Conversational role influences speech imitation

JENNIFER S. PARDO
Montclair State University, Montclair, New Jersey

ISABEL CAJORI JAY
Barnard College and Columbia University, New York, New York

AND

ROBERT M. KRAUSS
Columbia University, New York, New York

This study assessed the impact of a conscious imitation goal on phonetic convergence during conversational interaction. Twelve pairs of unacquainted talkers participated in a conversational task designed to elicit between-talker repetitions of the same lexical items. To assess the degree to which the talkers exhibited phonetic convergence during the conversational task, these repetitions were used to elicit perceptual similarity judgments provided by separate sets of listeners. In addition, perceptual measures of phonetic convergence were compared with measures of articulation rates and vowel formants. The sex of the pair of talkers and a talker's role influenced the degree of phonetic convergence, and perceptual judgments of phonetic convergence were not consistently related to individual acoustic–phonetic attributes. Therefore, even with a conscious imitative goal, situational factors were shown to retain a strong influence on phonetic form in conversational interaction.

One of the most important problems for accounts of speech production and perception is the enormous variability in phonetic form that has been observed in every setting of language use. Classic psycholinguistic treatments of phonetic variability focused on the effects of coarticulation during speech production, leading to the proposal that speech perception relies on mechanisms of speech production to resolve phoneme categories (Browman & Goldstein, 1991; Liberman, 1996). However, such accounts have neglected the wealth of descriptive data provided by sociolinguists that indicate that a great deal of phonetic variability results from the diverse expressive functions that language uses (Eckert, 2008; Eckert & McConnell-Ginet, 1999; Labov, 1974, 1986). Likewise, the literature on communication accommodation has demonstrated contextual influences on phonetic form by examining patterns of speech convergence and divergence in conversational interaction (Giles, Coupland, & Coupland, 1991; Shepard, Giles, & Le Poire, 2001).

Descriptive accounts are necessary for understanding the ecology of language use, but they fall short of providing an explanation of the cognitive mechanisms that support context-conditioned variation in individual talkers. Coarticulation during speech production is a challenge for speech perception, but there is a great deal of contextual phonetic variability that is not accounted for by classic psycholinguistic models (see Pardo & Remez, 2006). Thus, the present study attempts to elucidate the interplay of psy-

cholinguistic and social/situational factors in spoken communication by examining phonetic convergence between interacting talkers during conversational interaction.

Communication Accommodation

According to communication accommodation theory, interacting talkers employ speech convergence to accomplish various aims, which include expressing deference to dominant figures or maintaining social distance (Giles et al., 1991; Shepard et al., 2001). For example, Giles (1973) found that Bristol-accented speakers adjusted the degree of accentedness in their speech according to the accent of an interviewer. When interviewed by a mature experimenter who used received pronunciation (RP, a prestige British English accent), the speech of the Bristol talkers was less Bristol accented than when they had been interviewed by another Bristol talker who was a peer. In another study, Bourhis and Giles (1977) found that Welsh-accented talkers diverged from an RP interviewer, becoming more Welsh-accented after hearing the RP interviewer make a negative remark about their language. These patterns were attributed to attempts on the part of talkers to express deference to the RP interviewer under ordinary circumstances, but to express disdain and maintain social distance in other circumstances.

In the ensuing attempts to characterize the phonetic dimensions of accommodation, all of the studies examined relatively long stretches of speech, finding convergence

J. S. Pardo, pardoj@mail.montclair.edu

mainly in speaking rate or subvocal frequency/amplitude contours (e.g., Giles et al., 1991; Gregory & Webster, 1996; Putnam & Street, 1984; Shepard et al., 2001; Street, 1984). Moreover, convergence was found to occur both up and down social strata. For example, factory workers have been found to converge up to the speech patterns of superior foremen, and doctors have been observed to converge down to that of patients. However, until recently, none of these studies demonstrated that a talker becomes more or less similar in phonetic form to a particular interlocutor. The studies that employed an RP interviewer established patterns of code switching between dialects, but did not directly compare the speech of the Bristol or Welsh talkers with that of their interviewers. Likewise, sociolinguistic surveys provide detailed accounts of phonetic variables employed by individuals across multiple situations, but fall short of demonstrating the dynamics of their development in interpersonal interaction. If interacting talkers are sensitive to and adjust phonetic form to align with that of their interlocutor on a situational basis, then comprehensive accounts of speech production and perception must incorporate these constraints (see Pickering & Garrod, 2004).

Phonetic Convergence

In order to quantify the extent to which interacting talkers converge in phonetic form, Pardo (2006) recorded a set of talkers before, during, and after performance of a conversational task, the Human Communication Research Centre (HCRC) Map Task (see Anderson et al., 1991). The HCRC Map Task was designed to induce between-talker repetitions of the same lexical items by using paired maps with labeled iconic landmarks. For each pair of maps, one map contains a path drawn from a starting point, around various landmarks, and to a finish, and the corresponding map contains similar landmarks without a path. The goal of the task is for the person with the pathless map (designated the *receiver*) to draw the path contained on the corresponding map of the partner (designated the *giver*), according to the partner's spoken directions without seeing each other or each other's maps.

In completing the task, both talkers naturally repeated the landmark label phrases, providing a corpus of spontaneously produced between-talker repetitions of the same lexical items. Phonetic convergence of the conversational between-talker repetitions was assessed by comparing them with pretask and posttask versions of the same phrases in a perceptual similarity test with a separate set of listeners. Because the between-talker repetitions produced during the conversational task sounded more similar in pronunciation to each other than they did to those produced before the talkers had met, the talkers were found to converge in phonetic form during a single conversational interaction. Moreover, phonetic convergence was evident early in the conversational task (within the first 10–20 min), it increased over the course of the interaction, and it was still evident in the recordings sampled in the posttask session immediately after the conversation.

As expected from the literature on communication accommodation, the degree of phonetic convergence was influenced by a talker's role in the conversation (regardless of whether the receiver was converging toward the giver or vice versa). The most typical pattern found in the literature is for talkers to converge up toward an interlocutor of higher status, as in the Bristol talkers' convergence to an RP interviewer (Giles, 1973; Giles et al., 1991). However, Pardo (2006) found that givers converged more to receivers. This reversal could be due to constraints that are unique to the map task situation. Although upward convergence is the default in many situations, talkers are clearly sensitive to the demands of individual tasks and adjust accordingly.

Pardo (2006) also examined the influence of the sex of the pair of talkers, revealing surprising results. As suggested by a previous study on speech shadowing (Namy, Nygaard, & Sauerteig, 2002), it was expected that female talkers might display phonetic convergence to a greater degree than would male talkers. However, Pardo (2006) found that male pairs converged much more than did female pairs overall, and the influence of the talker's role was distinct across male and female pairs. In female pairs, givers converged to receivers and receivers did not converge at all, but in male pairs, both givers and receivers demonstrated phonetic convergence. More research is needed to determine how and why the sex of the pair of talkers would impact phonetic convergence.

The discrepancies with patterns reported in previous research could be due to a real difference in the effect of talker role and pair sex or to the difference in the measure of phonetic convergence. Recall that research on communication accommodation focused mainly on code-switching behavior or on measures of speaking rate, whereas phonetic convergence was assessed using perceptual similarity judgments comparing pretask speech with that produced during the interaction. It is possible that the convergence patterns evident in judgments of global perceptual similarity differ from those manifest in individual attributes such as speaking rate. For example, Bilous and Krauss (1988) reported both convergence and divergence in the same conversations across different indexes (e.g., average utterance length, frequency of pauses, and laughter). Because Pardo (2006) examined only global perceptual similarity, resolving these discrepancies requires assessment of multiple measures in tandem.

The Present Study

The present study builds on previous research by examining phonetic convergence both in terms of global perceptual similarity and also on measures of articulation rate and vowel spectra. To permit an assessment of phonetic convergence, a set of talkers was recorded before, during, and immediately after completion of a conversational task. In an article on the role of entrainment during conversational interaction, Wilson and Wilson (2005) argued that, because turn-taking is a highly coordinated activity, interlocutors should come to match rhythmic attributes of conversational interaction. A talker who intends to continue the conversation smoothly (without a prolonged delay) must be able to anticipate a turn transition point, which is governed by a number of factors, including articulation rate. The demand for close attention

to an interlocutor's articulation rate should lead naturally to entrainment on rate, and articulation rate is the most robust attribute reported in the communication accommodation literature. Moreover, some recent studies using nontraditional measures of vowel quality found that talkers converged to modeled vowel sounds (Delvaux & Soquet, 2007; Gentilucci & Bernardis, 2007; but see Pardo, 2010; Vallabha & Tuller, 2004). Vowels are likely candidates for convergence, due to their inherent productive and perceptual variability (Peterson & Barney, 1952; Pisoni & Tash, 1974) and to their participation in dialectal and indexical variation (Labov, 1974, 1986).

In addition, because phonetic convergence in ordinary conversational interaction was found to be subtle, variable, and susceptible to social/situational factors, it is necessary to assess whether phonetic convergence will become stronger when the focus of attention is placed on phonetic detail. In research on other forms of interactive behavior, participants who were asked to swing wrist pendulums simultaneously or to sit in rocking chairs within sight of each other exhibited entrainment in their oscillatory behavior. That is, they matched the rhythms of their oscillatory motions. Furthermore, coordination with the rhythmic movements of another person or an oscillating stimulus was stronger when participants were instructed to try to maintain coordination deliberately (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007; Schmidt, Richardson, Arsenault, & Galantucci, 2007).

Because deliberate attempts to reach entrainment fare better than does simple observation of a co-actor's behavior, a conscious intention to imitate a conversational partner ought to induce greater phonetic convergence, especially in terms of speaking rate. Furthermore, because female receivers had not been found to converge to givers in the unintentional setting, it is expected that the effect of the instruction to imitate should be greatest for female receivers. Likewise, because all givers had unintentionally converged to receivers, the givers who receive the instruction to imitate should converge more than do their receivers. Therefore, the instruction to imitate should induce the same overall pattern that was found in the unintentional case, but with givers converging to a greater degree when givers are instructed to imitate, and receivers converging to a greater degree when receivers are instructed to imitate. However, because female receivers had not been found to converge unintentionally, their level of convergence might not exceed that of givers. If convergence is influenced by this attentional manipulation, it is not completely driven by the nominal role imbalance or the sex of the talkers. Instead, the effects of talker role and sex that were observed in the unintentional case might be related to their influence on a talker's attention to a partner.

In order to permit comparison with previous research on phonetic convergence, we recorded a new set of talkers performing the modified version of the HCRC Map Task (Anderson et al., 1991). To assess the impact of an explicit attempt to match phonetic attributes, one member of each pair of talkers was instructed to try to imitate the speech of the other talker during the course of the conversation. A separate set of listeners provided measures of global perceptual similarity by judging excerpts from the conversations. In addition, measures of the talkers' articulation rates and pretask and posttask vowel formants assessed whether talkers converged in temporal and/or spectral attributes of speech. These indexes were chosen to permit a direct test of a model of conversational turn-taking proposed by Wilson and Wilson (2005), which predicts that interacting talkers should converge in articulation rate (see also Giles et al., 1991) and to permit comparison with recent studies that have reported imitation of vowel sounds (Delvaux & Soquet, 2007; Gentilucci & Bernardis, 2007; but see Pardo, 2010; Vallabha & Tuller, 2004).

METHOD

Participants

A total of 24 talkers (12 males, 12 females) provided speech samples and were paid at a rate of $10/h for their participation. In addition, 40 participants provided judgments in the similarity test and were compensated with course credit. All participants were native English speakers who reported normal hearing and speech and were drawn from the Columbia University population.

Materials

In the pretask and posttask recording sessions, each talker provided fluent speech samples prompted by a packet of printed sheets containing the landmark label phrases and a set of items that included nine vowels of American English (*heed, hid, head, had, hut, hot, caught, hood, hoot*) randomly ordered with filler words (Peterson & Barney, 1952). For the landmark label phrases, the top of the sheet instructed the talkers to say each phrase in the sentence context "Number *x* is the *phrase*," where *x* was a number printed next to each phrase. The vowel set items were to be embedded in the phrase "Say *word* again."

In the conversational task recording session, each pair of talkers received a printed instruction sheet attached to a packet of six modified map task maps. One member of each pair (the giver) received a set of maps containing landmarks with a path, whereas the other member (the receiver) received a corresponding set of maps without paths drawn on them. The instructions explained the task to the participants and were identical for each participant except for one paragraph. This additional paragraph instructed one member of each pair to try to imitate the speech of the other talker and to avoid letting the other member know about the instruction. The talker whose maps did not contain a path used a pencil with an eraser to draw the paths on the maps.

Procedure

Recording sessions. The talkers were recruited in advance, and members of a pair were unacquainted with each other prior to the experiment. The pairings were all same sex (6 pairs of female talkers and 6 pairs of male talkers) and were formed by placing talkers from different dialect regions together. Table 1 lists the place of origin for each talker, arranged by pair. Most of the talkers had resided in more than one location, and the location in which each talker resided for the longest duration is shown.

