On phonetic convergence during conversational interaction

Jennifer S. Pardo[a)]
Psychology Department, Barnard College, 3009 Broadway, New York, New York 10027

(Received 6 May 2005; revised 28 January 2006; accepted 30 January 2006)

Following research that found imitation in single-word shadowing, this study examines the degree to which interacting talkers increase similarity in phonetic repertoire during conversational interaction. Between-talker repetitions of the same lexical items produced in a conversational task were examined for phonetic convergence by asking a separate set of listeners to detect similarity in pronunciation across items in a perceptual task. In general, a listener judged a repeated item spoken by one talker in the task to be more similar to a sample production spoken by the talker's partner than corresponding pre- and postinteraction utterances. Both the role of a participant in the task and the sex of the pair of talkers affected the degree of convergence. These results suggest that talkers in conversational settings are susceptible to phonetic convergence, which can mark nonlinguistic functions in social discourse and can form the basis for phenomena such as accent change and dialect formation. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2178720]

PACS number(s): 43.70.Gr, 43.71.Bp, 43.70.Bk, 43.71.An [ARB] Pages: 2382–2393

I. INTRODUCTION

Speech is variable in its realization, both within and between talkers, despite apparent consistency in perception. Somehow, a listener is able to overcome the phonetically disparate productions of phonemes to arrive at what a talker intends to say. Recent studies of within and between-talker imitation have attempted to resolve the ongoing debate over whether speech perception yields acoustic, auditory, articulatory, gestural, or more abstract phonological parameters by positing a close connection between perception and production. In particular, such accounts propose that speech perception yields linguistically significant gestural parameters that automatically drive production, leading inevitably to imitation (such gestures are synergies of articulators; see Browman and Goldstein, 1991; Fowler *et al.*, 2003; Goldinger, 1998; Sancier and Fowler, 1997). A recent paper by Pickering and Garrod (2004) likewise proposes an automatic priming account for lexical, syntactic, and schematic parity in language use. The current study attempts to constrain such proposals by estimating the extent to which imitation in conversational settings is exact and automatic.

If the function that yields communicative parity is one in which perception of a linguistic unit automatically drives production of that unit, via priming or another cognitive mechanism, then imitative productions ought to match their models to some tolerance. There are at least two problems with such accounts. First, the question of matching depends crucially on which attributes are assessed, and in particular, the level of abstraction of the attributes. For example, when two talkers use the same word to designate an ambiguous figure, there is a match in relation to the lexicon; however, it is unlikely that two productions of the same word match at an acoustic-phonetic level. Indeed, phonetic imitation may be impossible to achieve in purely acoustic terms (Krauss and Pardo, 2004; see also Vallabha and Tuller, 2003, and Viechnicki, 2002). Even for a single talker, no two productions of the same phonetic segment are acoustically identical, nor do articulatory patterns match for the same phonemes under different speaking conditions (e.g., Perkell *et al.*, 2002). Therefore, at a phonetic level imitation is likelier to be graded and inexact, rather than perfectly compliant with the acoustic, articulatory, and phonetic detail of the spoken model for imitation. A satisfactory account of communicative parity must specify the particular dimensions that are relevant as well as the range of tolerance for mismatches to exact parity.

Second, an automatic mechanism of the kind discussed in the literature is characterized as an all-or-none obligatory function, with no processes intervening between perception and production. According to Chartrand and Bargh (1999), a "perception-behavior link posits the existence of a natural and nonconscious connection between the act of perceiving and the act of behaving, such that perceiving an action being done by another makes one more likely to engage in that same behavior" (p. 900). In the domain of spoken language, Pickering and Garrod (2004) claim that "as dialogue proceeds, interlocutors come to align their linguistic representations at many levels ranging from the phonological to the syntactic to the semantic. This interactive alignment process is automatic and only depends on simple priming mechanisms that operate at the different levels, together with an assumption of parity of representation for production and comprehension" (p. 188). That is, if perception yields parameters that automatically drive production, the straight line from perception to production is not modulated by intervening processes. Whatever parameters perception resolves, whether categorical or continuous, production must follow suit. In order to account for discrepancies between two instances of the same utterance (when they occur), such a model must assert that the differences occur as a result of error or noise in perception and/or production, and do not pattern according to other functions. At this point, such proposals do not adequately address the influence of processes

[a)]Electronic mail: jsp2003@columbia.edu

 0001-4966/2006/119(4)/2382/12/$22.50

outside of perception/production, the effects of which the current study attempted to evaluate. The next section reviews evidence of a close connection between perception and production, before moving to the main question addressed by this study: Whether ordinary circumstances of language use evoke phonetic convergence.

A. Perception-production link and imitation

Evidence for a close connection between speech perception and production spans a number of different time scales, from immediate shadowing of nonsense syllables through sentence syntax. At the briefest time scale, studies of simple versus choice response times in speech shadowing tasks show that speech perception yields targets for speech production very quickly (e.g., Fowler *et al.* 2003; Porter and Castellanos, 1980). In typical experimental investigations of response time, a participant must respond to a signal by performing a single response (in the simple condition) or by performing one of several alternative responses (in the choice condition). Fowler *et al.* note that choice response times exceed simple response times in similar tasks by 100 to 150 ms on average, and this difference is presumed to be due to an additional decision-making process in the choice condition. In studies of speech production response time, listeners hear a series of VCV utterances spoken by another talker that varies in the identity of the medial consonant. On each trial, a listener shadows the initial portion of the VCV utterance, and then responds to the consonant transition with either a single CV, or the CV that was produced by the model. Shadowers respond to the consonant transition in the choice condition almost as quickly as the simple condition (from 26 to 50 ms difference across different experiments; Fowler *et al.* 2003; Porter and Castellanos, 1980), presumably because the signal itself provides information for its articulation, obviating the need for a stage of decision-making requiring choice among abstract phoneme categories preliminary to articulating the response on each trial.

Fowler *et al.* (2003) took this paradigm a step further by showing that shadowers track subcategorical variability in consonant voicing by adjusting voice onset time (VOT) in their own shadowing responses toward those of the spoken models. However, it is important to note that shadowers did not match model VOTs precisely. Fowler *et al.* attribute this discrepancy to an influence of habitual action in speech production, and not to processes intervening between or superordinate to perception and production. A similar explanation was offered by Sancier and Fowler (1997) for a bilingual talker who varied her VOTs in both languages in the direction of her most recent language environment, but did not match the distribution in her second language. Accordingly, speech perception specifies gestural actions, and the direct link to speech production yields an imitative response with a moderate degree of fidelity. In contrast, Vallabha and Tuller (2004) explicitly asked talkers to imitate their own isolated steady-state vowels and measured the acoustic discrepancy between the sample and imitated versions. They found systematic biases (within each individual talker) in the discrepant repetitions that were not accounted for by models of random noise in perception or production. They pointed tentatively to dialectal differences among their talkers, which are not characterized by generic perceptual and productive models, as a likely source for the observed idiosyncrasies in the systematically biased imitations.

Although speech perception appears to specify gestural parameters that support rapid detailed shadowing, these studies find that talkers never match input signal properties, whether such signals were produced by others or by the talkers themselves. The shadowing paradigm is conducive to imitation, yet the acoustic output reflects both perceptual/productive limitations on fine-grain accuracy and the influence of other factors that induce directional biases in the discrepancies. As of yet, there is no account that explains how such factors intrude upon a presumably automatic and direct perception-production link.

