

Automated speaker verification for detecting identity anomalies and ensuring data integrity in multi-site CNS clinical trials

Kieran McVeigh¹, Jacob Ouyang¹, Vijay Yadav^{1,2}, Anzar Abbas¹, Michelle Worthington^{1,3}

¹ Brooklyn Health, ² University of New South Wales, ³ Yale University Department of Psychiatry

Introduction

- There is a critical need for objective, scalable, and automated solutions to identify data anomalies in clinical trials to ensure data and outcome integrity.
- Here, we developed and validated an automated speaker identification pipeline using an embedding-based speaker identification model to flag instances of voice similarity anomalies.

Methods

- 234 MADRS interview recordings (from 150 participants) from an open label depression study were used in these analyses.
- For each recording, clinician and participant speech was isolated using a transcription and diarization pipeline, then separated into a participant and clinician audio file (Figure 1).
- Embedding representations of participant and clinician voices were then extracted using a pre-trained open source voice identification model (ECAPA-TDNN; Figure 2A).
- The similarity of voice embeddings was then measured across interviews via cosine similarity, and compared to a fixed threshold (Figure 2B).

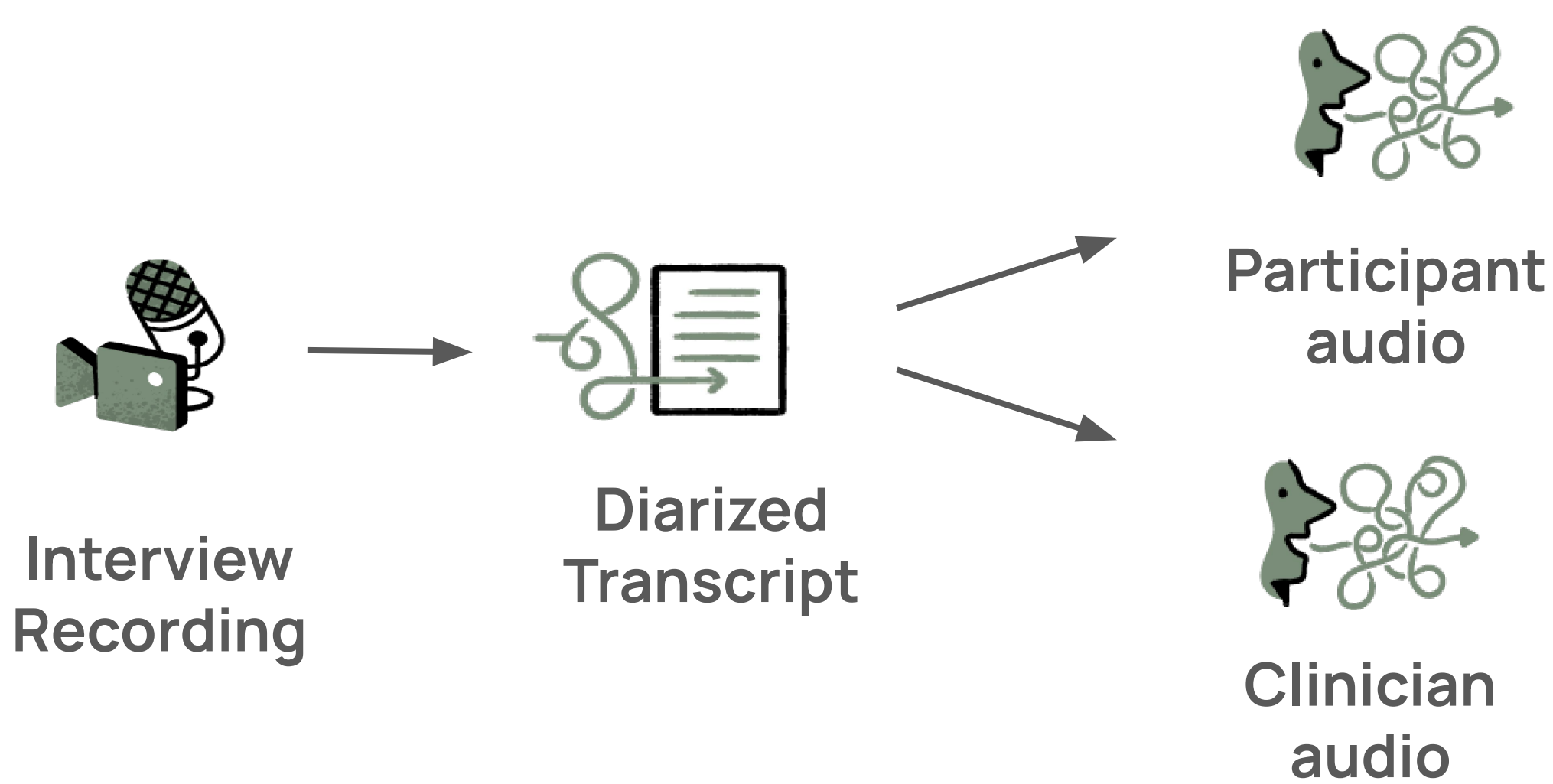


Figure 1. Audio preprocessing pipeline.

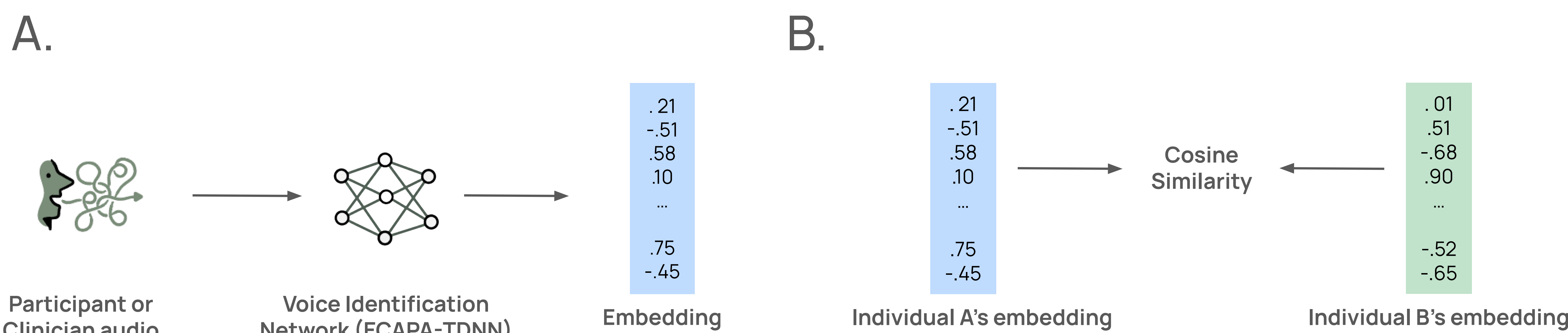


Figure 2. A) Extracting voice embeddings using a pretrained model. B) Comparing similarity of different voice embeddings with cosine similarity.

