7-1-2018

# A Comparison of Phonetic Convergence in Conversational Interaction and Speech Shadowing

Jennifer Pardo
*Montclair State University*, pardoj@montclair.edu

Adelya Urmanche
*Montclair State University*

Sherilyn Wilman-Depena
*Montclair State University*

Jaclyn Wiener
*Montclair State University*

Nicholas Mason
*Montclair State University*

*See next page for additional authors*

Journal of Phonetics 69 (2018) 1–11

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

# A comparison of phonetic convergence in conversational interaction and speech shadowing

Jennifer S. Pardo *, Adelya Urmanche, Sherilyn Wilman, Jaclyn Wiener, Nicholas Mason, Keagan Francis, Melanie Ward

*Department of Psychology, Montclair State University, USA*

ARTICLE INFO

*Article history:*
Received 9 May 2017
Received in revised form 29 March 2018
Accepted 2 April 2018
Available online 26 April 2018

*Keywords:*
Conversation
Phonetic convergence
Speech production

ABSTRACT

Phonetic convergence is a form of variation in speech production in which a talker adopts aspects of another talker's acoustic–phonetic repertoire. To date, this phenomenon has been investigated in non-interactive laboratory tasks extensively and in conversational interaction to a lesser degree. The present study directly compares phonetic convergence in conversational interaction and in a non-interactive speech shadowing task among a large set of talkers who completed both tasks, using a holistic AXB perceptual similarity measure. Phonetic convergence occurred in a new role-neutral conversational task, exhibiting a subtle effect with high variability across talkers that is typical of findings reported in previous research. Conversational phonetic convergence did not differ by talker sex on average, but relationships between speech shadowing and conversational convergence differed according to talker sex, with female talkers showing no consistency across settings in their relative levels of convergence and male talkers showing a modest relationship. These findings indicate that phonetic convergence is not directly compatible across different settings, and that phonetic convergence of female talkers in particular is sensitive to differences across different settings. Overall, patterns of acoustic–phonetic variation and convergence observed both within and between different settings of language use are inconsistent with accounts of automatic perception-production integration.



Phonetic convergence is a form of variation in speech production in which a talker adopts aspects of another talker's acoustic–phonetic repertoire. To date, this phenomenon has been investigated in non-interactive laboratory tasks extensively and in conversational interaction to a lesser degree (for recent reviews, see Pardo, 2017, and Pardo, Urmanche, Wilman, & Wiener, 2017). This imbalance is not surprising given the degree of effort associated with collecting and analyzing conversational speech relative to non-interactive speech shadowing tasks employed in most studies. An often implicit assumption contributing to this imbalance is that patterns and mechanisms revealed using non-interactive tasks will transfer to more naturalistic complex settings such as conversational interaction, with some adjustment for the nuances of social settings. Unfortunately, patterns of phonetic convergence observed across these settings challenge this assumption, and interpretations of such patterns are hindered by the use of different talkers, methods, and measures across different studies. To begin to address some of these concerns, the present study examines phonetic convergence in conversational interaction using a relatively large set of talkers in a new task, and explores the relationship between conversational and non-interactive speech shadowing convergence within the same set of talkers producing speech in both settings.

### Conversational phonetic convergence

Previous investigations of phonetic convergence in conversational interaction have involved a variety of settings, including interviews, goal-oriented interactive tasks, and free-form conversations. Research within the Communication Accommodation framework has focused on the influence of external social dynamics on patterns of convergence and divergence in conversational interactions (Gasiorek, Giles, & Soliz, 2015; Giles, Coupland, & Coupland, 1991; Shepard,

* Corresponding author at: Montclair State University, Psychology Department, 1 Normal Avenue, Montclair, NJ 07043, USA.
*E-mail address:* pardoj@montclair.edu (J.S. Pardo).

https://doi.org/10.1016/j.wocn.2018.04.001
0095-4470/

Giles, & Le Poire, 2001), and encompasses investigations of conversational convergence in a broad array of acoustic–phonetic parameters such as vocal intensity (Natale, 1975), speaking rate (Putman & Street, 1984; Street, 1982), subvocal acoustic structure (Gregory, 1990; Gregory, Dagan, & Webster, 1997; Gregory, Green, Carrothers, Dagan, & Webster, 2001; Gregory & Webster, 1996), accent (Bourhis & Giles, 1977; Giles, 1973), and individual phonological forms (Coupland, 1984). For example, talkers converged in accent toward interviewers of distinct dialects in a cooperative interview setting (Giles, 1973), but diverged from an insulting interviewer (Bourhis & Giles, 1977). Typically, talkers converged in acoustic–phonetic attributes toward those of higher social status to a greater degree than toward those of lower status, but some settings evoked the opposite pattern (Coupland, 1984; Gasiorek et al., 2015; Gregory & Webster, 1996).

Accordingly, patterns of accommodation have been interpreted as signals of affiliation and/or attraction that vary in relation to regulation and maintenance of social distance and social interaction (Byrne, 1971; Gallois, Giles, Jones, Cargiles, & Ota, 1995; Gasiorek et al., 2015; Street, 1982). In a recent study, Aguilar et al. (2016) compared phonetic convergence among individuals exhibiting high versus low levels of trait rejection sensitivity, which is defined as a disposition to anxiously expect rejection in social encounters. They found that individuals with high levels of trait rejection sensitivity in social settings converged more than individuals with low levels toward their conversational partners. Moreover, rejection-sensitive individuals felt less connected to partners who had converged less. Because rejection-sensitive individuals also tend to exhibit greater incidents of ingratiating behaviors, it is possible that phonetic convergence was a form of ingratiation on their part. Taken together, phonetic convergence might be one of a variety of strategies for promoting social affiliation, and individuals appear to be sensitive to a conversational partner's degree of reciprocity in phonetic convergence.

Investigations of conversational interaction in other research domains generally confirm observations of convergence on speaking rate in particular, but their findings reveal complexities in patterns of convergence that challenge a straightforward interpretation (e.g., Levitan & Hirschberg, 2011; Manson, Bryant, Gervais, & Kline, 2013; Pardo, Cajori Jay, & Krauss, 2010; Pardo et al., 2013; Schweitzer & Lewandowski, 2013; Staum Casasanto, Jasmin, & Casasanto, 2010). For example, Levitan and Hirschberg (2011) examined conversational convergence across multiple acoustic attributes in parallel, including intensity, pitch, voice quality, and speaking rate. They found that some attributes converged while others did not, and that measures of convergence differed across different scales of analysis—convergence was somewhat reliable and consistent when measured across conversational turns, but different patterns emerged at more macro-conversational levels. Likewise, Pardo et al. (2010) found that interacting talkers converged in a holistic measure of phonetic convergence, but that their speaking rates did not converge. Instead, rates differed according to the role of a talker in the conversation, with instruction givers speaking faster than receivers. Moreover, Pardo et al. (2013) found that speaking rates converged during some conversational epochs despite differences in conversational role and then diverged during others according to conversational role. These patterns are not readily accommodated by an interpretation based solely on social affiliation/distance strategies because a talker's degree of rate convergence varied with the same partner in a single conversation, and measures of different attributes showed both parallel and distinct patterns of convergence.

