmingly more frequent cases of convergence. Note that Gessinger et al. exposed their experimental subjects to synthesized speech; even voices that are clearly recognizable as being synthetic tend to trigger convergence, which may indicate that convergence is a preferred communicative style.

While patterns of variation in accommodation challenge automatic priming-based accounts of convergence grounded in a perception-production link (IAM), likewise, explanations rooted exclusively on talkers' strategies to create, maintain, or decrease social distance (CAT) fail to account for talkers

converging to a model talker in non-interactive settings, when instructed to avoid imitation, or without explicit instructions to imitate as documented in various papers of this collection (e.g. Kwon, 2021; Lin, Yao, & Luo, 2021; Wagner, Broersma, McQueen, Dhaene, & Lemhöfer, 2021; Ross, Lilley, Clopper, Pardo, & Levi, 2021). Both models also overlook effects of linguistic, personality, and cognitive factors in directing the accommodation behaviour of the speaker. Studies have demonstrated that the degree of convergence varies depending on lexical characteristics of utterances (Goldinger, 1998; Goldinger & Azuma, 2004; Nielsen, 2011), phonetic distance between interlocutors' language repertoires (Babel, 2012; Walker & Campbell-Kibler, 2015; Walters, Babel, & McGuire, 2013), vowel dispersion differences (Lewandowski & Nygaard, 2018), cognitive load involved in a task (Abel & Babel, 2017), and participants' openness and neuroticism (Lewandowski & Jilka, 2019; Yu, Abrego-Collier, & Sonderegger, 2013).

Given the complex nature of evidence for either account, it is not surprising that numerous hybrid approaches emerged. For example, one hybrid approach attempts to combine the automatic and social perspectives in a more prevailing view that integrates the effect of "extra-social" factors on accommodation (Babel, 2012; Pardo, 2012; Pardo et al., 2017). In this view, social, linguistic, cognitive, and personality factors are seen as modulators of accommodation in that they strengthen or weaken internal cognitive links between perception and production. Other hybrid views favor a *Conversational Synergy Account* (Fusaroli, Raczaszek-Leonardi, & Tylén, 2014; Fusaroli & Tylén, 2016) that is studied in detail in this issue by Olmstead, Viswanathan, Cowan, and Yang (2021).

## 2. The special issue

This special issue was inspired by the *Workshop on Accommodation in Speech Communication* held at the *University of Zurich* in 2018. The workshop brought together researchers from different theoretical fields in order to discuss topics and themes that were relevant to the understanding of human–human and human–machine accommodation. Consistent with the inclusive approach of the workshop, this special issue includes an assortment of contributions (not limited to workshop presentations) that examined the complexity of phenomenon from a rich array of theoretical, methodological, cognitive, socio-cultural and linguistic perspectives.

Some of the fundamental questions that this special collection aimed to cover concerned (1) the nature of vocal accommodation in terms of underlying mechanisms and social functions in human–human and human–computer interaction; (2) the effect of task-specific and talker-specific characteristics (gender, age, personality, linguistic and cultural background, role in interaction) on degree and direction of convergence towards human and computer interlocutors; (3) integration of articulatory, perceptual, neurocognitive, and/or multimodal data to the analysis of acoustic accommodation in interactive and non-interactive speech tasks; (4) 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; and (5) the implications of accommodation for human and machine

speaker recognition, language learning technologies, and speech rehabilitation.

The ten papers comprising this special edition cover different perspectives on vocal accommodation, ranging from acoustic adjustments to native and non-native speech and across dialects, to accommodation in interaction with a simulated spoken dialogue system. Accommodation is elicited in tasks along a continuum between more and less interaction between the study participants. On the less interactive end of the continuum, studies used a variety of tasks—imitation, repetition, shadowing, picture-naming, and word matching. On the more interactive end, there were picture matching tasks, mock police interviews, paired conversations, as well as the traditional map task. Convergence, divergence, or maintenance have been quantified primarily acoustically, with some integrating acoustic and perceptual measures of convergence. Production and perception data were collected from a variety of languages, the majority of which have a well-established tradition of research on accommodation, including L1 and L2 varieties of American English, regional varieties of British English, Dutch, and German. The collection also covers Semitic and Sino-Tibetan languages, Hebrew and Hong Kong Cantonese, that have received comparatively little attention, but can shed light on universals and cross-language and cross-cultural variation in accommodation. Orthogonal to this, acoustic and perceived convergence are analysed in language-dependent and/or independent features through a variety of statistical models (e.g., Linear Mixed Effect Models, Generalized Additive Model, and Machine Learning Classifiers) and discussed in terms of the function convergence may have served under the given experimental conditions and the theoretical accounts supported by the findings.

In line with the array of questions that this special issue aims to cover, the overview of papers starts with those dealing with questions related to the nature of accommodation in terms of its underlying mechanisms—forms that acoustic adaptations take on and their communicative relevance. The first paper by Ostrand and Chodroff (2021) calls directly into question the view of alignment offered by IAM as a unitary, automatic process. The authors address this question by examining whether (a) within the same dialogue alignment occurs synchronously at various linguistic levels (e.g. phonetic and syntactic), and (b) alignment in one phonetic feature primes alignment in other features within the same level. They collected conversational data from monolingual American English speakers playing a picture-matching game with partners having distinct phonetic profiles (Mandarin accented non-native speakers). Using individual and holistic measures of convergence, and combining traditional and machine learning methods, they found that speakers aligned on some linguistic levels and in some (but not all) features within the same level. They interpret these results as empirical support for the view that convergence does not happen automatically across linguistic levels. In line with view of Ostrand and Ferreira (2019) that conversational adaptations serve the function of communicative utility, Ostrand and Chodroff suggest that variability in alignment features depends on their relevance for a partner-specific linguistic profile and likeability in order to enhance communicative success during dialogue.