At a slightly greater time scale, shadowing tasks also provide evidence for imitation of lexical items (Goldinger, 1998; Namy, Nygaard, and Sauerteig, 2002). In these studies, talkers were recorded producing words prompted from a list, and these items were then used as models in a shadowing task with different talkers. In order to determine whether shadowers imitated the models in their use of phonetic variants, independent listeners were recruited to provide perceptual judgments of imitative fidelity in the shadowed utterances. The motivation behind this methodological innovation is that perception integrates across multiple acoustic-phonetic dimensions, thereby providing a more configural assessment of imitation than a selected acoustic measure would. The listeners performed an AXB task in which they heard three versions of the same lexical items and judged which item produced by the shadowing talker, A or B (taken from the pretask and shadowed sessions), sounded like a better imitation of (Goldinger) or was more similar to (Namy *et al.*) the model's sample item, X. Listeners chose a shadowed item more often than a previously produced item as a better imitation of a sample item; however, performance was variable, perhaps indicating inconsistent degrees of imitative fidelity or inconsistencies in perceptual judgments of imitation. Some of the variability was accounted for by factors related to episodic memory (Goldinger) or to talker sex (Namy *et al.*).

As these studies demonstrate, a linked perceptual-productive system might produce convergence, if not a perfect imitative match, in the acoustic-phonetic and sublexical domains. Studies of language use in broader settings and time scales support the proposed link, but also hint at other influences on convergence that are not readily encompassed by strict amalgamation of perceptual/productive mechanisms. A central phenomenon identified by such approaches is an increase in similarity among linguistic components, a process variously termed *convergence, accumulating common ground*, or *alignment*. Over longer stretches of speech, interlocutors are known to converge in speaking rate (Giles, Coupland, and Coupland, 1991), subvocal frequency/amplitude contour (Gregory, 1990), and vocal intensity (Natale, 1975); to establish and increase common ground to the exclusion of over-hearers (Schober and Clark, 1989); and to align description schemes (Garrod and Doherty, 1994) and

syntactic constructions (Branigan, Pickering, and Cleland, 2000). In addition, a talker will attenuate or accentuate regional dialect expression in response to an interviewer's expressed attitude toward regional dialect or use of formal or regional dialect in British English, but phonetic similarity between talkers in these settings was not assessed directly (Bourhis and Giles, 1977; Giles, 1973). Considering these phenomena theoretically, Labov (1976, 1984) suggests that dialect formation and change result largely from opportunities for direct social contact among talkers and are influenced by social relationships between interacting talkers. Missing from this literature is an assessment of how conversational settings promote or hinder a putatively ineluctable tendency to imitate at a fine phonetic grain.

B. Measuring phonetic convergence

In order to examine phonetic variability in social interaction, the current study collected a conversational speech corpus and performed perceptual measures of phonetic convergence. The goal of this experiment was to determine whether pairs of talkers converged in phonetic repertoire over the course of a single conversational interaction. Phonetic convergence is an increase in segmental and suprasegmental similarity of the speech of one talker to another. Although sublexical imitation/similarity increases when a talker simply shadows another's speech at short latency in the laboratory (Fowler *et al.*, 2003; Goldinger, 1998; Namy *et al.*, 2002), the current experiment examined a situation that evoked this process in more natural communicative contexts.

The first part of the experiment collected samples of speech before, during, and after pairs of talkers interacted in a conversational task. To permit assessment of phonetic convergence, the task had to elicit the same lexical items spoken across partners. Moreover, these items had to be identified prior to the conversation in order to collect preinteractive tokens. These constraints were satisfied by the map task, which was developed by the Human Communication Research Center at the Universities of Glasgow and Edinburgh, Scotland (Anderson *et al.*, 1991; see Appendixes A and B for a sample pair of map task maps). The map task uses paired schematic maps that contain labeled illustrated landmarks: One member's map includes a path drawn from a starting point, around various landmarks, to a finishing point, and the companion's map contains only a starting point and landmarks. The goal of the task is for the talkers to communicate effectively enough so that the path on the first map, which cannot be seen by the holder of the pathless map, can be duplicated on the second map. Completion of the task requires active involvement of both participants, and spoken samples of the landmark labels can be collected both before and after the conversational interaction, to compare to those that are produced by both participants over the course of conversation. In addition, the task permits assignment of different social roles—one member is the instruction giver and the other is the instruction receiver.

The repetition of landmark label phrases from the map task between talkers enables assessment of phonetic convergence in an AXB perceptual test. In addition to using a conversational corpus to provide materials, the AXB test for convergence differed from previous research on two other points. First, this study restricted perceptual judgments to assessments of similarity in pronunciation. Pronunciation was chosen as a precise and readily accessible concept for untrained listeners to use. The listeners were encouraged to focus specifically on the way that the talkers were articulating the consonants and vowels, and to decide which item sounded more similar to the sample item in pronunciation. This specification was introduced because a listener who is asked explicitly to judge imitation or simply to judge similarity may focus on other nonphonetic attributes, such as the melodiousness of the vocal quality, or the apparent emotionality in the voices, or any idiosyncratic dimension; and this study is particularly concerned with changes in a talker's phonetic repertoire while interacting with another talker, regardless of a conscious intention to imitate. Second, prior assessments of the persistence of imitation used items produced after a delay of a few seconds from sample presentation (Goldinger, 1998). The current study used items produced immediately after the conversational interaction and not directly prompted by another talker's utterance. Although this method induced a longer interval between a sample and its repetition, the talkers in the current study interacted in a coordinated social setting, which may have produced more robust and persistent convergence (see Pickering and Garrod, 2004).

Evidence for phonetic convergence in this experiment would consist of finding that a talker's speech became more similar in pronunciation to the partner's speech than it was before the interaction. If a talker converged toward his/her partner, then the landmark label phrases spoken in response to a partner's utterance should sound more similar to that utterance than an item produced by the talker before or after the task. If phonetic variation is not tied to a particular setting, then all variants should sound equally like or unlike that of a conversational partner's utterances. This study also investigated the time course of convergence by measuring convergence early and late in the conversational session, and by assessing whether talkers persisted in convergence into the post-task session. In addition, Giles' communication accommodation theory (Giles, Coupland, and Coupland, 1991; Shepard *et al.*, 2001) predicts that talkers may have varied in degree of convergence depending on conversational role. For example, a giver's more dominant role could have led to greater convergence on the part of a receiver. It is also worth considering that the sex of the pair of talkers might have influenced the degree of phonetic convergence, such that female talkers might have converged more than male talkers, as found by Namy *et al.* (2002).

II. METHOD

A. Materials

1. Corpus elicitation

For the conversational task, each participant received a packet of five 8.5 by 11-in. sheets of paper printed with map task maps. The instruction giver had a set of five map task

maps with paths, and the instruction receiver had a similar set of maps without paths and used a pencil to draw the path on his/her map. Some of the map task landmark labels were modified from the original set to adjust for differences between British and American-English naming conventions; however, none of the iconic drawings or paths was changed from the originals. The pre- and post-task sessions' items were prompted with a packet of printed sheets listing the map task landmark labels and other lexical items in the following order: (1) A numbered set of map task landmark labels each to be spoken in the carrier phrase, "Number_is the_;" (2) Five randomized repetitions of the complete American English vowel set embedded in hVt context and filler words, each to be spoken in the carrier phrase, "Say_again;" and (3) A second numbered set of map task landmark labels each to be spoken in the carrier phrase, "Number_is the_." In both the pre- and post-task sessions, the talkers were encouraged to produce the sentences fluently in a normal speaking style. The vowel samples were collected for a separate study and served as fillers between the sets of landmark label phrases.

