EFFECTS OF DIALECT PRIMING ON PHONETIC CONVERGENCE

Cynthia G. Clopper, Ellen Dossey, Roberto Gonzalez

Ohio State University
clopper.1@osu.edu, dossey.1@osu.edu, gonzalez.594@buckeyemail.osu.edu

ABSTRACT

Non-linguistic social information affects speech perception. The goal of the current study was to examine the effect on phonetic convergence of providing non-linguistic regional information about the model talker in a word shadowing task. Shadowers produced a set of target words in a baseline reading task and then repeated those words after a model talker from the Northern dialect of American English. Half of the shadowers were asked simply to repeat the model talker. The other half were told that the model talker is from the Northern United States and then asked to repeat her. Phonetic convergence was assessed perceptually using an AXB perceptual similarity task. The results revealed significant phonetic convergence overall, with more convergence by the shadowers who were told where the model talker is from than by those who were not. These results suggest that non-linguistic social information can prime phonetic convergence in a word shadowing task.

Keywords: phonetic convergence, word shadowing, dialect priming, Northern Cities Shift

1. INTRODUCTION

Priming social information about a talker affects how that talker's speech is perceived [1, 2, 3, 4, 5, 6, 7]. Effective primes range from movies of talking faces synchronized to the speech signal [7] to still visual images of faces presented synchronously with the speech signal [1, 3, 5, 6] to regional labels at the top of the response sheet [4] to casually-referenced stuffed animals associated with different regions [2]. The effects of these primes on speech perception include changes in perceptual category boundaries between phonemes [7], identification of best-matching synthetic vowels [2, 4], word and sentence recognition [1, 3, 5, 6], and accent ratings [1, 5].

Previous research examining social priming has further demonstrated that primes that accurately reflect the listeners' expectations about the talker's social background can facilitate speech processing, relative to primes that mismatch the listeners' expectations about the talker's social background [1, 3, 5]. For example, McGowan [3] found that speech intelligibility was higher for a Chinese-accented talker paired with a Chinese face than for the same talker paired with a white face. The goal of current study was to explore the effects of this kind of congruent social priming on phonetic convergence in a word shadowing task.

Word shadowing tasks have been shown to elicit phonetic convergence, even in the absence of explicit instructions to imitate the model talker [8, 9]. Although some aspects of the phonetic convergence observed in word shadowing tasks may therefore reflect automatic processes at the perception-production interface [8], more recent research has revealed considerable social selectivity in which features exhibit convergence. For example, the baseline phonetic distance between the model talker and the shadowers affects the magnitude of convergence, such that larger baseline differences allow for greater convergence [10]. As a result, phonetic convergence across dialects can be greater than phonetic convergence within dialects [11].

At the same time, however, social stereotypes appear to affect the magnitude of phonetic convergence to some dialects and linguistic variants. For example, more positive attitudes towards the talker [10, 12] and the talker's dialect region [13] lead to greater convergence. In addition, negative stereotypes associated with particular linguistic variants appear to block convergence to those forms [13, 14, 15]. Thus, even for a single model talker dialect, shadowers can exhibit varying degrees of phonetic convergence across variants, with greater convergence to non-stereotyped forms than to stereotyped forms [13, 14, 15]. For example, Clopper and Dossey [14] observed phonetic convergence in a word shadowing task to non-stereotyped Southern American English /u oʊ/ fronting, but not to stereotyped Southern American English /aɪ/ monophthongization.

The current study examined phonetic convergence to the Northern dialect of American English in a word shadowing task, in which social priming was manipulated across shadowers. The Northern dialect is characterized by the Northern Cities Shift, including the raising and fronting of /æ/, lowering and backing of /ɪ ɛ/, and fronting and lowering of /ɑ ɔ/ [16]. Although the Northern Cities Shift results in a phonetic vowel space that differs considerably from the ideological standard variety of American English, the Northern dialect is not socially stereotyped as a

unique dialect of American English (i.e., it is not enregistered), nor are any of the Northern Cities Shifted variants negatively stereotyped [17]. We therefore expected to observe robust phonetic convergence to the Northern Cities Shifted vowels of the model talker in the word shadowing task in the current study, consistent with previous work suggesting automatic phonetic convergence to non-salient forms [8, 13, 14, 15].