#### Speech shadowing phonetic convergence

A comprehensive survey of research on phonetic convergence in non-interactive settings to date reveals that patterns of convergence in speech shadowing tasks are no less complex (see review in Pardo et al., 2017). In a typical speech shadowing task, a talker first produces baseline pre-exposure utterances prompted by printed text, then speech shadowing utterances prompted by recordings of utterances from a model talker (also known as an auditory naming task). If a talker converged toward a model, their shadowed utterances should sound more similar to those of the model talker than their pre-exposure baseline utterances. Assessments of phonetic convergence in a speech shadowing task are generally assumed to reflect the activity of relatively fundamental internal cognitive processes connecting speech perception and speech production, but have also been shown to be modulated by social factors (e.g., Babel, 2010, 2012; Babel, McAuliffe, & Haber, 2013; Babel, McGuire, Walters, & Nicholls, 2014; Namy, Nygaard, & Sauerteig, 2002).

Integration of perception and production is a core feature of prominent accounts of speech perception and language comprehension. For example, perception-production integration plays a central role in Fowler's direct realist theory of speech perception, in the motor theory of speech perception, and in Pickering and Garrod's interactive alignment account of language use, and phonetic convergence is often cited as evidence to support a close connection (Fowler, 1986, 2014; Fowler, Brown, Sabadini, & Weihing, 2003; Fowler, Shankweiler, & Studdert-Kennedy, 2016; Goldstein & Fowler, 2003; Liberman, 1996; Pickering & Garrod, 2004, 2013; Shockley, Sabadini, & Fowler, 2004). That is, resolution of detailed phonetic form in articulatory terms is hypothesized to support and even goad phonetic convergence in production.

Pickering and Garrod (2013) provide a particularly elaborate account of perception-production integration, centered on a so-called Simulation route in language comprehension. In this account, a covert imitation process automatically generates speech production commands via inverse forward modeling during language comprehension. Covert imitation can become overt as phonetic convergence in production when consistent with situational demands. However, this account offers few testable predictions beyond proposing that interacting talkers might exhibit phonetic convergence given appropriate circumstances. Because covert imitation results from the same processes involved in self-regulation of speech production, yielding a listener's own motor commands, the only specific prediction offered is that overt imitation should be greatest among individuals who are already similar to each other. Furthermore, the automaticity of covert imitation entails that all acoustic–phonetic attributes should be resolved equally well and should be equally available for convergence. Once

resolved, other factors may intervene to modulate overall degree of phonetic convergence as covert imitation becomes overt (or not), but the account offers no mechanism for selecting one or another specific attribute.

Although a comprehensive assessment of automatic perception-production integration is beyond the scope of any single study, it is consequential to consider whether observed patterns of phonetic convergence are consistent with the tenets of such an account. Of particular concern is the observation that the inconsistency across measures of phonetic convergence that has been observed in conversational interaction occurs in speech shadowing tasks as well. In conversational interaction, such variation could be attributed to the complexity of the setting, in which talkers accomplish multiple communicative goals in parallel. For example, speaking rate and intonation vary in a conversation according to expressive goals, and related measurable attributes (e.g., duration and pitch) might not be as readily available for phonetic convergence in a conversational setting in comparison with other settings. Non-interactive speech shadowing tasks arguably reduce opportunities for such goals to affect these attributes of phonetic convergence, yet similar inconsistencies have been observed in speech shadowing studies, even when comparing measures of different attributes taken from the same items in a single study (e.g., Babel, 2012; Babel & Bulatov, 2012; Pardo, 2013). In two recent speech shadowing studies (Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013; Pardo et al., 2017), measures of duration, F0, F1, and F2 each exhibited a distinct, talker and item-dependent pattern of variation and convergence. Examination of each measure alone yielded a different pattern of results from that obtained in the other measures. For example, a talker might converge only in duration, or converge in duration of some items and vowel formants of other items. This well-established variation in convergence across particular acoustic–phonetic attributes offers one kind of challenge for automatic perception-production integration.

A related challenge centers on the consistency of phonetic convergence across different settings. Despite differences across settings, an automatic perception-production integration mechanism should yield phonetic patterns consistently so that talkers who converge to a greater degree in speech shadowing tasks should converge to a similar degree in conversational interaction relative to other talkers. That is, the demands of conversational interaction might temper overall levels of phonetic convergence, or impose different patterns on some acoustic–phonetic attributes, but talker-related variation in phonetic convergence should be consistent across both kinds of settings if convergence emerges from automatic perception-production integration. In order to determine the extent to which these kinds of inconsistencies across settings are related to the same underlying processes, it is necessary to directly compare phonetic convergence in conversational interaction and in speech shadowing using the same kind of measure in the same set of talkers.

A promising approach for addressing complexities in acoustic–phonetic convergence has been developed within the speech shadowing framework and extended to conversational interaction (Goldinger, 1998; Pardo, 2006). That is, rather than attempt to explain complex patterns of variation across multiple acoustic–phonetic attributes, a perceptual similarity task can provide a holistic measure of phonetic convergence that readily integrates across these attributes. Goldinger (1998) was the first to adapt a classic psychophysical AXB perceptual similarity paradigm to assess phonetic convergence in a speech shadowing task. In this adapted paradigm, a series of trials compare a talker's baseline pre-exposure utterances (A) and shadowed utterances (B) to the model talker's utterances that prompted the shadowing response (X). A separate set of listeners judge whether the pre-exposure or shadowed version of an utterance (A/B) was more similar in pronunciation to a model talker utterance (X) on each trial.

Adapting this paradigm to items from conversational tasks substitutes shadowed utterances with task or post-interaction utterances in comparison with partner utterances (Aguilar et al., 2016; Dias & Rosenblum, 2011; Kim, Horton, & Bradlow, 2011; Pardo, 2006; Pardo et al., 2010; Pardo, Cajori Jay, et al., 2013). If a talker converged toward a model or conversational partner, then their shadowed/post-interaction utterances should sound more similar to the model/-partner utterances than their pre-exposure baseline utterances. This paradigm yields a measure of phonetic convergence in terms of the proportion or percent of trials in which a listener chose a shadowed/post-interaction utterance as more similar than a pre-exposure baseline utterance. Previous studies have calibrated this paradigm, finding that AXB perceptual similarity measures of phonetic convergence reflect patterns of variation in multiple acoustic–phonetic attributes, without making an a priori commitment to an individual attribute that might not exhibit convergence consistently (Pardo, Jordan, et al., 2013; Pardo et al., 2017). Thus, assessment of phonetic convergence using a holistic measure is preferable to individual acoustic measures because acoustic–phonetic attributes vary inconsistently, there are no standards for selecting particular attributes to examine, and holistic appraisal reflects the multi-dimensionality of the phenomenon.