2. Convergence assessment

From the set of map task landmark labels that was repeated between both members of a pair, the listening test used samples of four items that were repeated across all six pairs of talkers—*abandoned monastery, green bay, walled city*, and *wheat field*. The selection of this set of items was guided by four design constraints: (1) Use of between-talker repetitions of conversational items that occurred within a relatively short period of time; (2) Exclusion of the first mentions of the items in the discourse, which have been found to be of longest duration, among other distinctions (Bard *et al.* 1991; Catchpole and Pardo, 2003; Fowler, 1988; Fowler and Housum, 1987; Fowler, Levy, and Brown, 1997; Krauss and Weinheimer, 1964); (3) Use of conversational items that were produced in clause- or sentence-final position, where the end of a clause was defined as the point where a breath or full pause occurred, in order to match the sentence-final position of the items produced in the pre- and post-task sessions; and (4) Use of landmark label phrases for items that were shared across the giver and receiver maps. As a result of these constraints, the items varied with respect to whether they were second or later mentions in the discourse (the factors distinguishing first from second and later mentions do not distinguish second from later mentions), and the items also varied in their discourse function (sometimes the repetitions were of the same type, sometimes of different types), but none of these considerations was confounded with the experimental variables considered in the analyses.

To test for differences in degree of convergence over the course of the interaction, for each pair, two of the repeated items were taken from early in the conversation and two were taken from later in the conversation. Items designated as early occurred prior to the halfway point in each pair's interaction, and late items occurred after that point. To assess effects of talker role, for each pair, two of the items were repeated from giver to receiver (GX repeated by receiver), and the other two were repeated from receiver to giver (RX repeated by giver; and also not confounded with timing). Finally, corresponding productions of the items from the pre- and post-task sessions of the appropriate talker served as competitors for the task repetitions in the AXB similarity test. In both nontask sessions, each talker produced the set of landmark labels twice—those items taken from the pretask session came from the second iteration of the set to use the most fluent and reduced productions possible, and those items taken from the post-task session came from the first iteration of the set to use productions from the briefest interval.

B. Procedure

1. Corpus elicitation

To provide speech samples, each talker sat at a desk in a sound-attenuated booth approximately 18 in. away from a desk-mounted dynamic microphone. All utterances were recorded onto analog cassette tapes via a Denon stereo cassette tape deck, which operated outside the booth. The utterances and biographical information collected in the pretask session were analyzed to determine pairings for the map task session. First, to avoid social dominance phenomena associated with mixed-sex pairs (see Bilous and Krauss, 1988; Namy *et al.*, 2002), this study employed same-sex pairs. Second, measures of average *F*0 in the hVd/t items were clustered to select pairs whose *F*0's were not exactly the same, but were proximal to each other in average *F*0. Goldinger's (1998) talkers were first compared in multidimensional scaling analyses of similarity ratings in order to maximize variability in his talker set. Although this method may be ideal for ensuring an evenly distributed set of talkers, the current study is concerned with more natural conversational settings, in which talker characteristics vary more freely. Therefore, the approach used here ensured that talkers differed in an acoustic attribute prior to contact, increasing the likelihood that some measurable difference would result without imposing a strict criterion on talker variability. Third, analyses of the filler words in the list verified that all talkers differentiated among vowels in the word sets, *marry/Mary* and *merry, cot* and *caught*, and *pen* and *pin*; however, the biographical information provided by the participants indicated that they were drawn from varied regional backgrounds. Thus, the talker set did not exhibit a homogeneous dialect; none of the talkers exhibited a strong regional accent; and, the pairs incidentally comprised individuals from different dialect regions. Finally, in none of the pairs were the members acquainted with each other prior to participating in the map task session.

The talkers returned between one and two weeks after the pretask session in same-sex pairs to participate in the map task session and to provide post-task speech samples. For the map task session, the talkers sat at identical desks in the same sound-attenuated booth and were separated by a plywood divider that prevented them from seeing each others' maps, bodies, and faces. To permit measurements of between-talker repetition latency, one talker's microphone recorded onto the left channel of a tape, and the other recorded onto the right channel. Prior to beginning, each talker

in a pair was assigned a role for the duration of the map task session—one talker was designated the instruction giver, and the other talker was the instruction receiver. However, both participants were encouraged to converse in order to complete the maps as accurately as possible. The ordering of the maps was varied across pairs. Once a session began, a pair of talkers performed the map task pairs in order until they completed all five; most pairs spent 30–60 min on the task session. At the end of a map task session, the talkers provided speech samples for the post-task session separately.

2. Convergence assessment

Because the perceptual measure of phonetic convergence requires lexical repetition, the recordings were first coded for timing of each instance of use of a map task label phrase by each talker. In addition, the accuracy and duration of completion of each map in the map task were measured. To determine accuracy, a similar technique was used as that of Anderson *et al.* (1991): A transparency of each of the giver's maps was created with a 1- by 1-cm grid superimposed, and then placed over the appropriate receiver's map. Accuracy was tallied as the proportion of the number of grid cells that a receiver's path duplicated a giver's path to the total number of grid cells on a giver's path. Descriptive analyses of these data from the map task session were conducted to confirm that there was effective communication during performance of the map task.

All items were digitized (44 kHz, 16 bit) from audiotape using a Denon stereo tape deck connected to a Power Macintosh 6100/60AV computer running SOUNDEDIT 16 (by Macromedia, Inc.). Minimal digital editing was required to remove all items from their contexts, either in running conversational speech or in sentences from the pre- and post-task sessions, and to remove infrequent artifacts produced by talker movements. Such noises were brief and were usually caused by the talker bumping into the desk or microphone. They were removed by excising the noise at zero crossings so that the editing was unnoticeable in the final tokens. The listening tests were conducted in quiet testing rooms and presented over Sennheiser HD280 Pro circumaural headphones connected to Macintosh G3 computers running PSYSCOPE (Cohen *et al.* 1993).

As shown in Fig. 1, the AXB similarity test consisted of a series of trials in which a listener heard three repetitions of the same landmark label phrase, and the phrase varied across trials. On each trial, a task sample item produced by either a giver or receiver (X) was flanked by two items at 200-ms ISI from the corresponding partner (A and B). In order to assess convergence within the conversation, the task repetitions were compared to an item from the pre- or post-task session. The post-task comparison condition was included to assess the persistence of convergence. In addition, another measure of persistence was taken by comparing pre- and post-task items directly to task sample items. If phonetic convergence does occur and extends beyond the conversation, post-task items ought to sound more similar to the task sample items than pretask items.

