Factors Affecting the Performance of Automated Speaker Verification in Alzheimer's Disease Clinical Trials

Malikeh Ehghaghi, Marija Stanojevic, Ali Akram, Jekaterina Novikova
Winterlight Labs, Toronto, ON, Canada
{malikeh, marija, aliakram, jekaterina}@winterlightlabs.com

1.Background

Detecting duplicate patient participation in clinical trials is a major challenge because repeated patients can undermine the credibility and accuracy of the trial's findings and result in significant health and financial risks. Developing accurate automated speaker verification (ASV) models is crucial to verify the identity of enrolled individuals and remove duplicates. However, there has been limited investigation into the factors that can affect ASV capabilities in clinical environments. In this work, we bridge the gap by conducting analysis of how participant demographic characteristics, audio quality criteria, and severity level of Alzheimer's disease (AD) impact the performance of ASV in clinical trials.

2.Methods

- Used the Alzheimer's Disease Clinical Trial (ADCT) dataset comprising 7,084 speech recordings of 659 English-speaking participants.
- Manually transcribed audio using CHAT[1] protocol and rated the quality of the recordings according to different quality criteria including background noise, clinician interference, participant accent, and participant clarity.
- Assessed the severity level of AD using the Mini-Mental State Examination[2] (MMSE) rating scale and categorized the participants into four levels of AD severity: Healthy Control (HC) (MMSE score > 26 points), Mild AD (MMSE score 21-26 points), Moderate AD (MMSE score 15-20 points), and Severe AD (MMSE score < 15 points).
- Collected the dataset every 12 weeks for a 48-week treatment period with recordings of participants performing a set of self-administered speech tasks, including picture description, phonemic verbal fluency, and semantic verbal fluency.
- Dataset Composition: Male: 43.4%, Female: 56.6%, Age range: 55-80, Average age: 69.7±6.7
- Utilized the TitaNet[3] model, which is a state-of-the-art end-to-end text-independent (TI) ASV model from the Nvidia NeMo toolkit, that had been pre-trained on an extensive collection of English speech data.
- Evaluated the performance of the TitaNet model on subsets of ADCT data based on genders, age groups, audio quality levels, and AD severity levels.
- Generated embeddings for audio files within each group and created positive and negative pairs of embeddings.
- Computed cosine similarity between the pairs of vector embeddings, adjusted a threshold value for each group to achieve equal true positive and true negative rates, and calculated equal error rate (EER).
- Considered pairs with cosine similarity above the threshold as belonging to the same speaker and pairs below the threshold as representing different speakers.

3.Results

Our results indicate that ASV performance:

- is slightly better on male speakers than on female speakers.
- Degrades with age.
- is comparatively better for non-native English speakers than for native English speakers.
- is negatively affected by clinician interference, noisy background, and unclear participant speech.
- tends to decrease with an increase in the severity level of AD.