#### Comparing speech shadowing and conversation convergence

In previous investigations using a holistic AXB perceptual similarity task, phonetic convergence in both speech shadowing and conversational interaction has been found to be very subtle on average and highly variable across talkers. Across studies using speech shadowing or conversational interaction tasks, average AXB estimates of convergence typically do not exceed 0.60 proportion post-exposure trials selected, with most estimates hovering around 0.56 (chance performance is 0.50). Higher average AXB estimates have been reported in 3 out of 13 total shadowing AXB studies to date (Dias & Rosenblum, 2016; Goldinger, 1998; Goldinger & Azuma, 2004) and in 2 out of 6 total conversational AXB studies to date (Dias & Rosenblum, 2011; Pardo, 2006). Within individual studies of convergence, AXB measures have varied greatly across talkers in both settings as well—from 0.46 to 0.72 for speech shadowing in Pardo et al. (2013) and from 0.33 to 0.75 for conversations in Pardo et al. (2010). Moreover, the extent that an individual model talker in a shadowing study elicits phonetic convergence from multiple shadowers also varies, from 0.50 to 0.64 in Pardo et al. (2017).

Thus, a survey of average performance levels and ranges of estimates for phonetic convergence yields similarities

across speech shadowing and conversational interaction, but it is unknown whether these similarities reflect an underlying similarity in the cognitive mechanisms that support phonetic convergence. If phonetic convergence in both kinds of settings reflects automatic integration of perception and production, then an individual talker who converges to a greater degree in speech shadowing ought to converge to a similarly high degree in conversational interaction. Conversely, while connections between speech perception and production might support phonetic convergence in one setting, other factors might influence the degree of phonetic convergence differently in different settings.

An additional concern arises when comparing studies of convergence across both settings with regard to effects of talker sex on phonetic convergence. Although talker sex effects are not consistent across the literature, one shadowing study found that females converged more than males (Namy et al., 2002), but the effect was completely driven by convergence to one out of four model talkers. Despite this limitation, some subsequent studies of phonetic convergence have cited this finding as a motivation for using only female talkers (Delvaux & Soquet, 2007; Dias & Rosenblum, 2016; Gentilucci & Bernardis, 2007; Sanchez, Miller, & Rosenblum, 2010; Walker & Campbell-Kibler, 2015). In contrast, one conversational study found that males converged more than females (Pardo, 2006), but this study used just six pairs of talkers. Although the discrepancy between these two studies could be due to the different settings (shadowing vs. conversation), other studies of both speech shadowing and conversational convergence have failed to find effects of talker sex in larger samples (Pardo et al., 2010; Pardo, Cajori Jay, et al., 2013; Pardo, Jordan, et al., 2013; Pardo et al., 2017). Finally, no previous studies have systematically compared convergence in same- and mixed-sex pairings of talkers in either setting, except the shadowing study that was conducted in conjunction with the present study. This is an important issue because some of the reported findings in the literature that low frequency words elicit greater phonetic convergence than high frequency words were obtained in studies that only used female shadowers (Babel, 2010; Dias & Rosenblum, 2016; Nielsen, 2011; but not Goldinger, 1998), and Pardo et al. (2017) found that the effect of word frequency on phonetic convergence only applied to female talkers. Because effects of talker sex on phonetic convergence are so inconsistent and variable across the literature, it is necessary to include a balance set of talkers in both same and mixed-sex pairings.

Given the degree of variation across talkers in the same holistic measure in both non-interactive and conversational settings, an investigation of the relationship between phonetic convergence in both settings among the same set of talkers is warranted. In order to provide a rigorous examination of a potential relationship, it is necessary to assess phonetic convergence among set of talkers who provide speech in both speech shadowing and in conversational interaction. Investigations of phonetic convergence using speech shadowing tasks operate under an implicit assumption that findings will generalize to more naturalistic settings of language use, such as conversational interaction, but no study to date has directly compared patterns of convergence across these settings. If patterns of phonetic convergence do not transfer across settings, then interpretations of these patterns must acknowledge this limitation.

In particular, phonetic convergence during speech shadowing should reflect the greatest influence of automatic perception-production integration, and while demands of conversational interaction might temper overall levels of phonetic convergence, each talker should demonstrate the same relative level of convergence in both settings. Furthermore, automatic perception-production integration should not vary according to talker sex, but talkers who are more similar to each other should converge more than those who are less similar, so that same-sex pairs of talkers should converge more than mixed-sex pairs.

To address these and other concerns, the present study investigates phonetic convergence during conversational interaction in a relatively large set of talkers (96; 48 female) placed into both same- and mixed-sex pairings who also provided utterances in a speech shadowing task. As reported in Pardo et al. (2017), each talker shadowed one of 12 model talkers (6 female) who were not part of the present study in either same- or mixed-sex pairing (analogous to their pairing in the present study). The shadowing study collected utterances in a pre-task phase prompted by text and in a speech shadowing phase prompted by model talker recordings of 80 mono- and 80 bi-syllabic words. Separate listeners completed AXB perceptual similarity tests that assessed phonetic convergence of shadowed utterances to model talker utterances (Shadowing AXB $N = 736$).

To evoke conversational speech in the present study, talkers completed the Montclair Map Task, which is a modified role-neutral version of the HCRC Map Task (Anderson et al., 1991). The goal of the Montclair Map Task is for talkers to reconcile differences across paired maps in the composition and location of iconic landmarks, making this a map matching task. This aspect of the task was inspired by the Diapix task used in the Wildcat corpus (Van Engen et al., 2010), but the current task permits a more precise measure of performance and more control over between-talker utterance repetitions. Thus, the Montclair Map Task neutralizes the explicit role manipulation in the original HCRC Map Task, focusses conversation on the landmarks, and evokes natural, more balanced conversational interaction while ensuring that a set of pre-determined phrases (landmark labels) will be repeated between talkers. To do so, each map in a pair contains five shared and five unique landmarks, and talkers are instructed to draw labeled markers indicating the location of the missing landmarks on their own maps as accurately as possible.

The amount of time that talkers spent completing the Montclair Map Task with six map pairs averaged 32 min ($sd = 10.68$ min), ranging from 16 to 62 min, with a high degree of accuracy in the landmark matching task itself. In the course of completing the task, talkers naturally repeated landmark label phrases that can be used to assess phonetic convergence in an AXB perceptual similarity test by comparing task and post-task utterances with their pre-task versions. A more detailed description and analysis of the Montclair Map Task Corpus appears in Pardo et al. (in press), and the corpus itself is available via an online data sharing repository (Pardo et al., 2018).

