Preservation of Spectral Modulations and Restoration of Binaural Loudness Perception to Improve Speech Intelligibility and Loudness Perception for Listeners with Hearing Loss

Merle Gerken[1], Dirk Oetting[2], Marc René Schädler[1]
[1] *Medical Physics and Cluster of Excellence Hearing4all, University of Oldenburg, Germany,*
e-mail: merle.gerken@uol.de, marc.rene.schaedler@uol.de
[2] *HörTech gGmbH, 26129 Oldenburg, Germany, e-mail: d.oetting@hoertech.de*

Introduction

Hearing loss is a common disease with a prevalence of about 11% of the worldwide population [1], leading to problems in daily communication situations. Listeners with impaired hearing have elevated absolute hearing thresholds which limits their ability to understand speech in calm environments. Additionally, they may suffer from supra-threshold hearing impairment [2], which limits their ability to understand speech in noisy environments. While the increased hearing thresholds in quiet can partially be compensated by an amplification of sounds with hearing aids, speech recognition is not necessarily improved. A further frequent problem are loudness complaints of hearing aid users [3], which may be related to the individually varying amount of binaural broadband loudness summation [4]. To restore normal loudness perception, Oetting et al. [5] developed the trueLOUDNESS hearing aid fitting which aims at improving the loudness perception of users with impaired hearing by binaural and broadband measurements. Furthermore, from experiments with automatic speech recognition it is assumed that spectral modulations are important for speech recognition [6], which led to the development of a new dynamic range compressor (DRC) (PLATT) which enhances the spectral contrast [7].

In this study, the influence of a combination of spectral modulation preservation and binaural broadband loudness compensation on speech intelligibility and loudness perception of participants with normal hearing and with impaired hearing was investigated. The empirical results were additionally compared to individually simulated predictions.

Methods

Hearing Aid Processing

The measurements and simulations of speech intelligibility and loudness perception were performed with and without hearing aid processing. The aided conditions included a multiband DRC combined with NAL-NL2 ("openMHA NAL") or trueLOUDNESS ("openMHA TL") gain prescription, and the new PLATT DRC combined with trueLOUDNESS, either without an expansion of spectral modulations ("PLATT-1 TL") or with an expansion factor of 4 ("PLATT-4 TL").

In contrast to the multiband DRC, PLATT DRC only compresses low spectral modulation frequencies and preserves or optionally expands higher modulation frequencies. To achieve this, PLATT conducts a spectro-temporal analysis of the input signal and calculates the spectral dynamics by a convolution of spectral values with Hanning windows of different widths and then taking the difference between these convolved results. The vector containing the lowest spectral modulation frequencies (pattern sizes above 16 equivalent rectangular bandwidths (ERB)) is mapped to a reduced output dynamic range, and then the vectors containing spectral modulation frequencies up to 4 ERB are added without compression. Finally, the remaining modulation frequencies with pattern sizes between 4 and 16 ERB are added with the constraint of not exceeding the acceptable output dynamic range of the signal. The dynamic contained in the second vector (pattern sizes of 2-4 ERB) is assumed to be most relevant for speech comprehension, so it is optionally expanded.

While the commonly used non-linear fitting rule NAL-NL2 is based on the pure tone audiogram and empirical data (e.g. age and gender of the user), trueLOUDNESS is based on binaural broadband loudness measurements to create binaural broadband compensating gains.

Speech Intelligibility Measurement

Speech recognition thresholds (SRTs) were measured using the German version of the matrix test [8] with a female speaker from the study [9], adapting towards a performance of 50% correctly understood words. The speech was played back in the presence of noise via Sennheiser HDA200 headphones in a soundproof booth. The noise conditions were varied between a fluctuating ICRA5-250 (by [10], with pause durations limited to 250 ms), a babble noise constructed by six overlays of the International Female Fluctuation Masker (IFFM) [11], and the stationary speech simulating masker IFnoise [12]. The speech intellibility measurements were conducted with speech and noise coming from 0° (S_0N_0), and with speech from 0° and noise from ±90° (S_0N_{90}) by convolving the signals with head-related impulse responses.