The view of conversational adaptations as a means to promote communicative success is also shared by Olmstead

et al. (2021). In a similar approach to that of Ostrand and Chodroff (2021), Olmstead et al. investigate speech adaptations in interactions between native and non-native speakers (e.g., clear speech, phonetic convergence, divergence, or maintenance) when communication success is not assured. Their study involves a word matching task in a group of Mandarin speakers of L2 English interacting with both Mandarin speakers and a native English speaker. The items include words that would engender miscommunication problems due to the presence of segments—medial vowel (/l/, /i/) and voicing of the final consonant (/t/, /d/)—which are realized differently between L1 and L2 English speakers. They find that the patterns of adjustments to indicate vowel identity or the voicing contrasts (only spectral, only duration, duration and spectral) are distinct between same and different language pairs, and are not reducible to either clear speech or the trichotomy direction of accommodation postulated by CAT. Interestingly, neither IAM, nor CAT, but the *Conversational Synergy Account* (Fusaroli, Raczaszek-Leonardi, & Tylén, 2014; Fusaroli & Tylén, 2016) is offered as a more effective model of the interactive, multidimensional phonetic change happening between L1 and L2 interlocutors.

Likewise, Tobin (2022) proposes that vocal accommodation is articulatorily, sociolinguistically, and temporally structured. His contribution examines the effects of linguistic background (mono- vs bilingualism) and ingroup/outgroup dynamics on patterns of voiceless VOT accommodation in Korean-English bilinguals, Spanish-English bilinguals, and English monolinguals shadowing an American English model talker. The paper offers an original perspective that ascribes cross-linguistic differences in VOT accommodation to the preference of a speakers' native language for stable/unstable articulatory coordination underlying VOT production (Spanish: stable, near-zero VOTs; Korean: unstable, long VOT; English: intermediate between Spanish and Korean). Methodologically, VOT times are analysed in two ways—differences from baseline-to-test and trial-specific difference from baseline. The results reveal that VOT accommodation is modulated by L1 laryngeal-oral coordination and phonetic distance, with speakers of Korean showing more VOT accommodation than Spanish. The study also finds that the time-course of accommodation varies between mono- and bilingual talkers, with the former converging only at the beginning of the task, while the latter exhibit immediate and sustained convergence over the time. The fact that the two bilingual groups (Korean-English and Spanish-English) showed greater convergence than the English monolinguals is interpreted as an indication that outgroup members converge, expressing their affiliation with their interlocutors more than the dominantly monolingual community. The results of Tobin's study speak in favor of a socio-experiential view of accommodation (aligning with CAT), in which automatic mechanisms play little to no role.

The effects of dialect-specific features and dialect familiarity on the extent and perceptibility of phonetic convergence further challenge integrated production-perception models of convergence, as demonstrated in the study by Ross et al. (2021). This study examines utterances from speakers of Mid-Atlantic and General American dialects who shadowed words with and without dialect distinguishing features, prompted by model talkers with comparable dialectal backgrounds. Listeners with dif-

ferent degree of exposure to either dialect rated the perceptual similarity of the shadowed words to the original productions of the model talkers in AXB tasks. Acoustic and perceptual assessments of convergence showed that dialect-specific features evoked more phonetic convergence. Dialect familiarity affects the magnitude of convergence towards dialect-specific variants, but this has little to no effect on the perceptibility of cross-dialectal convergence. The results of this study, highlighting the role of socio-linguistic factors on convergence, point to the need for hybrid models of accommodation which social and linguistic information driving variation in the magnitude of convergence.

Another contribution on accommodation that argues in favor of hybrid models of accommodation is that of Wagner et al. (2021). In this study, the aim is to examine (1) whether convergence towards non-native speech is possible in the absence of an interlocutor, and (2) to what degree the perceived strength of a non-native accent influences convergence. To elicit convergence, Dutch native speakers participated in a fake memory task consisting of repeating backwards word sequences produced by a Croatian-accented Dutch model speaker. Critical items contained segments which the model talker realized in a non-standard Dutch manner. Convergence was measured perceptually via an AXB task and acoustically in numerous dimensions by comparing baseline items and repeated items with that of the model speaker. Unlike the results of papers mentioned above emphasizing the communicative utility of acoustic adaptations in native vs non-native speaker setting, the results of the study by Wagner et al. show that convergence to non-native speech also occurs in non-interactive settings, in the absence of specific perceived conversational demands, but the amount of convergence is modulated by the strength of perceived accent. Likewise, a recent study by Gnevshva, Szakay, and Jansen (2021) finds that L2 speakers may have more malleable linguistic representations that lead to greater convergence toward L1 model talkers.

A more linguistic/phonological perspective on convergence is offered in the paper by Kwon (2021). In this study, convergence is conceptualized as the (presumably) unintentional changes in the direction of the model speech, and the focus is on the imitation pattern of non-primary cues in English monolinguals and Korean-English bilinguals. Specifically, Kwon poses the question of whether it is the phonology of the native language (Korean or English) or the language of the task (English) that mediates the spontaneous imitation of bilingual speakers, when an acoustic property plays a contrastive role in one language but not in the other. The methodology involves a shadowing task with English monolinguals and Seoul Korean-English bilinguals using English voiceless stops manipulated by either raising post-stop F0 (the primary cue for the Seoul Korean aspirated stops) or lengthening VOT (the primary cue for English voiceless stops). Convergence was assessed by comparing the stop VOT and post-onset F0 in baseline, shadowed, and post shadowed productions. Results reveal that English monolingual shadowers produced /t/ with similar VOT in both F0 and VOT manipulation conditions, while they enhanced the post-stop F0 only when the model speech had the particular property enhanced. The bilingual shadowers, on the other hand, imitated the manipulated cue phonetically without following their native language

phonology. In post-shadowing productions, instead, they did not retain the acoustic manipulations in the shadowed stimuli, but attended to the phonology of the language of the task. Although the study was not designed to examine the intentional or subconscious nature of bilingual imitation, Kwon suggests that stronger imitative changes in bilingual speakers' shadowed productions (i.e., the VOT lengthening in the long VOT condition) are also amenable to interpretations based on socio-motivational factors, i.e., the achievement of a native-like pronunciation.

A cross-cultural and cross-linguistic perspective on accommodation is offered in the paper by [Weise et al. \(2020\)](#). This study involves monolingual speakers of Hebrew, and offers itself as one of the few systematic studies examining acoustic-prosodic entrainment in dyadic interactions in this language. The focus here is on the role of gender and conversational role in degree and direction of entrainment, possibly reflecting fundamental differences in socialization and rigidity of role-assignment patterns between American and Israeli societies. The study prompted semi-spontaneous utterances from participants completing map tasks in which they alternated Instruction Giver and Receiver roles across same gender and mixed gender pairs. They find that Hebrew speakers accommodate similarly at turn exchanges as speakers of Indo-European languages do but, unlike the latter, Hebrew speakers are more prone to diverge, especially in mixed-sex speaker pairs. Speaker sex does not play a significant role in entrainment behaviour, while the role in conversation does to a greater extent, with information receivers tending to match their information giver more closely than vice versa. Discussing their data through a cross-linguistic and cross-cultural lens, the authors argue that factors other than the language (e.g., cultural, methodological differences between studies) can account for the observed differences between Hebrew speakers and speakers of Indo-European languages.

In addition to questions related to the nature of accommodation and the effect of talker-specific, task-specific, and language-specific information on convergence, another perspective on accommodation that is covered more extensively in this special collection is that of language or dialect in change, which is addressed by [Lin et al. \(2021\)](#). This paper examines whether an ongoing tonal merger in Hong Kong Cantonese (mid-level Tone 3 and low-level Tone 6) can be reversed upon imitation of a model talker with more conservative tonal productions. The authors collected data from shadowers with strong and weak merging trends in their baseline productions who repeated words after young and old model talkers. The degree of imitation was not consistently modulated by the age gap between shadowers and model talkers, nor by the perceived personality of the model talker nor attitudes towards ongoing changes in the language. More imitation was instead observed in speakers with more merged productions in the baseline and towards the speaker featuring greater tonal distinction. In line with Trudgill's hypothesis that short term accommodation can bring about language change if speakers accommodate frequently enough (Trudgill, 1986), the results of this study together with those obtained in [Ross et al. \(2021\)](#) suggest that phonetic convergence can create favourable conditions for the diffusion or the reversal of sound changes. However, given that dynamics of language variation

and change are intricately connected to verbal communication in face-to-face situations, further research exploring the contribution of short-term accommodation to long-term language change is warranted.

A remaining theme of this collection is the implications that accommodation has for human and machine speaker recognition. This topic is addressed by [Earnshaw's \(2021\)](#) paper, which investigates accommodation from a forensic phonetics perspective. By examining within-speaker variability in West Yorkshire FACE vowel (/ɛi/) production as a function of speaking styles and interlocutors, this study attempts to unravel the impact of accommodation on forensic speaker comparison, where the typical task is to decide whether the sample of a perpetrator and that of a suspect derived from the same or different speakers. Results from conversational speech data elicited via mock police interviews, paired conversations, and answer-phone messages show that, despite considerable amount of between-speaker and between-task variability in accommodation, for most participants, the changes in FACE (/ɛi/) productions were relatively small. These preliminary data suggest that accommodation might not be a tremendous threat in forensic speaker comparison. However, given the novelty and relevance of this topic for voice recognition technologies, more research is required to ascertain to what degree similarity in speech acoustics influences the similarity in speaker identities.

The papers discussed so far have addressed questions regarding the nature of accommodation and the effect of speaker-specific characteristics on the extent and perceptibility of accommodation in the domain of human–human interaction. Similar questions can also be examined in the domain of human–machine interactions. The paper by [Gessinger et al. \(2021\)](#) examines which features humans converge to when interacting with a machine, providing empirical evidence that can be implemented in the design of more effective spoken dialogue systems, as well as computer-assisted language learning technologies. Using a Wizard of Oz experimental paradigm, this study examines accommodation patterns of German native speakers interacting with a mock automatic spoken dialogue system simulating a virtual tutor for learning the German language. Of the three phonetic forms under examination, wh-question intonation, allophonic variations in the realization of the word ending <ig> ([ɪç] vs [ɪk]), and long vowel <-ä-> [ɛ:] vs [e:], only the first two evoked convergence. The pattern did not vary depending on the naturalness of the dialogue system's speech, but a higher degree of suprasegmental convergence was found for more neurotic speakers. Thus, human speakers appear to converge toward machines in similar ways to more natural utterances, with variation across different attributes and modulated by individual and social factors.