Figure 1: ASV Performance across Different AD Severity Levels

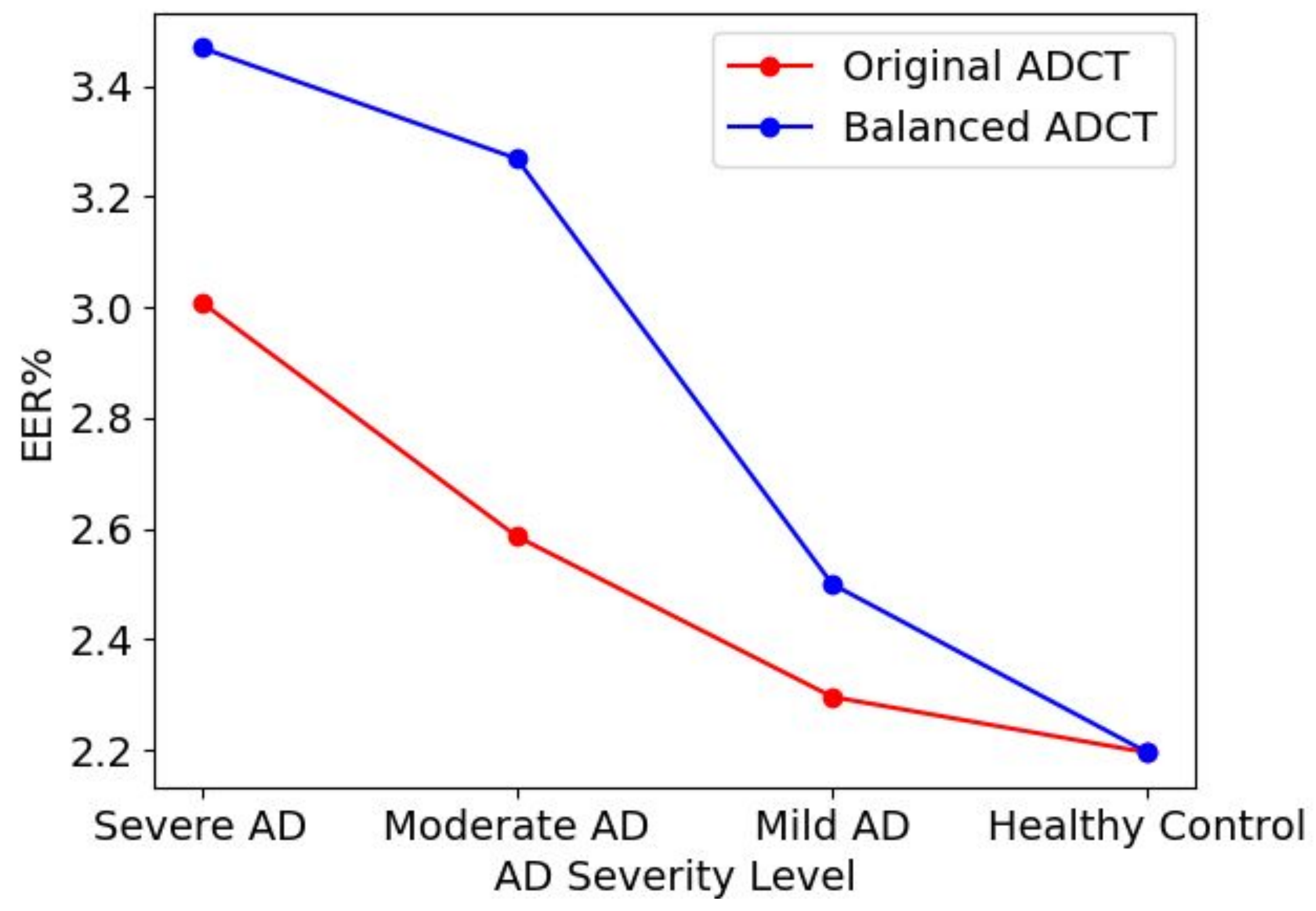
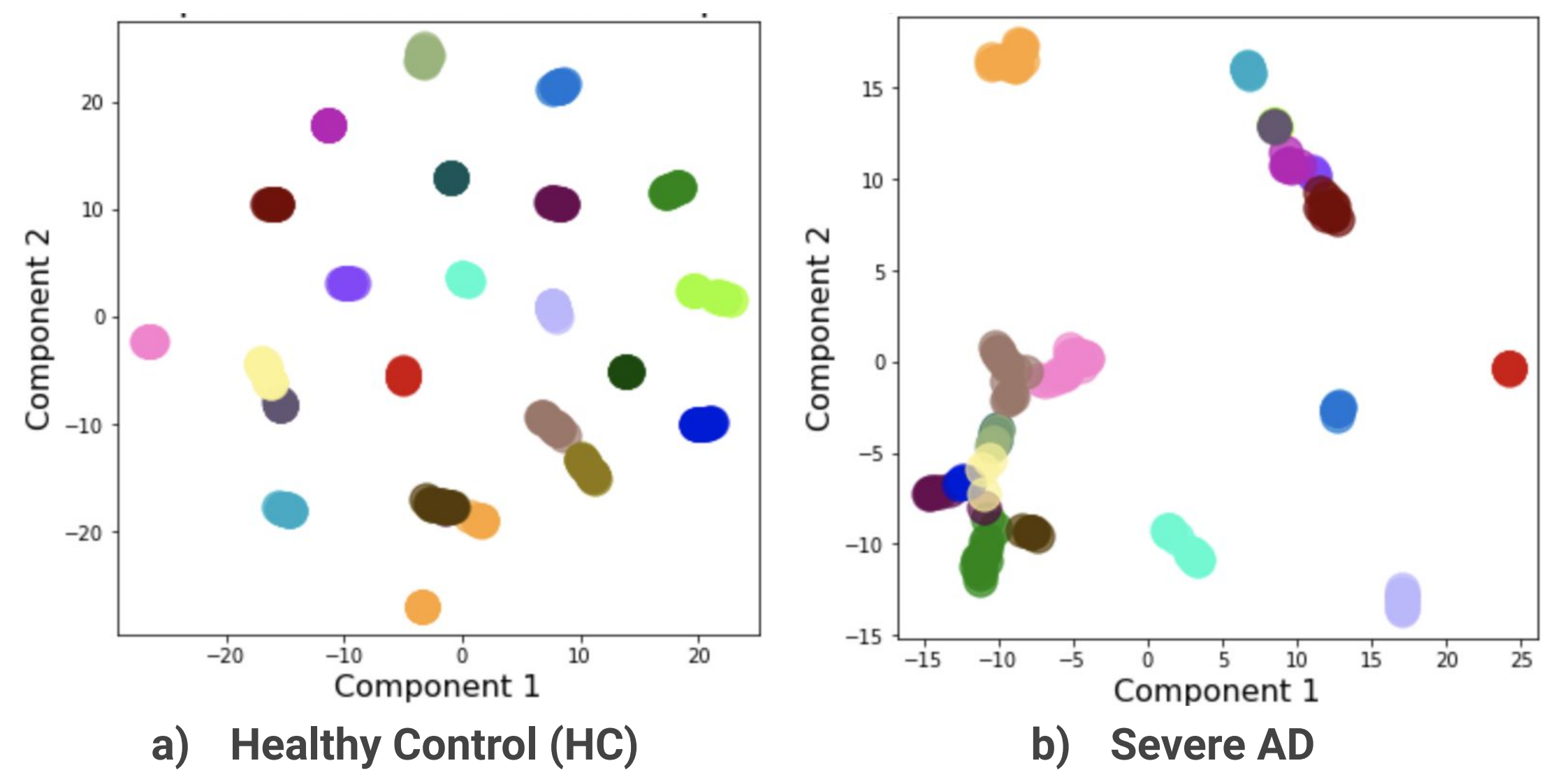