Loudness Measurement

Loudness perception was measured for a selection of 20 binaural and broadband natural signals [5]. The subjects rated the signals using the Adaptive Categorical Loudness Scaling (ACALOS) scale [13] with categories in the range of "inaudible" to "too loud" corresponding to 0-50 CU in steps of 5 CU.

Subjective Measurements

In the subjective measurements, one experienced pilot subject with impaired hearing at the age of 33 years (called "HI01" in the following) and two experienced self-reported normal-hearing pilot subjects at the age of 23 and 25 years (called "NH01" & "NH02" in the following) took part in the experiments. Subject NH02 is the author of this study, so it cannot be ruled out that there might be a bias in the data of this subject.

Speech Intelligibility Simulations

The individual SRTs were simulated using the framework for auditory discrimination experiments (FADE) [6], which is based on an automatic speech recognizer using a Hidden Markov Model and a Gaussian Mixture Model. This framework can take into account the individual hearing loss in terms of an attenuation and distortion component according to [2]. To model the individual hearing loss, tone in quiet detection thresholds of the left and right ear at the frequencies [250, 500, 1000, 2000, 4000, 6000] Hz as well as tone in noise detection thresholds of the left and right ear at [500, 1000, 2000, 4000] Hz were measured and utilized. The tone in noise detection thresholds were used to infer a supra-threshold parameter similar to Plomp's distortion [2] for the simulations.

Loudness Simulations

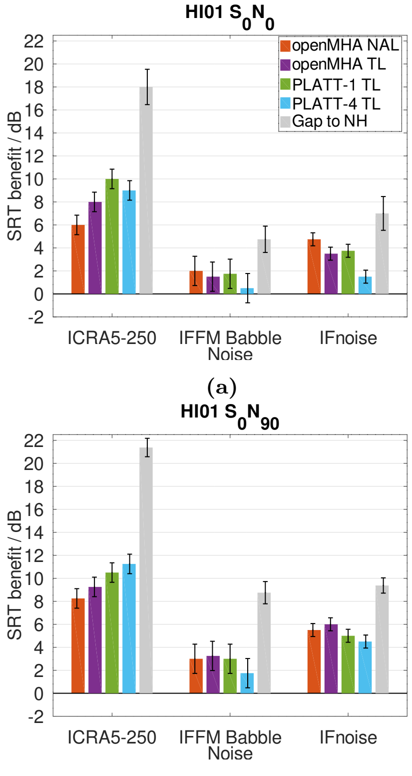
Individual loudness perception was simulated using the Dynamic Loudness Model (DLM) [14]. The natural sound signals, which were recorded for every test subject and condition for the eardrum position, were preprocessed before applying the model. The signals were first mixed down from two channels to one channel. The model assumes a free field signal at the input, so the test signals were converted to free field using a transformation from [15], where level offsets between free field and eardrum are defined for frequencies in the range of 200 to 8000 Hz. The output of the model was a time-dependent loudness in sone, from which the maximum value was computed and converted to CU using a transformation by [16].

Results

Speech Recognition Performance

The results of the speech recognition measurements for the pilot subject with impaired hearing are depicted in Figure 1 in terms of the difference between the unaided and aided measurements (benefit in dB), so a positive benefit corresponds to an improved speech recognition performance compared to the unaided condition. The error bars are based on noise-dependent test-retest accuracies from the study [17].

Positive benefits were measured in all conditions, with the highest values in ICRA5-250 noise, followed by IFnoise and babble noise. With PLATT-4 TL processing in babble noise, the error bar of the benefit in the S_0N_0 condition also reaches the negative value range. Comparing both spatial configurations, there are larger benefits in the S_0N_{90} condition, so the pilot subject benefited from the spatial separation of speech and noise, with effect sizes in the range of 0.5 to 3 dB. In ICRA5-250 noise, the performance with PLATT DRC was slightly higher than with the multiband DRC in openMHA. In babble and stationary noise, this effect could not be observed. For all conditions except for ICRA5-250 from ±90°, PLATT-1 performed slightly better than PLATT-4, so the expansion of spectral modulations did not seem to create a benefit for this pilot subject. In general, all aided results are below the gap to the median results of both normal-hearing participants ("Gap to NH").