Immediately before the conversational task, each member of a pair of talkers provided a set of baseline speech samples individually in a pretask recording session. Each landmark label phrase sentence was produced twice across the session, and each vowel set sentence was repeated five times across the session. To create a rough approximation of conversational speech, the talkers were instructed to generate the sentences quickly and to use a normal speaking voice (i.e., not citation form speech).

During the conversational task, each pair of talkers sat in the same room at individual tables separated by a 7-ft-high, 4-ft-wide barrier that prevented them from seeing each other or each other's maps. One of the members of each pair was randomly assigned to be the

Table 1
Location of Origin and Convergence Data for Individuals Participating in the Recording Sessions

	Place of Origin		Perceived Convergence of Giver to Receiver	Perceived Convergence of Receiver to Giver	Vowel Formant Change*
Participants	Giver	Receiver			
Female Pairs					
1	Wisconsin	California	52	37	23
2	Massachusetts	Indiana	46	54	0
3	Massachusetts	Texas	33	48	25
4	California	New York	51	59	−22
5	Vermont	Ohio	75	42	45
6	California	New Jersey	56	63	−61
Male Pairs					
1	Nebraska	California	61	42	19
2	New York	Canada	59	42	29
3	Maryland	Florida	56	51	78
4	New York	Illinois	47	60	−12
5	Washington	New Hampshire	65	55	−13
6	Ohio	Texas	52	63	20

Note—Recall that givers were instructed to imitate in Pairs 1–3 and receivers were instructed in Pairs 4–6.   *Vowel formant changes are in $F'_1 \times F'_2$ Hz space. The euclidean distances between paired talkers in the pretask session were subtracted from paired talker distances in the posttask session. Negative values indicate a reduction in paired vowel distances from the pretask to the posttask session.

instruction giver and received a set of modified map task maps with paths. The other member of each pair was assigned to be the instruction receiver and received a corresponding set of maps without paths. They were informed that the goal of the task was for the receiver to duplicate the paths that were drawn on each of the giver's maps. Despite the difference in nominal role, both members of each pair were encouraged to converse in order to complete the task.

In addition, one member of each pair was instructed to try to imitate the speech of the other talker without betraying the covert instruction (6 givers, 6 receivers). The imitation instruction appeared as an additional paragraph on the instruction sheet given to the talker. The instructions to their partners were identical, except that the paragraph mentioning imitation was absent. Appendix A contains the instructions that were provided to the imitating talkers.

Immediately after completing the conversational task, the talkers were separated to provide a final set of speech samples in the posttask session, which employed the same materials and procedures as the pretask session had. All recordings were collected via head-mounted AKG microphones and were digitized at a 44-kHz sampling rate.

Phrase selection. A set of landmark label phrases was selected to compose materials for a perceptual similarity test to be performed by separate sets of listeners. First, it was necessary to identify between-talker repetitions of landmark label phrases. As soon as all repetitions had been identified, the final set of phrases was selected according to three constraints: (1) The temporal distribution of the items within the conversation was balanced across conditions to ensure that no experimental condition was confounded with an item's timing of occurrence in the conversation (controlling for the increase in convergence over a session reported in Pardo, 2006); (2) all items were landmark label phrases in which one member of a pair produced the item, and it was repeated by the other member during the next conversational turn; and (3) none of the phrases was a first mention in the discourse.

To satisfy the last two constraints, the set of items included only those phrases in which one member of a pair produced the phrase a first time and then repeated the phrase again before the other member of the pair produced the item—only the noninitial utterance by the first talker and the repetition by the other talker were used. This constraint was necessary, because first mentions in discourse are known to be distinct in emphasis (Bard et al., 2000; Fowler & Housum, 1987; Fowler, Levy, & Brown, 1997).

To permit an assessment of the effect of a talker's role on convergence, the listening tests included items in which the giver provided the initial utterance that the receiver repeated and items in which the receiver provided the initial utterance that the giver repeated. For most of the conditions, this selection process yielded three different landmark label phrases to be used in the perceptual similarity test. Finally, the corresponding pretask and posttask versions of the landmark phrases were excised from their sentences. For these items, the first mentions in the pretask session were not used—only the second iterations of the landmark label phrases were taken. The complete set of landmark phrases used in the perceptual similarity tests appears in Appendix B.

AXB perceptual tests. The landmark phrases were presented to listeners in an AXB similarity test, in which the middle item, X, was produced by one talker in the conversation (giver or receiver), and the flanking items, A and B, were produced by the conversational partner (receiver or giver). For example, during one trial, a listener would hear three repetitions of the phrase *east lake*, produced by the talkers in female pair 1. The middle item would be the one produced by the giver during the map task, and the flanking items would be either the receiver's map task repetition and the receiver's pretask version or the receiver's pretask and posttask versions. The listeners were asked to judge which item, A or B, sounded most like X in the pronunciation of the consonants and vowels.

For half of the trials, the A and B items comprised the pretask (baseline) production of the item and the map task repetition of the same item. These trials provided assessments of convergence by the partner toward the talker (for item X) during the map task session. For the other half of the trials, the A and B items comprised the pretask and the posttask productions of the same items to determine whether phonetic convergence persisted beyond the conversation.

The order of presentation of the A and B items was counterbalanced, and each comparison type was presented three times in random order. There were 200-msec ISIs between AXB utterances, and each trial began 1,000 msec after a listener indicated a response. In order to simplify the task for the listeners, the trials were blocked by the speaker of the X items, and different sets of listeners performed the AXB similarity test for male and female talkers ($n = 20$ each). The materials were presented over Sennheiser headphones connected to Macintosh computers running PsyScope 1.2.5, and responses were collected via keyboard.

RESULTS

The talkers completed the map task session in an average of 45 min, with adequate performance in copying

the paths on the maps, and none of the imitated talkers guessed that their partner was attempting to imitate their speech. We assessed convergence by using the perceptual similarity judgments of landmark label phrases produced during the conversational task, measures of articulation rate during the conversational task, and measures of vowel spectra collected before and after the conversation. We analyzed the data to determine whether talkers converged in phonetic form during the conversation and to assess the impact of the sex of the pair of talkers, the role of the talker, and the instruction to imitate on convergence.

