

Digits-In-Noise Hearing Test using Text-To-Speech and Speech Recognition: Proof-of-Concept Study

M. Fatehifar, KJ. Munro, M. Stone, D. Wong, T. Cootes, J. Schlittenlacher
ManCAD, School of Health Science, University of Manchester, UK
Mohsen.fatehifar@manchester.ac.uk



Introduction

- Access to hearing tests is not always available.
- Remote hearing tests can make testing more accessible.
- Automatic speech recognition (ASR) can help create self-supervised and remote tests.
- Text-to-speech can reduce the cost of creating new stimuli.
- A digits-in-noise (DIN) test was developed that uses TTS to synthesise stimuli and ASR to capture participants' responses.

Research Questions

- Can the developed digits-in-noise (DIN) test be used to measure the speech reception threshold when no AI is involved?
- Is a DIN test with text-to-speech (TTS) and automatic speech recognition (ASR) valid and reliable?

Method

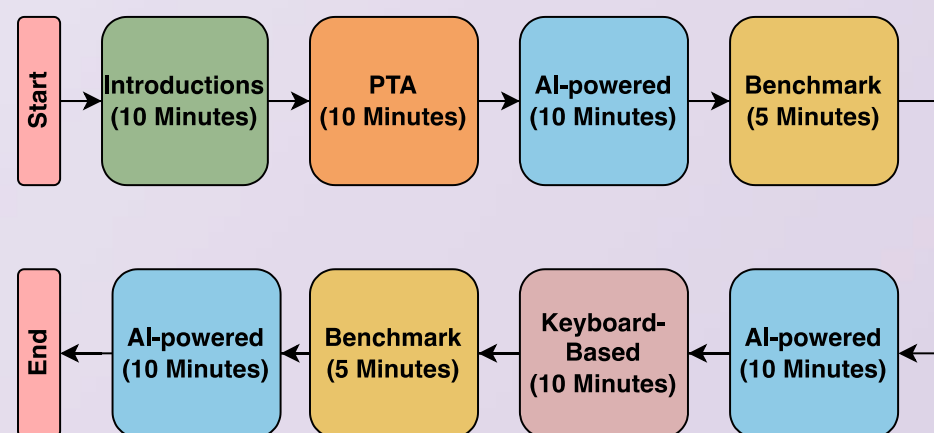


Fig 1. Steps in the testing session

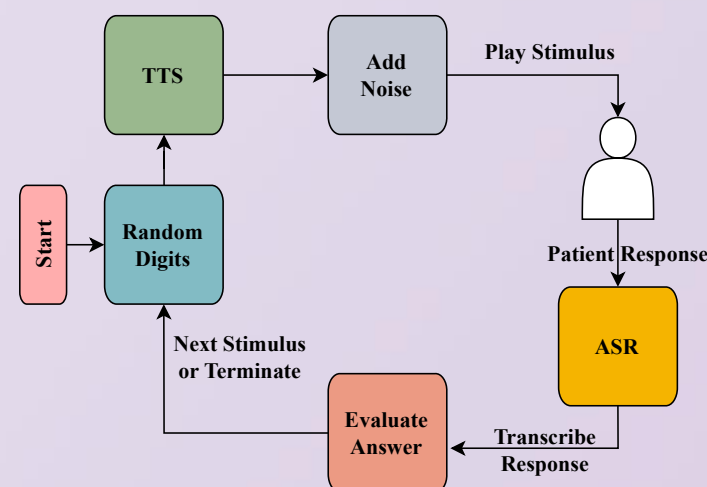


Fig 2. Flow of the AI-powered test

Participants

31 participants were recruited. Among them, 10 were hearing impaired, with an average PTA threshold of 36 ± 17 dB HL in their better ear for 0.5 to 4 kHz frequencies, and 21 were normal hearing (PTA <20 dB HL).

Equipment

Testing was carried out in a double-walled sound-proof booth. Stimuli were delivered diotically using a Scarlett Focusrite Solo 4th gen audio interface and Sennheiser HD650 headphones. Hearing thresholds were measured using a GSI Pello audiometer.

Tests

Three triple DIN tests with 1-up, 2-down staircase procedures are involved in this study. :

Benchmark test: A DIN test that used pre-recorded stimuli. This test was developed by a separate group of researchers at the University of Manchester. In this test, both noise and stimuli levels were adopted for constant overall level.

Keyboard-based test: This test used the same pre-recorded voice stimuli as the benchmark test. Its purpose was to check the developed software by comparing its results with those of the benchmark test. This test keeps the stimuli at 65 dB and adopts the noise level based on participants' responses.

AI-powered test: This test used the developed software to control the test, but instead of pre-recorded stimuli, it used TTS to generate the stimulus and ASR to capture participants' responses. This test also keeps the stimuli at 65 dB and adopts the noise level based on participants' responses. See Fig 2.

Removing Outliers

ASR was not able to accurately transcribe responses for participants with strong accents. To mitigate this and to see the potential of the AI-powered DIN test, participants with an ASR error rate of more than 40% were removed from the results. Doing this, removed 2 participants when measuring validity and 7 participants when measuring the reliability of the AI-powered test.

