00072

SCREENING AND DIAGNOSTICS

Application of Text-to-Speech Synthesis and Automatic Speech Recognition in speech audiometry

Inga Holube, Saskia Ibelings, Theresa Nuesse

Objectives

Text-to-speech synthesis systems (TTS systems) and automatic speech recognition (ASR) were investigated with regard to their applicability in speech audiometry.

Conclusions

- TTS systems reduce the development effort for new speech tests.
- ASR can be used to analyze verbal responses in speech tests and enables advanced analysis.

Text-to-Speech (TTS) Synthesis

- The use of TTS systems can simplify the development of speech tests by saving time for recordings and post-production.
- Behavioral evaluation of the synthesized speech material is still necessary.
- The measurement results of speech tests with synthesized speech material are comparable to those with natural speakers.
- Differences are of the same order of magnitude as between different natural speakers.

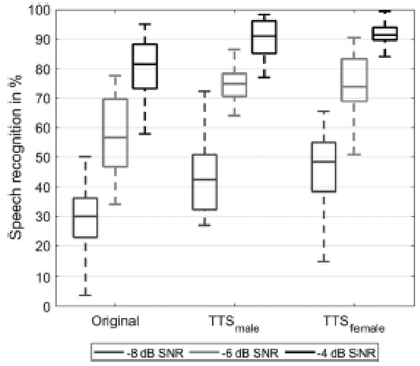

Figure 1 (left):
Speech-recognition scores for meaningful sentences
(Ibelings et al., 2022)

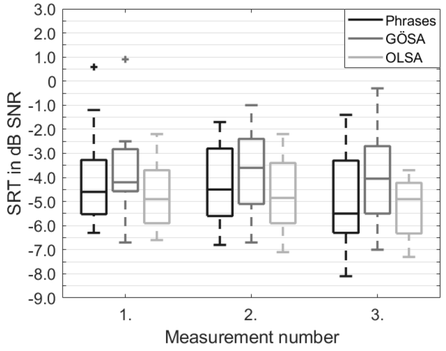

Figure 2 (right):
Comparison of phrases with sentence tests
(Ibelings et al., 2024)

Automatic Speech Recognition (ASR)

- The analysis of the participants' responses by ASR led to a high correlation with speech recognition scores rated by a human examiner.
- By analyzing the temporal structure of the participants' responses, a significant correlation between the verbal response times (VRT) and the self-reported listening effort could be demonstrated.

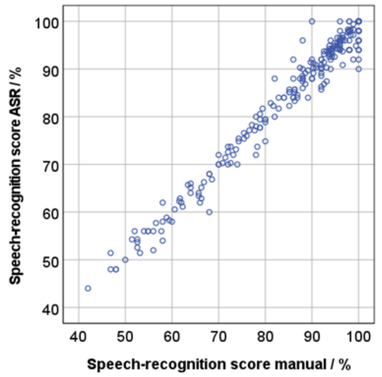

Figure 3 (left):
ASR and manual speech-recognition scores
(Holube et al., in print)

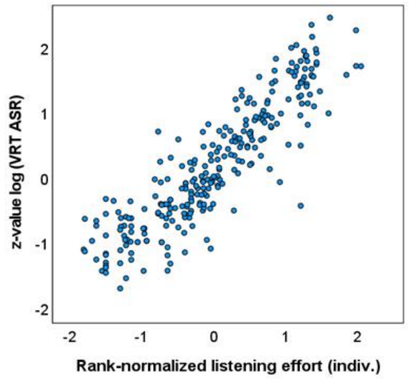

Figure 4 (right):
Verbal response time and self-reported listening effort
(Holube et al., in print)

Methods

- Listeners with and without hearing impairment, speech tests in noise
- Synthesis of speech materials of two existing German speech tests (matrix sentences of the OLSA and everyday sentences of the GÖSA)
- Development of a new synthesized speech test with 4-word phrases
- Analysis of audio recordings of the participant’s responses with ASR and comparison with the results of a human examiner
- Examination of the verbal response time between end of stimulus and begin of the participants’ responses).

References

- Holube I, Taesler S, Ibelings S, Hansen M, Ooster J (in print) Automated measurement of speech recognition, reaction time, and speech rate and their relation to self-reported listening effort for normal-hearing and hearing-impaired listeners using various maskers. Accepted for publication by Trends in Hearing
- Ibelings S, Brand T, Holube I (2022) Speech recognition and listening effort of meaningful sentences using synthetic speech. Trends in Hearing 26, 1-14.
- Ibelings S, Brand T, Ruigendijk E, Holube I (2024) Development of a phrase-based speech-recogntion test using synthetic speech. Trends in Hearing 28, 1-13.
- Nuesse T, Wiercinski B, Brand T, Holube I (2019) Measuring speech recognition with a matrix test using synthetic speech. Trends in Hearing 23, 1-14.

Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, Oldenburg, Germany. Contact: Inga.Holube@jade-hs.de

19-22 September 2024
Paris, France
CNIT Paris La Défense