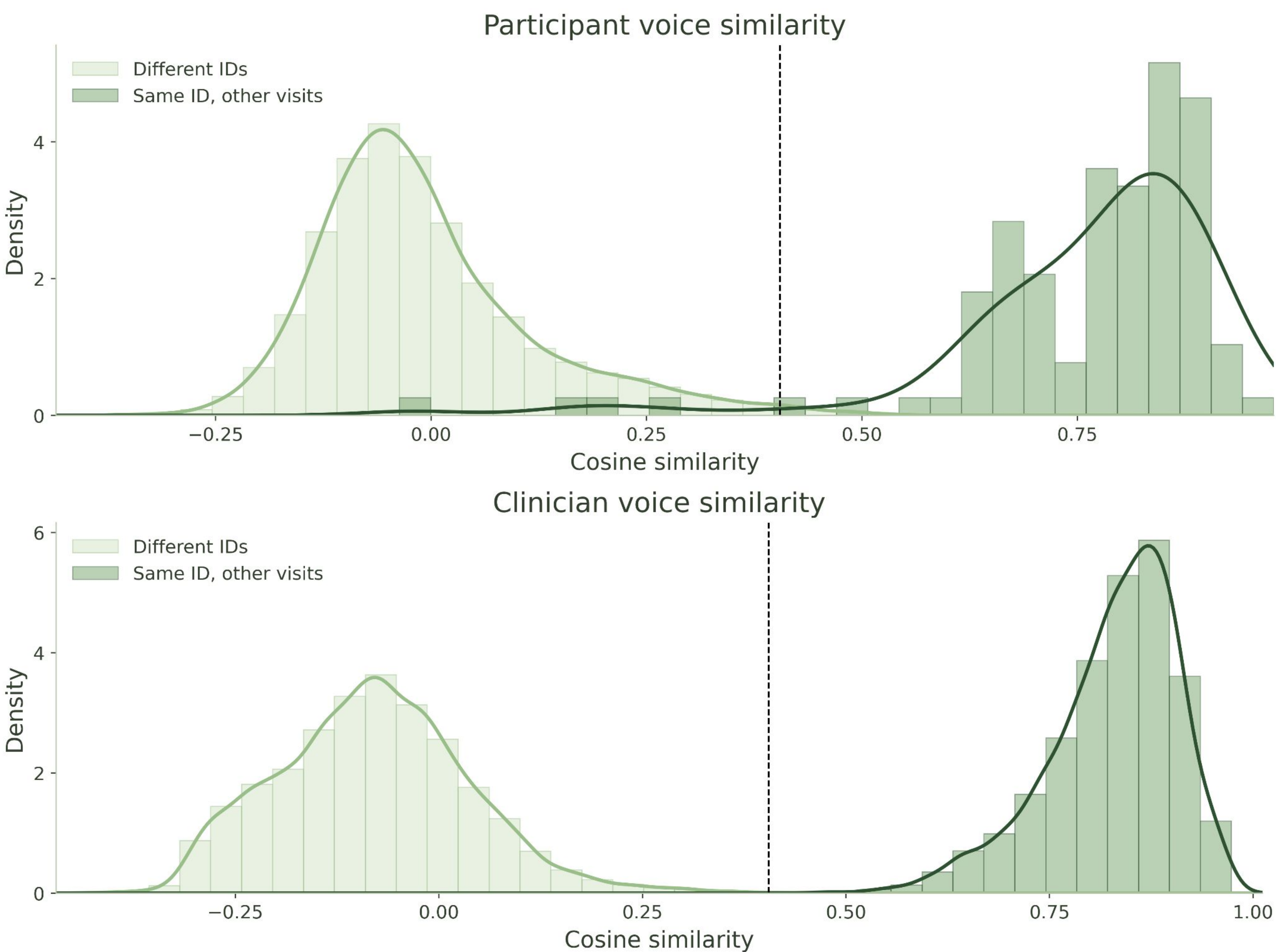


Figure 3. Plots of cosine similarity distributions for voice embeddings across participant IDs (light green, left) and within a participant ID across visits (dark green, right). Dashed vertical line is the threshold for anomalies.

Figure 4. Plots of cosine similarity distributions for voice embeddings across clinician IDs (light green, left) and within a clinician ID across visits (dark green, right). Dashed vertical line is the threshold for anomalies.

Results

- Across different participants IDs, 1% of visit-level voice embeddings were flagged as anomalously similar, indicating different IDs were associated with similar voice embeddings.
- Within a given participant ID, 3.7% of visit-level voice embeddings were flagged as anomalously dissimilar, indicating the same ID was associated with different voice embeddings.
- Across clinician IDs, 0% of visit-level voice embeddings were flagged as anomalously similar, indicating no different IDs were associated with similar voice embeddings.
- Within a given clinician ID, 0% of visit-level voice embeddings were flagged as anomalously dissimilar, indicating the same ID was not associated with different voice embeddings.

Discussion

- This automated speaker identification strategy accurately identified voices which were anomalously similar or dissimilar.
- This method provides a highly scalable, robust framework for automating voice identity anomaly flagging, essential for maintaining data integrity in global clinical trials.

References

Desplanques, B., Thienpondt, J., & Demuynck, K. (2020). Ecapa-tdnn: Emphasized channel attention, propagation and aggregation in tdnn based speaker verification. *arXiv preprint arXiv:2005.07143*.
Ravanelli, M., Parcollet, T., Plantinga, P., Rouhe, A., Cornell, S., Lugosch, L., ... & Bengio, Y. (2021). SpeechBrain: A general-purpose speech toolkit. *arXiv preprint arXiv:2106.04624*.

KM, JO, VY, AA, and MW are all current or former employees of and hold equity in Brooklyn Health, a private company that developed and indirectly benefits from the model described in this poster.

Presented at the 2026 ISCTM Autumn Conference, Philadelphia, PA.