We manipulated social priming in the current study by either providing or not explicit information to the shadowers about where the model talker was from. Given the positive effects of congruent social information on speech processing [1, 3, 5], we expected that providing explicit, authentic information about the model talker's region of origin would facilitate phonetic convergence to the Northern Cities Shift.

2. METHOD

Phonetic convergence was assessed perceptually in an AXB perceptual similarity task [8, 11, 14, 15, 18, 19]. This approach to assessing phonetic convergence was selected because it provides a holistic measure that is difficult to capture using individual acoustic distance measures [15, 18, 19]. In American English, vowel variation is observed across dialects in multiple acoustic dimensions, including midpoint F1 and F2, duration, and formant trajectories [16, 20]. The AXB perceptual similarity task in the current study therefore allowed for a robust test of phonetic convergence because this multidimensional vowel variation could be assessed simultaneously and holistically by the listeners, rather than separately in a series of acoustic analyses.

2.1. Participants

Forty-nine adult listeners (30 female, 19 male) were recruited from a local science museum in Columbus, OH, to participate in the AXB perceptual similarity task. The listeners were all native speakers of American English and ranged in age from 18-67 years old (M = 31 years).

2.2. Stimulus materials

The stimulus materials were produced in a word shadowing task and comprised the model talker utterances from the word shadowing task, as well as baseline read utterances and shadowed (i.e., repeated) utterances from 10 female shadowers. The model talker was a 22-year-old female lifetime resident of the Northern dialect region of American English, selected from the Indiana Speech Project corpus [21]. The 10 female shadowers were all lifetime residents of the Midwestern United States, including both the Midland and Northern dialect regions of American English. They ranged in age from 19-39 years old (M = 28 years).

The word shadowing task involved two blocks. In the first block, the shadowers read a set of 48 multisyllabic English words aloud one at a time from the computer screen. In the second block, the shadowers repeated the same set of 48 words after the model talker. Both the read and shadowed utterances were recorded. The set of 48 words included 24 target words containing a stressed vowel implicated in the Northern Cities Shift (/ɪ ɛ æ ɑ/), as well as 24 filler words containing other stressed vowels (/i oʊ u aɪ/). The 24 target words were used in the AXB perceptual similarity task in this study and are shown in Table 1.

Vowel	Target Words
ɪ	amphibian, aristocrat, conspicuous, imposition, liberator, precipitate
ɛ	clarinet, embezzle, epilepsy, legendary, obsession, silhouette
æ	appetizer, caterpillar, deactivate, evaporate, procrastinate, spatula
ɑ	hypnotic, octopus, rhinoceros, robin, roster, slobber

Table 1: Target words in the AXB perceptual similarity task.

Five of the shadowers were asked in the shadowing block to simply repeat the words aloud after the model talker. The other five shadowers were asked to repeat the words aloud after the model talker from Chicago, IL, a major, well-known city within the Northern dialect region of American English. Thus, the latter group of shadowers ("primed") were primed with the model talker's region of origin, whereas the former group of shadowers ("unprimed") were not.

An acoustic analysis of the stimulus materials was conducted to confirm the presence of the Northern Cities Shift in the model talker utterances and to assess the baseline phonetic distance between the model talker and the shadowers. Given the mixed residential histories of the shadowers, including both the Midland and Northern dialect regions, this acoustic analysis allowed us to confirm sufficient baseline distance between the model talker and the shadowers to ensure that we could observe potential phonetic convergence in the acoustic space [10]. First and second formant frequencies were estimated at the temporal midpoint of the stressed vowel in each target word for the model talker and for each shadower in the baseline block. A summary of these vowel spaces

is shown in Fig. 1, separately for the two shadower priming conditions.

