ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 3aSCa: Imitation, Accommodation, and Convergence in Speech Communication

3aSCa5. Reconciling diverse findings in studies of phonetic convergence
Jennifer Pardo*

*Corresponding author's address: Psychology Department, Montclair State University, 1 Normal Avenue, Montclair, New Jersey 10590, pardo@optonline.net

Phonetic convergence occurs both when individuals interact in conversation, and when listeners rapidly repeat words presented over headphones. Results from multiple studies examining phonetic convergence offer an array of often confusing and disparate findings. Reconciling such diverse findings is difficult without a clear rationale for engaging in one acoustic measure over another. The current paper proposes a paradigm that models perceptual and acoustic measures together. Measures of multiple acoustic-phonetic attributes were compared with a perceptual measure of phonetic convergence in a shadowing study. Although convergence was not significant in any acoustic measure alone, the combination of acoustic attributes predicted perceived phonetic convergence on an item-by-item basis. Because perceptual measures integrate across multiple acoustic-phonetic dimensions, future studies of phonetic convergence should use perceptual tasks to calibrate the relative contribution of individual acoustic-phonetic parameters.

Published by the Acoustical Society of America through the American Institute of Physics
RECONCILING DIVERSE FINDINGS IN STUDIES OF PHONETIC CONVERGENCE

Phonetic convergence occurs both when individuals interact in conversation and when listeners rapidly repeat words presented over headphones (e.g., Goldinger, 1998; Pardo, 2006). Moreover, the phonetic repertoire of individuals will change to become more similar to that of a new linguistic environment (Evans & Iverson, 2007; Sancier & Fowler, 1997), or to reflect changes in predominant speaking patterns (Labov, 1974, 1986). Results from multiple studies examining phonetic convergence offer an array of often confusing and disparate findings. Reconciling such diverse findings is difficult without a clear rationale for engaging in one acoustic measure over another. The current paper proposes a more standard paradigm for investigating phonetic convergence.

In Goldinger’s (1998) initial study of imitation during speech shadowing, the degree of imitative fidelity was influenced by word frequency, such that imitation of low frequency words was greater than high frequency words. Furthermore, imitation was greater in words that had been heard more times prior to shadowing. These effects even persisted after one week between presentation and shadowing (Goldinger & Azuma, 2004). Consequently, many subsequent studies of speech convergence have constructed word lists comprising only low frequency items (Babel, 2012; Miller, Sanchez, & Rosenblum, 2010; Namy, Nygaard, & Sauerteig, 2002; Shockley, Sabadini, & Fowler, 2004; but see Nielsen, 2011). Although this procedure might be more likely to evoke phonetic convergence in shadowing studies, its use impedes the generalizability of the findings to more natural settings of language use.

In terms of the underlying mechanisms that might support phonetic convergence, Goldinger (1998) used a modification of Hintzman’s MINERVA2 to demonstrate that imitation might result automatically from structural and functional characteristics of episodic memory systems. In such systems, frequency and repetition effects are an expected outcome, and the data patterns from shadowing imitation closely followed the predicted impact of frequency and repetition. However, it is clear to anyone who listens to speech that talkers maintain a variety of speaking styles despite a putative automatic tendency to imitate, whether that tendency is attributed to episodic memory systems, automatic priming mechanisms (Pickering & Garrod, 2004), or to mirror neurons (Rizzolatti & Craighero, 2004).

A series of studies by Pardo and colleagues have examined phonetic convergence in conversational interaction (Pardo 2006, 2010; Pardo, Cajori Jay, & Krauss, 2010; Pardo, Cajori Jay, Hoshino, Hasbun, Sowemimo-Coker, & Krauss, submitted). In these studies, unacquainted talkers complete a conversational task that is designed to induce natural inter-talker repetitions of the same lexical items. These items are then excised from the conversations to create an AXB perceptual similarity test that is presented to a separate set of listeners (adapted from Goldinger, 1998). In the AXB test, the X-items are those produced by one talker in the conversation, and the A & B items comprise the partner’s conversational repetition and a pre-task sample of the same items. The listeners simply select which item, A or B, sounds more similar in pronunciation to the middle item (X). The responses are scored as proportion or percentage of conversational repetitions selected as more similar to the partner’s model utterances. Perceptual assessment of phonetic convergence provides a measure that reflects global similarity across multidimensional aspects of acoustic-phonetic attributes simultaneously.

