Perceptual equivalence of test lists in a monosyllabic speech test with different hearing thresholds

ALEXANDRA WINKLER[1,3,*], BRITTA JENSEN[1,3], THOMAS BRAND[2,3], AND INGA HOLUBE[1,3]

[1] *Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, Oldenburg, Germany*

[2] *Medizinische Physik, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany*

[3] *Cluster of Excellence "Hearing4All"*

EN ISO 8253-3 (2012) describes the requirements and validation of speech material for speech audiometry. Although speech tests are typically applied to listeners with hearing impairment, the validation is conducted with listeners with normal hearing abilities. The aim of this study was to determine the effect of hearing thresholds on the validation results. Since hearing thresholds of listeners with hearing impairment show a large variability, groups of participants with normal hearing listened to the Freiburg monosyllabic speech test (Hahlbrock, 1953) preprocessed with two simulated homogenous hearing losses, as well as to the original speech material. Discrimination functions were fitted to the results and speech levels for speech recognition scores of 50% were determined. According to EN ISO 8253-3 (2012), the perceptual balance of the lists is given when the confidence interval of the speech levels is within 1 dB from the median across all lists. This criterion is not fulfilled for several test lists, which partly differed for the hearing-loss configurations. When taking the measurement accuracy of the experiment into account, consistent deviations are observed in four test lists. The results suggest that if perceptual balance is fulfilled for participants with normal hearing, this might not be valid for participants with hearing impairment. Predictions of speech recognition using the Speech Intelligibility Index could not replicate test list differences.

INTRODUCTION

The German monosyllabic speech test (Hahlbrock, 1953) is a standard test in hearing diagnostics and in the validation of hearing aid fitting. This test consists of 20 lists with 20 monosyllables. For comparison of different settings and/or hearing aids, speech material should be perceptually and phonemically balanced. EN ISO 8253-3 "specifies requirements for the composition, validation and evaluation of speech test materials" used in speech audiometry (EN ISO 8253-3, 2012, p. 1) for

*Corresponding author: alexandra.winkler@jade-hs.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

listeners with normal hearing. Nevertheless, speech audiometry is usually applied to listeners with hearing impairment, who show a large variety in hearing thresholds. Therefore, simulated hearing losses were used in this study to reduce the variability and to be able to resolve test list differences.

PARTICIPANTS

The requirements for the listeners according to EN ISO 8253-3 (2012) regarding age and hearing threshold were:

- Age: 18 - 25 years
- Hearing threshold ≤ 10 dB HL between 0.25-8 kHz and ≤ 15 dB HL at maximum two frequencies

These requirements were fulfilled by all listeners. In total, 120 listeners (80 female, 40 male, median age 23 years) participated in this study. The participants were separated into three groups of 40 listeners each (NH: 31 ♀, 9 ♂, SIM A: 25 ♀, 15 ♂ and SIM B: 24 ♀, 16 ♂). Median hearing thresholds for NH, SIM A, and SIM B are shown in Fig. 1. For group SIM B, discomfort levels were measured in addition to the pure tone thresholds. The discomfort level for this listener group had to be at least 90 dB HL at 500 Hz and 1 kHz to avoid too loud levels for the processed stimuli which were presented via headphones during the test.