AXB Perceptual Tests

To derive measures of perceived convergence, the listeners' responses in the AXB tests were scored as the percent of trials on which a map task repetition or posttask item was judged to be more similar to a map task sample item than it was to the pretask item. Values greater than 50% chance indicate phonetic convergence, in which a talker's speech became more similar to that of the partner during or immediately after the conversational task than it had been during the pretask session. The AXB data were submitted to a mixed-design ANOVA for the between-subjects factor of pair sex (male vs. female) and the within-subjects factors of role direction (receiver converging to giver vs. giver converging to receiver), session (task session vs. posttask session), and imitator role (giver instructed vs. receiver instructed). Unless otherwise indicated, for all percentages reported in the text, 95% confidence intervals from the analysis confirmed greater-than-chance performance.

Phonetic convergence. Analogous to the findings reported in Pardo (2006), interacting talkers converged in phonetic form during the conversation (53%) and both the sex of the pair of talkers and the role of the talker in the conversation affected the overall degree of convergence. Although female pairs did not converge at greater than chance levels (51%, n.s.), male pairs (54%) converged more than female pairs did [the main effect of pair sex was significant; $F(1,38) = 6.519$, $p < .02$, $\eta_p^2 = .146$]. Likewise, receivers did not converge to givers overall (51%, n.s.), but givers converged to receivers [54%; the main effect of role direction was significant; $F(1,38) = 9.344$, $p < .004$, $\eta_p^2 = .197$]. Finally, convergence occurred during the conversational task (53%) and persisted to the posttask setting [53%; the main effect of session was not significant, but both measures differed from chance; $F(1,38) = 1.119$, $p = .297$, $\eta_p^2 = .029$]. These findings replicate those reported in Pardo (2006), in which paired talkers had not been instructed to imitate. However, the degree of convergence detected in the present study was numerically lower, despite the instruction to imitate.

Imitation instruction. In order to understand the impact of the instruction to imitate, it is necessary to consider the role of the talker who received the instruction. Overall, the pairs in which receivers were instructed to imitate showed convergence (57%), but those in which givers were instructed to imitate did not show convergence [48%, n.s.; the main effect of imitator role was significant; $F(1,38) = 66.153$, $p < .001$, $\eta_p^2 = .635$; note that, of all the factors

analyzed, the role of the imitator has the highest effect size]. Moreover, the instruction to imitate evoked distinct patterns of convergence for male and female pairs of talkers, as is shown in Figure 1.

When receivers were instructed to imitate (panel A), listeners noticed convergence for all pairs of talkers, but female pairs showed a stronger effect of the role of the talker in the conversation; the pattern went in the opposite direction for female pairs than for male pairs. Again, this pattern roughly replicates that reported in the default, no-imitation instruction setting investigated in Pardo (2006). The main difference is that female receivers had not shown phonetic convergence until instructed to do so in the present study.

When givers were instructed to imitate (panel B), listeners noticed divergence for female pairs, regardless of the role of the talkers in the conversation; and in male pairs, givers converged to receivers (light bar), but receivers diverged from givers (dark bar). Thus, when givers
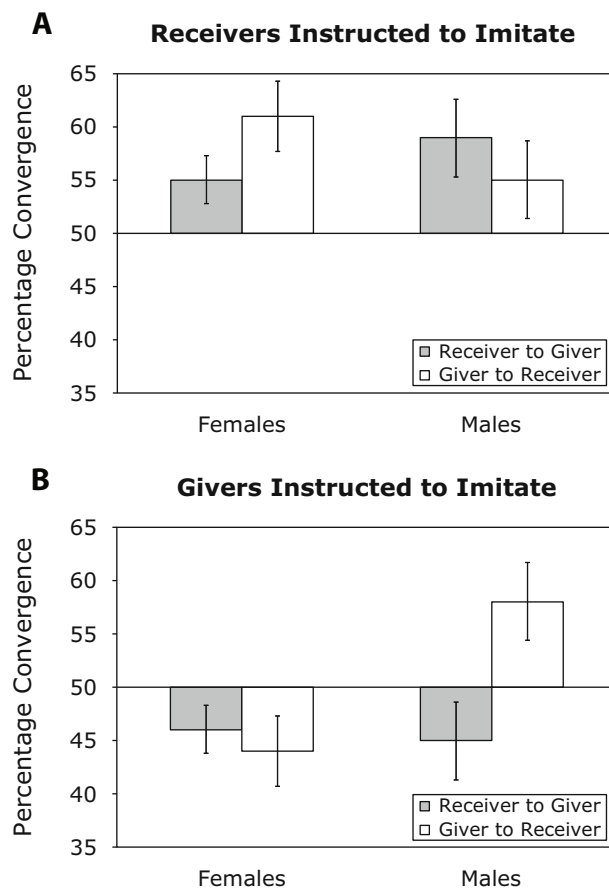

Figure 1. Perceived phonetic convergence as a function of the role of the imitator, the sex of the pair of talkers, and the role of a talker in the conversational task. Panel A displays the pattern of phonetic convergence detected for pairs in which receivers were instructed to imitate; panel B displays the pattern when givers were instructed to imitate. Dark bars indicate similarity of receivers' repetitions to givers' sample items, and light bars indicate similarity of givers' repetitions to receivers' sample items. Values above 50% indicate convergence, values below 50% indicate divergence, and error bars depict 95% confidence intervals.

were instructed to imitate, only male givers followed the instruction and all other talkers diverged. These patterns were verified by a significant interaction between pair sex, role direction, and imitator role [$F(1,38) = 71.015$, $p < .001$, $\eta_p^2 = .651$; error bars depict 95% confidence intervals]. These findings on the impact of the instruction to imitate are surprising, and the implications will be discussed after considering the analyses of articulation rates and vowel spectra.

Articulation Rates

To derive measures of articulation rate, the conversational recordings were first transcribed orthographically and marked for filled and unfilled pauses (silence intervals > 250 msec, as measured by inspection of the waveform in Sound Studio). For each minute of each conversation, we adjusted each talker's total time talking by subtracting the duration of their filled and unfilled pauses. Then, we calculated each talker's articulation rate by dividing the number of words produced by the adjusted total time talking (words/second; see Grosjean & Lane, 1976; Miller & Grosjean, 1981). This procedure yielded minute-by-minute measures of articulation rate that could be used to assess the degree of intertalker similarity in articulation rate over the course of the conversations. Overall, paired talkers did not converge in articulation rate during the conversational task. Rather, articulation rates were influenced by a talker's sex and role, but only when receivers had been instructed to imitate.

Rate convergence. To assess whether paired talkers increased or decreased articulation rates in concert, the measures were submitted to time-series cross-correlation analyses. Cross-correlation coefficients were computed for the paired minute-by-minute rate measures for each of the 12 pairs of talkers for up to seven steps of lag. Most of the coefficients did not approach significance; only five were significant, and, of these, three were negative, indicating divergence in articulation rate. The significant coefficients did not follow any meaningful patterns (female pair 2, lag 1 = −.31; female pair 4, lag 0 = −.34; male pair 1, lag 3 = .64; male pair 3, lag 0 = −.55; male pair 5, lag 0 = .44). Thus, articulation rates measured at 1-min intervals did not reveal consistent patterns of intertalker convergence or divergence over the course of the conversations.