### 3. Future directions & conclusions

The contributions in this collection raise new questions for research on the effects of similarity and naturalness between interlocutors' voices. In the context of native-nonnative communication, [Wagner et al. \(2021\)](#) find that non-native accents trigger convergence, but more extreme accentedness of a model speaker reduced the tendency of native shadowers to converge. In the context of human–machine interactions,

Gessinger et al. (2021) and other studies on the topic show that shadowing synthetic voices evokes convergence but to a lower degree than natural voices (Gessinger et al., 2017; Gessinger et al., 2018). Likewise, research on Computer Assisted Pronunciation Training has emphasized the importance of student/teacher voice similarity for the enhancement of pronunciation skills (Felps, Bortfeld, & Gutierrez-Osuna, 2009; Probst, Ke, & Eskenazi, 2002). A welcome line of research would thus involve studies testing similar predictions for phonetic convergence in both interactive and non-interactive settings. For example, one could test groups of interlocutors or shadowers and model talkers, paired based on a scale of increasing acoustic and perceived similarity in terms of voice quality characteristics, accentedness, or attractiveness, and examine whether the amount of convergence increases as similarity increases. Notions about the role of voice (dis)similarity in accommodation might help clarify the role of speakers' gender in accommodation. To date, there are very few studies on this topic, with inconsistent findings regarding whether greater between-talker similarity (Kim et al., 2011) or greater dissimilarity (Walker & Campbell-Kibler, 2015) is associated with greater convergence.

Individuals in accommodation are also interesting from other perspectives. In addition to different accommodation strategies in a multidimensional accommodation space, the situational and stationary vocal characteristics of a speaker play a central role in communication. Vocal utterances contain information about a speaker's state and individuality that are typically central for specifying the message (Dellwo, Huckvale, & Ashby, 2007) and do not seldom override linguistic information. The production of an utterance in a happy or sad voice, for example, can categorically change its meaning. But speakers also signal their individuality which, depending on the situation, can sometimes be more hidden (e.g., in voice disguise) or more exposed (e.g., in charismatic voices, Rosenberg & Hirschberg, 2005), and even to some degree controlled (Dellwo, Pellegrino, He, & Kathiresan, 2019). Applying a higher degree of vocal individuality to a signal, however, seems to stand in opposition to vocal accommodation between individuals. Indeed, there is new evidence from animal research showing that accommodation and individualisation in Marmoset vocalisations varies depending on the function of the call sign (Zürcher, Willems, & Burkart, 2021).

In humans and machines, it has been shown that vowel convergence does not have a detrimental effect on speaker discriminability (Pellegrino, Kathiresan, & Dellwo, *in press*), but deliberate forms of voice imitation can affect voice recognition (e.g. Farrús et al., 2010; López et al., 2013), thus revealing that individuality interacts with accommodation behaviour in a complex fashion. To summarize, vocal accommodation seems to stand in competition with a variety of other signalling functions of vocal utterances, in particular the function of signalling individuality. There are various ways in which such a dichotomy can be solved. One way might be to change signalling functions over time as observed in marmosets (Zürcher et al., 2021), i.e., to show more or less accommodation for certain utterances and individualising others (this may even be at a word level). Another way might be to make use of the many acoustic dimensions available in vocal signals that permit accommodation on some dimensions while individualising on

others. It is, of course, also possible that a mixture of the two approaches exist.

To quantify vocal accommodation, several of the papers in this special issue have used methods that evaluate the difference in distance (DID) between speech parameters of the interlocutors. DID estimates of convergence first compare a baseline/pre-exposure and model utterance parameter (pre-exposure difference), then the post-exposure and model utterance parameter (post-exposure difference), and then subtract the post-exposure difference from the pre-exposure difference, resulting in a positive estimate when convergence occurs (some studies have reversed the final difference calculation, deriving negative estimates to indicate convergence). Cohen Priva and Sanker (2019) have pointed out a number of limitations of this approach. For instance, DID estimates might indicate maintenance or divergence in cases of exaggerated/overshooting convergence; they might underestimate the degree of convergence when the initial pre-exposure distances are already small; and baseline utterance parameters might not be representative of a speaker's typical patterns and thus inflate the presumed effect of an interlocutor. In one of the papers of this special issue (Ross et al., 2021), the DID approach and the linear combination analysis were compared, revealing that DID estimates can be useful when integrating both approaches. In light of these concerns, studies of phonetic convergence should continue to incorporate multiple measures, including both holistic and fine-grain acoustic measures and multivariate analytical approaches.

Most of the research on vocal accommodation as well as several papers in this special issue have collected and analysed primarily acoustic-auditory data, avoiding visual contact between participants. In Weise et al. (2020), instead, participants were able to see each other and authors questioned whether the differences between their study findings and those obtained in other research using same methodology (e.g., Levitan & Hirschberg 2011; Pardo, 2006) could depend on the modality, audio-visual vs audio only, in which the task was performed. The question remained unanswered and authors called for future investigations since the limited research on the topic does not yet provide a clear picture about the role of modality on accommodation. In some studies, audio-visual information enhanced accommodation (Dias & Rosenblum, 2011, 2016; Miller et al., 2010; Sanchez et al., 2010). In a more recent study, however, the audio-only modality evoked more convergence (Savino et al., 2018).

