Switchboard speakers' vowels' covary, but they do not converge

Uriel Cohen Priva[1] and Chelsea Sanker[2]
[1]Brown University (USA) [2]Stanford University (USA)

Two effects are often grouped together broadly as convergence: (1) shifts of a speaker towards characteristics of an interlocutor or model talker [e.g. 4], and (2) covarying productions of a speaker and an interlocutor. The same cognitive process might underlie some aspects of both, but the latter can have a range of sources, e.g. lexical variation due to conversation topic, and conversation style (e.g. an argument). Few studies are set up to distinguish between the two effects [but see 3]. We find that the Switchboard corpus provides robust evidence for covariation in vowel formants, but formant convergence is weak or absent. These results suggest that covariation can be present even when two interlocutors are not impacted by each other's baseline behavior.

Using the Switchboard corpus of American English, we measured F1 and F2 in stressed monophthongs in words with no sonorant consonants. Formants were Lobanov normalized by speaker and averaged for each speaker in each conversation (one value per formant per vowel). In order to distinguish between covariation and convergence, we analyzed the data in two ways: (1) using the interlocutor's productions in the shared conversation as a predictor of the speaker's formants, and (2) using the interlocutor's productions from other conversations as a predictor. The latter allows us to ensure that only convergence is captured, while the former will find effects both of convergence and covariation. Separate models were used for each formant and vowel. We used the linear combination method for all analyses [2].

The following figure illustrates estimate distributions from models using the interlocutor's productions in other conversations (convergence) and the shared conversation (covariance) as factors predicting a speaker's formants; significant values are labeled. There is very limited evidence for convergence; the mode for these estimates is at zero. However, we found evidence for covariation for many vowel × formant combinations. In both sets of models, there was strong evidence for speakers' self-consistency; i.e., a speaker's formants in a particular conversation were predicted by their formants in other conversations.
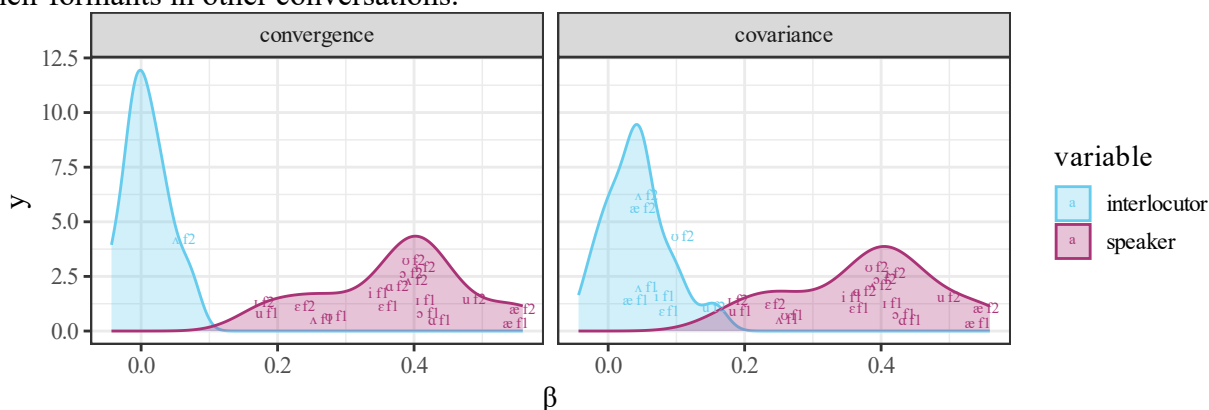
Distinguishing between convergence and covariation is important for establishing precisely what evidence is present in the data, particularly for characteristics which are sensitive to the conversational context. This distinction is important for understanding what underlies convergence in phonological representations. Adopting the phonetic characteristics of another speaker reflects a change in the target phonetic details, which can last beyond the conversation and could contribute to sound changes. In contrast, covariation might not reflect any change in the representation, and might be present in a range of contexts in which speakers do not converge in the characteristics of

interest.

We found no convergence in formant values, which might suggest that the variable targets for formants, in combination with the ambiguity caused by listeners' perceptual adjustments to handle speakers' different vowel space sizes, may result in convergence only to extreme formant differences. While prior work reports some significant convergence in formants [e.g. 1, 5], it is not clear that there is evidence for convergence in conversational contexts when covariation is excluded as a potential confound.

References

[1]   Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*. 40, 1 (2012), 177–189. DOI:https://doi.org/10.1016/j.wocn.2011.09.001.

[2]   Cohen Priva, U. and Sanker, C. 2019. Limitations of difference-in-difference for measuring convergence. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*. 10, 1 (Sep. 2019), 15. DOI:https://doi.org/10.5334/labphon.200.

[3]   Levitan, R. and Hirschberg, J.B. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of Interspeech* (Brisbane, 2011), 3081–3084.

[4]   Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics*. 39, (2011), 132–142. DOI:https://doi.org/10.1016/j.wocn.2010.12.007.

[5]   Schweitzer, A. and Lewandowski, N. 2014. Social factors in convergence of F1 and F2 in spontaneous speech. *Proceedings of the 10th international seminar on speech production, cologne* (2014), 391–394.