Speech intelligibility and loudness perception with the trueLOUDNESS fitting rule

F. Kramer[1], M. R. Schädler[1], V. Hohmann[1], D. Oetting[2], A. Warzybok[1]

[1] *Department of Medical Physics and Acoustics, Carl von Ossietzky University, Oldenburg, Germany*

[2] *HörTech gGmbH, Oldenburg, Germany*

Introduction

Two of the most common reasons for hearing aid users to reject wearing hearing aids are insufficient speech intelligibility (SI) and loudness perception using hearing aids (Jenstad et al., 2003). Especially in noisy conditions, for example in a typical cocktail party environment with background noises, hearing aid users often perceive the situation as too loud and effortful. Therefore, the normalization of loudness perception while also restoring speech intelligibility in everyday life is of great importance.

Here, the performance of the trueLOUDNESS fitting method (Oetting et al., 2018b) is investigated in realistic auditory scenes and laboratory test conditions concerning speech intelligibility and loudness perception. The widely used formula NAL-NL2 (Keidser et al., 2011) is used as reference.

Usually, the benefit from hearing aids is measured using speech intelligibility measurements in quiet with a fixed speech level. However, the outcomes of such measurements do not match the listeners' performance under acoustically challenging conditions including (non-) intelligible masking noise and/or reverberation (Bronkhorst, 2000). Some studies are restricted to the SI measurements in well controlled and rather simple acoustic condition with a single noise masker (Oetting et al., 2018b). Other studies investigated SI in more complex conditions including babble or cafeteria noise (Völker et al., 2015). Here, the assessment of the benefit from hearing aid processing is compared for simple well controlled laboratory conditions including stationary and modulated noises, and for realistic scenes represented by cafeteria ambience and a quiet natural environment.

For these realistic environments, dynamic binaural cues and varying background levels (Deike et al., 2014, Brungart et al., 2001) and thus varying effective signal-to-noise ratios are important factors that can influence the resulting speech reception thresholds (SRTs). A simple approach which aims to provide reliable SRT measurements in dynamic scenes is presented in this paper.

Listeners

Thirteen experienced hearing aid users (average age 71 years) and nine young normal hearing listeners participated in the study. The hearing-impaired listeners had symmetrical moderate or moderate to severe hearing losses corresponding to the N3 and N4 categories by Bisgaard et al. (2010).

The normal-hearing listeners had thresholds < 25 dB HL at all audiometric frequencies in agreement with the definition by the WHO (1998).

Methods

The speech intelligibility measurements were carried out with the Oldenburg sentence test (OLSA, Wagener et al., 1999). The realistic auditory scenes were implemented via the Toolbox for Auditory Scene Creation and Rendering (TASCAR, Grimm et al., 2019). Two scenes with a duration of approx. 90 seconds were used. The first one was the relatively loud "cafeteria" scene (Hendrikse et al., 2018) with a background level of approx. 72 dB SPL. The second one was the "nature" scene which has a lower background noise level of approx. 45 dB SPL. The cafeteria ambience includes the influence of reverberation and multiple intelligible disturbing talkers. The "nature" scene simulates a walk in a forest with acoustic presentation of walking on leaves and a runnel.

To ensure reliable SRT measurements in the realistic scenes, the short-term RMS levels were analyzed. Figure 1 shows the output levels for the left and right ear signal with no compression algorithm. Blocks of three seconds were selected corresponding to the maximum duration of OLSA sentences. For the "nature" scene, the levels were relatively stable across the time blocks with two noticeable valleys around the block indexes of 5 and 20. The marked part of 15 seconds was selected because it provided a relatively low level variation of 3.26 dB between selected blocks compared to 12 dB for the complete scene and also featured similar acoustic stimuli throughout. Overall, levels for left and right channel were almost identical for this scene.

For the "cafeteria" scene, the levels and binaural cues were more variable resulting from the presented conversation at the cafeteria table at the beginning of the scene and background music. Therefore, level variations were further analyzed in shorter windows of 200 ms. A sequence of approx. 11 seconds was selected which provided the lowest level variations up to 13 dB for the right channel in the shorter windows compared to 33 dB for the complete scene and a reduced range of binaural level differences below 3.8 dB (8.6 dB for the complete scene).

Test-retest measurements were carried out with the normal hearing listeners for the excerpts and complete scenes to evaluate the reliability of these measurements.

