

## Developing a challenging speaker discrimination test

*Andrea Fröhlich<sup>1,2,3</sup>, Volker Dellwo<sup>1</sup>, Meike Ramon<sup>3</sup>*

<sup>1</sup>Department of Computational Linguistics, University of Zürich, Switzerland

<sup>2</sup>Zurich Forensic Science Institute, Switzerland

<sup>3</sup>Applied Face Cognition Lab, University of Lausanne, Switzerland

andrea.froehlich@uzh.ch

volker.dellwo@uzh.ch

meike.ramon@unil.ch

In the field of vision sciences, so-called “Super-Recognizers” (SRs) - individuals with exceptional face identity processing abilities - have been described (Russell et al., 2009) and receive increasing interest from international law enforcement (Ramon, 2021). Whether superior abilities exist in the auditory domain remains an open question. We are currently developing highly challenging tests of voice processing.

Identifying voice SRs has potential impact for forensic phonetics, where practitioners are charged with increasing amounts of data. Traditional auditory and acoustic-phonetic methods are time-consuming and therefore not suitable to perform large numbers of speaker comparisons for police investigations where time is limited. Automatic voice comparison systems can rapidly complete such tasks. However, they are not reliable if the audio quality is poor or samples are of short duration. Therefore we seek to identify individuals with exceptional voice processing skills to assist forensic phoneticians in such scenarios.

An voice SR would excel at voice processing tasks such as discrimination and recognition. To find people with extraordinary voice processing skills we start by first creating a challenging discrimination task. We employed a pre-trained ECAPA-TDNN model (Ravanelli *et al.* 2021), and combined its embeddings with F0 deltas of the speaker pairs to find the most similar stimuli per trial. The pilot study was run with participants from different classes of police cadets using stimuli from the TEVOID-Corpus (Dellwo *et al.* 2012).

Our results show that in a discrimination task with randomly combined pairs, participants on average correctly classified 83.3% of the pairs, whereas in the challenging discrimination task only 69.5% of the pairs were classified correctly. These results show that we have successfully established a method to create challenging discrimination tasks to be used in the search for auditory SRs. We are currently exploring further test scenarios such as recognition and clustering.

[Abstract Word count: 297]

## References

- Dellwo, V., Leemann, A., & Kolly, M. J. (2012). Speaker idiosyncratic rhythmic features in the speech signal. *Interspeech Conference Proceedings*.
- Ramon, M. (2021). Super-Recognizers—a novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia*, *158*, 107809.
- Ravanelli, M. *et al.* (2021): SpeechBrain: A General-Purpose Speech Toolkit, *arXiv:2106.04624*.
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic bulletin & review*, *16*(2), 252-257.