Given that most social interactions are face-to-face, including those happening via video conferencing software, it is important to investigate multimodal accommodation. In view of the strong correspondence between facial movements and the acoustics of the corresponding voice (Rosenblum & Dorsi, 2021), it is important to examine to what extent accommodation is consistent across voice and facial information, or whether one channel is more susceptible to the influence of communication dynamics than the other. It also seems plausible that gestural or communicative posture information could enhance vocal accommodation. A study by Shockley, Santana, and Fowler (2003) found that participants who conversed directly exhibited similar postural sway patterns (see also Shockley, Baker, Richardson, & Fowler, 2007; Shockley, Richardson, & Dale, 2009). It would be interesting to test

whether different models, like gestural and vocal modes, co-occur and whether one can predict the other. It is possible, for example, that a speaker who applies postures of another speaker during a conversation also shows increased amounts of vocal accommodation during these intervals. Beyond opening a novel, integrative perspective in the study of accommodation, the results of research on mechanisms, forms, and factors of multimodal accommodation can enhance an understanding of communication and facilitate interactional exchange between human interlocutors and between humans and machines.

Future research should more rigorously assess aspects of the time course of accommodation—how much exposure is minimally necessary for convergence to occur, how much sustained exposure is necessary for maximal convergence, how long convergence persists beyond exposure, and how other factors interact with time to modulate the degree of initial and sustained convergence. The studies on convergence in shadowing tasks suggest that acoustic adjustments toward model speakers happen with very brief exposure, but there are other factors that modulate accommodation. Examples of modulating factors from this special issue include perceived accentedness of the model talker (Wagner et al., 2021), the acoustic distance between shadower and model speakers' productions (Lin et al., 2021), the interplay between the language of the shadowers and the phonetic-phonological characteristics of the lexical items (Kwon, 2021), and the presence of dialect-specific features (Ross et al., 2021).

Brief exposure, however, is not necessarily sufficient to trigger adjustments when accommodation is examined in interaction. While the two bilingual groups studied by Tobin (2022) showed convergence from the very beginning of the task as well as sustained convergence over the course of interaction (see also Pardo, 2006), in other studies, the short duration of the paired tasks has been listed among the factors explaining the limited effect of exposure time on convergence (Earnshaw, 2021), or the difference in the type of obtained convergence (only local vs local and global) in comparison with studies with similar research design (Weise et al., 2020). Although existing evidence of limited convergence after long-term exposure to a particular talker (e.g. Pardo et al., 2012) speaks against a prominent role of sustained exposure on accommodation, understanding the time needed for accommodation to occur and exert communicative or social benefits may be instead relevant in interactive contexts of mixed language/dialect pairs (also observed in this collection by Olmstead et al., 2021; Ostrand & Chodroff, 2021; and Tobin, 2022) where communicative success is not assured.

Finally, it is important to acknowledge that while some previous studies have indicated that talkers sometimes converge on suprasegmental prosodic attributes such as fundamental frequency (e.g. Babel & Bulatov, 2012) and duration/speaking rate (e.g., Pardo et al., 2017), more sophisticated approaches to prosody have not been applied to phonetic convergence (see e.g., Cho, 2022; Shattuck-Hufnagel & Turk, 1996). According to Cho (2022), prosodic encoding plays a part in speech production planning and influences segmental phonetic realization in a variety of ways at the level of phonetic encoding. Prosodic structure influences both suprasegmental and segmental features, but utterances that participate in sim-

ilar prosodic structures do not necessarily share the same phonetic features. This consideration reveals an additional source of variation that complicates measures of phonetic convergence. That is, the emphasis or prominence of a particular utterance, especially in more naturalistic conversational speech settings, affects realization of phonetic features in complex ways. Thus, measures of individual acoustic-phonetic attributes are susceptible to prosodic context in ways that might indicate divergence (such as when one talker produces an utterance in stressed prominence while their partner might de-emphasize the utterance), while other articulatory parameters are contextually convergence. This consideration could be a major reason for contrasting findings of convergence in duration and divergence in fundamental frequency in some previous studies, differences observed between conversational and shadowed speech convergence, and differences between holistic and individual acoustic measures of convergence (Pardo et al., 2017; Pardo et al., 2018). A fruitful avenue of future research could start with a comparison of convergence on items produced in same versus different prosodic structures.

Taken together, the findings of the studies in this special issue imply that the description and understanding of adjustments that speakers make while communicating with humans or machines, no matter how they are elicited and measured, must move beyond the dichotomy between automatic mechanisms and social factors. The results and discussions offered in this collection and beyond support hybrid accounts of accommodation that incorporate internal-cognitive mechanisms that, in turn, operate in connection with the multitude of factors (e.g., social, linguistic, personality) which have been shown to modulate convergence. As suggested in the accommodation literature and in this collection (see especially Gessinger et al., 2021), at the center of these factors is the speaker's idiosyncratic personality, cultural and linguistic background, discursive positioning, and manner of negotiating (linguistic-acoustic) identity and meaning in conversation. Recognizing the centrality of speaker individuality in speech communication can thus provide an interpretative key to the extreme variability in convergence documented in the accommodation literature and in this collection. Along this line, Olmstead et al. (2021) go a step further by recommending (a) focusing on the dyads, rather than individuals, as the unit of analysis of accommodation in interaction and (b) interpreting the variability in research findings in the light of the dynamic strategies that dyad members adopt to promote communication success, that do not necessarily coincide with convergence.

