Human-computer interaction: Convergence in allophonic contrasts

Iona Gessinger[1], Bernd Möbius[1], Eran Raveh[1] & Ingmar Steiner[2]
[1]Saarland University, [2]DFKI GmbH

gessinger@coli.uni-saarland.de

Humans in conversational interaction tend to become more similar to each other with respect to features pertaining to the phonetic domain (Pardo 2006). Since such phonetic convergence contributes to successful communication, the phenomenon is relevant for human-computer interaction (HCI) as well. A core question arising in the HCI context is whether users of spoken-dialogue systems (SDS) converge to the synthetic speech output. To shed light on this question, we consider reasons for converging behavior between humans assumed in the literature that apply to HCI as well: an automatic perception-production link and the belief of the speaker that their interlocutor, in HCI the computer, will benefit communicatively from the convergence (Branigan et al. 2010). The first should take effect independent of the interlocutor being a human or a computer, the second might be even more decisive in HCI than between humans. Given this background, we expect phonetic convergence on the part of the user to occur in HCI.

Studies exploring phonetic convergence in HCI are so far testing global features, such as speaking rate and pitch. We are targeting local features, specifically the German allophonic contrasts (1) [eː] vs. [ɛː] in a word like *K**ä**se* and (2) [ɪç] vs. [ɪk] as realization of the suffix *-ig* in a word like *biss**ig***. In a shadowing experiment with natural and synthetic stimuli, where participants repeated short German sentences containing these contrasts, a convergence effect was found for (1) and (2) in the case of the natural stimuli, but only for (2) in the case of the synthetic stimuli (Gessinger et al. 2017). A possible reason is insufficient perceptibility of fine phonetic detail in the synthetic stimuli, which makes it possible for the user to converge to the coarse, binary [ɪç] vs. [ɪk] contrast, yet prevents the same effect from occurring for the continuous [eː] vs. [ɛː] contrast.

The shadowing paradigm, although offering the advantage of high controllability, lacks a crucial element of real-life HCI: the dynamic, often goal-oriented, exchange of information. The latter is covered in a new corpus that we are collecting in a Wizard-of-Oz experiment. In this paradigm the user converses with a simulated intelligent SDS. The allophonic contrasts described above are elicited in a map task: The user asks the system about hidden target items on the map. The SDS is informed about the preferred allophone of the user and provides the missing information using the opposite allophone. The latter can then be adopted by the user in the following utterance.

References: • Pardo, J. 2006. On phonetic convergence during conversational interaction. *JASA* 119(4). 2382–2393. • Branigan, H. et al. 2010. Linguistic alignment between people and computers. *Journal of Pragmatic*s 42(9). 2355–2368. • Gessinger, I. et al. 2017. Shadowing synthesized speech – segmental analysis of phonetic convergence. *Interspeech.* 3797–3801.