Examining how the amount of training exposure affects recognition of voice identities.

Elise Kanber[1], Carolyn McGettigan[1]

[1]University College London

Elise.Kanber@ucl.ac.uk

Abstract (max 300 words)

In order to fully learn a voice, a listener must be able to recognise it in all its variations. Previous voice learning studies have found that listeners can accurately recognise new voice identities after a short amount of training, provided the conditions are roughly the same at training and test (Holmes et al., 2021). Yet recognition of lab-trained voices has been found to be fairly unstable, with discrimination and identification easily disrupted by changes in speaking style (Lavan et al., 2016; Lavan et al., 2019). Further, we recently showed a substantial disadvantage for recognising lab-trained voices compared to personally familiar (i.e., romantic partner) voices when they are acoustically modulated (Kanber et al., 2021). However, in our study, lab-trained familiarity was based on a relatively short amount of exposure to that identity during training – it may be that a greater degree of training might have allowed the listeners to form more stable identity representations. Here, I will present work that addressed this issue. We manipulated the amount of training listeners received when learning lab-trained voices and tested their subsequent abilities to recognise acoustically modulated and unmodulated samples of those voices. Participants were trained with either 20 or 80 stimuli per voice, and recognition was tested on excerpts with no modulation, or a range of adjustments to glottal pulse rate (GPR) and apparent vocal tract length (VTL). The results showed that increasing exposure can produce overall improved accuracy for acoustically modulated voices and reduce the reliance on low-level acoustic cues. This suggested that even a small amount of additional exposure may be beneficial for forming voice representations. Yet the findings also provided some evidence that task demands affected the precise patterns observed. We thus concluded that representations of lab-trained voices remain far inferior to those observed for personally familiar ones.

[Abstract word count: 298 words]

References:

Holmes, E., To, G., & Johnsrude, I. S. (2021). How long does it take for a voice to become familiar? Speech intelligibility and voice recognition are differentially sensitive to voice training. *Psychological Science*, *32*(6), 903-915. https://doi.org/10.1177/0956797621991137

Kanber, E., Lavan, N., & McGettigan, C. (2021). Highly accurate and robust identity perception from personally familiar voices. Journal of Experimental Psychology: General. Advance online publication. https://doi.org/10.1037/xge0001112

Lavan N., Scott S.K. & McGettigan C. (2016). Impaired generalization of speaker identity in familiar and unfamiliar voices. Journal of Experimental Psychology: General, 145(12), 1604-1614. https://doi.org/10.1037/xge0000223

Lavan N., Burston, L.F.K., Merriman, S.E., Ladwa P., Knight, S. & McGettigan, C. (2019). Breaking voice identity perception: Expressive voices are more confusable for listeners. Quarterly Journal of Experimental Psychology,72(9), 2240-2248. https://doi.org/10.1177/1747021819836890