Talking to a brick wall? Examining the role of a priori assumptions in computer-directed speech

Jurjen de Haan[1], Khiet P. Truong[1 2]
[1]Human Media Interaction, University of Twente, [2]Centre for Language Studies, Radboud University

Several studies have shown that speakers do not address computers in the same way as humans, leading to the identification of computer-directed speech as a distinct speech register (e.g., (Burnham et al., 2010; Cohn et al., 2022)). Findings of slower speaking rates (Cohn et al., 2024), increased loudness (Raveh et al., 2019), and vowel hyperarticulation (Burnham et al., 2010) resemble patterns of clear speech (Smiljanic, 2021; Smiljanić & Bradlow, 2009), and suggested reasons for adaptations commonly focus on speakers' mental models of computers as having limited communicative competence (Cohn et al., 2024; Cohn et al., 2022; Ibrahim & Skantze, 2021; Raveh et al., 2019; Zellou & Holliday, 2024).

However, there is a lack of evidence that speech adaptations toward computers are systematic patterns of variation associated with addressing a computer rather than a human or other broad category of interlocutor, as the term "computer-directed speech register" implies. Human and computer conditions in previous studies do not only differ in interlocutor category, but also in interlocutor-specific factors such as error rate and voice type, which may drive differences in speech production between conditions through hyperarticulation (Lindblom, 1990; Smiljanic, 2021) and phonetic alignment (Pickering & Garrod, 2004), respectively.

To address this problem, this study aimed to isolate the effect of participant's belief of their interlocutor's identity on speech production using a controlled experiment. Participants in the computer condition were instructed to perform an interactive task with a computer while those in the human condition were instructed to perform the task with another person (between-subjects factor). Human and computer conditions were controlled by the researcher in a Wizard-of-Oz style setup and were near-identical[1] in perceptual and interactional cues throughout the task, to control for the effects of interlocutor-specific factors on speech production. The task structure yielded original productions, confirmations, and error-repair utterances for analysis (within-subjects factor). The speech variables examined in this study include speaking rate, mean f0, f0 range, vowel space area, and vowel duration.

Beyond the effects of interlocutor identity and utterance type on speech production, this study additionally investigated whether potential speech differences in human and computer conditions are specifically driven by speakers' expectations of their interlocutor's ability to comprehend spoken language. To do this, participant ratings of how well they expect their interlocutor to understand them were collected before the task and were assessed as predictors of the speech variables.

Contrary to the hypothesis, interlocutor identity did not have an effect on overall speech style. Instead, speech differences were driven by local communicative demands, including repairing misunderstandings and confirming sentences that were correctly understood. Participants' expectations of their interlocutor's ability to understand them were not significantly associated with the speech variables. The findings contribute to our understanding of the sources of computer-directed speech and motivate further research on the relative contributions of a priori assumptions, perceptual differences, and local communicative demands.

[1] Human and computer conditions differed in a name presented on the task screen throughout the experiment. Otherwise, the conditions were perceptually and interactionally identical.

References

Burnham, D., Joeffry, S., & Rice, L. (2010). Computer- and human-directed speech before and after correction. In *Australasian International Conference on Speech Science and Technology* (pp. 13-17): ASSTA.

Cohn, M., Barreda, S., Graf Estes, K., Yu, Z., & Zellou, G. (2024). Children and adults produce distinct technology- and human-directed speech. *Scientific Reports*, *14*(1), 15611. https://doi.org/10.1038/s41598-024-66313-5

Cohn, M., Segedin, B. F., & Zellou, G. (2022). Acoustic-phonetic properties of Siri- and human-directed speech. *Journal of Phonetics*, *90*, 101123. https://doi.org/https://doi.org/10.1016/j.wocn.2021.101123

Ibrahim, O., & Skantze, G. (2021, 2021). *Revisiting robot directed speech effects in spontaneous Human-Human-Robot interactions* Human Perspectives on Spoken Human-Machine Interaction, https://doi.org/10.6094/UNIFR/223822http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-312486

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403-439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169-190. https://doi.org/10.1017/S0140525X04000056

Raveh, E., Steiner, I., Siegert, I., Gessinger, I., & Möbius, B. (2019, 2019/3). *Comparing phonetic changes in computer-directed and human-directed speech* Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2019, https://www.essv.de/pdf/2019_42_49.pdf

Smiljanic, R. (2021). Clear Speech Perception. In *The Handbook of Speech Perception* (pp. 177-205). https://doi.org/https://doi.org/10.1002/9781119184096.ch7

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass*, *3*(1), 236-264. https://doi.org/https://doi.org/10.1111/j.1749-818X.2008.00112.x

Zellou, G., & Holliday, N. (2024). Linguistic analysis of human-computer interaction [Review]. *Frontiers in Computer Science*, *Volume 6 - 2024*. https://doi.org/10.3389/fcomp.2024.1384252