(a)

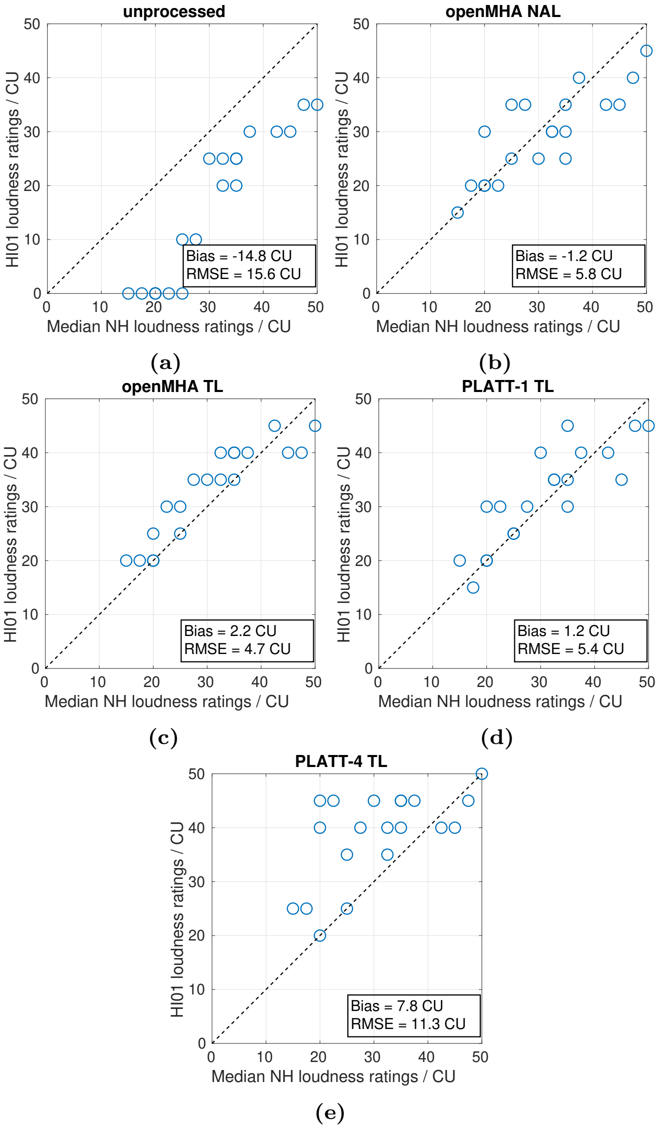
(b)

Figure 1: Measured speech recognition threshold benefit (difference between unprocessed and processed sound presentation) in different background noise types and spatial configurations for different processings of the pilot subject HI01. The gap to the speech recognition threshold mean value of both normal-hearing listeners is given additionally. The error bars are based on noise-specific test-retest differences from literature [17] and are calculated with Gaussian error propagation.

Loudness Perception

Figure 2 visualizes the loudness ratings of the pilot subject with impaired hearing in comparison to the median results of both normal-hearing listeners. The normal-hearing loudness ratings are given for the unaided condition, while those of the listener with impaired hearing are shown for the different processing types.

For the unprocessed condition, the loudness ratings of the listener with impaired hearing were below normal-hearing loudness perception, with a bias of -14.8 CU. For all types of hearing aid, the ratings were approaching normal-hearing loudness perception. The conditions with openMHA and PLATT-1 dynamic compression led to similar results, and the smallest root-mean-square error (RMSE), i.e. spread of data around the diagonal, was obtained with the multiband DRC combined with trueLOUDNESS, followed by PLATT-1 combined with trueLOUDNESS. The smallest bias was reached with openMHA NAL and PLATT-1 TL, and in contrast to NAL-NL2 fitting, the data with trueLOUDNESS fitting led to higher-than-normal loudness ratings, i.e. a positive bias, in case of this listener with impaired hearing. When spectral modulations were expanded with a factor of 4 (PLATT-4 TL), there was a much higher-than-normal loudness perception and a larger spread of data points around the diagonal.