In summary, a set of 96 talkers were recruited to complete two separate tasks, a speech shadowing task in which they

shadowed speech from one of 12 model talkers, and a paired conversational interaction while completing the Montclair Map Task. The order of completion of speech shadowing and map task sessions was counterbalanced across talkers, and the shadowing task employed 80-item mono- and bisyllabic word sets, while the map task permitted phonetic convergence assessment in up to 79 distinct landmark label phrases. Talkers were paired in same- and mixed-sex pairings consistently across both shadowing and conversational interaction settings. Items from both settings were presented to separate listeners for AXB perceptual assessments of phonetic convergence, to permit comparisons across settings within the same talkers using the same holistic measure. The present study first establishes phonetic convergence in conversational interaction in this set of talkers with a new task, and then compares conversational convergence to convergence in speech shadowing.

## Method

### *Participants*

*Talkers:* A total of 96 native English speakers (48 female) from the Montclair State University community completed the Montclair Map Task in 16 same-sex female, 16 same-sex male, and 16 mixed-sex pairs. None of the paired talkers were acquainted with each other prior to participating in the experiment. All talkers reported normal hearing and speech, provided IRB-approved informed consent, and received $20 compensation for their time.

*Listeners:* A total of 564 native English speakers from the Montclair State University community completed AXB perceptual similarity tests assessing phonetic convergence in the map task. All listeners reported normal hearing and speech, provided IRB-approved informed consent, and received course credit for their participation.

*Materials*: To complete the Montclair Map Task, each member of a pair of talkers received a packet of six iconic maps printed on $8.5'' \times 11''$ sheets of paper that corresponded with their partner's map packet. Each map in a corresponding pair had five shared and five unique landmarks, and both maps in a pair contained an identical path drawn from a starting point, around various landmarks, to a finish. The set of landmarks and the shape of the map path varied across map pairs in a packet, and the order of map pairs varied across participant pairs. The full set of maps contained 79 unique landmark label phrases, with some appearing on more than one map. Appendix A displays a sample map task map pair, and Appendix B lists the full set of landmark label phrases printed on the maps.

### *Procedures*

*Map Task Conversations:* All recordings took place in an *Acoustic Systems* sound proof booth. Talkers wore AKG head-mounted microphones connected to a *Macintosh* computer that recorded each talker onto an individual channel in a time-aligned 2-channel audio file using *SoundStudio* software (Felt Tip).

Each pair of talkers completed the Montclair Map Task with six pairs of maps and provided pre-task and post-task recordings of landmark label phrase utterances for comparison with the map task utterances in AXB perceptual similarity tests. During the map task session, each talker sat at an individual table separated from their partner by a room divider and used a pencil to draw on the maps as they completed the task. Instructions informed talkers that there were five shared and five unique landmarks on each map, and that their goal was to discover the five unique landmarks printed on each of their partner's maps and to draw markers on their own maps indicating the location and identity of the missing landmarks. To collect pre-task and post-task utterances, all 79 landmark label phrases were printed on a single $8.5''$ by $11''$ sheet, and talkers produced each landmark label phrase embedded in the carrier sentence, "Number *X* is the *landmark phrase*." First, talkers provided individual pre-task recordings of the landmark label phrases (going through the list twice), then they completed the Montclair Map Task paired conversations, followed by individual post-task recordings of landmark label phrases. Instructions encouraged talkers to produce the pre-task and post-task utterances in their normal fluent speaking voice.

*AXB Perceptual Similarity Tests:* Assessment of phonetic convergence in conversational interaction entailed comparing utterances of landmark label phrases that were produced before, during, and immediately after the map task session. In an AXB perceptual similarity test, a listener hears three versions of the same landmark label phrase on each trial. An utterance from one member of a pair of talkers serves as the standard X-item, and the A/B flanking items are corresponding utterances from their task partner. Instructions inform a listener to decide on each trial which of the two flanking A/B items from one talker sounds more similar in pronunciation to the middle X-item from the other talker. If phonetic convergence had occurred, then those items produced by a partner during or after the map task session should sound more similar to a talker's X-item than those produced before the map task session. Phonetic convergence was assessed in two sets of comparisons, one that focused on landmark label phrases that were repeated between talkers during the conversational interaction and another that included all landmark label phrases produced in the pre-task and post-task sessions.

The first set of comparisons was designed to permit assessment of phonetic convergence to each member of a pair during the map task session. In this case, a custom set of task session items for each talker was derived to serve as standard X-items in AXB tests. In the course of completing the Montclair Map Task, nearly all landmark label phrases were mentioned by at least one talker, with each member of a pair having occasions in which they were the first talker to produce a particular landmark label phrase. If their partner produced the same phrase soon afterwards, that pair of items could be used to assess convergence of the partner to the talker who introduced the item. Thus, the full set of landmark label phrases that were repeated between talkers in a pair formed two subsets—one set for each member of each pair composed of the items that they produced first in the conversation. These talker-led task session X-items were then compared with a partner's corresponding pre-task, task utterance, and post-task versions of each phrase.

This selection procedure resulted in distinct sets of task session X-items that differed with respect to the number of

phrases that could be used for assessing phonetic convergence to each talker, ranging from 4 to 31 items across talkers, with an average of 17 items. Each distinct set of talker-led task session X-items was presented with the partner's pre-task, task utterance, and post-task versions such that a pre-task item was compared with a task utterance on half of the trials and with a post-task item on the other half of trials. Order of presentation of pre-task items (A/first versus B/last) was counterbalanced across trials, and items were sampled randomly and repeated in order to total 150 trials for each talker's task session X-items (half with pre-task items first, and half comparing pre-task with task utterances). Each of the resulting 47 AXB tests for conversational convergence assessed phonetic convergence of both members of a pair by including a block of 150 trials for one pair member's task session X-items followed by a block of 150 trials for the other member's X-items (with block order counterbalanced across listeners). Each pair's AXB test was presented to four listeners ($N = 188$).

In order to assess phonetic convergence in the post-task session alone, a second set of AXB tests was composed using only pre-task and post-task items (this condition also addresses any concerns regarding the idiosyncratic item sets for each talker in the task session item condition). In this case, each AXB test used all 79 post-task items from each talker as X-items in comparison with pre-task and post-task items from their partner (counterbalancing pre-task item order across trials). For these post-task X-item tests, each AXB test assessed convergence to one member of a pair of talkers, and each item was repeated twice in both orders, yielding 316 total trials for each talker's AXB test in the post-task X-item condition. Each talker's post-task AXB test was presented to four listeners ($N = 376$).

