Speech intelligibility prediction for normal hearing and hearing impaired persons – MCHI-S

L. Goenner, J. Haubold

ciAD, 44287 Dortmund, Germany, Email: {goenner,haubold}@audioquality.de

Introduction

It is well known that a successful hearing aid fitting can be characterized by the following three properties:

- An improvement in speech understanding must be obtained, especially in the presence of background noise.
- The loudness impression needs to be adequate.
- Finally, the sound quality plays an important role in customer satisfaction with their hearing device.

These three requirements will need to be fulfilled in a wide range of hearing situations. Obviously, a prediction of these parameters is highly desirable.
These items are addressed by "MCHI – Model for the comfort of hearing impaired subjects" developed by ciAD [2]. MCHI predicts loudness, timbre, auditory quality and listening effort.
Validation experiments with hearing impaired subjects revealed that the speech intelligibility predictions were also possible and acceptable in the case of speech in quiet, whereas the important case of speech in noise was non-satisfactory. Therefore, an extension to the MCHI model, named MCHI-S, was developed which predicts speech intelligibility for both, speech in quiet and speech in noise. In the following, MCHI-S, will be described. A summary of validation results will be given as well.

MCHI-S structure

The MCHI-S model inputs are audiogram data, a reference sound and a processed sound (e.g., such as processed by a hearing aid, a mobile phone or in a car).
The MCHI-S outputs are speech intelligibility in percentage and verbal categories for speech intelligibility.

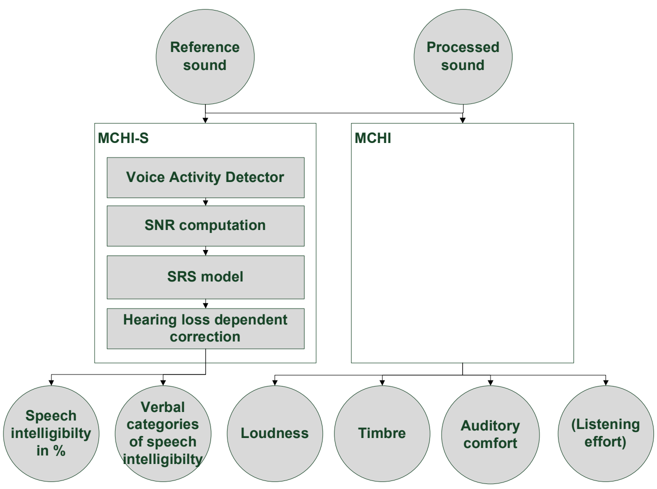

Figure 1: Overview of MCHI-S model structure.

First investigation

As a first step during model development, listening tests with normal hearing persons were conducted. Eighteen speech sounds were rated by twenty-three listeners for subjective speech intelligibility impressions. The sounds consisted of two talkers in the presence of real-life background noise, such as car noise, cocktail party babble or bird sounds.
The results were compared to three speech understanding prediction models: Articulation Index (AI), Speech Intelligibility Index (SII) and Speech Recognition Sensitivity Index (SRS) [4]. It turned out that the SRS outperformed the AI and the SII. Rank correlation values (Spearman's rho) were 0.899 for SRS, 0.736 for SII and 0.674 for AI. Results are displayed in Figure 2 .

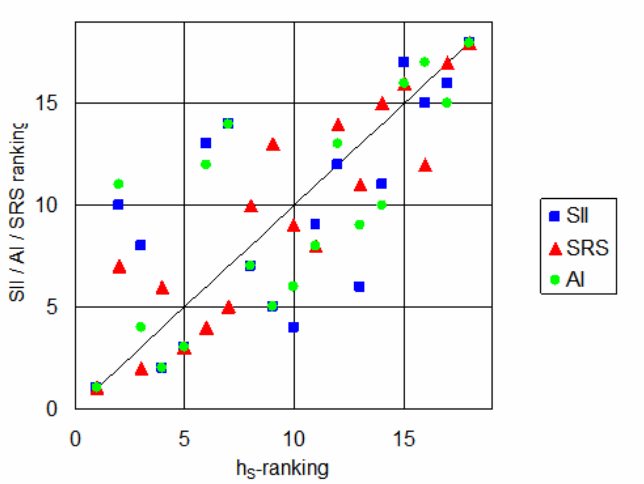

Figure 2: Ranked subjective speech intelligibility compared to SII, SRS and AI.