Rate variability. To assess whether aspects of the conversational setting influenced articulation rates, we analyzed the measures to assess the influence of map order, the sex of the pair of talkers, the role direction, and the role of the imitator, in a mixed-design ANOVA testing for the between-subjects factors of pair sex (female vs. male), talker role (giver vs. receiver), and imitator role (giver instructed vs. receiver instructed), and the within-subjects factor of map (1–6). Map order had no effect on articulation rates overall (range = 3.556–3.672 wps) or in interaction with any of the other factors, indicating that average articulation rates did not change consistently over the course of the interaction (all $ps > .05$; $\eta_p^2 < .08$). On average, male talkers produced faster articulation rates than female talkers did [3.716 vs. 3.550 wps; $F(1,40) =$

7.269, $p = .010$, $\eta_p^2 = .154$], and givers produced faster articulation rates than receivers did [3.724 vs. 3.541 wps; $F(1,40) = 11.054$, $p = .002$, $\eta_p^2 = .217$].

As with perceptual similarity, the role of the talker receiving the imitation instruction impacted variability in articulation rates. Overall, instructing receivers to imitate led to faster articulation rates than did instructing givers to imitate [3.718 vs. 3.547 wps; $F(1,40) = 7.697$, $p = .008$, $\eta_p^2 = .161$]. For those pairs in which receivers were instructed to imitate, males talked faster than females did [3.880 vs. 3.557 wps; $F(1,40) = 6.455$, $p = .015$, $\eta_p^2 = .139$] and givers talked faster than receivers did [3.889 vs. 3.548 wps; $F(1,40) = 8.151$, $p = .007$, $\eta_p^2 = .169$]. However, when givers were instructed to imitate, there were no differences between male and female talkers (3.552 vs. 3.542 wps) or between givers and receivers (3.560 vs. 3.534 wps).

These patterns do not align with those observed in the perceptual assessments of phonetic convergence. Instructing receivers to imitate led to the greatest distinctions in articulation rates but also showed the greatest perceived convergence. The exact opposite patterns were found when givers were instructed to imitate. Those pairs with no distinction in articulation rates were not heard as converging. In order to illustrate the distinction in articulation rates, paired talker differences (giver minus receiver) were calculated and are presented in Figure 2. In most cases, givers spoke faster than receivers did, but, when givers were instructed to imitate (bars on left), female receivers showed a trend toward talking faster than did female givers [error bars in the figure depict 95% confidence intervals; $F(1,40) = 9.459$, $p = .004$, $\eta_p^2 = .191$]. Overall, it appears that female pairs were more susceptible to situational factors, in terms of both articulation rates and perceived convergence.

In contrast to the hypothesis, paired talkers did not converge on articulation rates, as shown by the lack of cross-
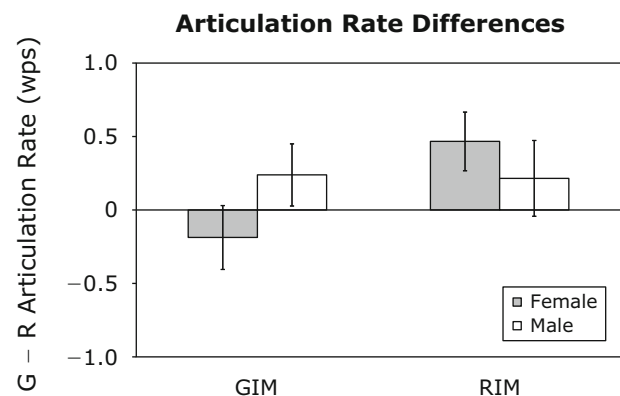

Figure 2. Articulation rate differences between paired talkers (givers' rates minus receivers' rates) as a function of the role of the imitator and the sex of the pair of talkers. The bars on the left show the data for pairs in which givers had been instructed to imitate (GIM), and the bars on the right show the data for pairs in which receivers had been instructed to imitate (RIM). Dark bars indicate differences for female pairs, light bars indicate differences for male pairs, and error bars depict 95% confidence intervals.

correlation between paired talker rates and the observed maintenance of a rate distinction between givers and receivers during the task. When givers were instructed to imitate, listeners heard divergence in female talker pairs, yet articulation-rate differences did not differ from zero for these pairs. When receivers were instructed to imitate, listeners heard convergence for female talker pairs, yet articulation-rate differences were at their greatest. For male pairs of talkers, givers converged to receivers when instructed to do so, yet their articulation rates were faster, and the articulation-rate differences were smallest when receivers were instructed to imitate, aligning with the perceptual data. This lack of complete correspondence between the perceptual data and rate-difference measures indicates that the listeners were sensitive to other acoustic–phonetic attributes when judging similarity between paired talkers.

Vowel Spectra

Because phonetic convergence was found to persist into the posttask session, the next set of analyses compared measures of vowel formants from the pretask and the posttask sessions. Although it would be useful to have a measure of vowels produced during the map task session, the sampling of vowel tokens provided in these spontaneous conversations is limited, leading to potentially unreliable measures. Therefore, the present analyses were undertaken with the complete and balanced sets of items produced during the pretask and posttask recording sessions: five repetitions of each of the nine English vowels produced in consistent consonant frames in randomized sentences.

All vowels were measured by visual inspection of spectrographs produced by the default linear prediction coding analyses available in Praat (www.praat.org; the automatic formant tracking utility provided with Praat was not used due to occasional estimation error). For each item, formant center frequency was estimated at the amplitude maxima of the first and second formants (F_1, F_2) at the midpoint of the vowel. To eliminate the influence of anatomical differences between talkers on vowel spectra, we normalized the formant measures by using an algorithm that preserved dialectal and ideolectal differences between talkers, but scaling every talker's acoustic values into a common acoustic space (see Labov, 2006; Labov, Ash, & Boberg, 2006; Nearey, 1989). All data for all talkers were scaled together via an online utility hosted by the North Carolina State University Linguistics Program using the Labov ANAE (*Atlas of North American English*) extrinsic setting (Thomas & Kendall, 2007; Thomas, Kendall, Yeager-Dror, & Kretzschmar, 2007).