Because one talker's pre-task session items were unusable due to a noisy recording, AXB tests assessed phonetic convergence in the remaining 47 pairs of talkers (yielding 16 same-sex female, 16 same-sex male, and 15 mixed-sex pairs). Thus, a total of 188 listeners provided AXB perceptual similarity judgments for the task session X-item condition (4 per pair, blocked by talker), and 376 listeners judged the post-task X-item condition (4 per talker). All AXB perceptual similarity tests were presented to listeners in quiet testing rooms over *Sennheiser Pro* headphones via iMac computers running *SuperLab 4.5* (Cedrus).

#### Results

If talkers converged phonetically during the conversation, then utterances produced by a partner during and after the map task session should sound more similar in pronunciation to a talker's X-items than those produced before the map task session. Accordingly, descriptive data reported in the text reflect proportion that a listener selected a task/post-task item as more similar to an X-item than a pre-task item. For statistical analyses examining phonetic convergence, all AXB trials were collated to form a single dataset, and logistic mixed-effects modeling assessed the likelihood that a listener chose a partner's task/post-task item (1) versus their pre-task item (0) as more similar to a talker's task/post-task X-item. These procedures are more appropriate than analysis of variance for this dataset because they treat binary data as such (without conversion), and they enable inclusion of all three crossed random sources of variance without collapsing (talkers, phrases, and listeners). For all model fits, 2-level fixed-effects factors were contrast coded (-0.5, 0.5), continuous factors were z-scale normalized, and chi-square tests assessed whether inclusion of a reported factor improved model fit relative to a model without the factor (lme4 version 1.1-12, R version 3.3.2; Baayen, 2008; Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013; Bates, Mächler, Bolker, & Walker, 2015; Bates et al., 2016; Quené & van den Bergh, 2008; R Development Core Team, 2016).

Overall, the proportion of trials in which a listener detected convergence averaged 0.54, and a base logistic mixed-effects regression model with only random intercepts (by talkers, phrases, and listeners) confirmed a significant difference from chance [*Intercept* = 0.168 (0.023), $Z = 7.36$, $p < 0.0001$]. Convergence levels varied across talkers, ranging from 0.44 to 0.73, and across phrases, ranging from 0.46 to 0.60. With respect to timing in the conversation, convergence was detected for items produced during the first map, which averaged six minutes in duration, and was consistent across maps (no significant differences across Maps 1–6: 0.54, 0.53, 0.56, 0.53, 0.55, 0.55). In contrast, AXB perceptual similarity was greater for tests using map task session X-items (0.55) versus those using post-task X-items (0.53) [fixed effect of condition $\beta = 0.080$ (0.017), $Z = 4.77$, $p < 0.0001$; condition model versus base model $X^2(1) = 22.30$, $p < 0.0001$].

A closer examination of the relationship between these conditions reveals that AXB perceptual similarity in post-task session tests was moderately correlated with map task session tests when collapsed by talkers [$r(92) = 0.43$, $p < 0.00001$] and weakly correlated when collapsed by landmark label phrases [$r(75) = 0.27$, $p < 0.017$; Bonferroni corrected alpha level for correlation analyses is $p < 0.01$]. Examining the same correlation with a subset of post-task phrases that were consistent with those used in the map task session yields nearly the same estimate by talkers [$r(92) = 0.47$, $p < 0.000001$]. These correlation analyses indicate that persistence of convergence by talkers into the post-task session was moderately related to their initial convergence levels in the map task session, and that inclusion of the full set of phrases in post-task AXB tests yielded a similar estimate of this relationship to that obtained with consistent item sets. However, this relationship is completely driven by male talkers [$r(45) = 0.59$, $p < 0.0000 1$] as opposed to females [$r(45) = 0.19$, $p = 0.2$ ns].

Thus, talkers converged within the first map in the map task session, they maintained initial convergence levels throughout the course of the map task session, and convergence persisted into the post-task session, albeit slightly reduced and somewhat distinct from convergence in the map task session. Convergence levels did not differ on average between males and females in either session (Task: 0.55, 0.54; Post-Task: 0.53, 0.53), but the relationship between convergence during the map task and in the post-task session was not consistent across female talkers.

In order to determine whether AXB assessment of convergence for an individual talker is independent of their partner's convergence level, a second set of analyses examined the relationship between paired talkers' estimates of phonetic convergence. A number of factors could contribute to a

relationship between paired talker phonetic convergence levels, including actual equivalence of convergence within talker pairs, the use of the same listeners to assess convergence across members of a pair in the map task session, and/or the use of the same items to assess convergence in the post-task session. However, correlations between paired talkers' estimates of phonetic convergence revealed their independence in both the map task session condition [$r(45) = 0.19$, $p = 0.19$] and in the post-task session condition [$r(45) = 0.03$, $p = 0.82$]. Thus, a talker's level of phonetic convergence did not depend on that of their partner, even when assessing convergence using the same listeners in the task session or the same sets of items in the post-task session.

Additional analyses examined effects of talker sex (female 0.537 vs. male 0.538) and pair sex (same 0.537 vs. mixed 0.539), but neither factor yielded significant differences, nor did they interact with each other or with AXB testing condition. Given the somewhat lower average level of detected convergence in the current study relative to previous studies, a median split of the data explored whether any differences might emerge among a subset of talkers with higher levels of convergence. This split resulted in a subset of 47 higher-converging talkers (25 female) with an average of 0.58, ranging from 0.53 to 0.73. Even within this subset of higher-converging talkers, effects of talker and pair sex failed to reach significance. Thus, the lack of sex effects in this study cannot be attributed to a floor effect or to lack of power in the dataset, rather, talker sex bears a weak and unreliable relationship with phonetic convergence.

### Convergence in conversations & speech shadowing

A final set of analyses examined the relationship between phonetic convergence in conversational interaction and in speech shadowing. As mentioned previously, talkers in the current study also provided utterances in a speech shadowing task in which they repeated words prompted by model talkers. Four talkers failed to keep their shadowing session appointments, leaving a total of 90 talkers in the shadowing condition who each shadowed speech from one of twelve model talkers in either same or opposite-sex pairing (analogous to their pairing in the Montclair Map Task with 32 same-sex female, 30 same-sex male, and 28 mixed-sex pairings remaining). The shadowed utterances were assessed for phonetic convergence in AXB perceptual similarity tests that compared shadower pre-task and shadowing (A/B) items to model talker X-items. As reported in Pardo et al. (2017), shadowing convergence averaged 0.56 and did not differ by talker or pair sex. However, there was an effect of word type, such that convergence to bisyllabic words (0.57) exceeded convergence to monosyllabic words (0.55). Despite this difference, phonetic convergence to mono- and bisyllabic words in the shadowing task was related in this talker set and the relationship was equivalent across male and female talkers [All $r(88) = 0.68$, $p < 0.0001$; Female $r(44) = 0.70$, $p < 0.0001$; Male $r(44) = 0.69$, $p < 0.0001$]. Note that this relationship is somewhat stronger than that between convergence in the map task and post-task sessions of the current study, despite having been assessed across different items by different listeners.