### Acknowledgments

We are extremely grateful for the patience and efforts of the Journal Editor, Dr. Taehong Cho, who provided essential guidance as we navigated the unique challenges presented over the course of the past two years. We also gratefully acknowledge the dedication and persistence of the contributors and reviewers involved in bringing about this special edition. This project was partially supported by the URPP Language and Space and the NCCR Evolving Language, Swiss National Science Foundation Agreement #51NF40\_180888, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 232722074 SFB 1102 and MO 597/6-2.

## References

- Abel, J., & Babel, M. (2017). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech*, 60(3), 479–1402. <https://doi.org/10.1177/0023830916665652>.
- Aguilar, L. J., Downey, G., Krauss, R. M., Pardo, J. S., Lane, S., & Bolger, N. (2016). A dyadic perspective on speech accommodation and social connection: Both partners' rejection sensitivity matter. *Journal of Personality*, 84, 165–177. <https://doi.org/10.1111/jopy.12149>.
- Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39, 437–456. <https://doi.org/10.1017/S0047404510000400>.
- 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.jwocn.2011.09.001>.
- Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language & Speech*, 55, 231–248. <https://doi.org/10.1177/0023830911417695>.
- Babel, M., McAuliffe, M., & Haber, G. (2013). Can mergers-in-progress be unmerged in speech accommodation. *Frontiers in Psychology*, 4, 653.
- Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, 5. <https://doi.org/10.1515/lp-2014-0006>.
- Bailly, G. & Martin, A. (2014). Assessing objective characterizations of phonetic convergence. *Interspeech 2014 - 15th Annual Conference of the International Speech Communication Association*, Sep 2014, Singapore, 19–9.
- Borrie, S. A., & Liss, J. M. (2014). Rhythm as a coordinating device: Entrainment with disordered speech. *Journal of Speech, Language, and Hearing Research*, 57, 815–824. [https://doi.org/10.1044/2014\\_JSLHR-S-13-0149](https://doi.org/10.1044/2014_JSLHR-S-13-0149).
- Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity, & intergroup relations* (pp. 119–135). London: Academic Press.
- Cappella, J. N., & Planalp, S. (1981). Talk and silence sequences in informal conversations III: Interspeaker influence. *Human Communication Research*, 7, 117–132. <https://doi.org/10.1111/j.1468-2958.1981.tb00564.x>.
- 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>.
- Cho, T. (2022). Linguistic functions of prosody and its phonetic encoding with special reference to Korean. In K. Horie, K. Akita, Y. Kubota, D. Y. Oshima and A. Utsugi (Eds.), *Japanese/Korean Linguistics 29* (pp. 1–24).
- 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).
- Cohen Priva, U., & Sanker, C. (2019). Limitations of difference-in-difference for measuring convergence. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 10, 1–29. <https://doi.org/10.5334/labphon.200>.
- Dellwo, V., Huckvale, M., & Ashby, M. (2007). How is individuality expressed in voice? An introduction to speech production and description for speaker classification. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4343 LNAI(Chapter 1), 1–20. [https://doi.org/10.1007/978-3-540-74200-5\\_1](https://doi.org/10.1007/978-3-540-74200-5_1).
- Dellwo, V., Pellegrino, E., He, L., & Kathiresan, T. (2019). The dynamics of indexical information in speech: Can recognizability be controlled by the speaker? *Acta Universitatis Carolinae Philologica* 2, 2019(2), 57–75. <https://doi.org/10.14712/24646830.2019.18>.
- Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64, 145–173.
- Dias, J. W., & Rosenblum, L. D. (2011). Visual influences on interactive speech alignment. *Perception*, 40, 1457–1466. <https://doi.org/10.1068/p7071>.
- Dias, J. W., & Rosenblum, L. D. (2016). Visibility of speech articulation enhances auditory phonetic convergence. *Attention, Perception, & Psychophysics*, 78, 317–333. <https://doi.org/10.3758/s13414-015-0982-6>.
- Earnshaw, K. (2021). Examining the implications of speech accommodation for forensic speaker comparison casework: A case study of the West Yorkshire FACE vowel. *Journal of Phonetics*, 87. <https://doi.org/10.1016/j.jwocn.2021.101062>.
- Edlund, J., Heldner, M., & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. In *INTERSPEECH-2009*.
- Felps, D., Bortfeld, H., & Gutierrez-Osuna, R. (2009). Foreign accent conversion in computer assisted pronunciation training. *Speech Communication*, 51(10), 920–932.
- Fusaroli, R., Raczaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32, 147–157.
- Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive Science*, 40(1), 145–171.
- Gasiorek, J., Giles, H., & Soliz, J. (2015). Accommodating new vistas. *Language & Communication*, 41, 1–5. <https://doi.org/10.1016/j.langcom.2014.10.001>.
- Gessinger, I., Möbius, B., Le Maguer, S., Ravesh, E., & Steiner, I. (2021). Phonetic accommodation in interaction with a virtual language learning tutor: A Wizard-of-Oz study. *Journal of Phonetics*, 86. <https://doi.org/10.1016/j.jwocn.2021.101029>.
- Gessinger, I., Ravesh, E., Le Maguer, S., Möbius, B., & Steiner, I. (2017). Shadowing synthesized speech – segmental analysis of phonetic convergence. In 3801 (pp. 3797). Stockholm: Interspeech. <https://doi.org/10.21437/Interspeech.2017-1433>.
- Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, 15, 87–109.
- Giles, H. (2016). *Communication accommodation theory: Negotiating personal and social identities in context*. Cambridge: Cambridge University Press.
- Giles, H., Coupland, J., & Coupland, N. (1991). Chapter 1: Accommodation Theory: Communication, context, and consequence. In *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge: Cambridge University Press.
- Gnevshcheva, K., Szakay, A., & Jansen, S. (2021). Phonetic convergence across dialect boundaries in first and second language speakers. *Journal of Phonetics*, 89. <https://doi.org/10.1016/j.jwocn.2021.101110>.
- Gold, E., French, P., & Harrison, P. (2013). Clicking behavior as a possible speaker discriminant in English. *Journal of the International Phonetic Association*, 43(3), 339–349.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.
- Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming Retrieved from. *Psychonomic Bulletin & Review*, 11, 716–722 [http://www.public.asu.edu/~sgolding/docs/pubs/Goldinger\\_Azuma\\_PBR\\_04.pdf](http://www.public.asu.edu/~sgolding/docs/pubs/Goldinger_Azuma_PBR_04.pdf).
- 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, 1231–1240.
- 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.
- Kwon, H. (2021). A non-contrastive cue in spontaneous imitation: Comparing mono- and bilingual imitators. *Journal of Phonetics*, 88. <https://doi.org/10.1016/j.jwocn.2021.101083>.
- Levitán, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *INTERSPEECH-2011* (pp. 3081–3084). Italy: Florence.
- Lewandowski, E. M., & Nygaard, L. C. (2018). Vocal alignment to native and non-native speakers of English. *Journal of the Acoustical Society of America*, 144, 620–633. <https://doi.org/10.1121/1.5038567>.
- Lewandowski, N. (2012). *Talent in nonnative phonetic convergence*. Ph.D Stuttgart.
- Lewandowski, N., & M. Jilka (2019). "Phonetic convergence, language talent, personality & attention". In: *Frontiers in Communication* 4.18. doi: 10.3389/fcomm.2019.00018
- Lin, Y., Yao, Y., & Luo, J. (2021). Phonetic accommodation of tone: Reversing a tone merger-in-progress via imitation. *Journal of Phonetics*, 87. <https://doi.org/10.1016/j.jwocn.2021.101060>.
- Manson, J. H., Bryant, G. A., Gervais, M. M., & Kiine, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution & Human Behavior*, 34, 419–426. <https://doi.org/10.1016/j.evolhumbehav.2013.08.001>.
- Matarazzo, J. D., Weitman, M., Saslow, G., & Wiens, A. N. (1963). Interviewer influence on durations of interviewee speech. *Verb Learn Verb Behaviour*, 1, 451–458.
- Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics*, 72, 1614–1625. <https://doi.org/10.3758/APP.72.6.1614>.
- Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2013). Is speech alignment to talkers or tasks? *Attention, Perception, & Psychophysics*. <https://doi.org/10.3758/s13414-013-0517-y>.
- 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>.
- Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132–142. <https://doi.org/10.1016/j.jwocn.2010.12.007>.
- 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. *Journal of Phonetics*, 87. <https://doi.org/10.1016/j.jwocn.2021.101054>.
- Ostrand, R., & Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. *Journal of Phonetics*, 88. <https://doi.org/10.1016/j.jwocn.2021.101074>.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119, 2382–2393. <https://doi.org/10.1121/1.2178720>.
- Pardo, J. S. (2010). Expressing oneself in conversational interaction. In E. Morsella (Ed.), *Expressing Oneself/Expressing One's self: Communication, Cognition, Language, and Identity* (pp. 183–196). London: Taylor & Francis.
- Pardo, J. S. (2012). Reflections on phonetic convergence: Speech perception does not mirror speech production. *Language & Linguistics Compass*, 6, 753–767. <https://doi.org/10.1002/linc.3.367>.
- Pardo, J. S., Cajori Jay, I., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., & Krauss, R. M. (2013). The influence of role-switching on phonetic convergence in conversation. *Discourse Processes*, 50, 276–300. <https://doi.org/10.1080/0163853X.2013.778168>.
- Pardo, J. S., Cajori Jay, I., & 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., 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.jwocn.2011.10.001>.
- 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 & Language*, 69, 183–195. <https://doi.org/10.1016/j.jml.2013.06.002>.
- Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, 79, 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. *Journal of Phonetics*, 69, 1–10.
- Pellegrino, E., Kathiresan, T. & Dellwo, V. (in press). Towards understanding the effects of vowel convergence on speaker discrimination in humans and machines: Evidence from Swiss German dialects. *International Journal of Speech, Language and the Law*.
- Pentland, A. (2008). *Honest signal: How they shape our world*. Cambridge: MIT Press.
- Phillips, S., & Clopper, C. G. (2011). *Perceived imitation of regional dialects*. Proceedings from Proceedings of Meetings on Acoustics.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02), 169–190.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral & Brain Sciences*, 36, 49–64. <https://doi.org/10.1017/S0140525X12003238>.
- Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge: Cambridge University Press.