The model talker's vowel space (black symbols) is shifted down and to the left overall relative to the shadowers' vowel spaces (gray symbols), making direct acoustic comparisons between the model talker and the shadowers' baseline vowels difficult. However, the phonetic features of the Northern Cities Shift are evident in the acoustic approximation of /ɛ æ/ and in the relative position of /ɛ æ/ in the F2 dimension for the model talker, in comparison to the unshifted shadowers' baseline utterances. The shadowers' baseline utterances in the two priming conditions are similar overall, except that /ɑ/ is lower and more similar to the model talker's /ɑ/ for the shadowers in the unprimed condition (left panel) than in the primed condition (right panel).

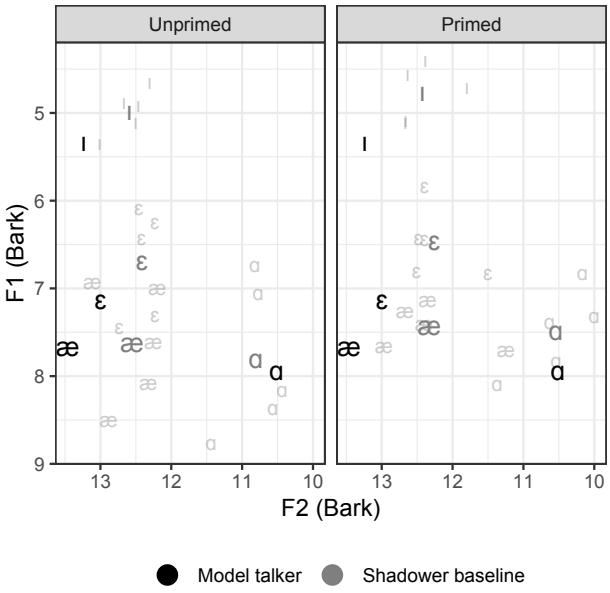

Figure 1: Mean vowel formant frequencies for the model talker (black) and for the shadower baseline utterances (gray), separately by priming condition. Large gray symbols show shadower grand means. Small gray symbols show individual shadower means.

Together, the limited selection of four target vowels and the considerable variation in the relative positions of the selected vowels across the model talker and shadowers make normalization of the acoustic vowel measures inappropriate [20]. This lack of acoustic normalization further motivates the use of a perceptual assessment of phonetic convergence. In particular, we assumed that the listeners in the AXB perceptual similarity task would normalize for overall talker differences and attend to the phonetic properties of the Northern Cities Shift in making their similarity judgments.

2.3. Procedure

The listeners in the AXB perceptual similarity task were seated at personal computers equipped with headphones and button boxes in a quiet lab space in the science museum. On each trial, they were presented with three tokens of the same word: the baseline read utterance from one of the shadowers, the model talker utterance, and the shadowed utterance from the same shadower. The model talker's utterance was always presented second (X). The order of the shadower's baseline and shadowed utterances (A/B) were counterbalanced within and across experimental lists. The three stimulus tokens were separated by 100 ms inter-stimulus intervals. Listeners were asked to press the leftmost button on the button box if the first utterance was more similar to the second utterance and to press the rightmost button on the button box if the third utterance was more similar to the second utterance.

Each listener heard all 24 target words for each of five shadowers, either from the unprimed condition or the primed condition, for a total of 120 trials. Prime condition was therefore a between-listener variable. For each prime condition, two lists were constructed so that across lists, the baseline and shadowed utterances of each word from each shadower appeared in both A/B orders. Trial order was randomized separately for each listener. Twenty-five listeners assessed phonetic convergence by the unprimed shadowers and 24 listeners assessed phonetic convergence by the primed shadowers.

3. RESULTS

Phonetic convergence is assumed to be perceptible when listeners in an AXB perceptual similarity task identify the shadowed utterance as more similar to the model talker utterance than the baseline utterance [18, 19]. The mean proportion of shadowed utterances identified as more similar to the model talker utterance in each of the two priming conditions is shown in Fig. 2. As is typical for perceptual assessments of phonetic convergence [8, 11, 14, 15, 18, 19], the proportions are near 0.50, which corresponds to chance performance in a two-alternative forced-choice task. Despite this low performance overall, performance was higher in the primed condition than in the unprimed condition, as expected.