Table 1 displays correlation coefficients from comparisons between AXB phonetic convergence estimates obtained in the current experiment (combined, map task session items, and post-task session items) and in both conditions of the shadowing experiment (bisyllabic and monosyllabic words), all collapsed by talker. As shown in the table, convergence for the map task combined measure (averaging across map task and post-task estimates) and in both conditions individually was weakly related to convergence in each set of shadowing items, except for convergence of map task session items with monosyllabic shadowing items. However, these relationships interacted with talker sex, indicating that the weak relationship observed overall was driven by convergence patterns in male talkers as discussed below. Finally, measures of convergence in the map task session items do not appear to contribute as strongly to relationships with shadowing convergence as measures from the post-task session.

**Table 1**
Correlation coefficients for relationships between conversational and shadowing convergence.

| | Combined | Map Task Session | Post-Task Session |
|---|---|---|---|
| Shadowing Bisyllabic | 0.31 | 0.21 | 0.30 |
| Shadowing Monosyllabic | 0.23 | 0.16 ns | 0.24 |

All significant at $p < 0.05$ with df = 88, except where indicated ns.

As shown in the scatterplot in Fig. 1, the relationship between conversational and shadowing convergence in female talkers was not significant [$r(44) = 0.13$, $p = 0.37$], while male talkers showed a modest relationship between convergence in conversational interaction and in shadowing of bisyllabic words [$r(42) = 0.47$, $p < 0.001$]. This distinction between male and female talkers was consistent across pairings with same or mixed-sex partners/models. Although it appears that male talkers were more variable in convergence during the map task session while females were more variable in phonetic convergence during speech shadowing, these differences are due mainly to convergence of just four talkers with the highest convergence levels in the two tasks. Elimination of these four participants' data from the dataset did not alter the overall pattern of results, and a review of demographic information did not reveal anything distinctive about these talkers.

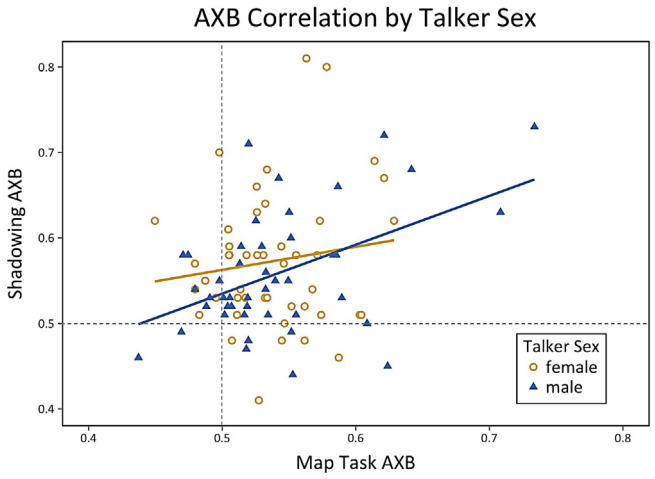



**Fig. 1.** Correlations between combined Map Task AXB phonetic convergence and bisyllabic word Shadowing AXB phonetic convergence for female and male pairs of talkers. Female talkers are plotted with open orange circles, and male talkers are plotted with filled blue triangles. Lines represent linear fits, but the relationship for female talkers was not significant.

Logistic mixed-effects modeling assessed these patterns by using bisyllabic Shadowing AXB levels (z-scaled) as continuous predictors of individual trials in the Map Task AXB test dataset. Model parameters confirmed a significant prediction for Shadowing AXB levels [$\beta = 0.071$ (0.021), $Z = 3.41$, $p < 0.0007$] and a significant interaction with talker sex [$\beta = 0.099$ (0.042), $Z = 2.38$, $p < 0.02$; interaction model versus simple shadowing model $X^2(2) = 5.89$, $p = 0.05$]. An additional model examined whether this pattern differed by map task condition (task session items versus post-task session items), but the interaction was not significant. Thus, the relationship between Shadowing AXB and Map Task AXB did not differ whether assessing map task convergence using phrases from the map task session or from the post-task session. These analyses confirmed that phonetic convergence in speech shadowing was related to that in conversational interaction, but the relationship was limited to male talkers and did not depend on map task testing condition.

## Discussion

The present study established phonetic convergence in conversational interaction in a relatively large set of talkers using a holistic measure in a new role-neutral task, the Montclair Map Task. Average levels and patterns of variation in phonetic convergence observed in the present study align with those observed in other studies, but did not differ by talker sex and did not differ between same- and mixed-sex pairs of talkers. Moreover, average levels of phonetic convergence in conversation and speech shadowing in the same set of talkers are comparable, 0.54 in the map task session and 0.56 in speech shadowing, with overlapping distributions across tasks. Correlation analyses between conversational and shadowing convergence reveal a weak relationship across all talkers, but this relationship differed by talker sex such that male talkers showed a modest relationship while females showed no relationship. Thus, levels of convergence of female talkers in a shadowing task were not consistent with those in conversational interaction, while those of male talkers aligned to a moderate degree. Additional analyses ruled out an impact of same- versus mixed-sex pairings, therefore, talker sex effects were not due to the sex of their partners.

An examination of patterns of talker sex effects in previous studies reveals that phonetic convergence of female talkers might be more susceptible in general to a variety factors than that of male talkers. For example, in the conversational interactions examined in Pardo (2006), males converged more than females on average, but the difference was due to an interaction with task role—female instruction receivers in the task did not converge, while female instruction givers converged to the same degree that male talkers converged. Subsequent conversational studies using the same task with larger samples did not replicate sex differences in average convergence levels, but talker sex interacted with other variables in those studies as well (Pardo et al., 2010; Pardo, Cajori Jay, et al., 2013). In the shadowing study that reported greater convergence of female talkers, the effect was completely driven by greater convergence of females to just one out of four talkers, thus, convergence of females might also be more sensitive to differences between talkers (Namy et al., 2002). This kind of sensitivity could have contributed to the lack of relationship between convergence to a talker in a shadowing task and to a different talker in conversational interaction in the present study.

In the shadowing study conducted in conjunction with the present study, talker sex interacted with word frequency, such that female talkers converged more to low than to high frequency words while male talkers showed no difference. Likewise, female talkers converged more to bisyllabic than to monosyllabic words, while males showed a smaller difference (Pardo et al., 2017). Thus, phonetic convergence of female talkers has been found to be more sensitive to lexical factors and to talker role than that of males, and the current study has revealed yet another factor influencing females more than males—task setting and/or talker identity. In this case, females converged to equivalent degrees overall, but convergence of individual female talkers in conversational interaction was unrelated to that in the post-task session and in speech shadowing.