Over the course of these studies, phonetic convergence has been found to be highly variable. The range of scores have been as low as 33% up to a high of 78% perceived convergence (using AXB task) in conversational interaction. Some of this variability is due to the impact of conversational role and talker sex on phonetic convergence. The conversational task employed in these studies involved an instruction Giver who directed a Receiver to follow a path around various landmarks on a map. In Pardo (2006), Givers converged to Receivers more so than the reverse, and pairs of male talkers converged more than pairs of female talkers. In a subsequent study, one member of each pair was instructed to try to imitate their partner, and phonetic convergence was impacted by the role of the talker who was given the instruction. In this case, instructing Receivers to imitate led to greater convergence (Pardo et al., 2010).

It is noteworthy that measures of individual acoustic attributes did not align with the perceptual measures of phonetic convergence. Neither item duration, speaking rate, nor vowel spectra in these studies produced consistent patterns of convergence, nor were the patterns consistent with the perceptual data (see also Pardo, Cajori Jay, et al., submitted). Finally, a long-term study of college roommates likewise failed to find consistency in convergence when comparing acoustic measures of vowel spectra with perceptual assessment of phonetic convergence over the course of an academic year (Pardo, Gibbons, Suppes, & Krauss, 2012).

Across the research literature, most studies examining phonetic convergence use either a repetition or exposure task without conversational interaction, focus on acoustic measures of convergence, and report data for only a single
Acoustic-phonetic attribute. Although multiple acoustic attributes are represented across these studies, this practice does not provide a comprehensive assessment of convergence. There are many reasons to prefer perceptual similarity tests over acoustic measures. First, talkers likely converge on multiple attributes simultaneously. Second, talkers might converge on some attributes at the same time that they diverge on other attributes (Bilous & Krauss, 1988; Pardo, 2010; Pardo et al., 2010; Pardo et al., 2012; Pardo, Cajori Jay, et al., submitted). Third, talkers might converge on one set of attributes for one set of items or talkers and another set of attributes for other items or talkers (Bilous & Krauss, 1988). This last point is particularly relevant when word frequency is at play—for example, talkers might converge in duration of low frequency items but fundamental frequency of high frequency items. If only one of these attributes is measured, then convergence in the other attribute would be missed. If both are measured, then the results would appear discrepant or difficult to interpret.

In order to derive a valid, comprehensive assessment of acoustic-phonetic convergence, future studies should employ perceptual assessment of phonetic convergence in concert with measures of multiple acoustic dimensions. On the one hand, perceptual tasks provide a global measure of convergence that is grounded by what might be accessible to individuals during conversational interaction. On the other hand, acoustic-phonetic measures provide information about the detailed attributes employed by talkers and potentially detected by listeners. These attributes can then be tested in regression models as predictors of the perceptual data. Regression models not only provide information about which acoustic-phonetic attributes are relevant to listeners, but also the relative contribution of each attribute.

A recent study demonstrates the utility of this approach (Pardo, Jordan, et al., submitted). In this study, perceptual measures of shadowed speech convergence were compared to measures of convergence in duration, F0, and vowel spectra. None of the acoustic attributes yielded significant convergence, but convergence was detected in the perceptual similarity task. However, a regression model that included all three acoustic measures of convergence provided a good fit to the perceptual data, indicating that the perceptual data reflected acoustic-phonetic variability of multiple attributes simultaneously. Moreover, convergence in duration was the strongest predictor of the perceptual data, followed by F0 and vowel spectra.

The resurgence of interest in speech accommodation has resulted in a diverse set of findings across multiple measures. However, there is no clear rationale in favor of any particular acoustic-phonetic attribute, and there is little consistency across these measures. The current paper proposes a more comprehensive approach to the study of phonetic convergence—future studies should include both perceptual and acoustic measures of phonetic convergence. In doing so, such studies can provide well-calibrated estimates of the full set of acoustic-phonetic parameters at play in phonetic convergence during language use.

ACKNOWLEDGMENTS

Completion of this work was funded in part by NSF PAC award #BCS-1229033. The author thanks Molly Babel and Kuniko Nielsen for their invitation to contribute to this special session.

REFERENCES