The pre- and post-task direct comparison condition also alleviates a potential confound in the other two comparison types. When a listener chooses the task repetition as more similar to the task sample item than the pre- or post-task items, the decision could be based on putative differences between spontaneous and prompted speech. However, most studies demonstrating accurate perceptual classification of read and spontaneous speech have used stretches of discourse longer than the short phrases used here. Blaauw (1994) cites an unpublished finding that classification is above chance for items that are 4 to 6 syllables long, but also notes that performance is much lower than for longer stretches of speech. Even for longer stretches of speech, Blaauw found that classification of "the spontaneous-read distinction is not strictly categorical, but seems to be of a more continuous character" (p. 366). Therefore, read and spontaneous speech samples vary in the degree to which they support perceptual classification as such, and the ability to make the distinction is susceptible to the duration of the sample used to elicit the judgment. At this point, a listener's ability to judge whether a short phrase was read or spontaneously produced is uncertain, and the current protocol attempted to minimize the kinds of prosodic markers that have been found to support perception of the difference by encouraging fluent production of the items in the prompted sessions, and by using items sampled from the second iteration of the map task phrase list in the pretask session. Furthermore, the AXB comparison condition that compared pretask and post-task items directly is not susceptible to a spontaneous-read distinction because both items were prompted from lists.

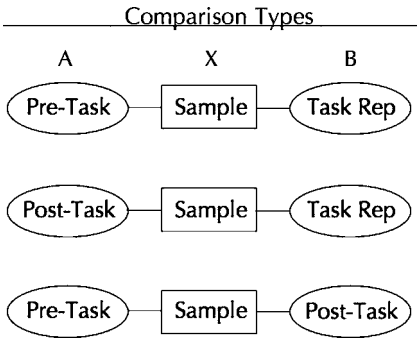


FIG. 1. AXB trial structure. On each trial, a listener heard the same landmark label phrase three times. The first and last items were to be compared to the middle item for similarity in pronunciation. The sample item was a giver's or receiver's task session production, and the task repetition was the corresponding partner's repetition of the same phrase. One-third of the time, the phrase was compared to the partner's production of the item from the pretask session, or to an item from the post-task session, or a pretask and a post-task item were compared. Half the trials used items that were repeated from giver to receiver, and the other half, from receiver to giver.

For each triplet, a listener decided as quickly as possible which item, the first or the last, sounded more similar to the middle item in its pronunciation. Listener responses were collected via the number 1 (first item) and 0 (last item) keys on the keyboard. Each successive trial began 1000 ms after a listener indicated a response. The order of presentation of flanking items was counterbalanced, each trial type was presented three times in mixed random order, and the effects of *timing, persistence/comparison type, talker role*, and *pair sex* were all tested within subject (blocked by *talker role*, to keep

the speaker of the task sample item the same throughout a block).

C. Participants

The talkers were six men and six women from the undergraduate population of Yale University who were paid for their participation. All participants reported that they were native speakers of American English with no speech or hearing disorders.

A total of 30 listeners participated in the AXB similarity test. All were native American English speakers from the student population of Columbia University who reported normal hearing and received Introductory Psychology course credit or were paid for participation.

III. RESULTS

A. Map task performance

The accuracy of path duplication in the map task (percentage of path grid cells on a giver's map that were duplicated on a receiver's map) was 85% overall, and there were no significant differences in performance across the dataset. The average time spent on each map was 8.84 min, and the average total amount of time spent on all five maps was 44.22 min (47.30 min for female pairs and 41.13 min for male pairs, with overlapping ranges). Overall, performance on the map task was good with a moderate amount of conversation, indicating that talkers were communicating effectively throughout the map task session.

The conversations also yielded enough repetitions across partners to compose the AXB test of phonetic convergence. From the total set of 24 pairs of items used in the AXB test, nine of the repetitions occurred less than 4 s after the sample, ten of the repetitions occurred between 4 and 11 s after the sample, and the remaining five repetitions occurred between 19 and 83 s after the sample. The length of time between repetitions was not confounded with any experimental factor. Although Goldinger (1998) found that shadowed tokens produced after a 3-s delay did not sound like imitations of their samples, there were not enough items in this corpus to satisfy both the current constraints and such a short interitem repetition interval. Given that a finding of convergence despite longer interitem intervals would constitute a more conservative test, the timing constraint here was more relaxed than Goldinger's findings would prescribe.

B. Perceptual assessment of convergence

Responses in the AXB test were scored as the percentage of trials on which a task repetition was chosen as more similar to the task sample item than a pre- or post-task item, or a post-task item was chosen as more similar to a task sample than a pretask item. The data were submitted to a repeated measures ANOVA to test for the effects of *timing* (early vs late), *persistence/comparison type* (pretask versus task, post-task versus task, and pretask vs post-task comparison conditions), *talker role* (giver X repeated by receiver vs receiver X repeated by giver), and *pair sex* (females vs males).

TABLE I. Interaction between comparison type and role.

	R to GX	G to RX
Pretask vs task[a]	62	68
Post-task vs task	56	57
Pretask vs post-task[a]	58	65

[a]95% confidence intervals verified that means comparisons differed across these two rows and all measures were different from chance.

In the AXB similarity test, listeners detected increased similarity in pronunciation between talkers during conversational interaction (percent task repetition chosen vs pretask items, 65%, vs post-task items, 57%, and pretask vs post-task, 62%. Unless otherwise noted, for all percentages reported in the text or in the table, 95% confidence intervals confirmed that performance was significantly above chance, which was 50%). This similarity was detected for items produced early in the conversation (59%) and was greater for items produced later in the conversation [63%; main effect of *timing*, $F(1,29)=9.88$, $p<0.004$]. In addition, the similarity persisted beyond the conversation, as indicated by the finding that post-task items were judged more similar to task sample items than the pretask items (62%). Persistence of convergence was also reflected in the difference between the pretask vs task and post-task vs task comparison conditions [65% vs 57%; main effect of *persistence/comparison type*, $F(2,58)=27.24$, $p<0.001$]. Presumably the persistence of phonetic convergence in the post-task session was strong enough to allow a listener to resolve convergence in the task session items.

In contrast, the effects of *talker role* and *pair sex* went in the opposite directions than predicted: task repetitions produced by receivers were less similar to givers' task sample items than givers' task repetitions were to receivers' task sample items (GX 59% < RX 63%); and female pairs' task repetitions were less similar to task sample items than male pairs' task repetitions [females 55% < males 68%; main effect of *talker role*, $F(1,29)=12.92$, $p<0.001$, main effect of *pair sex*, $F(1,29)=63.55$, $p<0.001$].

Table I shows the interaction between *persistence/comparison type* and *talker role*. The effect of *talker role* was significant for those trials in which the task repetitions were compared to the pretask productions, but not for trials in which the task repetitions were compared to the post-task productions [interaction between *persistence/comparison type* and *talker role*, $F(2,58)=5.48$, $p<0.007$; 95% confidence intervals were used to establish the differences for the means comparisons in the first and third rows of the table]. Therefore, givers converged to receivers more than receivers converged to givers, but only when considered against the items produced before the interaction. After the interaction, the residual convergence during the post-task session neutralized the talker role asymmetry displayed during the task session, arguably because the post-task items reflected a similar degree of convergence that was evoked during the conversational setting. The data in the bottom row show that the role-governed asymmetry in similarity was detected in the trials comparing a talker's pretask and post-task items to their partner's task items. In this condition, both comparison

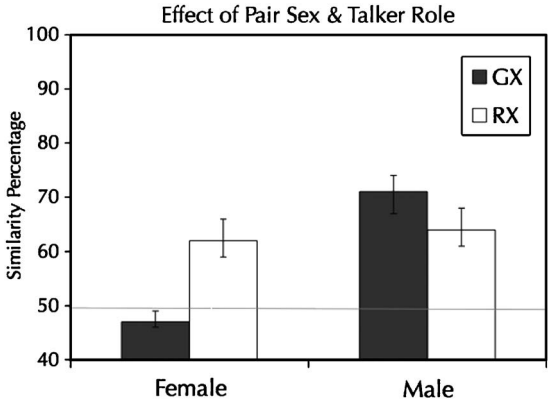


FIG. 2. AXB test interaction between *pair sex* and *talker role* with 95% confidence intervals. The dark GX bars correspond to the convergence of receivers' items to givers' sample items, and vice versa for the RX bars.