The normalized formant measures were used to calculate paired intertalker euclidean distances in $F_1' \times F_2'$ Hz space for each vowel token during the pretask session and the posttask session. To derive a measure of vowel convergence from the pretask to the posttask session, the intertalker distances in vowel spectra during the pretask session were subtracted from those of the posttask session for each token of each of the nine vowels of each pair of talkers. For these measures, a negative difference would

indicate paired-talker convergence in normalized vowel spectra from the pretask to the posttask session. We conducted a mixed-design ANOVA on the vowel-convergence measures, testing for the between-items factor of vowel (nine vowels) and the within-items factors of pair sex (male vs. female) and imitator role (giver instructed vs. receiver instructed). Overall, talkers did not converge in vowel spectra, and instructing givers to imitate led to divergence in vowel forms.

The instruction to imitate led to an overall increase in the intertalker differences between vowels from the pretask to the posttask session (i.e., divergence) when givers were instructed (21.362 in $F_1' \times F_2'$ Hz), but no real change in the intertalker differences between vowels when receivers were instructed (−5.042 in $F_1' \times F_2'$ Hz). The main effect of imitator role was significant [$F(1,36) = 4.527$, $p = .04$, $\eta_p^2 = .112$]; 95% confidence intervals confirmed that only the giver-instructed condition differed from zero. This finding indicates that instructing givers to imitate led to divergence in vowel formants between paired talkers. In contrast to the perceptual similarity and articulation rate data, there were no significant main effects of the sex of the pair of talkers on vowel-convergence measures ($p > .24$; $\eta_p^2 < .10$).

In relation to the data presented in Figure 1, the pretask and posttask vowel changes appear to align with the perceptual data from the similarity tests. Instructing givers to imitate led to perceived divergence in the similarity test and an increase in intertalker vowel differences from the pretask to the posttask sessions. Instructing receivers to imitate led to perceived convergence in the listening test and a trend toward a decrease in intertalker vowel differences from the pretask to the posttask sessions.

In order to compare the vowel data with the perceptual data directly, we collapsed the pretask to posttask vowel-convergence measures across vowel, which yielded a global vowel-convergence index for each pair of talkers. These data were then compared with the individual pairs' perceptual similarity data, both overall and split by role direction (giver converging to receiver, receiver converging to giver). Table 1 includes the place of origin for each talker, as well as the measures of perceived convergence and vowel formant changes. Vowel convergence was not significantly related to overall perceived similarity [$r(10) = −.331$, $p > .05$] or perceived similarity of givers to receivers [$r(10) = .133$, $p > .05$]. The only significant relationship was between vowel-convergence measures and perceived similarity of receivers to givers [$r(10) = −.590$, $p < .05$], such that reduced intertalker vowel distances were related to greater perceived similarity of receivers to givers. Therefore, the patterns of phonetic convergence detected by ordinary listeners align only partially with measured vowel formants and do not align at all with articulation rates.

DISCUSSION

The present study aimed to assess the impact of a conscious imitation goal on phonetic convergence during conversational interaction. In addition, three measures

of phonetic convergence were compared: global perceptual similarity judgments, articulation rates, and vowel formants. A set of unacquainted talkers participated in a conversational task designed to elicit between-talker repetitions of the same lexical items. To assess the degree to which the talkers exhibited phonetic convergence during the conversational task, these repetitions were used to elicit perceptual judgments provided by separate sets of listeners. The main finding was that phonetic convergence did not occur when givers (except for male givers) were instructed to imitate and occurred primarily when receivers were instructed to imitate. Furthermore, the sex of the pair of talkers and a talker's role influenced the degree of phonetic convergence. Finally, talkers did not show articulation rate entrainment or faithful imitation of vowel formants, and perceptual judgments of phonetic convergence were not consistently related to these individual acoustic–phonetic attributes.

The patterns of phonetic convergence found in the present study both align and contrast with those reported in Pardo (2006), and the main difference between these two studies was the introduction of the instruction to imitate. Male talkers converged more than female talkers did in both studies, and this persistent finding is not readily interpretable, neither by appeal to principle nor by comparison with empirical precedent. Women have been found to converge more than men in a shadowing task (Namy et al., 2002), but other researchers have failed to find a difference in convergence between men and women (e.g., Thomson, Murachver, & Green, 2001).

Namy et al. (2002) posited that, as perceivers, women might be more sensitive to indexical features of talkers, which could have led to the greater convergence that they observed. In studies of speech production, women have been found to produce more formal and prestige forms than men do, to exceed men in the adoption of markers of dialect change, and to differ from men in the underlying perceptual dimensions that they elicit from perceivers (Byrd, 1994; Clopper, Conrey, & Pisoni, 2005; Labov, 2006; Labov et al., 2006). If women show a greater tendency to resolve differentiating phonetic features of talkers and to adopt them more readily than men do, then the finding that men converged more than women in conversational interaction is interesting and merits further investigation. However, participants in these studies were not engaged in an interactive conversational task with a role imbalance. Therefore, it is necessary to consider the impact of a social setting on a talker's phonetic form (see Eckert, 2008; Eckert & McConnell-Ginet, 1999).

In contrast with previous findings and with predictions from a recently proposed model of interactional synchrony (Giles et al., 1991; Jungers & Hupp, 2009; Putnam & Street, 1984; Street, 1984; Wilson & Wilson, 2005), the present study found inconsistent evidence of entrainment on articulation rate. The cross-correlation analyses of the time-series data revealed significant positive coefficients for only two pairs of talkers and negative coefficients for three pairs of talkers. The remaining seven pairs showed no tendency toward cross-correlation of articulation rates for lags of up to 7 min. Analyses of the effects of pair sex, talker role, and imitator role on articulation rates revealed that these situational factors influenced talkers' rates more than did any potential tendency toward rate convergence or divergence. The overall pattern was for givers to maintain a faster articulation rate than receivers, especially among male pairs or when receivers had been instructed to imitate.

Perceived talker confidence, intelligence, and potency has been found to increase with increased speaking rate (Apple, Streeter, & Krauss, 1979; Brown, Strong, & Rencher, 1974), and the present findings could be due to an attempt on the part of givers to assert expertise or to express confidence in their role. A complete understanding of situational influences on articulation rates awaits the outcome of future investigations, but it is clear that an interlocutor's articulation rate is not automatically yoked to the rate of a speaking partner, even in a cooperative task setting with an explicit imitative goal. It is possible that the imitation instruction in the present study was too vague to induce rate entrainment, and it will be necessary to examine the impact of more specific instructions on interlocutor rate entrainment.