Consequently, we chose the SRS approach as a fundament for our model.

Implementation of the SRS model

The SRS model developed by Müsch/Buus [4] is based on statistical decision theory. The required input data are the signal-to-noise ratio, SNR, over frequency, the linguistic entropy of the speech material (a value that differs for nonsense-syllables and for complete sentences, for instance) and the number of response alternatives.
A distinctive feature of the SRS model is its capability of modeling interaction between different frequency bands. Finally, the model output is the percentage of correct responses.

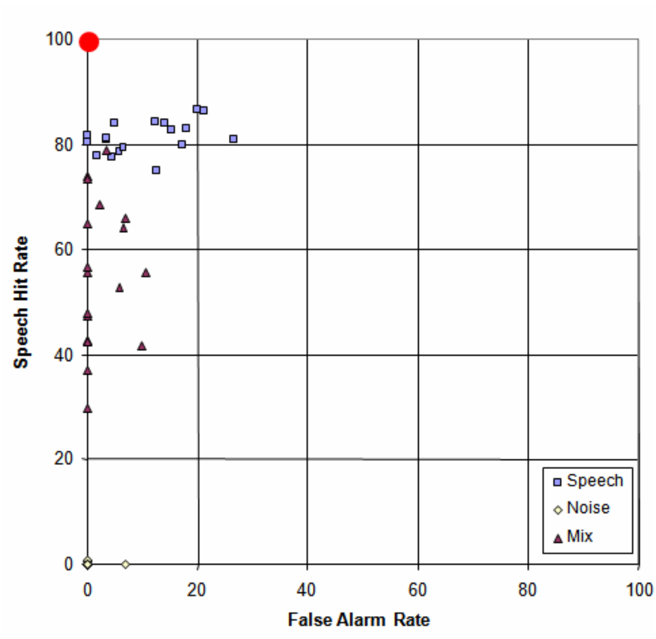

Figure 4: VAD results. Note: The red dot in the upper left-hand corner marks ideal results.

SNR calculation with Voice Activity Detection

As a method to calculate SNR values, a Voice Activity Detector (VAD) was integrated into MCHI-S. In an assessment of existing methods for Voice Activity Detection, the approach formulated by Marzinzik and Kollmeier [3] performed best. However, the speech hit rates obtained with real-life sounds were in need of improvement. With the goal of obtaining better classification results, a new VAD algorithm was developed which uses a specific envelope tracking. For the speech detection, the envelope dynamics as well as properties of the modulation spectrum are taken into account. The variables which were found to be fundamental for classification are the spectral centroid, the frequency of maximum energy and the spectral power distribution.

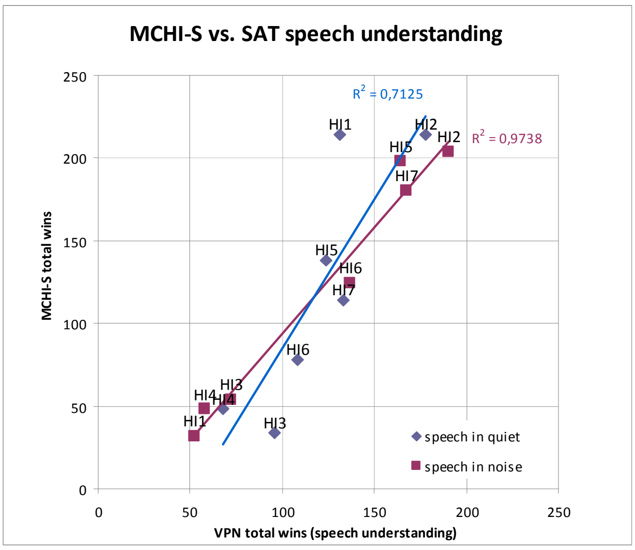

Figure 3: Validation results: Pair comparison of seven hearing instruments (shown: number of times each hearing aid was prefered).

For evaluation purposes, the algorithm's classification of real-life speech sounds was compared to a manual labeling of the respective speech sections. Results for speech in quiet show high speech hit rates (75% ... 87%) and low false alarm rates (0 … 27%). In case of speech in noise, speech hit rates are somewhat lower – depending on the amount of noise –, but false alarm rates remain low. Finally, false alarm rates for pure noise signals are very low. See Figure 4 for details. The Voice Activity Detector provides a method to compute SNR values in critical bands and in each analysis time window.