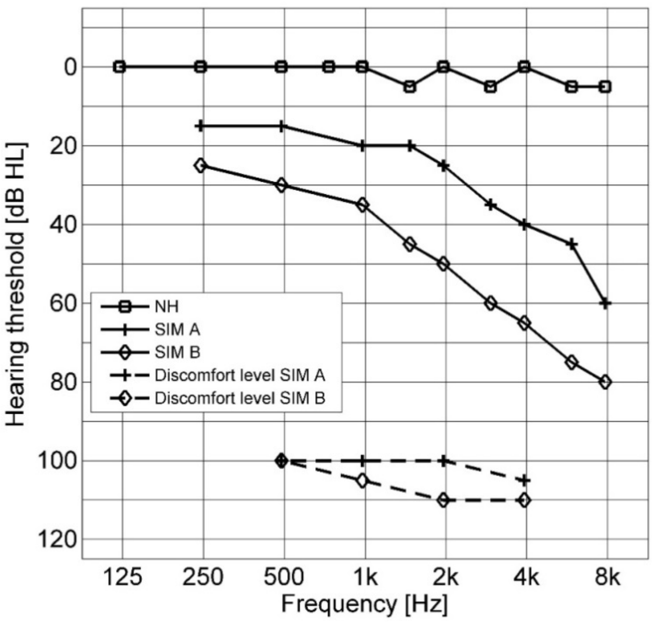

Fig. 1: Median hearing thresholds for NH (squares) and simulated hearing thresholds with uncomfortable loudness levels for SIM A (crosses) and SIM B (diamonds).

In this study the software SIM PRO (HörTech gGmbH) was used to simulate the hearing thresholds and to filter the original speech material. SIM PRO is based on the Master Hearing Aid (Grimm *et al.*, 2006). Multichannel dynamic signal processing and spectral smearing modify test signals using expansion instead of compression, to simulate different hearing losses. For the purpose of presenting sounds to a participant, the original speech material is filtered according to a pure tone audiogram and a discomfort level. The simulated hearing loss SIM A (see Fig. 1) was based on average hearing thresholds of a population aged 65 to 74 years from von Gablenz and Holube (2015). The discomfort levels for SIM A were specified according to Pascoe (1988). For SIM B (see Fig. 1), customer data acquired by German hearing aid acousticians (Nüsse *et al.*, 2014) were used. Those data included hearing thresholds as well as levels of discomfort and were selected to meet speech recognition scores between 30 and 80% for standard diagnostic test levels. All test lists for SIM A and B were processed separately, depending on the presentation level.

TEST SIGNALS

Freiburg monosyllables (recordings from 1969) according to DIN 45626-1 (1995a) and DIN 45621-1 (1995b) were presented monaurally via headphones (Sennheiser HDA200). The levels are overall sound pressure levels (SPL) measured in an ear simulator. All 20 lists were tested with all participants, five lists at four levels each:

- NH: Original speech material presented at 17.5 dB SPL, 23.5 dB SPL, 29.5 dB SPL, and 35.5 dB SPL.
- SIM A: Filtered speech material at 39.5 dB SPL, 45.5 dB SPL, 51.5 dB SPL, and 57.5 dB SPL.
- SIM B: Filtered speech material at 65 dB SPL, 80 dB SPL, 90 dB SPL, and 95 dB SPL.

SPEECH INTELLIGIBILTY INDEX (SII)

The SII (ANSI S3.5, 1997) estimates speech recognition based on the amount of speech contained in each frequency band. Hearing thresholds or different speech material can be used as input for the model. In this study, normal hearing ability was assumed and the band important function of the Northwestern University Auditory Test No. 6 (NU6-monosyllables) was chosen. The speech material for the SII prediction was the same as for the experiments.

RESULTS

Speech recognition curves

Based on the speech recognition scores for the different presentation levels, discrimination functions for NH and SIM A were fitted to the data of all 20 lists separately (Brand and Kollmeier, 2002). For SIM B, linear interpolation between data points above and below 50% speech recognition per list was used, because recognition scores were well below 100%, even for very high presentation levels.

Fig. 2 shows all 20 discrimination functions for NH, SIM A, and SIM B together with the expected data for SIM B. The expected data represent the range of the recognition scores of the customer data (Nüsse *et al.*, 2014) from whom the hearing thresholds and levels of discomfort were selected.

The fitted functions were used to estimate the level for 50% speech recognition (L_{50}). The median L_{50} was 25.2 dB for NH, 50.2 dB for SIM A, and 73.5 dB for SIM B. The range between the lowest and highest L_{50}, i.e. the easiest and the most difficult list, respectively, varied between 4.2 dB (NH), 7.8 dB (SIM A), and 16.5 dB (SIM B).

