Predicting speech intelligibility in fluctuating noise

Ralf Meyer, Thomas Brand, Birger Kollmeier

Carl-von-Ossietzky Universität Oldenburg, Medizinische Physik, Oldenburg

Summary

Different versions of a model based on the Speech Intelligibility Index (SII, ANSI S3.5-1997) are investigated with respect to their ability to predict speech reception thresholds (SRTs) of hearing impaired subjects in fluctuating noise. The different versions consider fluctuations of the input signals in different ways. The first version is the standard SII. The second version is an extension of the original SII based on Brand (DGA, 2003). The third version is based on a publication by Rhebergen et al. (JASA, 2005). The fourth version is an extension of the model from Rhebergen et al. Each version requires additional complexity and takes into account a larger amount of temporal information. On the other hand it is not clear if this additional complexity yields better predictions or if the SII concept is overextended by the increasing deviation from the standard. The predictions from the different versions are evaluated using speech intelligibility data from an audiological database. Correlations between predicted and observed SRTs ranged between r=0.48 for the standard SII and about r=0.7 for the other three versions.

Model versions

I.) SII (ANSI S3.5-1997): The starting point is the standard SII which is based on the long-term spectra of speech and noise. The audibility in 21 frequency bands is calculated and the weighted sum (band importance function depending on test material) over all bands is calculated. Consequently, the original version of the SII is insensitive to temporal fluctuations of the input signals, as the standard is based on the long-term spectra only.

II.) Frequency independent fluctuations of the noise. (Brand et al., DGA, 2002): In a first step towards a short-term SII, a version proposed by Brand is used. Now, fluctuations of the noise are considered. However, only fluctuations of the overall level of the noise are taken into account (i.e. the frequency spectrum of the noise is regarded as constant). For every level occurring in the noise level-histogram an SII value is calculated. Finally the weighted (with the rate in the level-histogram) mean over all SII values is calculated. A sketch of this model version is shown in Fig. 1 (left panel).

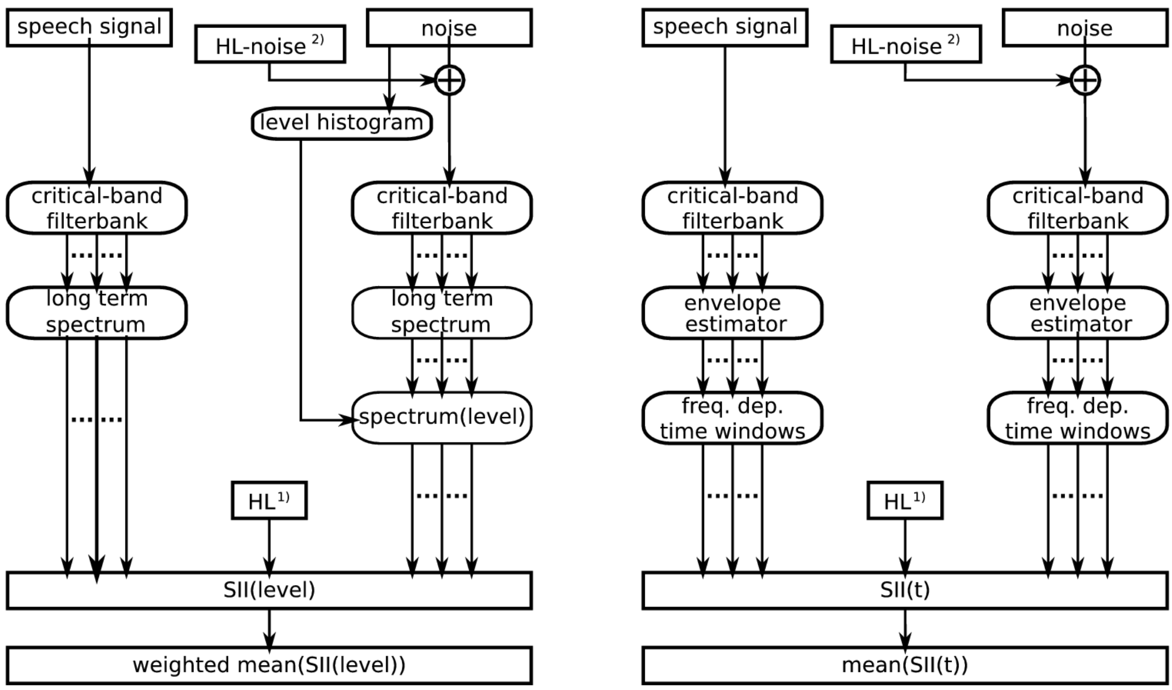

Figure 1: Diagram of two model versions used. Left: 'short-term' SII calculation scheme according to Brand et al. (2002). Right: 'short-term' SII calculation according to Rhebergen et al. (2005). In each version the hearing threshold can be included in two ways: as a parameter for the SII ('HL[1)]') or as a threshold simulating noise added to the noise signal ('HL-noise[2)]',not used here)

III.) Frequency dependent fluctuations of the noise (Rhebergen et al. JASA, 2005): In the second step, also the frequency dependency of the fluctuations of the noise are considered. This is done by using the model proposed by Rhebergen et al.. This model proposes a pre-processing of the input signals. First the signals are filtered into 21 frequency bands. In every frequency band the envelope is estimated via the Hilbert-transform. In frequency dependent time windows the instantaneous intensity is estimated. At last the mean over all SII values is calculated. A noise with the long-term spectrum of speech is used as representation of the speech signal, as it was done by Rhebergen et al.. Since this speech simulating noise shows no fluctuations, this approach does not take fluctuations of the speech into account. A sketch of this model version is shown in Fig. 1 (right panel).

