Phonetic convergence as an optimality-guided process

Benjamin Elie[1], Juraj Simko[2], and Alice Turk[1]

1. LEL; PPLS; The University of Edinburgh; Edinburgh; United Kingdom
2. Faculty of Arts; University of Helsinki; Helsinki; Finland

Interlocutors' production can converge (see Pardo *et al.* 2022, Pellegrino *et al.* 2024 for a detailed review). This convergence occurs on many levels, involving extra-linguistic features such as facial expression, body movements, as well as linguistic features, including lexical choice and acoustic correlates of contrast and prosodic structure.

To date, there is no computational model for these effects. In this paper, we propose a simple computational implementation of the Interactive Alignment Model (Pickering and Garrod 2004). Our focus is on the alignment of mental representations that relate phonetic characteristics to phonological categories. We adopt the Phonetic Planning component of XT/3C (Turk & Shattuck-Hufnagel 2020, Turk e*t al.* 2025). It assumes articulatory movements are chosen to satisfy conflicting requirements, such as maximal acoustic intelligibility, minimal articulatory effort, and minimal duration. Our computational model assumes that planned movements minimise a cost function that account for these requirements. Using previous versions of XT/3C (Elie *et al.* 2024), we model acoustic intelligibility as the likelihood of identifying a phoneme based on auditory characteristics learned through experience. We assume that the auditory representation of each phoneme is flexible and is updated during multi-speaker interactions each time a speaker hears productions of others. Assuming that a listener/speaker can interpret what they hear in terms of what they would have done to produce it, we hypothesise that the update of auditory representations can change the optimal solution to the phonetic planning task and can eventually lead to phonetic convergence in conversation.

We present simulations of interactions of pairs of virtual speakers with distinct pretrained auditory representations of vowels, trained on a formant database. We show that our approach replicates phonetic convergence. We also show that the alignment of auditory representations increases mutual intelligibility while reducing articulatory effort, thus optimising communication efficiency, as illustrated in Figure 1. This computational model can be used to understand mechanisms that may explain variation in convergence, including divergence.

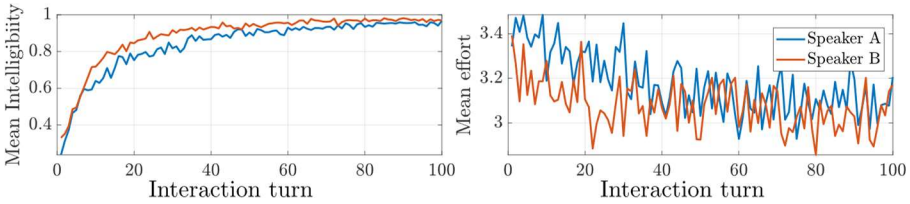

Figure 1: Intelligibility score and articulatory effort of a pair of speakers

References

- Elie, B., Šimko, J., & Turk, A. (2024). Optimization-based planning of speech articulation using general Tau Theory. *Speech Communication*, *160*, 103083.
- Pardo, J. S., Pellegrino, E., Dellwo, V., & Möbius, B. (2022). Vocal accommodation in speech communication. *Journal of Phonetics*, *95*, 101196.
- Pellegrino, E., Dellwo, V., Pardo, J. S., & Möbius, B. (2024). Forms, factors and functions of phonetic convergence. *Speech Communication*, *165*, 103142.
- Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, *27*(2), 212-225.
- Turk, A., & Shattuck-Hufnagel, S. (2020). *Speech Timing: Implications for Theories of Phonology, Phonetics, and Speech Motor Control.* Oxford University Press.
- Turk, A., & Shattuck-Hufnagel, S., Elie, B., & Šimko, J. (2025). From Phonological Symbols to Articulation: A new 3-Component Model of Speech Production. In *Speech Sciences Entries. Speech Prosody Studies Group.* https://gepf.falar.org/entries/73.