Voice identity as predicted from the acoustic properties of fillers

Emily O'Hara[1], Alessandro Vinciarelli [1], Phil McAleer [1]

[1]Affiliation(s) University of Glasgow

e.ohara.1@research.gla.ac.uk

Abstract (max 300 words)

The human voice contains a multitude of social information, allowing listeners to rapidly infer speaker identity as well as impressions of characteristics such as personality and affective state (Belin *et al.,* 2011). Recently, perception research has moved towards more naturalistic representations of voices, with greater emphasis on within-speaker variability and socially relevant stimuli (Lavan *et al*., 2019). In this current study we focused on conversational fillers i.e., filled pauses such as 'uhm' (əm) and 'uh' (ʌh). These are highly common but understudied aspects of natural speech with minimal linguistic cues. We aimed to investigate if even these brief non-verbal utterances contained sufficient information for the formation of stable percepts. Little previous research has used fillers, but Kanber *et al.* (2020) found a strong effect of familiarity on recognition accuracy, illustrating that it is currently unclear in fillers which cues are necessary for recognition and perception. As a starting point, we sought to establish this on a basic acoustic level, by performing k-means clustering to determine whether it was possible for the algorithm to predict identity from acoustic properties. Using 10 fillers each from 93 speakers, we ran an acoustic analysis extracting information on 19 features for each filler. Entropy was then calculated on each participant's clustering result, which determined the amount of information contained within said result, based on the probability of its occurrence. Results yielded a mean entropy score of 0.426, SD = 0.091, which was statistically significant ($p < 0.001$) from a random distribution of 1. These results show that algorithmically it is possible to predict identity from fillers using their underlying acoustics. Furthermore, due to successful extraction of acoustic properties considered in the voice perception literature to be essential cues, the results suggest that stable personality and affective percepts should also be obtainable from fillers.
[Abstract Word count: 298]

References

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding Voice Perception: Understanding voice perception. *British Journal of Psychology*, *102*(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Kanber, E., Lavan, N., & McGettigan, C. (2022). Highly accurate and robust identity perception from personally familiar voices. *Journal of Experimental Psychology: General, 151*(4), 897–911. https://doi.org/10.1037/xge0001112

Lavan, N., Burton, A. M., Scott, S. K., & McGettigan, C. (2019). Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bulletin & Review*, *26*(1), 90–102. https://doi.org/10.3758/s13423-018-1497-7