- Probst, K., Ke, Y., & Eskenazi, M. (2002). Enhancing foreign language tutors – In search of the golden speaker. *Speech Communication*, 37, 161–173.
- 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.
- Rosenberg, A., & Hirschberg, J. (2005). *Acoustic/Prosodic and lexical correlates of charismatic speech*. Academic Commons: Columbia University.
- Rosenblum, L. D. (2021). Primacy of multimodal speech perception for the brain and science. In J. S. Pardo, L. C. Nygaard, R. E. Remez, & D. B. Pisoni (Eds.), *The handbook of speech perception* (second ed., New York: Wiley).
- 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. *Journal of Phonetics*, 86. <https://doi.org/10.1016/j.wocn.2021.101041>
- Ruch, H., Zürcher, Y., & Burkart, J. M. (2018). The function and mechanism of vocal accommodation in humans and other primates. *Biological Reviews*, 93, 996–1013.
- Sanker, C. (2015). Comparison of phonetic convergence in multiple measures. In *Cornell Working Papers in Phonetics and Phonology*, 60–75.
- 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.
- Sanker, C. (2015). Comparison of phonetic convergence in multiple measures. In *Cornell Working Papers in Phonetics and Phonology*, 60–75.
- Schweitzer, A., & Lewandowski, N. (2013). *Convergence of articulation rate in speech*. Proceedings from in INTERSPEECH-2013.
- Schweitzer, K., Walsh, M., & Schweitzer A. (2017). To see or not to see: interlocutor visibility and likeability influence convergence in intonation. *Proceedings of Interspeech 2017* (Stockholm), pp. 919–923. <https://doi.org/10.21437/Interspeech.2017-1248>.
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193–247.
- Shepard, C. A., Giles, H., & Le Poire, B. A. (2001). Communication Accommodation Theory. In W. P. Robinson & H. Giles (Eds.), *The New Handbook of Language & Social Psychology* (pp. 33-56). New York: Wiley.
- Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception & Performance*, 33, 201–208. <https://doi.org/10.1037/0096-1523.33.1.201>.
- Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, 1, 305–319. <https://doi.org/10.1111/j.1756-8765.2009.01021.x>.
- Shockley, K., Santana, M., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 326–332. <https://doi.org/10.1037/0096-1523.29.2.326>.
- Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Attention, Perception, & Psychophysics*, 66, 422–429.
- Solanki, V. J. (2017). *Brains in dialogue: Investigating accommodation in live conversational speech for both speech and EEG data*. Ph.D: University of Glasgow.
- Soliz, J., & Giles, H. (2014). Relational and identity processes in communication: A contextual and meta-analytical review of Communication Accommodation Theory. In E. Cohen (Ed.), *Communication yearbook 38* (pp. 106–143). Thousand Oaks, CA: Sage.
- Staum Casasanto, L. S., Jasmin, K., & Casasanto, D. (2010). *Virtually accommodating: Speech rate accommodation to a virtual interlocutor*. Proceedings from Proceedings of the 32nd Annual Conference of the Cognitive Science Society, Austin, TX.
- Tobin, S. J. (2022). Effects of native language and habituation in phonetic accommodation. *Journal of Phonetics*, 93. <https://doi.org/10.1016/j.wocn.2022.101148>
- Wagner, M. A., Broersma, M., McQueen, J. M., Dhaene, S., & Lemhöfer, K. (2021). Phonetic convergence to non-native speech: Acoustic and perceptual evidence. *Journal of Phonetics*, 88. <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. *Frontiers in Psychology*, 6, 546.
- Walters, S. A., Babel, M. E., & McGuire, G. (2013). The role of voice similarity in accommodation. *Proceedings of Meetings on Acoustics*, 19(1) 060047.
- Weise, A., Silber-Varod, V., Lerner, A., Hirschberg, J., & Levitan, R. (2020). Entrainment in spoken Hebrew dialogues. *Journal of Phonetics*, 83. <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. *Journal of Phonetics*, 94. <https://doi.org/10.1016/j.wocn.2022.101173>
- 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, e74746.
- Zürcher, Y., Willems, E. P., & Burkart, J. M. (2021). Trade-offs between vocal accommodation and individual recognisability in common marmoset vocalizations. *Scientific Reports*, 11(1). <https://doi.org/10.1038/S41598-021-95101-8>.
- Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same-and mixed-gender dyads. *Language and Communication*, 8(3/4), 183-194.
- Gessinger, I., Schweitzer, A., Andreeva, B., Raveh, E., Möbius, B., & Steiner, I. (2018). Convergence of pitch accents in a shadowing task. In *International Conference on Speech Prosody. Poznań* (pp. 225–229). SpeechProsody.2018-46.
- López, S., Riera, P., Assaneo, M. F., Eguía, M., Sigman, M., & Trevisan, M. A. (2013). Vocal caricatures reveal signatures of speaker identity. *Scientific Reports*, 3, 3407. <https://doi.org/10.1038/srep03407>.
- Farrús, M., Wagner, M., Erro, D., & Hernando, J. (2010). Automatic speaker recognition as a measurement of voice imitation and conversion. *The International Journal of Speech, Language, & the Law*, 17, 119–142.
- Savino, M., Lapertosa, L., & Refice, M. (2018). Seeing or Not Seeing Your Conversational Partner: The Influence of Interaction Modality on Prosodic Entrainment. In A. Karpov, O. Jokisch, & R. Potapova (Eds.), *Speech and Computer. SPECOM 2018. Lecture Notes in Computer Science* (vol 11096). [https://doi.org/10.1007/978-3-319-99579-3\\_59](https://doi.org/10.1007/978-3-319-99579-3_59).