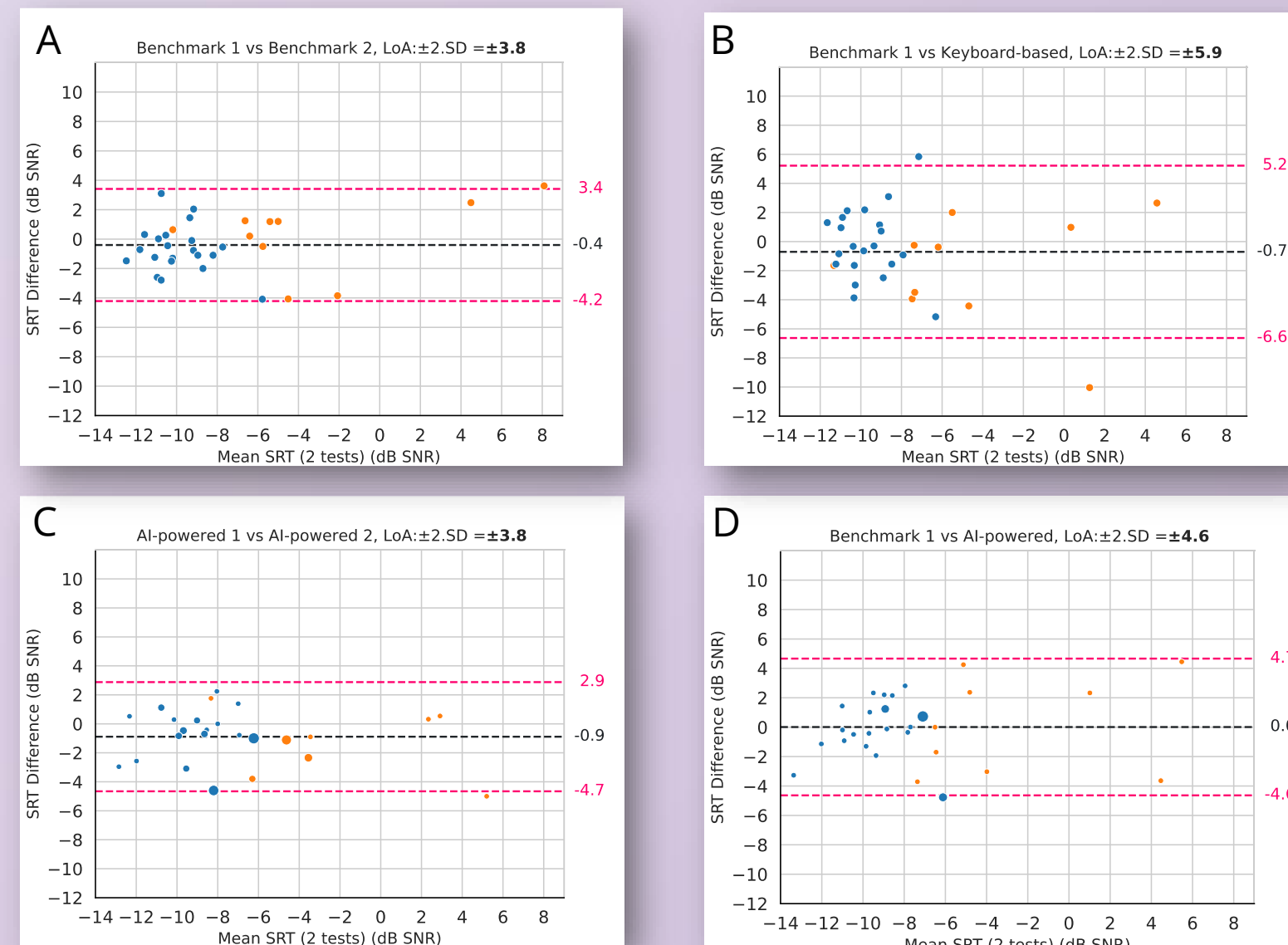


Fig 3. Bland-Altman plots showing validity and reliability

Evaluation

Validity: For each person, the keyboard-based run and the AI-powered run with the lowest ASR error rate were compared to the first run of the benchmark test.

Reliability: The results of the two runs of the benchmark test were compared to each other, and the first two runs of the AI-powered test were compared to each other.

Results

Keyboard-based test: Results of keyboard-based and the benchmark test have limits of agreement of ± 5.9 (panel B), while it is ± 3.8 (panel A) for two runs of the benchmark test.

AI-powered test: Results of AI-powered and the benchmark test have limits of agreement of ± 4.6 (panel D) and two runs of the AI-powered test have limits of agreement of ± 3.8 (panel A).

This shows that the results of both keyboard-based and AI-powered tests are aligned with the benchmark test.

Summary

- The developed software is working as expected and the results are in accordance with the benchmark test.
- Addition of TTS and ASR does not have a negative impact on the results of the DIN test when the ASR error rate was not exceptionally high. In fact, AI-powered test is closer to the benchmark test compared to the keyboard-based test.
- The promising results suggest that TTS and ASR can also be used for more complex speech-in-noise tests (e.g. sentences-in-noise).

Acknowledgments

This work is supported by the MRC-DTP, with grant reference number MR/W007428/1.

KJ. Munro is supported by the NIHR Manchester Biomedical Research Centre.