The analyses of vowel formants from the pretask to the posttask setting aligned with the perceptual measures to some degree: Instructing givers to imitate led to divergence or no change in vowel formants, and vowel convergence occurred only in two female pairs in which receivers had been instructed to imitate. Because the findings reported here are limited in scope, any conclusions about vowel variability would be pure speculation. Rather, the present findings can serve as a starting point for more extensive investigations of vowel changes resulting from conversational interaction. It will be useful to employ a more rigorous sampling of different dialect regions, more extensive measures of vowel variability, and additional interactive contexts.

The prevailing pattern is that of a lack of correspondence—or of even a contrastive relationship—between measures of individual acoustic–phonetic attributes and measures of global perceptual similarity. Paired talker articulation rates did not entrain, despite empirical precedent and theoretical predictions (Giles et al., 1991; Jungers & Hupp, 2009; Putnam & Street, 1984; Street, 1984; Wilson & Wilson, 2005). Task constraints related to the role-distinction-induced patterns of articulation rate variability that were not tied to those of a partner. Yet, other acoustic–phonetic attributes were employed for convergence by all male givers and by those talkers in pairs in which receivers had received the instruction to imitate.

Finally, the role of the talker impacted phonetic convergence, both in individual pairs and across groups of pairs in which the giver or the receiver had been subject to the imitation instruction. Contrary to the current predictions, instructing givers to imitate virtually eliminated the levels of phonetic convergence observed in the pairs in which receivers had been instructed and in the pairs reported by

2262 PARDO, JAY, AND KRAUSS

Pardo (2006), in which talkers interacted more naturally. Among the pairs in which the givers had been instructed to imitate, only the male givers converged in this experiment, and the level of convergence was not greater than had been found for these pairs in the unintentional setting of Pardo (2006). This finding is especially puzzling, given the consistency in the interpersonal coordination literature that intentional coordination is stronger than unintentional coordination (Richardson et al., 2007; Schmidt et al., 2007). It appears that instructing givers to imitate interfered with whatever process female givers had used to converge; consequently, it induced receivers in these pairs to diverge as well. Among the pairs in which the receivers had been instructed to imitate, all talkers converged, and it is noteworthy that the prediction that female receivers in these pairs would converge was borne out. Although these patterns are not readily interpretable, they contribute to an emerging portrait of the landscape of phonetic variation in natural conversational interaction.

With respect to underlying cognitive mechanisms, speech imitation has been attributed to direct perception of speech gestures (Fowler, Brown, Sabadini, & Weihing, 2003; Shockley, Sabadini, & Fowler, 2004), automatic priming of perceived forms (Pickering & Garrod, 2004, 2007), processes entailed in episodic memory systems (Goldinger, 1998), and mirror neurons (Gentilucci & Bernardis, 2007; see also Rizzolatti & Craighero, 2004). All of these investigations have employed or relied on findings from shadowing paradigms, in which a talker hears an utterance and rapidly repeats it at short delay. They all share a common premise—that perception automatically resolves speech in sufficient detail to support imitative production. Although detailed perceptual resolution is necessary to support convergent production, automatic perceptual resolution is evidently not sufficient to induce phonetic convergence on all occasions (Kraljic, Brennan, & Samuel, 2008; Kraljic, Samuel, & Brennan, 2008; Mitterer & Ernestus, 2008). Rather, a talker adjusts both the resolution and production of phonetic form, in accordance with situational demands, leading to convergence, divergence, or other kinds of phonetic variability. A more comprehensive account of speech production and perception must include an explanation of these patterns.

Although detailed patterns remain difficult to predict, it is clear that a talker's role has profound consequences for acoustic–phonetic variability in conversational interaction. In light of the present findings, a comprehensive analysis of every acoustic–phonetic dimension is unlikely to yield a single attribute that shows reliable accommodation during conversational interaction. Talkers employ acoustic–phonetic variables ad hoc—according to situational constraints—and the resulting patterns of convergence and divergence will be evident only in the multidimensional perceptual space that listeners resolve. Although many previous studies have reported convergence on individual attributes, concurrent variability in unmeasured attributes is equally important. Global perceptual assessments are necessary in order to understand the net impact of diverse situational factors on acoustic–phonetic variability.

REFERENCES

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., et al. (1991). The HCRC Map Task Corpus. *Language & Speech*, **34**, 351-366.

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality & Social Psychology*, **37**, 715-727. doi:10.1037/0022-3514.37.5.715

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory & Language*, **42**, 1-22.

Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language & Communication*, **8**, 183-194. doi:10.1016/0271 -5309(88)90016-X

Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity and intergroup relations* (pp. 119-135). London: Academic Press.

Browman, C. P., & Goldstein, L. (1991). Gestural structures: Distinctiveness, phonological processes, and historical change. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 313-338). Hillsdale, NJ: Erlbaum.

Brown, B. L., Strong, W. J., & Rencher, A. C. (1974). Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *Journal of the Acoustical Society of America*, **55**, 313-318. doi:10.1121/1.1914504

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, **15**, 39-54. doi:10.1016/0167-6393(94)90039-6

Clopper, C. G., Conrey, B., & Pisoni, D. B. (2005). Effects of talker gender on dialect categorization. *Journal of Language & Social Psychology*, **24**, 182-206. doi:10.1177/0261927X05275741

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, **64**, 145-173.

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, **12**, 453-476.

Eckert, P., & McConnell-Ginet, S. (1999). New generalizations and explanations in language and gender research. *Language in Society*, **28**, 185-201.

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 & Language*, **49**, 396-413.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory & Language*, **26**, 489-504. doi:10.1016/0749 -596X(87)90136-7

Fowler, C. A., Levy, E., & Brown, J. M. (1997). Reductions of spoken words in certain discourse contexts. *Journal of Memory & Language*, **37**, 24-40. doi:10.1006/jmla.1996.2504

Gentilucci, M., & Bernardis, P. (2007). Imitation during phoneme production. *Neuropsychologia*, **45**, 608-615.

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

Giles, H., Coupland, J., & Coupland, N. (Eds.). (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**, 251-279.

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 & Social Psychology*, **70**, 1231-1240. doi:10.1037/0022-3514.70.6.1231

GROSJEAN, F., & LANE, H. (1976). How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 538-543. doi:10.1037/0096-1523.2.4.538

JUNGERS, M. K., & HUPP, J. M. (2009). Speech priming: Evidence for rate persistence in unscripted speech. *Language & Cognitive Processes*, **24**, 611-624. doi:10.1080/01690960802602241

KRALJIC, T., BRENNAN, S. E., & SAMUEL, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, **107**, 54-81.

KRALJIC, T., SAMUEL, A. G., & BRENNAN, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, **19**, 332-338.