For the comparison of the speech recognition performance in realistic and laboratory conditions, SRTs were also measured with the test-specific, stationary and the modulated $ICRA5_{250}$ (Dreschler et al., 2001) noise maskers representing typical maskers used in laboratory measurements. Two presentation levels of background noise were considered: 45 dB SPL and 70 dB SPL. In all conditions, the target speech was presented at a distance of 1.5 m. In the "nature" scene, the target speech was placed to the left of the listener to simulate a realistic conversation while walking. In all other conditions, the target speaker was placed in front of the listener.

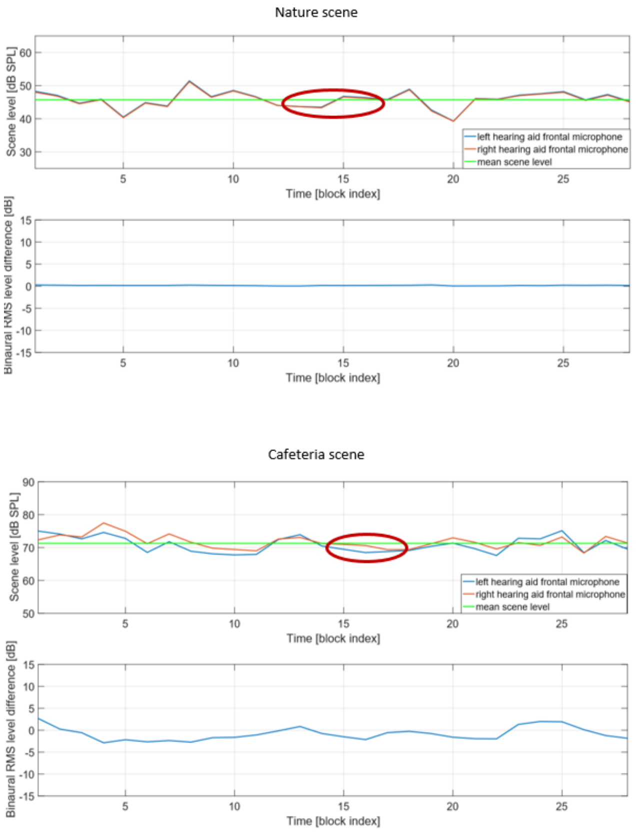

Figure 1: Level analysis of the realistic scenes in blocks of three seconds. The marked areas indicate the selected parts of the scenes.

Two test lists (each with 20 sentences) were used for training the participants. SRTs at 50% intelligibility were measured adaptively with a fixed noise level and open-set response format.

For the loudness measurements, the 20 everyday signals including for example flute and jackhammer recordings used by Oetting (2018a) with levels from 40 to 90 dB SPL were used. In addition, two short examples of both realistic scenes used in the SI measurements with a duration of three seconds were included. The signals were presented to the listeners in a random order and rated on the ACALOS-scale (Brand and Hohmann, 2002). The training consisted of binaural presentations of the stationary IFNoise (Holube et al., 2010) with levels corresponding to those of the realistic signals, e.g. levels ranging from 40 to 90 dB SPL.

Comparison of hearing aid fittings

For the trueLOUDNESS fitting method, the monaural narrowband loudness perception is first determined by measuring adaptive categorical loudness scaling (ACALOS) with uniformly exciting noises (UEN) for different frequency bands. The derived narrowband gains are then applied, and the loudness perception is determined for a speech-like binaural broadband signal (IFNoise) and binaural broadband gains are derived for 50 dB SPL, 65 dB SPL and 80 dB SPL. In contrast, NAL-NL2 uses a large database to derive gain prescriptions based on the audiogram, experience and gender of the hearing aid user. Whereas trueLOUDNESS has not yet been optimized for speech intelligibility, NAL-NL2 uses a modified Speech Intelligibility Index (SII, ANSI (1997)) and a loudness model to optimize both loudness and speech intelligibility.

Both fitting rules lead to similar median gains at 65 dB SPL for the hearing impaired listeners in this study from 1 kHz to 4 kHz. At lower frequencies, the median gain for trueLOUDNESS was up to 7 dB higher. This results from gain reduction at lower frequencies for NAL-NL2 to reduce upward spread of masking. trueLOUDNESS also provides higher median gains at 50 dB SPL overall.

Test setup

The test setup is illustrated in figure 2. It consisted of two computers. One ran the Oldenburg Measurement Application (OMA) and provided the target speech signals and anechoic maskers as well as the adaptive procedure to determine the SRTs.

The second computer ran the realistic scene rendering and hearing aid processing. To closely simulate the use of hearing aids, hearing aid impulse responses of the front microphone of a BTE device were used (Kayser et al., 2009). These hearing aid input signals for both ears were then processed with the Master Hearing Aid (MHA, Grimm et al., 2006). A simple dynamic compression algorithm with nine ¾-octave bands ranging from 177 Hz to 11314 Hz with τ_{attack} = 20 ms and $\tau_{release}$ = 100 ms was used. For NAL-NL2, the gains were thus derived for a fast compression speed.

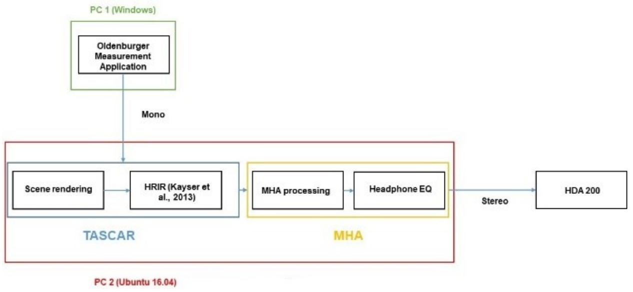

Figure 2: Schematic illustration of the test setup

For both fitting rules, the gains for a speech signal of 50 dB SPL, 65 dB SPL and 80 dB SPL were used as basis for the gain table. These gains were then interpolated towards the mid frequencies of the used compression algorithm. In a second step, the standard speech spectrum for male speech (Byrne et al., 1994) was regarded and these band levels were extrapolated towards the bandwidths of the used compression algorithm, ensuring the application of the correct gains. For headphone equalization, the coupler gains of the HDA 200 headphones were compensated with a FIR filter within the MHA. A hard clipping stage at 100 dB was included to avoid levels over the restriction of the ethics grant. For all listeners, the same setup was used, e.g. the normal hearing listeners were also presented signals convolved with the BTE impulse responses. The "cafeteria" scene was also used for the anechoic listening conditions where the reverberation and all background noises were then switched off. The same approach was used for the presentation of everyday signals in the loudness measurements.

Results

Figures 3 and 4 show the results for the speech intelligibility measurements in the laboratory scenes and realistic auditory scenes, respectively. It can be seen that the normal hearing performance is not reached for both hearing aid fittings (Wilcoxon rank sum test, $p<0.01$), especially in the modulated masker conditions. However, the comparison of both fitting rules shows lower median SRTs at lower levels of background noise for trueLOUDNESS compared to NAL-NL2. The maximum difference between fitting rules can be observed for the “nature” scene with approx. 3 dB difference in median SRT. In addition, significant differences between aided conditions were found for the stationary masker (“olnoise”) at 45 dB SPL (Wilcoxon signed rank test, $p<0.05$). At higher levels, both fitting rules perform similarly.

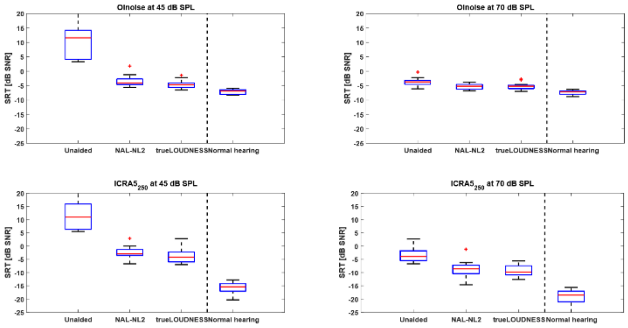

Figure 3: Boxplots of the SRTs in the anechoic listening conditions. The blue box represents the lower and upper quartiles; the red line shows median values. The whiskers indicate values inside 1.5 times the interquartile range.

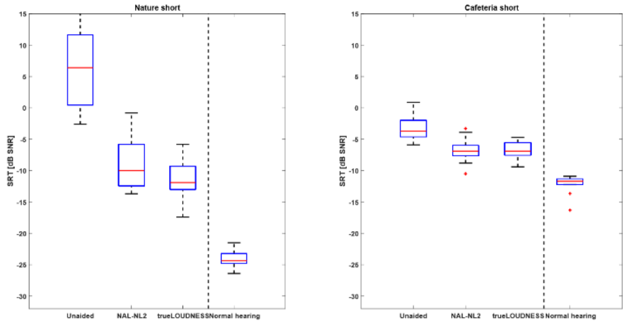

Figure 4: Boxplots of the SRTs in the realistic scenes. The blue box represents the lower and upper quartiles; the red line shows median values. The whiskers indicate values inside 1.5 times the interquartile range.