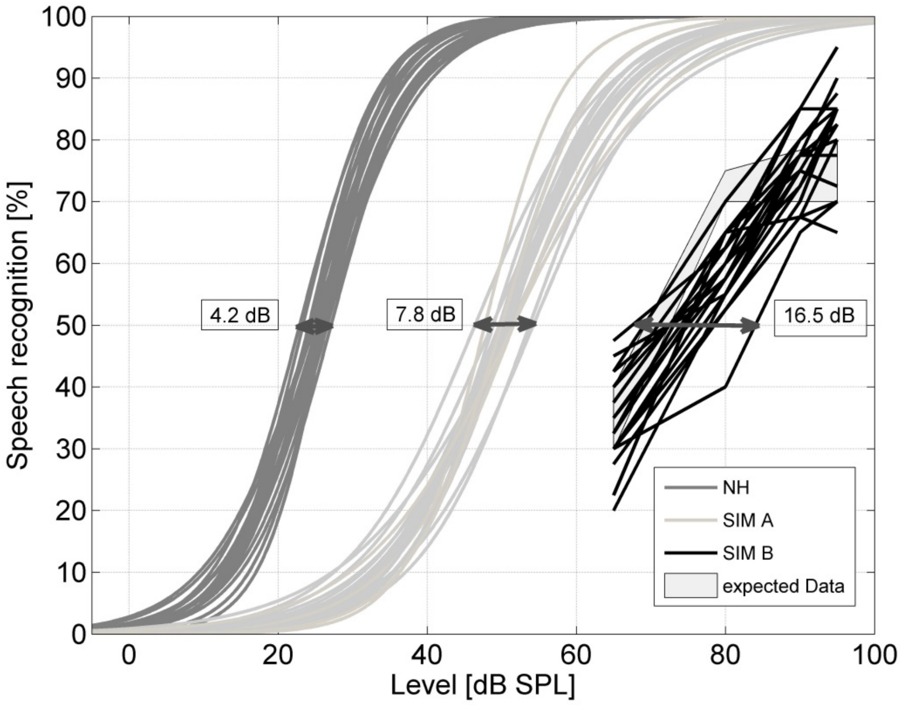

Fig. 2: Discrimination functions for all 20 lists: NH (dark grey), SIM A (gray), SIM B (black), and expected data in light gray. Ranges between the lowest and highest L_{50} are marked by arrows.

Measurement inaccuracy and perceptual balance across lists

Binomial distribution and Gaussian error propagation leads to an inaccuracy in the estimated L_{50} (σ in Fig. 3) according to equation 4 in Brand and Kollmeier (2002). For the calculation of σ for each list, speech recognition scores for one level below and one level above L_{50}, as well as the slope at L_{50} and the number of data points, were required. The number of data points was given by 20 words per list and ten listeners (per list) each. For every subject group, the highest value for σ of all 20 test lists was selected. This led to inaccuracies of from 1.4 to 7.5 dB and therefore of more than 1 dB as predefined in EN ISO 8253-3 (2012, clause 4.9). Nevertheless, several lists showed consistent deviations larger than σ for at least two subject groups (marked lists in Fig. 3).

