

Using the visual world paradigm to explore voice identity processing

Leah Bradshaw¹, Chiara Tschirner¹, Lena Jäger¹ and Volker Dellwo¹

¹Department of Computational Linguistics, University of Zurich, Zürich, Switzerland
leah.bradshaw@uzh.ch

Abstract (max 300 words)

Voice recognition and identification tasks are an experimental technique used regularly in forensic phonetic research to explore lay-listener identification capabilities of personally familiar or trained-to-familiar voices. Multiple studies have used them to explore the influence of missing acoustic information on naïve speaker recognition abilities, e.g. glottal-waveform, fundamental frequency and formant modifications (Lavner, Gath & Rosenhouse, 1999), or noisy/degraded signals such as telephone speech (Foulkes & Barron, 2000), as well as the influence of acoustic feature adaptations, e.g. whisper voice (Foulkes & Sóskuthy, 2017) and sweeping harmonics (Dellwo *et al.*, 2018).

The Visual World Paradigm (VWP – Allopenna *et al.*, 1998) is a popular eye-tracking experimental technique in psycholinguistic and phonetic research domains, used to explore online processing of various linguistic information. Typically, it involves participants being presented with a visual scene while listening to speech. Where and when participants visual attention shifts to a given object in the visual world is taken to reflect their current interpretation of the audio stimulus. Although typically used to explore online speech processing, there are limited but promising findings showing its utility for assessing online processing of speaker recognition (Schindler & Reinisch, 2015).

Given its capacity to assess online processing, the VWP represents a viable technique for exploring, for instance, the timing of voice recognition, namely how long it takes for the target to be selected following stimulus onset and the sequence in which voices are considered. Further, it could also be useful for assessing exactly what role voice similarity plays in decision-making. This project presents a first-of-its-kind experimental technique combining a VWP and voice recognition task as a method for exploring voice identity processing.

[Abstract Word count: 271]

References

- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, 38(4), 419–439. <https://doi.org/10.1006/jmla.1997.2558>
- Dellwo, V., Kathiresan, T., Pellegrino, E., He, L., Schwab, S., & Maurer, D. (2018). Influences of Fundamental Oscillation on Speaker Identification in Vocalic Utterances by Humans and Computers. *Interspeech 2018*, 3795–3799. <https://doi.org/10.21437/Interspeech.2018-2331>
- Foulkes, P., & Barron, A. (2000). Telephone speaker recognition amongst members of a close social network. *International Journal of Speech, Language and the Law*, 7(2), 180–198. <https://doi.org/10.1558/sll.2000.7.2.180>

- Foulkes, P., Smith, I., & Sóskuthy, M. (2017). Speaker Identification in Whisper. *Letras de Hoje*, 52(1), 5. <https://doi.org/10.15448/1984-7726.2017.1.26659>
- Lavner, Y., Gath, I., & Rosenhouse, J. (2000). The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Communication*, 30(1), 9–26. [https://doi.org/10.1016/S0167-6393\(99\)00028-X](https://doi.org/10.1016/S0167-6393(99)00028-X)
- Schindler, C. & **Reinisch, E.** (2015). Tracking the temporal relation between speaker recognition and processing of phonetic information. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK.