Journal of Phonetics 69 (2018) 1–11

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

A comparison of phonetic convergence in conversational interaction and speech shadowing

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

Department of Psychology, Montclair State University, USA

ARTICLE INFO

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

Keywords:
Conversation
Phonetic convergence
Speech production

ABSTRACT

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

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

Conversational phonetic convergence

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

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

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

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

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

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

Speech shadowing phonetic convergence

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

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

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