(a) (b) (c) (d) (e)

Figure 2: Loudness ratings of listener HI01 plotted over the median loudness ratings of both normal-hearing listeners NH01 and NH02. The normal-hearing responses are given for the unprocessed natural stimuli, while the hearing-impaired responses are shown for the different processing conditions: a) unprocessed, b) openMHA NAL, c) openMHA TL, d) PLATT-1 TL, e) PLATT-4 TL. In the bottom right boxes, the bias and the RMSE are given.

Speech Intelligibility Simulations

The SRTs simulated with FADE for all three participants and all conditions are compared to the corresponding measured results in Figure 3. For the predicted data, a simulation accuracy of 1 dB was assumed. On average, the SRTs were overestimated by the simulations, leading to a bias of 1.3 dB. A high prediction accuracy with an RMSE of 3.0 dB SNR was obtained, and larger deviations were achieved for the results of both normal-hearing listeners in fluctuating noise.

Loudness Simulations

The loudness ratings simulated with the DLM are shown in Figure 4, related to the corresponding empirical results for all three listeners and all conditions. On average, the loudness predictions were slightly overestimated with a bias of 0.5 CU. In case of the subject with impaired hearing, the upper loudness range between 40 and 50 CU was not utilized by the DLM. The predicted results provide a first guess of the loudness perception with an RMSE of 5.5 CU.

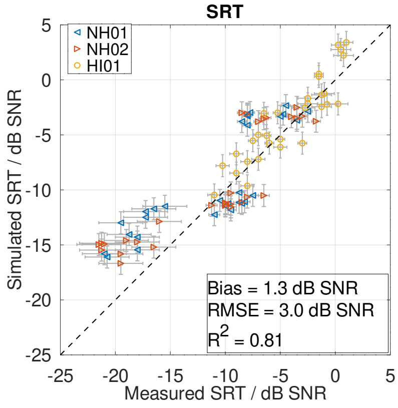


Figure 3: Results of the speech recognition simulations plotted against the corresponding results of the subjective experiments in terms of the SRT in dB SNR. The results of the three different subjects are indicated by different markers. The vertical error bars indicate the assumed simulation accuracy, the horizontal error bars are based on test-retest accuracies from [17].

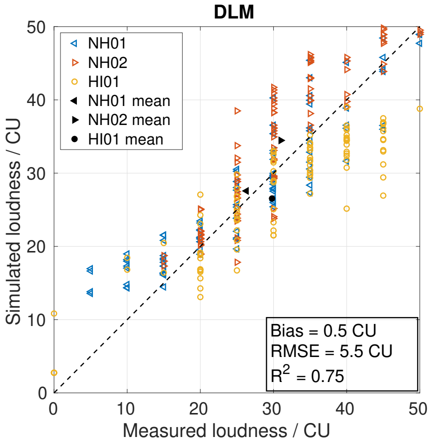


Figure 4: Simulated loudness in CU of all three subjects and all conditions as a function of the measured loudness in CU, shown for the dynamic loudness model (DLM). The mean results per subject are visualized by black markers.

Discussion and Conclusions

The effect of combining a DRC algorithm preserving spectral modulations with a hearing aid fitting rule restoring binaural broadband loudness perception was measured and predicted in terms of SRTs and loudness perception.

It has to be noted that with the current number of participants, the data at hand is not generizable. For the tested pilot subject, a processing benefit was observed in all conditions, including the new spectral modulation preserving DRC PLATT combined with the trueLOUDNESS fitting procedure. In fluctuating noise, PLATT was performing slightly better than the reference DRC, but no difference was observed in stationary and babble noise. Using trueLOUDNESS fitting, the normal-hearing loudness perception could partially be restored, but the use of PLATT with an expansion factor of 4 yielded a higher-than-normal loudness perception and no improvements in speech recognition in comparison to the other methods. Instead, the expansion factor 4 led to the lowest SRT benefits in stationary and babble noise. The speech recognition performance was accurately predicted with an RMSE of 3.0 dB, and the loudness perception simulation yielded promising first results with an RMSE of 5.5 CU, which requires validation with a larger dataset.

