The impact of voice recognition skills on earwitness testimony

Sascha Schäfer, Paul Foulkes

University of York

sascha.schaefer|paul.foulkes@york.ac.uk

Abstract

Eliciting reliable testimony from earwitnesses has been a long-standing endeavour in the forensic speech science community. Most recent efforts to do so focused on the improvement of a particular procedure, the voice parade (VP), by finding optimal settings for variables that can be controlled by an investigator ("system variables"), such as the quality (McDougall, 2021; Smith et al., 2019) and presentation (Smith et al., 2020) of the stimuli.

The present study complements these findings with an analysis of inter-listener differences in voice recognition, which cannot be controlled by an investigator ("estimator variables"). Psychological tests for assessing voice recognition have already shown participant performances ranging from developmental phonagnosia to 'super recognition' (Aglieri et al., 2017; Mühl et al., 2018). It is, however, unclear whether these results translate to earwitnesses, as the stimuli were created from isolated sounds/syllables rather than naturalistic speech. The present study addresses this problem:

100 British participants (50 male, mean age = 36, SD = 13.8) took part in an AX discrimination task hosted on Pavlovia. For the stimuli, two 10s-long recordings were taken from 48 speakers of the DyVis corpus (Nolan et al., 2009). Three stimulus lists of comparable difficulty were created based on the f0-difference between speakers. Participants were assigned to one of the stimulus lists and provided a same/different rating for 32 voice pairs (16 same), while reaction times were measured. They also reported their confidence (6pt-scale). Participants differed markedly in recognition accuracy (range 50-93.8%, mean =75%, SD = 9.1%), including two 'super-recognisers' (>= 2 SDs above mean) and four participants at the opposite end of the spectrum (<= 2 SDs below mean). The index *d prime* revealed high differences in listener discriminability (range 0-2.94, mean = 1.38, SD = 0.57). The results indicate that earwitnesses might not be equally suited for a standardised VP.

[Abstract Word count: 300]

References

Aglieri, V., Watson, R., Pernet, C., Latinus, M., Garrido, L., & Belin, P. (2017). The Glasgow Voice Memory Test: Assessing the ability to memorize and recognize unfamiliar voices. *Behavior Research Methods*, *49*(1), 97–110. https://doi.org/10.3758/s13428-015-0689-6

McDougall, K. (2021). Ear-catching versus eye-catching? Some developments and current challenges in earwitness identification evidence. *Proceedings of XVII AISV*. https://www.phonetics.mmll.cam.ac.uk/ivip/

Mühl, C., Sheil, O., Jarutytė, L., & Bestelmeyer, P. E. G. (2018). The Bangor Voice Matching Test: A standardized test for the assessment of voice perception ability. *Behavior Research Methods*, *50*(6), 2184–2192. https://doi.org/10.3758/s13428-017-0985-4

Nolan, F., McDougall, K., de Jong, G., & Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech Language and the Law*, *16*(1), 31–57. https://doi.org/10.1558/ijsll.v16i1.31

Smith, H. M. J., Baguley, T. S., Robson, J., Dunn, A. K., & Stacey, P. C. (2019). Forensic voice discrimination by lay listeners: The effect of speech type and background noise on performance. *Applied Cognitive Psychology*, *33*(2), 272–287. https://doi.org/10.1002/acp.3478

Smith, H. M. J., Bird, K., Roeser, J., Robson, J., Braber, N., Wright, D., & Stacey, P. C. (2020). Voice parade procedures: optimising witness performance. *Memory*, *28*(1), 2–17. https://doi.org/10.1080/09658211.2019.1673427