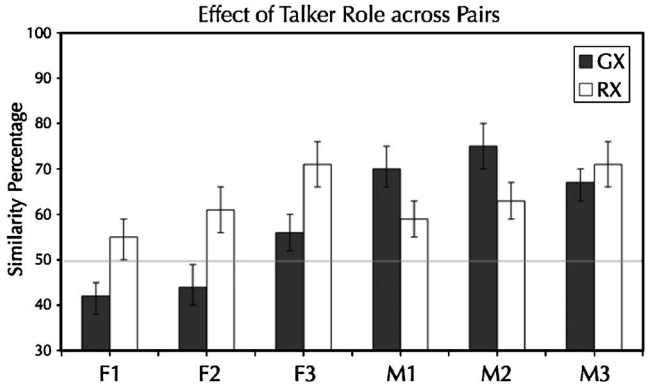


FIG. 3. AXB test interaction between *pair* and *talker role* with 95% confidence intervals. The dark GX bars correspond to the convergence of receivers' items to givers' sample items, and vice versa for the RX bars.

items from the talker were prompted by a list, yet the post-task items were produced immediately after the conversational setting, and exhibited persistent phonetic convergence in the same pattern of role-governed asymmetry that was found to be present in the task repetitions. An appeal to a distinction between spontaneous and read speech does not apply to this condition, yet parallel results obtained. This assertion was confirmed by a separate analysis of variance on the pretask vs post-task comparison trials' data: The main effects of *talker role* and *pair sex* were significant, as well as the interaction between *talker role* and *pair sex*, all showing similar patterns to the full dataset [*talker Role*: GX 58% $<$RX 65%, $F(1,29)=17.21$, $p<0.001$; *pair sex*: females 57% $<$males 65%, $F(1,29)=18.93$, $p<0.001$; *interaction*: females GX 50% and RX 65%, males GX 66% and RX 65% $F(1,29)=18.20$, $p<0.001$].

Figure 2 shows the interaction between *pair sex* and *talker role*, collapsing across all comparison conditions. Female talkers exhibited the overall pattern, greater similarity of givers to receivers, while male talkers exhibited the opposite pattern, greater similarity of receivers to givers. Comparing across the whole data set, the similarity of givers to receivers was comparable, but the receivers were not similar to givers for the female pairs, while the receivers were more similar to the givers for the male pairs [$F(1,29)=118.48$, $p<0.001$; error bars depict 95% confidence intervals]. It appears that the male talkers followed giver-dominated convergence, and the female talkers exhibited receiver-dominated convergence. Figure 3 shows the effect of *talker role* across individual pairs of talkers. The group behavior appears to be more consistent across the female than the male pairs of talkers [$F(2,58)=5.80$, $p<0.005$; error bars depict 95% confidence intervals].

Overall, the AXB similarity data indicate that phonetic convergence occurred during the map task conversation, carried through to speech produced immediately after the conversation, and was greater when a receiver provided the sample utterance that a giver repeated.

IV. DISCUSSION

This study found robust phonetic convergence between conversational participants. Despite the fact that partners were unacquainted prior to interaction, there was evidence for early convergence that increased over the course of conversation and persisted beyond the conclusion of the interaction. In the AXB listening tests, there were effects of the sex of the pair and the role of the talker in the conversation. Overall, male talkers converged more than females, and givers converged more than receivers. In female pairs, givers exhibited convergence to receivers, but receivers did not converge to givers. In male pairs, the opposite pattern was found—male receivers converged to male givers more than the reverse. Therefore, this study establishes the existence of a relatively rapid process of phonetic convergence between interacting talkers influenced by a talker's role and sex, and persisting beyond the conversation that induces it.

The effects of talker sex and role did not follow the predictions suggested by previous research on accommodation, in which female talkers converged more than male talkers and less dominant talkers converged more than more dominant talkers. In a study of accommodation that used a lexical shadowing task, Namy *et al.* (2002) found that female shadowers converged to their models more than male shadowers, and that female listeners detected convergence more readily than male listeners. Citing earlier research finding that women were more accurate in identifying talkers (Nygaard and Queen, 2000), Namy *et al.* suggested that female listeners detected convergence more readily due to greater perceptual sensitivity or attention to indexical features of talkers. If female listeners are generally more sensitive to indexical features of talkers, then the female talkers in the current study should have resolved their partner's phonetic forms in more detail than the males, leading to greater convergence in response to their partner's speech. Because the current study found greater convergence on the part of male talkers, it is more likely that attention rather than absolute perceptual sensitivity is moderating these effects. Men and women might sustain habitual attentional sets that vary across different circumstances, leading to differences in the grain of perceptual analysis and subsequent shadowing or conversational convergence.

Previous research on the influence of status or dominance on accommodation phenomena found that the implied dominance relationship among members of the pairs did not

always determine the direction of convergence (Bilous and Krauss, 1988; Bourhis and Giles, 1977; Giles, 1973; Gregory and Webster, 1996). In the current study, female receivers did not converge to female givers, but male receivers converged more than male givers, indicating that the interpretation of dominance is not a straightforward function of nominal role. The interaction with talker sex hints that men and women may interpret role labels differently, leading to distinct patterns of convergence. If such interpretations are not mandated by a talker's sex *per se*, it should be possible to influence the direction of the effect in different settings or instructional conditions. These findings do not support an interpretation based solely on differences in perceptual sensitivity; rather, functions outside the domain of perception appear to be influencing the degree of phonetic convergence.

Although perceptual sensitivity to indexical or phonetic features is necessary for accommodation, it is not sufficient to evoke phonetic convergence. Attention may adjust the grain of perceptual resolution, but additional processes influence a talker's phonetic form. The current study cannot provide a clear explanation for the direction of the observed effects of talker sex and role, but the reliability of these effects challenges an account of convergence based solely on a direct link between speech perception and speech production. If automatic priming is the mechanism by which phonetic convergence operates, then the sex-and role-governed patterns found here must be due to processes within speech perception and/or production that are susceptible to extralinguistic factors. There is currently no account of priming that incorporates the sex or role of a talker as a modulator of the degree of priming. Priming is mainly influenced by factors internal to lexical representation, such as semantic relatedness, word frequency, and lexical neighborhood density (see Luce and Pisoni, 1998). In the current experiment, the differences in phonetic convergence do not hinge on such factors because the same lexical items were counterbalanced across all talker role and sex conditions.