The test-retest measurements with the normal hearing listeners showed reliable SRTs with RMSEs below 1.7 dB for both excerpts and full realistic scenes.

A comparison of the SRT benefits in the laboratory conditions with those in the realistic scenes (figure 5) shows high coefficients of determination ($R^2 \geq 0.77$) for the lower level conditions, i.e. at 45 dB SPL. This is not the case for higher level conditions ($R^2 \leq 0.31$).

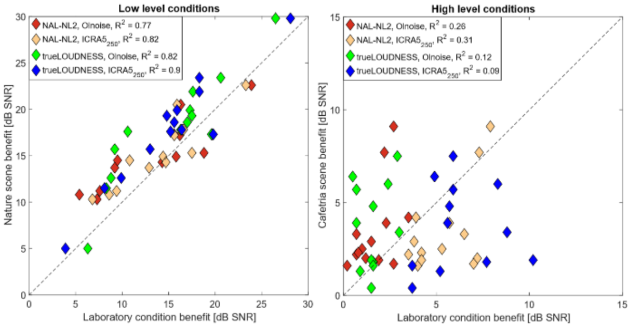

Figure 5: Comparison of SRT benefits in realistic scenes with those in laboratory conditions for both anechoic maskers at different levels (left panel: 45 dB SPL; right panel: 70 dB SPL).

The median values for the loudness ratings of the natural signals and realistic scenes are shown in Figure 6. For the scene stimuli (filled symbols), the trueLOUDNESS fitting provides normal median loudness perception. In contrast, NAL-NL2 shows lower median loudness ratings by approx. 5 CU for most scene stimuli. Fitting functions through the median values in Figure 7 further illustrate this trend for the natural signals and show similar functions for trueLOUDNESS and normal hearing listeners up to signal levels of 75 dB SPL. At higher levels, trueLOUDNESS can produce increased loudness perception compared to normal hearing listeners, however the difference is approx. 2 CU on average, e.g. below one unit on the ACALOS scale.

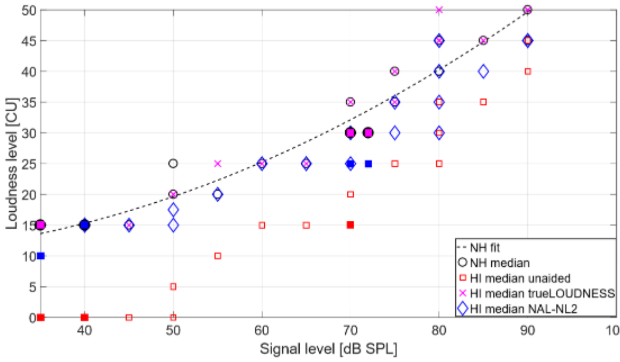

Figure 6: Median loudness ratings for the everyday signals and excerpts of the realistic scenes. The filled circles indicate excerpt of the realistic scenes. The black dotted line represents a 2nd order polynomial fitted through the normal hearing median values.

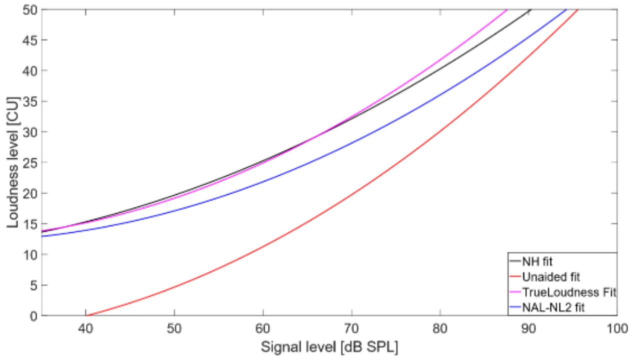

Figure 7: 2nd order polynomials fitted through the median loudness ratings for the everyday signals and excerpts of the realistic scenes for different aided and unaided conditions.

Discussion

The results show better speech intelligibility at lower input levels for trueLOUDNESS compared to NAL-NL2 throughout. This could result from higher gains at lower levels and lower frequencies for trueLOUDNESS. The lower frequency gain could be especially useful in the “nature” scene where the background has more high-frequency characteristics compared to the anechoic maskers.