Extension of the SRS model

The next step inside MCHI-S model is the calculation of the speech recognition sensitivity (SRS). As mentioned above, the SRS model yields good predictions in the case of normal hearing listeners. However, the case of hearing impairment and the influence of the presentation level are not considered in the original SRS model. Therefore a psychoacoustic pre-processing to consider hearing loss information was integrated. This pre-processing is based on the dynamic loudness model (DLM) according to Chalupper [1]. The loudness statistics obtained from the DLM is evaluated in order to obtain a correction for the SRS model. This correction is designed such that the standard curves for speech intelligibility are realized. In further developments the SRS prediction was additionally improved by implementing the standard curves for the SNR as a limiting factor for speech intelligibility predictions.

Validation results

In a validation study, twenty hearing impaired subjects rated 7 individually fitted hearing aids in a pair comparison test. Test sounds were speech sounds in quiet and in noise. Results show a high correlation between listeners' rating and MCHI-S prediction for both, speech in quiet and speech in noise, and are shown in Figure 3.
As a further validation experiment, speech audiograms measured with hearing-impaired listeners were compared to MCHI-S speech intelligibility predictions. Due to the known high inter-subject variability in speech recognition tests [5] MCHI-S predicts individual speech audiograms with an acceptable accuracy. Four examples are shown in Figure 5.

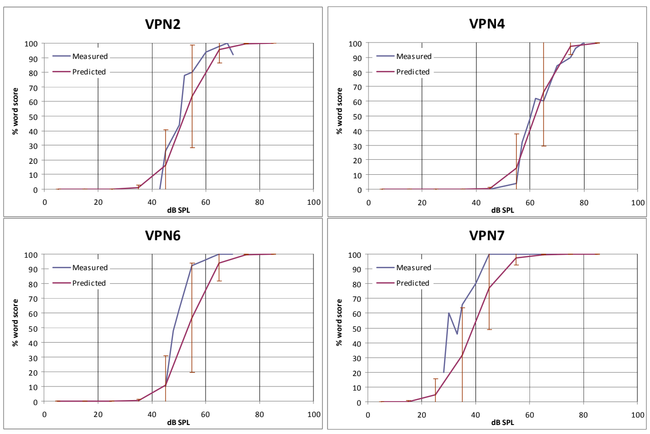

Figure 5: Measured speech audiograms compared to MCHI-S predictions.

Conclusion

The MCHI-S model predicts speech intelligibility for normal and hearing impaired persons. The prediction is based on a specifically developed Voice Activity Detector, which evaluates dynamical and spectral properties of the signal envelope, and SNR calculations, on which the adapted SRS model is applied.
Validation results show a high correlation between listeners' rating and MCHI-S speech intelligibility prediction for both, speech in quiet and speech in noise. The prediction of individual speech audiograms are of a somewhat lower quality.
Further developing steps will include a real-time optimization of the MCHI-S model. As an additional refinement, listeners' preference will be modeled on the basis of speech intelligibility and auditory comfort.

References

[1] Chalupper, J.: Perzeptive Folgen von Innenohrschwerhörigkeit: Modellierung, Simulation und Rehabilitation. Diss., TU München 2002.

[2] Fröhlich, M., Haubold, J., Schmalfuß, G.: Vorhersage von Klangqualitätsunterschieden für verschiedene Hörgeräte mittels MCHI. Fortschritte der Akustik - DAGA 2005, 89-90.

[3] Marzinzik, M. and Kollmeier, B.: Speech Pause Detection for Noise Spectrum Estimation by Tracking Power Envelope Dynamics. IEEE T. Speech Audi. P. **10** (2), 109-118 (2002).

[4] Müsch, H. and Buus, S.: Using statistical decision theory to predict speech intelligibility. I. Model structure. J. Acoust. Soc. Am. **109** (6), 2896-2909 (2001).

[5] Studebaker, G., Gray, G., Branch, W.: Prediction and Statistical Evaluation of Speech Recognition Test Scores. J Am. Acad. Audiol. **10**, 355-370 (1999).