Figure 2: Speaker cluster visualizations of HC and Severe AD Groups



a) Healthy Control (HC)   b) Severe AD

4.Conclusion

- We explored the influence of participant demographic characteristics, audio quality, and AD severity level on ASV performance in a clinical trial.
- Our results suggest that variations in ASV performance can be attributed to inherent voice characteristics of different subgroups.
- Our results emphasize the need to reassess the ASV technology to mitigate biases towards certain subgroups and ensure fairness.
- Our results highlight the importance of quality assurance for speech recordings during trials.

5.References

(1) Brian MacWhinney. 2014. The CHILDES project: Tools for analyzing talk, Volume II: The database.Psychology Press.

(2) Carsten Henneges, Catherine Reed, Yun-Fei Chen, Grazia Dell'Agnello, and Jeremie Lebrec. 2016. Describing the sequence of cognitive decline in alzheimer's disease patients: results from an observational study. Journal of Alzheimer's Disease, 52(3):1065–1080.

(3) Nithin Rao Koluguri, Taejin Park, and Boris Ginsburg. 2022. Titanet: Neural model for speaker representation with 1d depth-wise separable convolutions and global context. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 8102–8106. IEEE.

Table 1: Gender-level Analysis of ASV Performance

Gender	EER(%)	#Spkrs	#Smpls	Avg #Smpls per Spkr	Avg Age	Avg MMSE Score
Female	5.13	170	2,735	16.09±3.94	69.53±6.72	17.33±4.37
Male	**4.98**	170	2,671	15.72±4.02	69.41±6.96	17.45±4.45

Table 2 : Age-level Analysis of ASV Performance

Age	EER(%)	#Spkrs	#Smpls	Avg #Smpls per Spkr	Gender	Avg MMSE Score
Age <= 70	**3.62**	197	3,235	16.42±3.86	M + F	17.09±4.72
Age > 70	4.20	195	3,022	15.50±4.07	M + F	17.57±4.11

Table 3 : Quality-level Analysis of ASV Performance

Audio Quality Criterion	EER(%)	#Spkrs	#Smpls	Avg #Smpls per Spkr	Gender	Avg Age	Avg MMSE Score
Background Noise - No Issue	**2.90**	125	426	3.40±1.45	M + F	69.60±6.72	16.94±5.83
Background Noise - Minor to Major Issue	3.54	125	511	4.08±2.08	M + F	69.21±6.45	16.78±5.58
Participant Clarity - No Issue	**2.85**	112	481	4.29±1.70	M + F	69.80±6.38	16.81±5.62
Participant Clarity - Minor to Major Issue	3.41	112	432	3.85±1.83	M + F	69.23±6.81	16.04±5.54
Clinician Interference - No Issue	**2.90**	103	659	4.30±2.08	M + F	69.40±6.77	17.65±5.61
Clinician Interference - Minor to Major Issue	3.38	103	399	3.87±1.86	M + F	69.43±6.84	14.77±5.22
Participant Accent - Native	2.97	188	901	4.79±2.82	M + F	68.63±6.89	17.22±5.01
Participant Accent - Non-Native	**2.01**	188	594	3.16±1.54	M + F	70.45±6.32	17.19±4.56
All	3.10	659	7,084	10.70±7.00	M + F	69.55±6.75	17.32±4.44

Presented at the Fifth Clinical Natural Language Processing (ClinicalNLP) Workshop at ACL 2023