The cross-task inconsistency in convergence patterns of female talkers [$r(44) = 0.13$, $p = 0.37$] is unlikely to be due solely to the use of different lexical items across tasks because convergence of female talkers was consistent when comparing across the different mono- and bisyllabic items used in the shadowing task [$r(44) = 0.70$, $p < 0.0001$]. Moreover, convergence of female talkers in the post-task setting was inconsistent with that during the map task despite using the same landmark label phrases [$r(45) = 0.19$, $p = 0.2$ ns]. While convergence patterns of female talkers differed across the shadowing task, the map task, and the post-task session of the map task, convergence of male talkers was moderately consistent across these settings [Mono- to Bisyllabic Speech Shadowing $r(44) = 0.69$, $p < 0.0001$; Shadowing to Map Task $r(42) = 0.47$, $p < 0.001$; Map Task to Post-Task $r(45) = 0.59$, $p < 0.00001$].

Because talker sex interacts with multiple factors in phonetic convergence, including the setting of speech production, it is important that future investigations use both male and female talkers in balanced designs with large samples. More importantly, findings from non-interactive tasks, like speech shadowing, might transfer to conversational settings when considering average levels of convergence, but the lack of consistency for individual talkers across settings indicates that other talker-modulated factors might not transfer across settings. For example, because phonetic convergence is not consistent across task settings (or only moderately so for male talkers), a finding that social biases influence phonetic convergence of talkers in a shadowing task might not bear any relation to convergence patterns observed in conversational interaction, especially for female talkers.

Across conversational and non-interactive settings in the same group of talkers, phonetic convergence did not differ on average. However, talkers varied in the extent that each setting evoked phonetic convergence, with only a modest relationship among male talkers and greater overall inconsistency among female talkers. These findings are difficult to reconcile with a proposal that phonetic convergence reflects automatic perception-production integration (Pickering & Garrod, 2013). According to such an account, the underlying integrative mechanism derives motor commands automatically and consistently, whether their effects are ultimately manifested in production or not. Thus, covert imitation is an automatic consequence of perception that varies only according to degree of between-talker

similarity, while overt imitation (i.e., phonetic convergence) depends on a variety of factors. External social factors and communicative goals may temper overall levels of convergence, but each talker should converge to an analogous degree across settings relative to other talkers because automatic covert imitative motor commands provide the foundation for convergence in all settings of language use. Instead, it appears that phonetic convergence in each kind of setting is equally susceptible to a variety of factors that interfere with proposed linkages between perception and production. Thus, it is likely that the fundamental connection between perception and production varies across settings, apart from a subsequent connection between covert and overt motor commands.

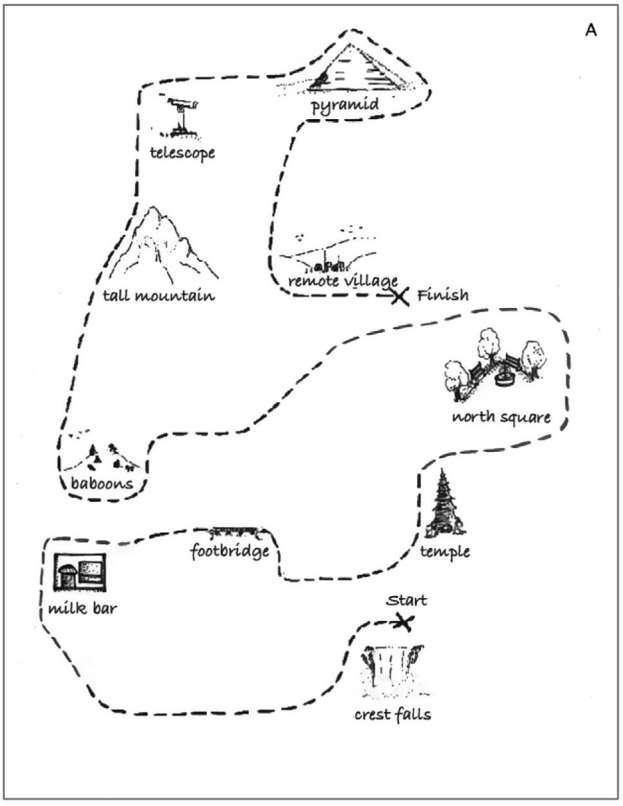
The present study constitutes a rigorous examination of phonetic convergence in conversational interaction using a holistic measure that integrates across multiple acoustic–phonetic attributes. Moreover, this protocol has proven effective in providing measures of phonetic convergence that are independent across members of an interacting pair. Overall, phonetic convergence in both non-interactive speech shadowing and in conversational interaction is very subtle and highly variable. The talker-related variability in phonetic convergence across settings was moderately consistent for male talkers, but inconsistent in female talkers. This greater sensitivity of phonetic convergence in female talkers to differences in experimental conditions is likely responsible for some of the inconsistencies across different studies and complicates comparisons between studies and interpretations of studies with unbalanced designs. This pattern of results warrants further investigation to determine why phonetic convergence of females is less consistent across speech shadowing and conversational interaction than that of males who engaged in the same tasks. Ultimately, patterns of acoustic–phonetic variation and convergence observed both within and between different settings of language use are inconsistent with accounts of automatic perception-production integration.

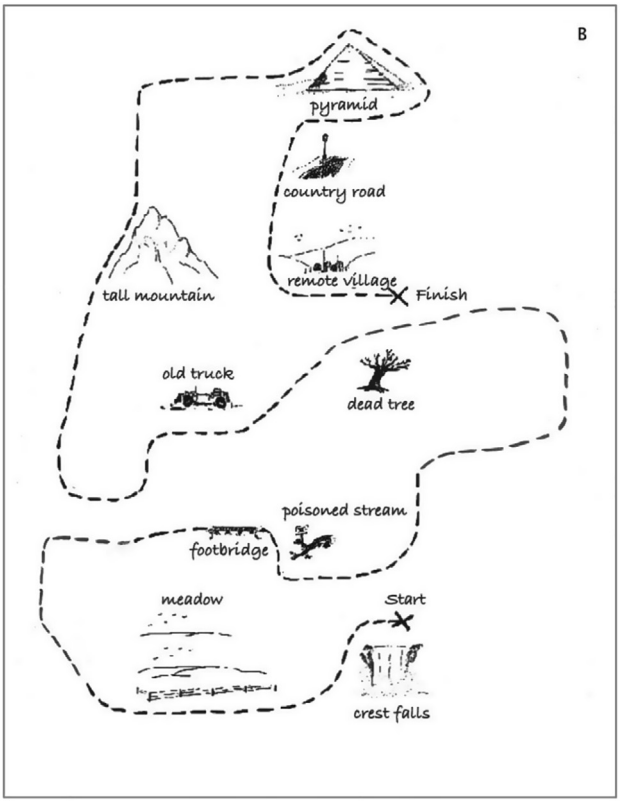
**Author note**

This work was supported by the National Science Foundation [BCS-1229033]. The authors are grateful to Hannah Gash, Alexa Decker, Nishi Patel, Erica Celis, Cassandra Sardo, Tailor Tighe, and Corey Ford-Graham for their assistance with data collection, and to the action editor and three anonymous reviewers for their assistance with narrative development.