If speech perception and production are not labile to superordinate influences—if they are modular (Fodor, 1983; Liberman and Mattingly, 1985)—then these findings suggest three alternative possibilities: (1) a loosening of the tie between perception and production; (2) a process intervening prior to speech perception; or (3) a process intervening between speech production planning and execution. It could be the case that a talker's role or sex induces a more or less focused attentional set, which in turn specifies an appropriately detailed phonetic representation that drives production. A report by Goldinger and Azuma (2003) provides evidence that talkers and listeners can be biased to produce and to respond to relatively finer- or coarser-grained aspects of speech signals, depending on a simple instructional manipulation. Therefore, processes outside speech perception and production enter the system at some point, but a detailed elaboration of the mechanisms and acoustic-phonetic attributes awaits further investigation. The patterned variability in phonetic convergence demonstrates that convergence is not a result of automatic priming—the social setting of language use modulates the degree to which a talker's phonetic repertoire converges on that of a conversational partner. How does a talker incorporate aspects of an interacting partner's phonetic repertoire?

A. Episodic memory

Goldinger (1998) provided compelling evidence that lexical imitation is a natural consequence of an episodic memory system. Indeed, some of the factors that influenced the degree of lexical imitation found in his shadowing experiments coincided with predictions derived from an episodic memory model, namely Hintzman's (1986) MINERVA2. This model begins with an assumption that every perceptual episode leaves a trace in long-term memory, and warrants that every time a new episode is encountered, all traces that are similar to the original episode are activated and compared directly to the sample token. Based on the outcome of comparison, the episodic system generates a composite representation, an echo, of the activated traces and the sample token. Goldinger proposed that this composite serves as the model for the shadowed productions. Because the multitude of traces integrated into an echo has different effects depending on the number of exemplars that are available, word-frequency effects on imitation were predicted and found both in perceptual judgments and MINERVA2-based modeling. Common words have more traces in memory that attenuate the prominence of distinctive aspects of a new token in the echo. Echoes from rarer words, in contrast, will integrate a recent occurrence with fewer competing traces, leading to greater imitative fidelity.

Crucially, for most items repeated after a 3-s delay in Goldinger's study, imitation was not detected. The absence of delayed imitation was attributed to dilution of distinctive aspects of a sample token while holding the item in working memory: Analogous to effects found for common words, as an item is rehearsed in working memory, its distinctive attributes are lost through reiterative comparison with traces in long-term memory. Therefore, the relative influence of long-term memory on echo-based imitation increases with both greater numbers of similar traces and greater amounts of time between a sample and its repetition, leading to a decrease in imitative fidelity.

The current study found evidence for a form of delayed imitation: Phonetic convergence in conversational settings persisted into a post-task session. An episodic memory system containing detailed lexical episodes could not form the basis for this effect because the delay would increase the influence of long-term memory traces on repeated words, leading to reduced phonetic convergence. Moreover, the observed increase in phonetic convergence over the course of conversational interaction occurred across different lexical items, providing evidence that the change in phonetic repertoire was not tied to specific lexical forms. At a much broader time scale, Sancier and Fowler (1997) found that a bilingual talker shifted pronunciation of consonants in both languages as a result of recent language experience—even consonant VOTs in the unused language were affected. This finding is unlikely to be based on lexical episodes, as the episodes in question did not match the vocabulary of the

affected language. Finally, any account relying on episodic memory must be mute with respect to the effects of talker sex and role—both talkers in a given pair heard the same lexical items over the course of the conversational setting, yet one talker produced more convergent forms. Like an automatic priming mechanism, an episodic memory system fails to explain the effect of a transient social factor on phonetic convergence. From this evidence, it appears that the functional circumstances of language use induce a kind of phonetic convergence that is not found in individual echoes of mere exposures. A more appropriate conceptualization of these findings might be found in the literature on entrainment, which incorporates relative dominance in dynamical systems without appeal to automatic priming or episodic memory traces.

B. Entrainment

The principles of entrainment were initially identified in von Holst's (1937/1973) early research on endogeneous rhythmicity, in which the complex motions of oscillating fish fins were readily described in terms of the superimposition of sinusoidal functions. Later, von Holst's principles were found to scale up to the dynamics of more glamorous organisms, like pairs of humans swinging legs or wrist pendulums (see Turvey, 1990). Accordingly, the *magnet effect* is the tendency for a more dominant or stable oscillator to pull a less dominant oscillator into synchrony. Absolute coordination or entrainment is a rare phenomenon in which both the phase relationship and the frequency of oscillation match, and only occurs with rigid coupling of systems that have identical intrinsic dynamics (Schmidt and Turvey, 1989). Lacking rigid coupling, interpersonal entrainment typically exhibits only relative coordination. Relative coordination demonstrates another principle of entrainment, the *maintenance tendency*. Despite the pull to entrain to a coupled oscillator, the manifest pattern exhibits a latent influence of the original intrinsic dynamics, presumably because the external oscillator's pattern is superimposed onto the internal oscillator's pattern rather than supplanting it.

These properties of entrainment in coordinated dynamical systems provide a ready model of the integration of internal and external forces in human behavior. Beek, Turvey, and Schmidt (1992) framed the relation between internal organization of coordinated activities and information from external sources in terms of dynamical systems theory, proposing that external information is an embedded forcing function on internal dynamics. For example, externally derived information might influence speech production by first indirectly participating in a separate perceptual-memory system or by directly specifying phonetic forms. Beek *et al.* propose that external information acts as an embedded forcing function on internal dynamics, inducing changes in the overall pattern of activity that push the activity to different values in its intrinsic range. Research on self-regulation of speech production, in particular the Lombard sign (Lane and Tranel, 1971) and perceptual-productive adaptation of vowel formants, speaking fundamental frequency, and consonant spectra (Houde and Jordan, 2002; Jones and Munhall, 2000, 2003) shows that talkers can incorporate auditory feedback of their own productions to adjust subtle aspects of speech at short latencies. If perception of another talker's speech yields detailed phonetic forms, such forms could influence subsequent production under circumstances, such as the demands of conversational interaction, that promote coupling between talkers.

For the kinds of phenomena examined in the current study, the direction and form of phonetic convergence is difficult to predict. Viewed one way, the giver is dominant by providing the information to be copied, and the task requires the receiver to comply; on the other hand, the receiver must ensure that the giver provides adequate instructions, therefore, the receiver might set the tone for interaction. Either organization is feasible, and potentially idiosyncratic to different pairs. In future investigations, it will be useful to manipulate relative dominance by explicitly instructing one talker to imitate the other talker. Perhaps the instruction to imitate will override the nominal effects of talker role in setting up the dominance hierarchy. More significant, however, dominance is irrelevant for entrainment if the systems are not coupled. With looser coupling, there is likely to be less convergence, as is generally the case with informationally coupled systems, such as interacting talkers (see Schmidt and Turvey, 1989). With respect to the current findings, an account that acknowledges varied degrees of coupling and dominance in between-talker interaction is a better fit to the findings than an autonomous account of automatic priming from speech perception to speech production.

As an instance of relative entrainment, phonetic convergence may be analogous to other forms of alignment phenomena between talkers. Pickering and Garrod (2004) claim that discrepancies from alignment are attributed to indirect secondary processes that monitor comprehension and adjust output when comprehension fails. However, sometimes a talker will diverge from an interlocutor without a failure of comprehension (Bilous and Krauss, 1988; Bourhis and Giles, 1977). Although speech perception resolves the detailed aspects of phonetic form that would be necessary for phonetic convergence with any degree of fidelity, a talker is not automatically driven to imitate those forms. What sorts of nonlinguistic functions might phonetic convergence or divergence serve for a talker?