LABOV, W. (1974). Linguistic change as a form of communication. In A. Silverstein (Ed.), *Human communication: Theoretical explorations* (pp. 221-256). Hillsdale, NJ: Erlbaum.

LABOV, W. (1986). Sources of inherent variation in the speech process. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 402-425). Hillsdale, NJ: Erlbaum.

LABOV, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics*, **34**, 500-515.

LABOV, W., ASH, S., & BOBERG, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change.* New York: Mouton de Gruyter.

LIBERMAN, A. M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.

MILLER, J. L., & GROSJEAN, F. (1981). How the components of speaking rate influence perception of phonetic segments. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 208-215. doi:10.1037/0096-1523.7.1.208

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

NAMY, L. L., NYGAARD, L. C., & SAUERTEIG, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language & Social Psychology*, **21**, 422-432.

NEAREY, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, **85**, 2088-2113.

PARDO, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, **119**, 2382-2393.

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). New York: Psychology Press.

PARDO, J. S., & REMEZ, R. E. (2006). The perception of speech. In

M. Traxler & M. A. Gernsbacher (Eds.), *The handbook of psycholinguistics* (2nd ed., pp. 201-248). New York: Academic Press.

PETERSON, G. E., & BARNEY, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, **24**, 175-184.

PICKERING, M. J., & GARROD, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral & Brain Sciences*, **27**, 169-226.

PICKERING, M. J., & GARROD, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, **11**, 105-110.

PISONI, D. B., & TASH, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, **15**, 285-290.

PUTNAM, W. B., & STREET, R. L., JR. (1984). The conception and perception of noncontent speech performance: Implications for speech-accommodation theory. *International Journal of the Sociology of Language*, **46**, 97-114.

RICHARDSON, M. J., MARSH, K. L., ISENHOWER, R. W., GOODMAN, J. R. L., & SCHMIDT, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, **26**, 867-891. doi:10.1016/j.humov.2007.07.002

RIZZOLATTI, G., & CRAIGHERO, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, **27**, 169-192.

SCHMIDT, R. C., RICHARDSON, M. J., ARSENAULT, C., & GALANTUCCI, B. (2007). Visual tracking and entrainment to an environmental rhythm. *Journal of Experimental Psychology: Human Perception & Performance*, **33**, 860-870. doi:10.1037/0096-1523.33.4.860

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 and social psychology* (pp. 33-56). New York: Wiley.

SHOCKLEY, K., SABADINI, L., & FOWLER, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, **66**, 422-429.

STREET, R. L., JR. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, **11**, 139-169.

THOMAS, E. R., & KENDALL, T. (2007). NORM: The vowel normalization and plotting suite. [Online resource: http://ncslaap.lib.ncsu.edu/tools/norm/]

THOMAS, E. R., KENDALL, T., YEAGER-DROR, M., & KRETZSCHMAR, W. (2007). *Two things sociolinguists should know: Software packages for vowel normalization, and accessing linguistic atlas data*. Paper presented at Workshop at New Ways of Analyzing Variation (NWAV) 36. Philadelphia: University of Pennsylvania.

THOMSON, R., MURACHVER, T., & GREEN, J. (2001). Where is the gender in gendered language? *Psychological Science*, **12**, 171-175. doi:10.1111/1467-9280.00329

VALLABHA, G. K., & TULLER, B. (2004). Perceptuomotor bias in the imitation of steady-state vowels. *Journal of the Acoustical Society of America*, **116**, 1184-1197.

WILSON, M., & WILSON, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, **12**, 957-968.

APPENDIX A
Map Task Recording Instructions

One person is the direction giver and the other is the direction receiver.

The giver has a packet of maps with paths drawn around various iconic landmarks, and the receiver's packet has similar maps without paths. The maps are matched and numbered in order. However, you will find that there are some differences between the maps in the composition of landmarks, as if they were drawn by two different explorers—most of the landmarks match, but some are missing from either map.

Your task is to communicate verbally so that the receiver can duplicate the giver's path on his/her map, *without looking at each other or each other's maps*. The receiver should draw the path on each map with the pencil provided, and erasures are permitted if necessary.

Also, we have found that if 1 talker imitates the other talker without their knowledge, the performance on this task improves. *Please try to imitate your partner's speech as you perform the task, but do not mention to him/her that you are imitating.* By imitation, we mean making your speech sound like your partner's speech.

Go through the packet of 6 map pairs in order, taking as long as you like for each map. When you have finished with all 6 maps, the person closest to the door should get the experimenter without additional conversation. You will be given an opportunity to compare maps after completing the final recording phase of the experiment.

Please begin as soon as you both are ready.

APPENDIX B
Items Used in AXB Listening Tests of Phonetic Convergence

The table below lists the items used to assess phonetic convergence for each cell of the design. There were a total of six pairs of female talkers and six pairs of male talkers, each split into groups of three pairs. For Pairs 1–3, the givers were instructed to imitate receivers, and for Pairs 4–6, the receivers were instructed to imitate givers. Moreover, on some occasions during the conversation, a giver produced an item that a receiver repeated (receiver convergence to giver), whereas on other occasions, a receiver produced at item that a giver repeated (giver convergence to receiver). Once the conversational repetitions were identified, the corresponding utterances from the pretask and posttask sessions were sampled and used as flanking items. Using these items, the AXB listening tests assessed phonetic convergence by pair sex, talker role, and role of imitator.

Instructions		Female Pair 1	Female Pair 2	Female Pair 3
Givers instructed to imitate	Receivers repeat givers	east lake walled city winter garden	flowing river green bay walled city	blacksmith cattle ranch
	Givers repeat receivers	baboons farmed land old truck	baboons east lake tall pine	graveyard wishing well
		Female Pair 4	Female Pair 5	Female Pair 6
Receivers instructed to imitate	Receivers repeat givers	east lake green bay parked van	camera shop green bay monastery	large cottage farmed land remote village
	Givers repeat receivers	flowing river walled city west lake	east lake farmed land pyramid	east lake marshland winter garden
		Male Pair 1	Male Pair 2	Male Pair 3
Givers instructed to imitate	Receivers repeat givers	blacksmith meadow wheat field	large cottage fallen rocks winter garden	crest falls east lake walled city
	Givers repeat receivers	country road east lake fallen rocks	dead tree sandy shore walled city	diamond mine picket fence teepees
		Male Pair 4	Male Pair 5	Male Pair 6
Receivers instructed to imitate	Receivers repeat givers	east lake marsh land old truck	east lake fallen rocks graveyard	camera shop east lake monastery
	Givers repeat receivers	baboons dead tree telephone booth	dead tree meadow wheat field	farmed land golf course