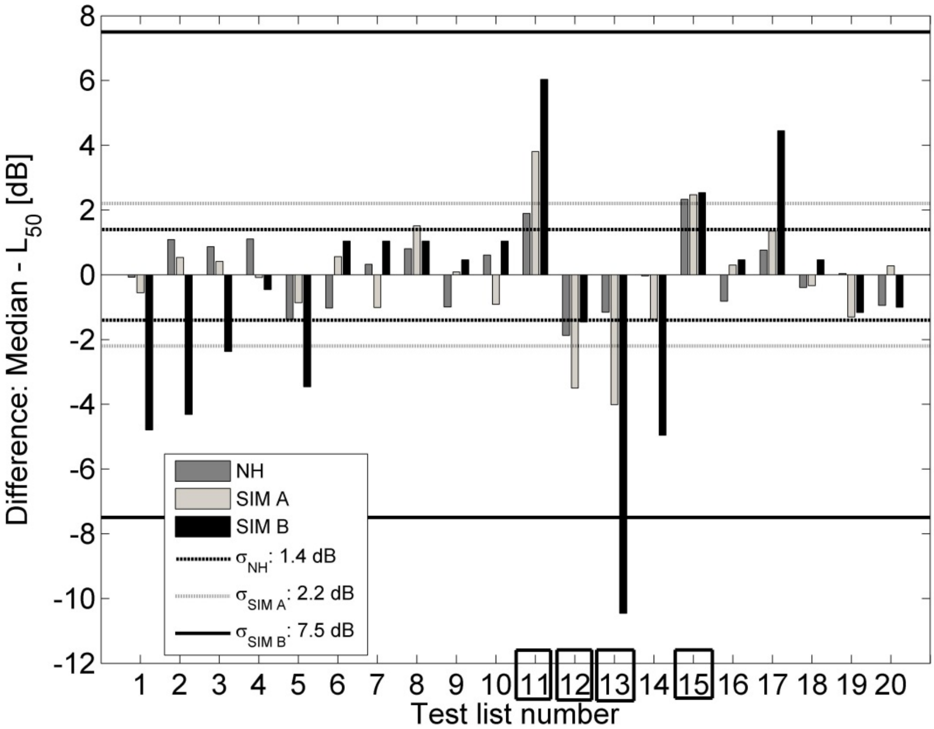

Fig. 3: Deviation of L_{50} from median L_{50} for all lists including inaccuracy (σ) of L_{50}. Marked test list numbers (11, 12, 13, and 15) are above σ for at least two groups of listeners.

Speech Intelligibility Index (SII)

The SII was applied to analyse whether the differences in L_{50} can be predicted by spectral deviations in the test lists for the three subject groups. Frequency analysis (third octave bands) of the lists revealed minor spectral differences for all thresholds and presentation levels. Fig. 4 shows an example for SIM A at 45.5 dB SPL presentation level.

The SII was calculated for each test list, subject group, and presentation level. The SII values were converted to predict speech recognition scores by using the average discrimination function of the original speech material for group NH. Then, linear interpolation was used to calculate the predicted L_{50} for every test list. These values were compared to the L_{50} estimation of the measured speech recognition scores. The correlation between estimated and predicted L_{50} are given in Table 1. The correlation coefficient for SIM A is larger than for the other groups, but none of the correlations are significant ($p > 0.05$).

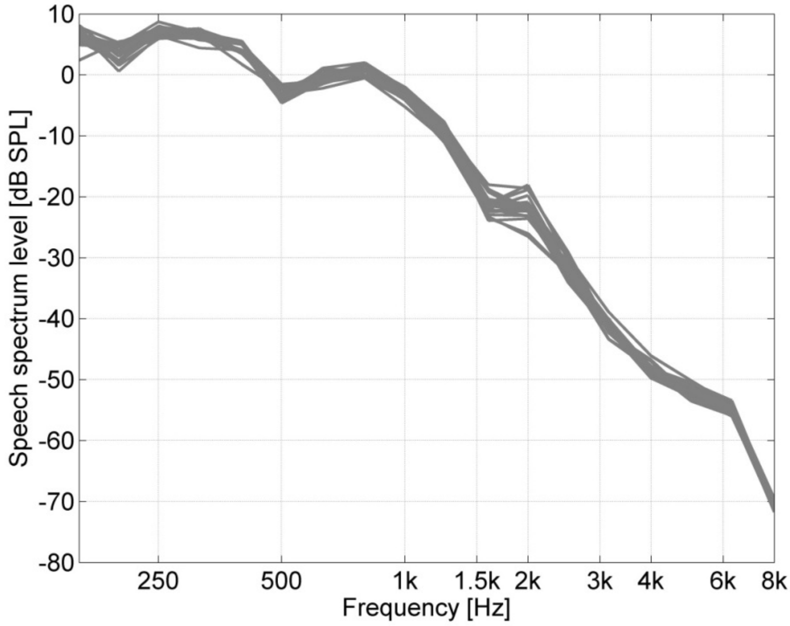

Fig. 4: Third octave spectra of 20 lists for SIM A at 45.5 dB SPL.

NH	SIM A	SIM B
r = 0.23 p = 0.335	r = 0.42 p = 0.066	r = 0.02 p = 0.428

Table 1: Pearson's correlation between estimated and predicted L_{50}.

