Speaker Verification For Remote Self-Assessments in Clinical Trials

Meepegama U[1], Skirrow C[1], Ropacki MT[2], Fristed E[1], Weston J[1]

[1]Novoic Ltd, London, UK; [2]Strategic Global Research & Development, Temecula, California, USA

OBJECTIVE

To evaluate a novel analytical approach: using a person's voice and speech data for identity verification in longitudinal clinical research.

INTRODUCTION

Remote self-administration of clinical and cognitive assessments have become increasingly common, but present a unique challenge: verifying a participant's identity.[1] Ensuring data is uniquely attributed to the correct individual is important for maintaining the integrity, accuracy, and reliability of trial results. Patient identification methods to avoid 'professional patients' in traditional in-person clinical trials have been described.[2] However, identity verification methods in remote digital trials in general, and Alzheimer's in particular, remain to be established.

The current study evaluates the performance of an automated speaker verification system to detect responses from the same individual across different assessments, devices, environments and over time.

METHODS

Participants: 197 adults confirmed as cognitively unimpaired (N=93), or with mild cognitive impairment or mild Alzheimer's disease (N=104) from the AMYPRED-UK (NCT04828122) and AMYPRED-US (NCT04928976) studies.[3]

Audio recorded assessments:
Baseline assessment: Participants underwent supervised cognitive assessments via zoom or in-person, including the Automated Story Recall Task[4] (ASRT stories L1 and L2, immediate recall) and Category Fluency tasks (CAT). Assessments were recorded on zoom or via a dictaphone.

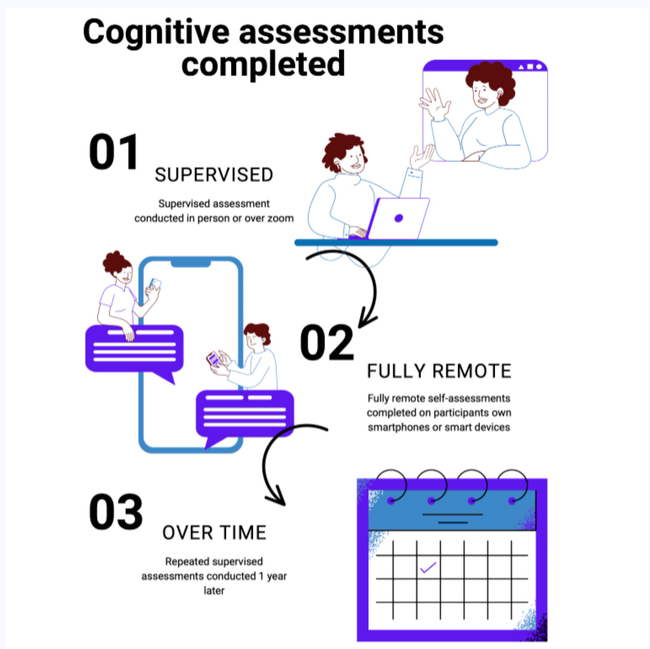
Remote assessments: A subsample completed the ASRT L1 story (N=110) or CAT (N=129) which were self-administered remotely on their smart devices in the week following baseline assessments.

1-Year follow-up: 102 participants (N=42 MCI/Mild AD, N=60 CU) re-enrolled in AMYPRED FUTURE in which supervised ASRTs were repeated 1 year later.

Speaker verification model: Representations from a pre-trained deep learning model were used for speaker verification, extracting a 192-dimensional vector from each audio recording. Vectors were compared with cosine similarity distance, producing an output from -1 to 1 with a higher score indicating greater similarity.

Performance of the speaker verification system was evaluated for different tasks (ASRT, CAT) in different settings (supervised, self-administered remote), at different time points (baseline and +1y follow-up)

Model performance was assessed using Receiver Operating Curve analyses to evaluate Area Under the Curve (AUC) and Equal Error Rate (EER) in the full sample and in male and female only sub-samples.



RESULTS

High performance of the speaker verification system was seen across all assessment contexts (table 1, fig 1 & 2). When restricting the analyses within male and female groups, performance of the speaker verification system remained high, with all AUCs≥0.950.

Assessment pairings: times, tasks and settings compared			AUCs (EER) for speaker verification		
Comparison type	Time, Task, setting (N=196)	Time, Task, setting (N with data)	Full Sample	Females only	Males only
Related tasks, same setting	Baseline, ASRT L1, supervised	Baseline, ASRT L2, supervised (195)	0.997 (0.011)	0.997 (0.006)	0.993 (0.050)
Different tasks, same setting	Baseline, ASRT L1, supervised	Baseline, CAT, supervised (197)	0.992 (0.033)	0.993 (0.035)	0.984 (0.035)
Same task, different settings & devices	Baseline, ASRT L1, supervised	Baseline, ASRT L1, remote (110)	0.993 (0.019)	0.985 (0.027)	0.994 (0.017)
Different tasks, settings & devices	Baseline, ASRT L1, supervised	Baseline, CAT, remote (129)	0.998 (0.024)	0.998 (0.025)	0.997 (0.020)
Same task, same setting, 1 year apart	Baseline, ASRT L1, supervised	1-year follow-up, ASRT L1, supervised (102)	0.973 (0.049)	0.981 (0.027)	0.950 (0.068)
Related tasks, same setting, 1 year apart	Baseline, ASRT L1, supervised	1-year follow-up ASRT L2, supervised (103)	0.996 (0.028)	0.989 (0.051)	0.999 (0.029)

Table 1: AUCs and EERs for speaker verification technology across tasks, settings and time, in the full sample and in female and male only subgroups.

