

Can voice recognizability be controlled by speakers? A study on identity marked speech

Valeriia Perepelytsia¹, Leah Bradshaw¹, Volker Dellwo¹

¹Department of Computational Linguistics, University of Zurich, Zürich, Switzerland
valeriia.perepelytsia@uzh.ch

Abstract (max 300 words)

Previous studies have shown that voice recognizability can be controlled by speakers. For example, they can deliberately change individuality cues to disguise their voices (Eriksson, 2010; Eriksson & Wretling, 1997) by either manipulating their own individuality cues (e.g., Hove & Dellwo, 2014), or by imitating those of other speakers (e.g., Kitamura, 2008). Therefore, speakers have some intuition regarding which cues are best manipulated to reveal less information about their voice identity. However, the evidence is limited regarding whether speakers are also capable of altering their vocal properties to reveal *more* information about their identity. One study by Dellwo et al. (2019) asked participants to speak to a mock automatic speech recognition system, eliciting more intelligible speech (*clear speech*) and to a mock automatic speaker recognition system, eliciting more recognizable speech (*identity-marked speech*). They found preliminary evidence that speech within these two styles differed and argued that speakers exhibit some control over cues to their voice identity and ability make themselves more recognizable.

This study offers an expansion of the work by Dellwo et al. (2019), in which we will further explore the phenomenon of *identity-marked speech*. We will present an updated experimental technique with improved ecological validity to make the task of talking to a mock speech and speaker recognition system more realistic. We will collect speech data in three speaking styles: read, clear, and identity-marked speech, which will be used to examine whether human and computer recognition performance benefits from identity-marked speech compared to clear and read speech. We will also present acoustic analyses of these speaking styles, as well as results of data clustering using t-distributed Stochastic Neighborhood Embedding method (van der Maaten & Hinton, 2008). These findings will contribute to the discussion of the extent to which speakers are able to manipulate their indexical vocal cues.

[Abstract word count: 299/300]

References

- Dellwo, V., Pellegrino, E., He, L., Kathiresan, T. (2019). The dynamics of indexical information in speech: Can recognizability be controlled by the speaker? *AUC PHILOLOGICA*, 2019(2), 57-75.
- Eriksson, A. (2010). The disguised voice: imitating accents or speech styles and impersonating individuals. In C. Llamas & D. Watt (Eds.), *Language and Identities* (pp. 86–96). Edinburgh University Press.
- Eriksson, A., Wretling, P. (1997). How flexible is the human voice? A case study of mimicry. *Fifth European Conference on Speech Communication and Technology*.
- Hove, I., Dellwo, V. (2014). The effect of voice disguise on f0 and on the formants. *Proceedings of IAFPA*.
- Kitamura, T. (2008). Acoustic analysis of imitated voice produced by a professional impersonator. *Proceedings of Interspeech 2008*, 813-816.

van der Maaten, L., Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.