**Appendix A. Sample map task map pair**

The images below comprise a single pair of maps used in the Montclair Map Task. Each map contains five shared landmarks and five unique landmarks. For example, both maps contain a *pyramid*, but Map 1A on the left has a *telescope* at the upper left corner, while Map 1B is missing this landmark. Instructions informed participants about the disparity in the composition of the landmarks, and asked them to discover the five missing landmarks that appear on their partner's map and draw a marker for each landmark in its corresponding location on their map. All talker pairs completed the map task with a total of six pairs of maps like the ones below.



## Appendix B. Landmark label phrases

A listing of the full set of 79 landmark label phrases used in the map task.

| | | |
|---|---|---|
| baboons | green bay | small island |
| bakery | headstone | stone cliffs |
| bear cave | huge nuclear plant | stony creek |
| blacksmith | land parcels | tall mountain |
| buffalo | large house | tall pine |
| cactus | large cottage | tavern |
| camera shop | lighthouse | teepees |
| caribbean palm | marsh land | telephone booth |
| cattle ranch | meadow | telescope |
| cement roof house | milk bar | temple |
| chapel | monastery | totem pole |
| cottages | monument | tower |
| country road | mud hut | trailer park |
| crest falls | museum | train crossing |
| dead tree | north square | train bridge |
| diamond mine | oily rag | walled city |
| east lake | old truck | west lake |
| fallen rocks | old mill | wheat field |
| farmed land | orange car | winter garden |
| flat rocks | parked van | yacht club |
| flowing river | picket fence | |
| footbridge | pine forest | |
| forked stream | pirate ship | |
| fortress | poisoned stream | |
| garage | pyramid | |
| ghost town | remote village | |
| golf course | round hills | |
| graveyard | sandy shore | |
| greasy wash water | small forest | |

## References


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.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., ... Weinert, R. et al. (1991). The H.C.R.C Map Task Corpus. *Language & Speech, 34*, 351–366. Retrieved from https://catalog.ldc.upenn.edu/LDC93S12.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

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.wocn.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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory & Language, 68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D. M., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., et al. (2016). lme4: Linear mixed-effects models using Eigen and S4 classes. R package version 1.1-12.

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*. https://doi.org/10.18637/jss.v067.i01.

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.

Byrne, D. (1971). *The attraction paradigm*. New York, NY: Academic Press.

Coupland, N. (1984). Accommodation at work: Some phonological data and their implications. *International Journal of the Sociology of Language, 46*, 49–70.

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.*. https://doi.org/10.3758/s13414-015-0982-6.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*, 3–28.

Fowler, C. A. (2014). Talking as doing: Language forms and public language. *New Ideas in Psychology, 32*, 174–182. https://doi.org/10.1016/j.newideapsych.2013.03.007.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language, 49*(3), 396–413. https://doi.org/10.1016/S0749-596X(03)00072-X.

Fowler, C. A., Shankweiler, D., & Studdert-Kennedy, M. (2016). "Perception of the speech code" revisited: Speech is alphabetic after all. *Psychological Review, 123*, 125–150. https://doi.org/10.1037/rev0000013.

Gallois, C., Giles, H., Jones, E., Cargiles, A. C., & Ota, H. (1995). Accommodating intercultural encounters: Elaboration and extensions. In R. Wiseman (Ed.), *Intercultural communication theory* (pp. 115–147). Newbury Park, CA: Sage.

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.

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

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics, 15*, 87–109.

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge: Cambridge University Press.

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. *Psychonomic Bulletin & Review, 11*, 716–722.

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In A. S. Meyer & N. O. Schiller (Eds.), *Phonetics & phonology in language comprehension & production: Differences & similarities* (pp. 1–53). Berlin: Mouton.

Gregory, S. W. (1990). Analysis of fundamental frequency reveals covariation in interview partners' speech. *Journal of Nonverbal Behavior, 14*, 237–251.

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

Gregory, S. W., Green, B. E., Carrothers, R. M., Dagan, K. A., & Webster, S. (2001). Verifying the primacy of voice fundamental frequency in social status accommodation. *Language and Communication, 21*, 37–60.

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.

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *INTERSPEECH-2011, Florence, Italy*, 3081–3084.

Liberman, A. M. (1996). *Speech: A Special Code*. Cambridge, MA: MIT Press.

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, 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.

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

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology, 32*, 790–804.

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

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. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology: Cognitive Science, 4*, 559.

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., Urmanche, A., Gash, H., Wiener, J., Mason, N., Wilman, S., et al. (2018). The Montclair Map Task Corpus of Conversations in English. Linguistic data corpus available at Montclair State University's data repository at http://digitalcommons.montclair.edu/mmt_corpus.

Pardo, J. S., Urmanche, A., Gash, H., Wiener, J., Mason, N., Wilman, S., et al. (2018). The Montclair Map Task: Balance, efficacy, and efficiency in conversational interaction. *Language & Speech*.

Pardo, J. S. (2017). Parity and disparity in conversational interaction. In E. Fernández & H. S. Cairns (Eds.), *The handbook of psycholinguistics* (pp. 131–152). New York: Wiley-Blackwell.

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.

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.

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.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory & Language, 59*, 413–425.

R Core Development Team. (2016). R: A language and environment for statistical computing, version 3.3.2. Retrieved from http://www.R-project.org/.

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.

Schweitzer, A., & Lewandowski, N. (2013). *Convergence of articulation rate in speech*. Proceedings from in INTERSPEECH-2013.

Shepard, C. A., Giles, H., & Le Poire, B. A. (2001). Communication Accommodation Theory. In W. P. Robinson & H. Glles (Eds.), *The New Handbook of Language & Social Psychology* (pp. 33–56). New York: Wiley.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Attention, Perception, & Psychophysics, 66*(3), 422–429.

Staum Casasanto, L. S., Jasmin, K., & Casasanto, D. (2010). *Virtually accommodating: Speech rate accommodation to a virtual interlocutor.* Proceedings of the 32nd Annual Conference of the Cognitive Science Society, Austin, TX.

Street, R. L. (1982). Evaluation of noncontent speech accommodation. *Language & Communication, 2*, 13–31.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language & Speech, 53*, 510–540. Corpus available at http://groups.linguistics.northwestern.edu/speech_comm_group/wildcat/.

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology, 6*, 546.