SUMMARY AND DISCUSSION

In summary, the results of the measurements showed that the Freiburg monosyllables are not perceptually balanced for participants with normal hearing and hearing impairment. This conclusion was drawn although the hearing impairment was simulated and it is questionable whether the recognition scores are similar to those for "real" hearing impairments. Nevertheless, recognition scores for SIM B are well within the range of data from customers of hearing aid acousticians and the conclusion of perceptual imbalance is drawn not only from the results of one subject group.

The L_{50} of several test lists deviate by more than 1 dB from the median values as defined by EN ISO 8253-3 (2012). On the other hand, the shallow slopes of the discrimination functions led to a measurement inaccuracy of up to 7.5 dB. Even for NH, the measurement inaccuracy was calculated to be up to 1.4 dB, which could only be improved to 1 dB by increasing the number of participants from 40 to 80.

Despite this measurement inaccuracy, lists 11, 12, 13, and 15 still deviate noticeably for at least two groups of listeners each and should be avoided in future applications of the Freiburg monosyllabic speech test.

To further analyze the test list deviations, the SII was used as an objective measure based on the spectrum of the speech material. Unfortunately, none of the exceptional lists (11, 12, 13, and 15) differ or shows larger variation compared to the other lists of the test (s. Fig. 4). Hence, the predicted L_{50} of the different lists within one group was very similar. Therefore, there seems to be no direct relation between the measurement and the prediction, even though there was a tendency towards correlation of predicted and measured L_{50} for the mild hearing loss (SIM A).

Other approaches to explain observed test list deviations might be, for example, a possible phonemic imbalance of the test lists – even though the phoneme distribution was taken into account by Hahlbrock (1953) – or possible differences in word popularity or knowledge. These criteria will need further examination.

ACKNOWLEGDEMENTS

We thank Daniel Berg (HörTech gGmbH), Lena Haverkamp, Sascha Bilert for support and data collection. This project was funded by Ph.D. program Jade2Pro. English language support by www.stels-ol.de.

REFERENCES

Deutsches Institut für Normung **(1995a)**. "DIN 45626-1; Tonträger mit Sprache für Gehörprüfung, Teil 1: Tonträger mit Wörtern nach DIN 45621-1," Beuth Verlag, Berlin.

Deutsches Institut für Normung **(1995b)**. "DIN 45621-1; Sprache für Gehörprüfung. Teil 1: Ein- und mehrsilbige Wörter," Beuth Verlag, Berlin.

American National Standards Institute **(1997)**. "ANSI S3.5-1997; Methods for calculation of the Speech Intelligibility Index (SII)," Acoustical Society of America, New York.

Brand T. and Kollmeier B. **(2002).** "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am., **111**, 2801-2810.

European Committee for Standardization **(2012)**. "EN ISO 8253-3; Acoustics – Audiometric test methods – Part 3: Speech Audiometry," Beuth Verlag, Berlin.

Grimm G., Herzke T., Berg D., and Hohmann V. **(2006)**. "The master hearing aid: a pc-based platform for algorithm development and evaluation," Acta Acust. United Ac., **92**, 618-628.

Hahlbrock K.H. **(1953)**. "Über Sprachaudiometrie und neue Wörterteste", ArchivOhr-usw. Heilk. u. Z.Hals- usw. Heilk., Bd. **162**, 394-431.

HörTech gGmbH: http://www.hoertech.de (accessed on 31st of July 2015).

Nüsse T., von Gablenz P., Holube I., and Bonsel H. **(2014)**. "Analyse tonaudiometrischer Daten von Kunden einer Hörgeräteakustikerkette," Hörakustik, **10**, 8-13.

Pascoe, D.P. **(1988)**. “Clinical measurements of the auditory dynamic range and their relation to formulas for hearing aid gain,” in Proc. 13th Danavox symposium: *Hearing Aid Fitting: Theoretical And Practical Views*. Ed. J. Jensen. Copenhagen, Denmark.

von Gablenz, P. and Holube, I. (**2015**). “Hearing threshold distribution and effect of screening in a population-based German sample,” Int. J. Audiol., 1-16, doi: 10.3109/14992027.2015.1084054