A logistic mixed-effects model predicting shadowed responses from priming condition, target word stressed vowel, and their interaction was constructed. Both predictor variables were sum-contrast coded. The maximal random effect structure that converged included only a random by-listener

intercept. The model revealed a significant positive intercept (Est. = 0.10, $z = 3.64$, $p < 0.001$), confirming significant perception of convergence overall. The main effect of priming condition was also significant (Est. = -0.07, $z = -2.48$, $p = 0.013$), confirming greater perception of phonetic convergence in the primed condition than the unprimed condition, as expected. Although the model returned a significant slope estimate for /ɑ/ (Est. = -0.10, z = -2.13, p = -0.034), suggesting less convergence to /ɑ/ than to the overall mean, no by-vowel pairwise comparisons using estimated marginal means were significant. The interaction was also not significant.

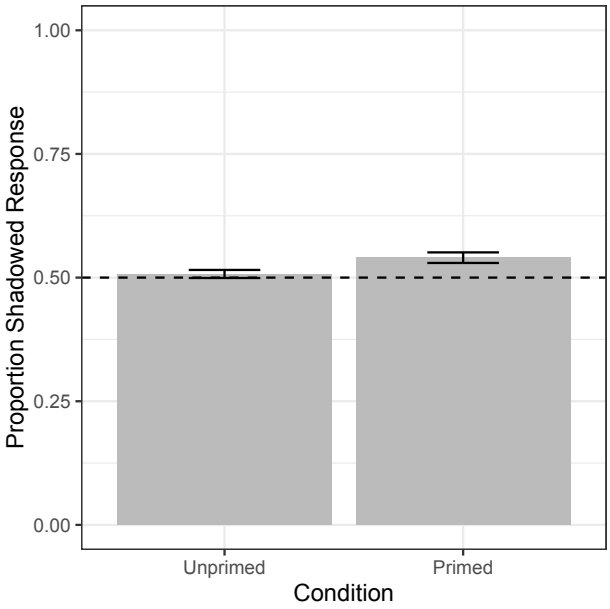

Figure 2: Mean proportion shadowed responses in the AXB task for the two priming conditions (unprimed, primed). Error bars are standard error of listener means.

4. DISCUSSION

The AXB perceptual similarity task revealed significant convergence to the Northern model talker by the Midwestern shadowers, as well as more robust convergence in the primed condition than in the unprimed condition. Consistent with previous work [11, 14, 15, 19], the effect sizes for both the intercept (i.e., overall convergence) and priming condition were small (see Fig. 2), reflecting the difficulty of the AXB perceptual similarity task.

The significant overall convergence is consistent with previous research demonstrating cross-dialect convergence in word shadowing tasks [10, 11, 13, 14, 15]. As shown in Fig. 1, the model talker produced phonetic features of the Northern Cities Shift and the shadowers' baseline exhibited sufficient acoustic distance from the model talker for phonetic convergence to be observed. In addition, the lack of social stereotyping of the Northern dialect and its vowel variants led to convergence for all four vowels, as expected based on previous work [10, 11, 13, 14, 15]. These results therefore provide further evidence for cross-dialect phonetic convergence when social stereotypes are not present to block it.

The stronger perception of phonetic convergence in the primed condition than in the unprimed condition suggests that providing shadowers with accurate information about the model talker's region of origin can facilitate phonetic convergence. That is, simply being told where the model talker is from, without any further instruction to imitate her, led to more robust phonetic convergence than when no information about the model talker was provided. This result parallels previous findings that accurate social information can facilitate speech processing [1, 3, 5] and extends these findings to include phonetic convergence.

However, this result critically differs from a parallel study involving a Southern American English model talker, for whom phonetic convergence did not differ in unprimed and primed conditions [14]. Given that Southern American English is negatively stereotyped in the U.S. [22], this difference across studies suggests that the blocking of phonetic convergence due to negative social stereotypes may outweigh any priming effects arising from providing social information about the model talker. That is, the negative effects on phonetic convergence of social stereotypes related to Southern American English in the previous study appear to have been stronger than the positive effects of social priming that we observed for the Northern American English model talker in the current study. As a result, no positive effects of social priming were observed in the earlier study with a Southern American English model talker.

