

F0 convergence in a sentence repetition task: Testing the role of dialectal similarity and perceived social characteristics of the talker

Jessamyn Schertz, Noemi Caponi, and Reem Mahgoub¹

¹*University of Toronto Mississauga*

Introduction: While it is well-established that speakers align to the phonetic characteristics of their interlocutors, the mechanisms underlying phonetic convergence are not well understood. Previous studies report greater convergence to talkers sharing the same speech variety [2,5], and to talkers perceived as having more positive social attributes [1,4]. However, these findings have not always been consistently replicated, leading to mixed results [6]. We use a paradigm which circumvents a common issue in quantifying convergence, the effect of initial phonetic distance [3], to examine the extent of f0 convergence to three different varieties of English (UK, US, and Indian English) by speakers of two of these varieties (US and UK). We test the following questions: *1) Do participants show greater convergence to their own variety? and 2) Is convergence modulated by social perceptions of the talkers?*

Methods: 24 US and 25 UK English speakers completed two online tasks. First, in a shadowing task, they heard and repeated trivia sentences spoken by different talkers; no explicit instructions to imitate were given. They then completed a social rating task in which they rated a subset of these sentences on social attributes including friendliness, socioeconomic status, and intelligence [1]. Target sentences were created from natural recordings of six native speakers of English, two from each of three English varieties: US Midland (US), South East England (UK), and India (IN). Two versions of each naturally-produced utterance were created by artificially increasing and decreasing the natural f0 range by 50%, resulting in 96 target stimuli (8 sentences * 2 f0 versions (HI vs. LO variability) * 3 varieties * 2 talkers per variety). We measured f0 of the shadowed productions and used the interquartile range as a proxy for f0 variability. Convergence was quantified as the difference between participants' f0 range in the HI and LO versions of each sentence, with larger values representing more convergence. Mixed-effects models were used for statistical analysis.

Results: To test for the presence of f0 convergence, and whether extent of convergence differed across varieties and participant groups, we examined how f0 range in shadowed productions was predicted by shadowing variety (UK, US, IN), participant group (UK, US), f0-version (HI, LO) (all interactions and the maximal random effects structure were included in the model). As shown in Fig. 1, F0 ranges were significantly higher in the HI vs. LO condition, indicating overall convergence. US participants showed overall less convergence (albeit still a significant amount) than UK participants for all varieties; however, there were no differences in extent of convergence to the three varieties, either within or across participant groups, indicating that participants did not show greater convergence to their own dialect.

As shown in Fig. 2a, results of the rating task showed that UK talkers were perceived, by both UK and US participants, as higher in intelligence and socioeconomic status than the talkers from the other two varieties, but there were no differences in friendliness ratings across varieties or talker groups. However, there was substantial individual variability along all rating dimensions, allowing us to test whether participants showed greater convergence to talkers they rated highly (Fig. 2b). Numerically, participants' individual extent of convergence showed a positive correlation in all rating scales, but none reached significance.

In sum, we found that in a controlled task where variability in initial distance from the target was removed, there was equally robust f0 convergence regardless of the variety being spoken. We found no evidence that perceived social characteristics of the talker influenced levels of convergence. We speculate that individual speakers' proclivity toward convergence may play a greater role than characteristics of the exposure speech in predicting the extent of convergence.

F0 range in shadowed productions of sentences with high vs. low f0 variability

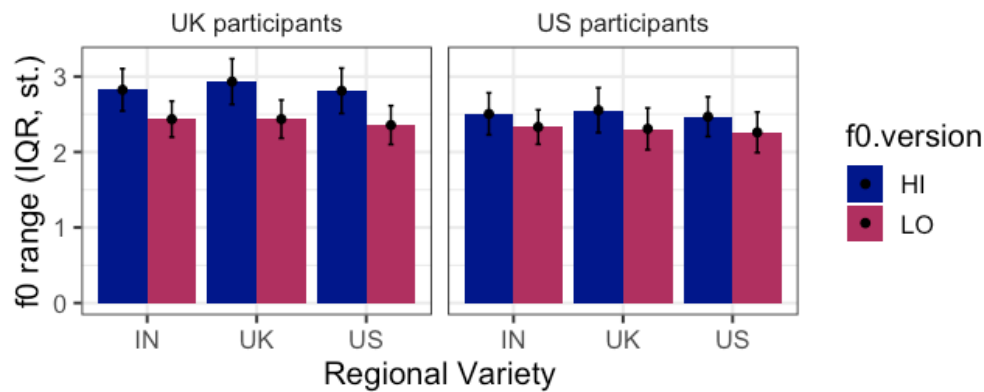
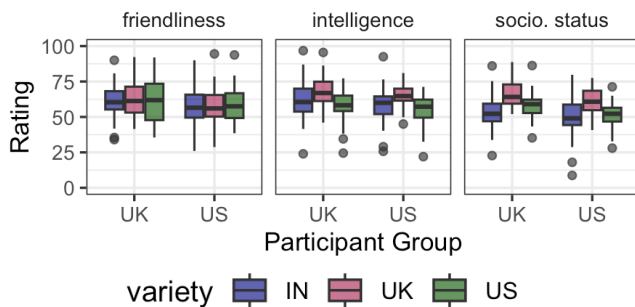


Fig. 1. F0 ranges (as quantified by the interquartile range of f0 across the utterance, measured in semitones) for UK and US participants shadowing sentences spoken by speakers from India (IN), South East England (UK), and Midland United States (US), for sentences with high (blue) or low (red) f0 range. Error bars show 95% confidence intervals based on the distribution of by-participant means.

a. Ratings by Participant Group and Talker Variety



b. Ratings by extent of convergence

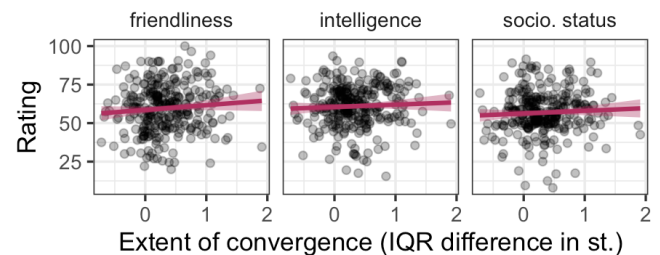


Fig. 2. Ratings for friendliness, intelligence, and socioeconomic status by participant group and stimulus variety (a) and by extent of convergence (b). Group-level results (a) show UK talkers rated higher in intelligence and socioeconomic status by both US and UK listeners. For individual-level results (b), each point represents a pair of shadowed productions (HI and LO versions of the same sentence) by one participant; extent of convergence represents the difference in IQR between the HI and LO versions, and Rating represents that participant's average rating for the relevant talker in friendliness, intelligence, or socioeconomic status.

References

- [1] Dodd, N., Cohn, M., & Zellou, G. (2023). Comparing alignment toward American, British, and Indian English text-to-speech (TTS) voices: influence of social attitudes and talker guise. *Frontiers in Computer Science*, 5.
- [2] Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2(1), 125–156.
- [3] MacLeod, B. (2021). Problems in the Difference-in-Distance measure of phonetic imitation. *Journal of Phonetics*, 87, 101058.
- [4] Michalsky, J., & Schoormann, H. (2017). Pitch convergence as an effect of perceived attractiveness and likability. In *Proceedings of Interspeech 2017*, 2253-2256.
- [5] Ross, J., 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. *Journal of Phonetics*, 86, 101041.
- [6] Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, 6, 546.