C. Convergence and social function

Currently, accounts of language use in social interaction emphasize the social situation in which speech occurs, as opposed to individual factors in speech production. Thus, language is not produced by isolated individuals in order "to generate grammatical strings," (Krauss, 1987, p. 97); rather, in addition to the production of linguistic forms, speech projects social categories (Giles, Scherer, and Taylor, 1979; Shepard, Giles, and Le Poire, 2001), is used to accomplish mutual goals (Clark, 1996; Clark and Wilkes-Gibbs, 1986), and/or to align representations (Garrod and Doherty, 1994; Pickering and Garrod, 2004). Communication is more than a matching process; a talker expresses more than a sequence of phonemes, and a listener uses the speech signal to under-

stand the talker as well as the message. If some part of phonetic variability is communicative, i.e., it serves some purpose for the talker, then a listener who resolves the phonetic detail can project this into a perception of the talker's connotation, apart from lexical access. Among interacting talkers, phonetic convergence might contribute to mutual comprehension and/or rapport through a decrease in social distance (Shepard, Giles, and LePoire, 2001).

The current study establishes the existence of a process of rapid phonetic convergence that emerges in conversational settings, providing a link between the laboratory studies of nonsocial shadowing imitation and community-level linguistic change. However, the constraints on convergence in conversational settings appear to differ somewhat from those of shadowing studies. Labov (1974, 1986) suggested that linguistic change is motivated by the need to add emphasis to expression, and that new forms are adopted as a result of interactive conversation. At the same time that talkers share a common phonetic ground, each talker maintains some distinction through novelty, perhaps as Labov (1974) suggests, "to signal a stronger meaning than the older form; to display the speaker's membership in a local group; and to demonstrate greater intimacy than an older form." (p. 253) The current study also provides some evidence that interactive changes persist beyond a particular social situation, perhaps to be carried to the next interaction.

Garrod and Doherty (1994) provided insight into the development of conventions in paired and community-level conversational interactions. They proposed that greater overall stability might arise across a community as opposed to individual pairs because of the greater initial variability of viewpoints afforded by communities. This variability induces competition among different concepts, leading to a greater likelihood that a more stable form will survive across a community of talkers. In the context of the current findings, it may be possible to trace phonetic convergence among a community of talkers to determine more precisely the kinds of changes that are durable. If the observed effects of talker role serve a pragmatic purpose, they ought to endure community interaction.

Although these findings extend convergence in conversational interaction to the phonetic domain, interacting talkers do not match on all acoustic-phonetic dimensions. If perception yields goals that drive production, then an adequate account of this relation requires an explanation of the lack of perfect correspondence (Pardo and Remez, in press). Because some of the disparities between talkers in this study patterned according to talker role and sex, it is likely that the discrepancy is due to more than noisy perceptual resolution or productive output of gestural goals (see also Vallabha and Tuller, 2004). Individual talkers in social settings have communicative goals that go beyond mere intelligibility (but see Lindblom, 1990), and these goals must be an integral part of the perceptual and productive system that creates the spoken message (see also Bradlow, 2002). Perhaps such goals act as weights on embedded forcing functions that incorporate perceived phonetic structures into produced phonetic forms.

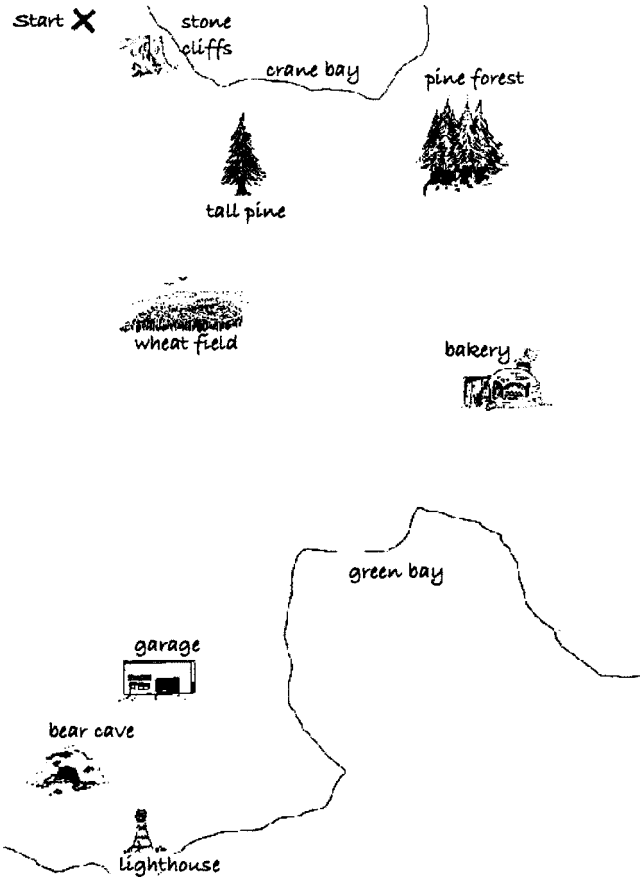
This study attempts to broaden an understanding of speech perception and production to include social function in language use. The link between perception and production in spoken communication is not automatic; it is subject to situational constraints that influence the direction and magnitude of phonetic convergence in conversational interaction. Future investigations might question whether these effects are durable enough to extend across community interactions, perhaps because of a broader cooperative function in social discourse. Many theorists propose some form of cooperative principle in social interaction (e.g., Clark 1996), yet few have examined the operation of the principle at the level of phonetic variability, and fewer still have attempted a rigorous exposition of the likely structural factors that evoke or attenuate the cooperative principle (e.g., Giles, Coupland, and Coupland, 1991). Although phonetic convergence varies across talkers in social interaction, an individual is not ruled completely by circumstance, but each implements convergence as the situation warrants.

ACKNOWLEDGMENTS

This research was supported in part by an NRSA training grant from the National Institutes of Mental Health to Jennifer Pardo and Robert Krauss at Columbia University. Portions of this project were conducted in partial fulfillment of the requirements for the degree of Doctor of Philosophy to J. S. Pardo at Yale University. The author thanks Robert Crowder, Carol Fowler, Wendel Garner, Raquel Gardner, Robert Krauss, Robert Remez, George Ton, Rebecca Treiman, and many anonymous reviewers for their help in conceptualizing and completing this project.

APPENDIX A: SAMPLE MAP TASK MAP FOR A GIVER

APPENDIX B: SAMPLE MAP TASK MAP FOR A RECEIVER.



Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., and Weinert, R. (**1991**). "The HCRC Map Task corpus," Lang Speech **34**, 351–366.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (**2000**). "Controlling the intelligibility of referring expressions in dialogue," J. Mem. Lang. **42**, 1–22.

Beek, P. J., Turvey, M. T., and Schmidt, R. C. (**1992**). "Autonomous and nonautonomous dynamics of coordinated rhythmic movements," Ecological Psychol. 4, 65–95.

Bilous, F. R., and Krauss, R. M. (**1988**). "Dominance and accommodation in the conversational behaviors of same- or mixed-gender dyads," Lang. and Commun. **8**, 183–194.