To conclude, the outcomes with PLATT and the reference DRC are similar. Yet, the algorithm needs to be tested with a larger population for finding significant effects. Further open questions concern the expansion factor, i.e., factors between 1 and 4 might provide higher benefits than currently observed, and listening effort and preference of the algorithms have not yet be examined.

References

[1] Stevens, G., Flaxman, S., Brunskill, E., Mascarenhas, M., Mathers, C. D., & Finucane, M. (2013). Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries. The European Journal of Public Health, 23(1), 146-152.

[2] Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. The Journal of the Acoustical Society of America, 63(2), 533-549.

[3] Jenstad, L. M., Van Tasell, D. J., & Ewert, C. (2003). Hearing aid troubleshooting based on patients' descriptions. Journal of the American Academy of Audiology, 14(7), 347-360.

[4] Oetting, D., Hohmann, V., Appell, J. E., Kollmeier, B., & Ewert, S. D. (2016). Spectral and binaural loudness summation for hearing-impaired listeners. Hearing research, 335, 179-192.

[5] Oetting, D., Hohmann, V., Appell, J. E., Kollmeier, B., & Ewert, S. D. (2018). Restoring perceived loudness for listeners with hearing loss. Ear and hearing, 39(4), 664-678.

[6] Schädler, M. R., Warzybok, A., Ewert, S. D., & Kollmeier, B. (2016). A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. The journal of the acoustical society of America, 139(5), 2708-2722.

[7] Schädler, M. R. (2021). Thoughts on the potential to compensate a hearing loss in noise. arXiv preprint arXiv:2102.12397.

[8] Wagener, K., Brand, T., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache. I-III: Design, Optimierung und Evaluation des Oldenburger Satztests. Zeitschrift für Audiologie (Audiological Acoustics), 38, 4-15.

[9] Hochmuth, S., Jürgens, T., Brand, T., & Kollmeier, B. (2015). Talker-and language-specific effects on speech intelligibility in noise assessed with bilingual talkers: Which language is more robust against noise and reverberation?. International journal of audiology, 54(sup2), 23-34.

[10] Dreschler, W. A., Verschuure, H., Ludvigsen, C., & Westermann, S. (2001). ICRA Noises: Artificial Noise Signals with Speech-like Spectral and Temporal Properties for Hearing Instrument Assessment. Audiology, 40(3), 148-157.

[11] Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an international speech test signal (ISTS). International journal of audiology, 49(12), 891-903.

[12] Holube, I., Blab, S., Fürsen, K., Gürtler, S., Meisenbacher, K., Nguyen, D., & Taesler, S. (2009). Einfluss des Maskierers und der Testmethode auf die Sprachverständlichkeitsschwelle von jüngeren und älteren Normalhörenden. Zeitschrift für Audiologie, 48, 120-127.

[13] Brand, T., & Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. The Journal of the Acoustical Society of America, 112(4), 1597-1604.

[14] Chalupper, J., & Fastl, H. (2002). Dynamic loudness model (DLM) for normal and hearing-impaired listeners. Acta Acustica united with Acustica, 88(3), 378-386.

[15] Bentler, R. A., & Pavlovic, C. V. (1989). Transfer functions and correction factors used in hearing aid evaluation and research. Ear and Hearing, 10(1), 58-63.

[16] Heeren, W., Hohmann, V., Appell, J. E., & Verhey, J. L. (2013). Relation between loudness in categorical units and loudness in phons and sones. The Journal of the Acoustical Society of America, 133(4), EL314-EL319.

[17] Wagener, K. C., & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. International journal of audiology, 44(3), 144-156.