



**TECHNISCHE
HOCHSCHULE
LÜBECK**

Bachelor's Thesis:

**Development and Modeling of a Speech Intelligibility
Test in a Simulated Everyday Environment Using
Virtual Acoustics and Synthetic Speech**

Jonas Stohlmann

First Examiner: Prof. Dr. rer. nat. Tim Jürgens
Second Examiner: M.Sc. Fabian Hettler
Department: Angewandte Naturwissenschaften
Degree Program: Hörakustik B.Sc.
Submission Date: September 29, 2025

I a Abstract

Although numerous speech intelligibility tests exist, the need for speech audiometric tests successfully simulating acoustic everyday environments still persists. In this study, a speech intelligibility test is developed using a dialog between a feminine and a masculine synthetic voice. The new test consists of ten key sentences embedded between carrier phrases; each key sentence contains two disyllabic keywords with ten variants each. The test is modeled as a matrix test using FADE and, for comparison, measured with 24 normal-hearing participants, 4 of whom formed a pilot study. Six conditions – two fixed SNRs, two voice providers (acapela and Google), and test-retest – were tested. The ambient noise was a 4th order HOA signal of a food court. For presentation, a multi-channel loudspeaker system was used. The FADE models exhibited ceiling effects at $\approx 70\%$ speech intelligibility. FADE's predicted curves possessed higher SRTs and shallower slopes ($SRT_{\text{acapela}} = -6.1$ dB, $m_{\text{acapela}} = 5.1 \frac{\%}{\text{dB}}$; $SRT_{\text{Google}} = 0.7$ dB, $m_{\text{Google}} = 3.6 \frac{\%}{\text{dB}}$) than the measurements with human listeners ($SRT_{\text{acapela}} = -9.1$ dB, $m_{\text{acapela}} = 15.2 \frac{\%}{\text{dB}}$; $SRT_{\text{Google}} = -9.5$ dB, $m_{\text{Google}} = 9.7 \frac{\%}{\text{dB}}$). The FADE models require improvement through optimization of modeling parameters to the speech material. For normal-hearing listeners, the measured psychometric functions can be used as reference curves. The acapela voices were rated less natural than Google voices ($p < 0.01$) and exhibited learning effects between test and retest ($p < 0.001$) despite one training run; therefore, the Google voices are better suited for the new test. Currently, the new test does not meet all criteria set by the ISO 8253-3:2022.

Keywords: Ambisonics, ecological validity, FADE, speech audiometry, speech intelligibility

I b Zusammenfassung

Obwohl zahlreiche Sprachverständlichkeitstests existieren, werden weiterhin sprachaudiometrische Tests benötigt, die erfolgreich akustische Alltagssimulationen simulieren. In dieser Studie wird ein Sprachverständlichkeitstest als Dialog zwischen einer maskulinen und einer femininen synthetischen Stimme entwickelt. Der neue Test besteht aus zehn Schlüsselsätzen, eingebaut zwischen Ankündigungssätzen; jeder Schlüsselsatz beinhaltet zwei zweisilbige Schlüsselwörter mit je zehn Varianten. Der Test wird mittels FADE als Matrixtest modelliert und zum Vergleich mit 24 normalhörenden Proband:innen gemessen, davon 4 als Pilotstudie. Sechs Konditionen – zwei feste SNRs, zwei Stimmanbieter (acapela und Google), und Test-Retest – wurden untersucht. Das Hintergrundgeräusch war ein HOA-Signal 4. Ordnung von einem Essbereich. Für die Wiedergabe wurde ein Multikanal-Lautsprechersystem verwendet. Die FADE-Modelle weisen Deckeneffekte bei $\approx 70\%$ Sprachverständlichkeit auf. Die von FADE vorhergesagten Kurven besitzen höhere SRTs und geringere Steigungen ($SRT_{\text{acapela}} = -6,1$ dB, $m_{\text{acapela}} = 5,1 \frac{\%}{\text{dB}}$; $SRT_{\text{Google}} = 0,7$ dB, $m_{\text{Google}} = 3,6 \frac{\%}{\text{dB}}$) als die Messungen mit Proband:innen ($SRT_{\text{acapela}} = -9,1$ dB, $m_{\text{acapela}} = 15,2 \frac{\%}{\text{dB}}$; $SRT_{\text{Google}} = -9,5$ dB, $m_{\text{Google}} = 9,7 \frac{\%}{\text{dB}}$). Die FADE-Modelle benötigen weitere Optimierung der Modellierungsparameter an das Sprachmaterial. Für Normalhörende können die gemessenen psychometrischen Funktionen als Referenzkurven verwendet werden. Die acapela-Stimmen wurden als weniger natürlich als Google bewertet ($p < 0.01$) und zeigten trotz einer Trainingsmessung Lerneffekte zwischen Test und Retest ($p < 0.001$); folglich sind die Google-Stimmen besser für den neuen Test geeignet. Aktuell erfüllt der neue Test nicht alle Kriterien der ISO 8253-3:2022.

Schlüsselwörter: Ambisonics, ökologische Validität, FADE, Sprachaudiometrie, Sprachverständlichkeit