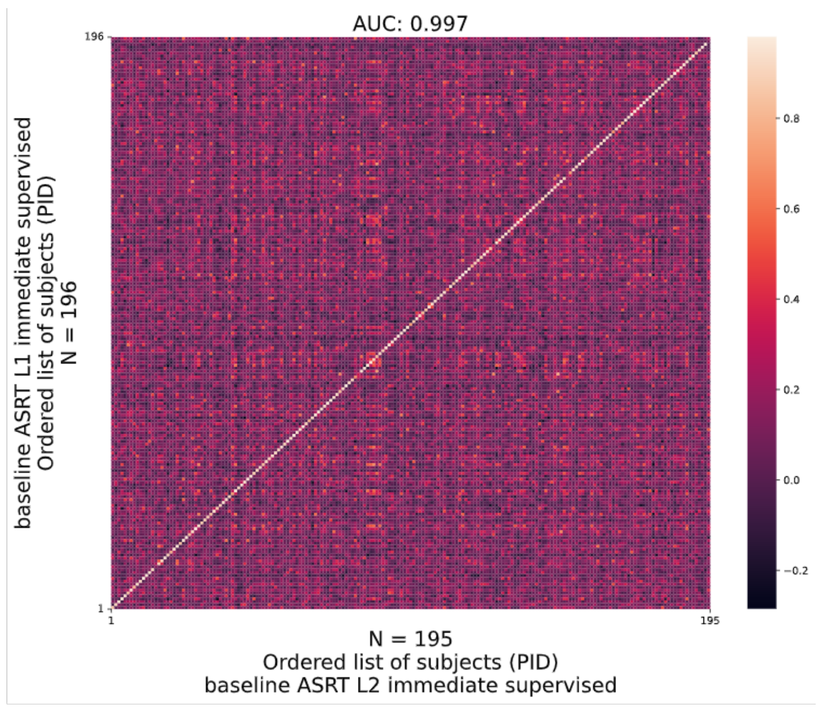


Figure 1: Heatmap visualisation of speaker verification across similar tasks (ASRT L1 and L2 immediate recall) in the same setting (supervised). The lighter colours denote greater similarity between the associated speaker tuples, with this representing the same participant on the diagonal.

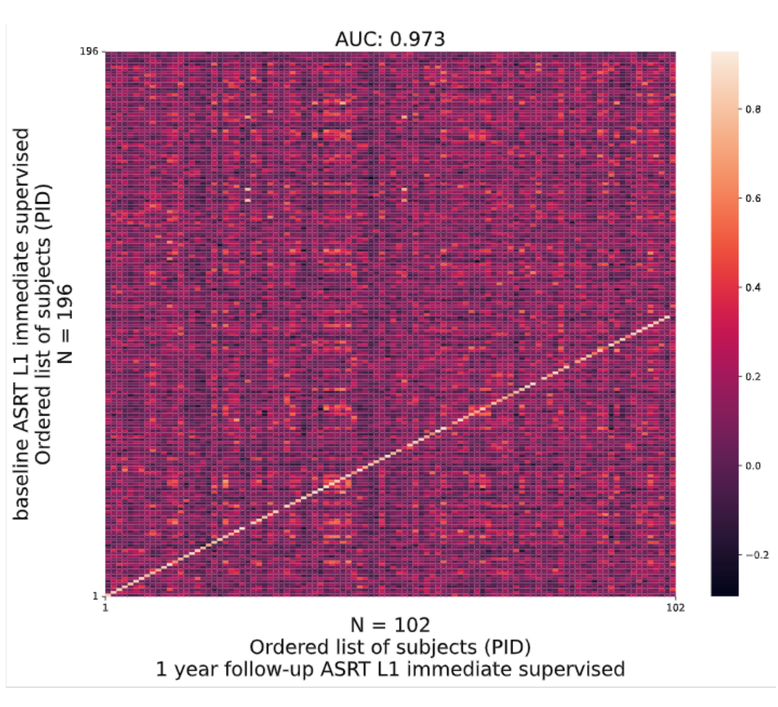


Figure 2: Heatmap visualisation of speaker verification across the same task (ASRT L1 immediate recall), in the same setting, one year apart. The lighter colours denote greater similarity between the associated speaker tuples, with this representing the same participant on the diagonal.

CONCLUSION

The speaker verification system is effective for confirming participant identity directly from audio verbal data, and is robust to changes in tasks, environments, and devices. The system shows high performance longitudinally, even in participants with a progressive neurodegenerative condition.

Audio recordings and speaker verification technologies would assist data collection in decentralized and hybrid clinical trials, ensuring the identity of the trial participant and confirming they completed the measures independently, as well as increasing confidence in collected data. Finally, any issues detected by the system would automatically trigger additional review and adjudication to ensure minimization of error in all collected trial data.

REFERENCES

1. Van Patten (2021). *J Clin Exp Neuropsychol*, 43(8), 767–73.
2. Resnik & McCann (2015). *NEJM,* 373(13), 1192–3
3. Fristed et al. (2022). *Brain Commun,* 4(5), fcac231
4. Skirrow et al. (2022) *JMIR Aging,* 5(3), e37090

DISCLOSURES/CONTACT

One or more authors report potential conflicts which are described in the program.

Novoic

udeepa@novoic.com
www.novoic.com