Blaauw, E. (**1994**). "The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech," Speech Commun. **14**, 359–375.

Bourhis, R. Y., and Giles, H. (**1977**). "The language of intergroup distinctiveness," in *Language, Ethnicity and Intergroup Relations*, edited by H. Giles (Academic, London), pp. 119–135.

Bradlow, A. R. (**2002**). "Confluent talker- and listener-oriented forces in clear speech production," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, New York), pp. 241–273.

Branigan, H. P., Pickering, M. J., and Cleland, A. A. (**2000**). "Syntactic co-ordination in dialogue," Cognition **75**, B13–25.

Browman, C. P., and Goldstein, L. (**1991**). "Gestural structures: Distinctiveness, phonological processes, and historical change," in *Modularity and the Motor Theory of Speech Perception*, edited by I. G. Mattingly and M. Studdert-Kennedy (Erlbaum, Hillsdale, NJ), pp. 313–338.

Catchpole, C., and Pardo, J. S. (**2004**). "Articulatory shortening in repeated noun phrases is affected by participant role," Architectural Mechanisms for Language Processing, Universite de Provence, 16–18, September, 2004.

Chartrand, T. L., and Bargh, J. A. (**1999**). "The chameleon effect: The perception-behavior link and social interaction," J. Pers Soc. Psychol. **76**, 893–910.

Clark, H. H. (**1996**). *Using Language* (Cambridge University Press, Cambridge).

Clark, H. H., and Wilkes-Gibbs, D. (**1986**). "Referring as a collaborative process," Cognition **22**, 1–39.

Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (**1993**). "PSYSCOPE: An interactive graphical system for designing and controlling experiments in the psychology laboratory using Macintosh computers," Behav. Res. Methods Instrum. Comput. **25**, 257–271.

Fodor, J. A. (**1983**). *The Modularity of Mind* (MIT Press).

Fowler, C. A. (**1988**). "Differential shortening of repeated content words produced in various communicative contexts," Lang Speech **28**, 47–56.

Fowler, C. A., and Housum, J. (**1987**). "Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction," J. Mem. Lang. **26**, 489–504.

Fowler, C. A., Levy, E., and Brown, J. (**1997**). "Reductions of spoken words in certain discourse contexts," J. Mem. Lang. **37**, 24–40.

Fowler, C. A., Brown, J., Sabadini, L., and Weihing, J. (**2003**). "Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks," J. Mem. Lang. **49**, 396–413.

Garrod, S., and Doherty, G. (**1994**). "Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions," Cognition **53**, 181–215.

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

Giles, H., Coupland, J., and Coupland, N. (**1991**). *Contexts of Accommodation: Developments in Applied Sociolinguistics* (Cambridge University Press, Cambridge).

Giles, H., Scherer, K. R., and Taylor, D. M. (**1979**). "Speech markers in social interaction," in *Social Markers in Speech*, edited by K. R. Scherer and H. Giles (Cambridge University Press, Cambridge, England), pp. 343–375.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Goldinger, S. D., and Azuma, T. (**2003**). "Puzzle-solving science: The quixotic quests for units in speech perception," J. Phonetics **31**, 305–320.

Gregory, D., and Webster, S. (**1996**). "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status predictions," J. Pers Soc. Psychol. **70**, 1231–1240.

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

Hintzman, D. L. (**1986**). "'Schema abstraction' in a multiple trace memory model," Psychol. Rev. **93**, 411–428.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech. I. Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of *F*0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Jones, J. A., and Munhall, K. G. (**2003**). "Learning to produce speech with an altered vocal tract: The role of auditory feedback," J. Acoust. Soc. Am. **113**, 532–543.

Krauss, R. M. (1987). "The role of the listener: Addressee influences on message formulation," J. Lang. Soc. Psychol. **6**, 81–98.

Krauss, R. M., and Pardo, J. S. (**2004**). "Is alignment always the result of priming?," Behav. Brain Sci. **27**, 203–204.

Krauss, R. M., and Weinheimer, S. (**1964**). "Changes in the length of reference phrases as a function of social interaction: A preliminary study," Psychonomic Sci. **1**, 113–114.

Labov, W. (**1974**). "Linguistic change as a form of communication," in *Human Communication: Theoretical Explorations*, edited by A. Silverstein (Erlbaum, Hillsdale, NJ), pp. 221–256.

Labov, W. (**1986**). "Sources of inherent variation in the speech process," in *Invariance and Variability in the Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 402–425.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht, Netherlands), pp. 403–439.

Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The

neighborhood activation model," Ear Hear. **19**, 1–36.
Namy, L. L., Nygaard, L. C., and Sauerteig, D. (**2002**). "Gender differences in vocal accommodation: The role of perception," J. Lang. Soc. Psychol. **21**, 422–432.
Natale, M. (**1975**). "Convergence of mean vocal intensity in dyadic communication as a function of social desirability," J. Pers Soc. Psychol. **32**, 790–804.
Nygaard L. C., and Queen J. S. (**2000**). The role of sentential prosody in learning voices. Paper presented at the meeting of the Acoustical Society of America, Atlanta, GA.
Pardo, J. S., and Remez, R. E., "The perception of speech," in *The Handbook of Psycholinguistics*, 2nd ed., edited by M. Traxler and M. Gernsbacher (Elsevier, Cambridge, MA, in press).
Perkell, J. S., Zandipour, M., Matthies, M. L., and Lane, H. (**2002**). "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues," J. Acoust. Soc. Am. **112**, 1627–1641.
Pickering, M. J., and Garrod, S. (**2004**). "Toward a mechanistic psychology of dialogue," Behav. Brain Sci. **27**, 169–190.
Porter, R., and Castellanos, F. (**1980**). "Speech production measures of speech perception: Rapid shadowing of VCV syllables," J. Acoust. Soc. Am. **67**, 1349–1356.
Sancier, M. L., and Fowler, C. A. (**1997**). "Gestural drift in a bilingual speaker of Brazilian Portuguese and English," J. Phonetics **25**, 421–436.
Schmidt, R. C., and Turvey, M. T. (**1989**). "Absolute coordination: An ecological perspective," in *Perspectives on the Coordination of Movement*, edited by S. A. Wallace (Elsevier Science, New York), pp. 123–156.
Schober, M. F., and Clark, H. H. (**1989**). "Understanding by addressees and overhearers," Cogn. Psychol. **21**, 211–232.
Shepard, C. A., Giles, H., and Le Poire, B. A. (**2001**). "Communication accommodation theory," in *The New Handbook of Language and Social Psychology*, edited by W. P. Robinson and H. Giles (Wiley, New York), pp. 33–56.
Turvey M. T. (**1990**). Coordination. American Psychologist **45**, 938–953
Vallabha, G. K., and Tuller, B. (**2004**). "Perceptuomotor bias in the imitation of steady-state vowels," J. Acoust. Soc. Am. **116**, 1184–1197.
Viechnicki, P. D. (**2002**). "Composition and granularity of vowel production targets," Diss. Abstr. Int., C **63**(4-A), 1320, UMI.
von Holst, E. (**1937/1973**). "On the nature of order in the central nervous system," in *The Behavioral Physiology of Animal and Man: The Collected Papers of Erich von Holst* (University of Miami Press, Miami, FL).