Taken together, the results of this study suggest robust cross-dialect phonetic convergence to the non-stereotyped Northern dialect of American English, consistent with previous work on cross-dialect phonetic convergence [10, 11, 13, 14, 15]. In addition, the results provide novel evidence for enhancement of convergence when the model talker's region of origin is explicitly provided. This effect of social priming on phonetic convergence advances our understanding of the role of non-linguistic social information in speech processing [1, 2, 3, 4, 5, 6, 7]. In particular, the results of the current study suggest that social priming extends beyond speech perception to include speech production, consistent with a very tight connection between speech perception and production processes [8, 23].

5. ACKNOWLEDGMENTS

This research was conducted at the Buckeye Language Network Language Sciences Research Lab at the Center of Science Industry in Columbus, Ohio, USA. We would like to thank Laura Beebe, Emily Behm, Sarah Chilson, Ben Elsbrock, Chance Goodwin, Conor Higgins, Rachel Monnin, and McKenna Reeher for assistance with data collection and analysis.

6. REFERENCES

[1] Babel, M., Russell, J. 2015. Expectations and speech intelligibility. *J. Acoust. Soc. Am.* 137, 2823–2833.

[2] Hay, J., Drager, K. 2010. Stuffed toys and speech perception. *Linguistics* 48, 865–892.

[3] McGowan, K. B. 2015. Social expectation improves speech perception in noise. *Lang. Speech* 58, 502–521.

[4] Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *J. Lang. Social Psychol.* 18, 62–85.

[5] Rubin, D. L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Res. High. Educ*. 33, 511–531.

[6] Staum Casasanto, L. 2008. Does social information influence sentence processing? In: *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, 799–804.

[7] Strand, E. A. 1999. Uncovering the role of gender stereotypes in speech perception. *J. Lang. Social Psychol.* 18, 86–99.

[8] Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279.

[9] Shockley, K., Sabadini, L., Fowler, C. A. 2004. Imitation in shadowing words. *Percept. Psychophys.* 66, 422–429.

[10] Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J. Phon.* 40, 177–189.

[11] Walker, A., Campbell-Kibler, K. 2015. Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Front. Psychol.* 6(546), 1–18.

[12] Yu, A. C. L., Abrego-Collier, C., Sonderegger, M. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and 'autistic' traits. *PLoS One* 8(e74746), 1–13.

[13] Babel, M. 2010. Dialect divergence and convergence in New Zealand. *Lang. Soc*. 39, 437–456.

[14] Clopper, C. G., Dossey, E. 2020. Phonetic convergence to Southern American English: Acoustics and perception. *J. Acoust. Soc. Am.* 147, 671–683.

[15] Ross, J. P., Lilley, K. D., Clopper, C. G., Pardo, J. S., Levi, S. V. 2021. Effects of dialect-specific features and familiarity on cross-dialect phonetic convergence. *J. Phon.* 86(101041), 1–23.

[16] Labov, W., Ash, S., Boberg, C. 2006. *Atlas of North American English*. Mouton de Gruyter.

[17] Campbell-Kibler, K. 2012. Contestation and enregisterment in Ohio's imagined dialects. *J. Eng. Linguist.* 40, 281–305.

[18] Pardo, J. S. 2013. Measuring phonetic convergence in speech production. *Front. Psychol*. 4(559), 1–5.

[19] Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., Lewandowski, E. 2013. Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *J. Mem. Lang.* 69, 183–195.

[20] Clopper, C. G., Pisoni, D. B., de Jong, K. 2005. Acoustic characteristics of the vowel systems of six regional varieties of American English. *J. Acoust. Soc. Am*. 118, 1661–1676.

[21] Clopper, C. G., Carter, A. K., Dillon, C. M., Hernandez, L. R., Pisoni, D. B., Clarke, C. M., Harnsberger, J. D., Herman, R. 2002. The Indiana Speech Project: An overview of the development of a multi-talker multi-dialect speech corpus. In: *Research on Spoken Language Processing Progress Report No. 25*. Speech Research Laboratory, Indiana University, 367–380.

[22] Niedzielski, N., Preston, D. R. 2003. *Folk Linguistics*. Mouton de Gruyter.

[23] Pickering, M. J., Garrod, S. 2013. An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347.