The comparison of the benefits in SRT in the realistic scenes with those from standard laboratory conditions indicate that measurements in realistic environments at levels close to the hearing threshold did not provide additional information. In contrast, supra-threshold deficits as well as informational masking resulted in lower comparability of laboratory and realistic scenes at higher levels.

The loudness ratings by the hearing aid users indicate a normalized median loudness perception with the trueLOUDNESS fitting. However, four test persons showed higher-than-normal loudness perception for trueLOUDNESS for most signals and only one for NAL-NL2. The lower loudness perception for NAL-NL2 might be caused by gain reductions compared to its predecessor, NAL-NL1, which resulted from hearing aid users perceiving NAL-NL1 as too loud. It has to be noted that preference of hearing aid fitting was not tested in this study. However, the hearing aid users with higher-than-normal loudness perception might benefit from better speech intelligibility with the individual loudness compensation with trueLOUDNESS.

Literature

American National Standards Institute (1997). Methods for calculation of the Speech Intelligibility Index, ANSI S3.5–1997. New York: Acoustical Society of America.

Bisgaard, N., Vlaming, M. S. M. G. and Dahlquist, M. (2010). Standard Audiograms for the IEC 60118-15 Measurement Procedure. Trends in Amplification, 14(2), 113–120.

Brand, T., and Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. J. Acoust. Soc. Am. 112, pp. 1597–1604.

Bronkhorst, A.W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acta Acustica united with Acustica, 86(1), pp. 117-128.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. 0001-4966.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., and Kiessling, J. (1994). An international comparison of long-term average speech spectra. J. Acoust. Soc. Am. 96(4), pp. 2108–2120.

Deike, S., Denham, L. and Sussman E. (2014). Probing auditory scene analysis. Frontiers in Neuroscience. Volume 8, p. 29.

Dreschler, W. A., Verschuure, H., Ludvigsen, C. and Westermann, S. (2001). ICRA Noises: Artificial Noise Signals with Speech-like Spectral and Temporal Properties for Hearing Instrument Assessment. International Journal of Audiology 40.3 (2001): 148-157.3.

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. Acust. 92, pp. 618–628.

Grimm, G., Luberadzka, J., & Hohmann, V. (2019). A toolbox for rendering virtual acoustic environments in the context of audiology. Acta Acustica United with Acustica, 105(3), pp. 566-578.

Hendrikse, M., Llorach, G., Grimm, G. and Hohmann, V. (2018). Realistic virtual audiovisual environments for evaluating hearing aids with measures related to movement behavior. The Journal of the Acoustical Society of America. 143.

Holube, I., Fredelake, S., Vlaming M. and Kollmeier, B. (2010). Development and analysis of an International SpeechTest Signal (ISTS). International Journal of Audiology, 49, pp. 891-903.

Jenstad, L. M., Van Tasell, D. J. and Ewert, C. (2003). Hearing Aid Troubleshooting Based on Patients' Descriptions. Journal of the American Academy of Audiology, Volume 14, Number 7, July 2003, pp. 347-360(14).

Keidser, G., Dillon, H., Flax, M., Ching, T. and Brewer, S. (2011). The NAL-NL2 prescription procedure. Audiology Research, 1(1), e24.

Oetting, D., Hohmann V., Appell, J.-E., Kollmeier B. and Ewert S. D. (2018a). Restoring perceived loudness for listeners with hearing loss, Ear and Hearing. Volume 39. Issue 4. pp. 664-678.

Oetting, D., Rennebeck S., Wiercinski, B. and Schulte, M. (2018b). Binaurale breitbandige lautheitsbasierte Anpassung trueLOUDNESS im Vergleich zu NAL-NL2. 21. Jahrestagung der Deutschen Gesellschaft für Audiologie. Halle.

Völker, C., Warzybok, A., and Ernst, S. M. (2015). Comparing binaural pre-processing strategies III: Speech intelligibility of normal-hearing and hearing-impaired listeners. Trends in Hearing, 19.

Wagener, K., Kühnel, V. and Kollmeier, B. (1999). Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test. Z. Audiol. 38.

World Health Organisation. WHO (1998). Occupational exposure to noise: evaluation, prevention and control. 8 - Hearing Measurements.

Acknowledgement

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330 A5.