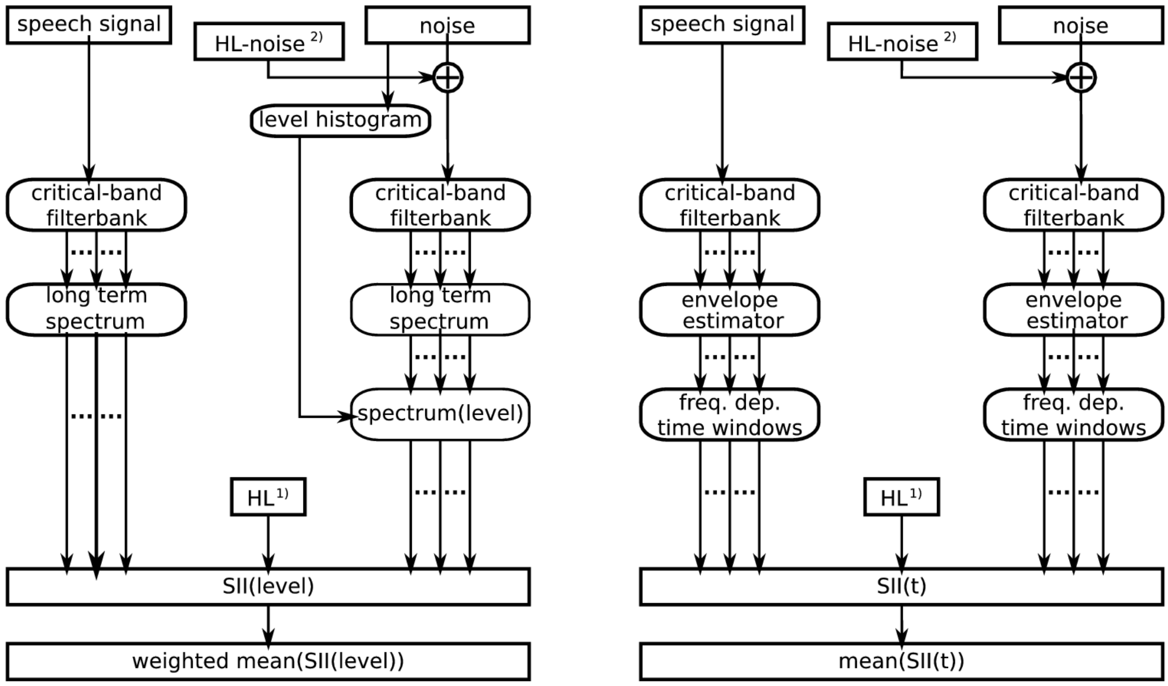

Figure 1: Diagram of two model versions used. Left: 'short-term' SII calculation scheme according to Brand et al. (2002). Right: 'short-term' SII calculation according to Rhebergen et al. (2005). In each version the hearing threshold can be included in two ways: as a parameter for the SII ('HL[1)]') or as a threshold simulating noise added to the noise signal ('HL-noise[2)]',not used here)

IV.) Frequency dependent fluctuations of speech and noise: In the last step, also the fluctuations of the speech are considered. This is achieved by taking real speech signals (sentences from the sentence test) as input. For every speech signal the SRT is calculated with the model according to Rhebergen et al. and then the mean over all SRTs is calculated. This requires much more computation time than the other versions of the model. A sketch of this model version is shown in Fig. 1 (right panel). The only difference to model version III) is that speech signals are used as input and that the averaging takes part across much more speech samples.

In each version of the model a resulting SII value is transformed into an intelligibility. The speech level is then adjusted to achieve an SII of 0.133. This SII value corresponds to the Speech Reception Threshold (SRT), i. e. the signal to noise ratio which corresponds to an intelligibility of 50%. The subject's hearing-loss is included in the SII as described in the standard.

Database

An audiological database (Brand et al. 2002) is used to evaluate the predictions of the different model versions. The SRT values were acquired using the Oldenburg Sentence Test (Wagener et al., 1999) in noise. An adaptive procedure and wordscoring was used to determine the SRT. The Oldenburg Sentence Test (OlSa) is a matrix test, i. e. the syntactic structure of each sentence is the same ('name – verb – digit - adjective – noun'). All measurements were performed monaurally in a sound isolated booth via headphones (Sennheiser HDA 200).

The database contains data from 113 normal-hearing and hearing-impaired subjects with different kinds of hearing-losses. The subjects age ranges from 26 till 85 years. The fluctuating ICRA5-250 noise is used as noise. This noise is derived from the ICRA5 noise, which simulates the long-term spectrum and the modulation properties of one male speaker. The ICRA5-250 noise used in this study includes only silent periods with a maximum length of 250ms (Wagener et al.

2006), whereas the original ICRA5 noise includes silent intervals of up to two seconds duration. The noise levels varied between 65 an 85 dB SPL, depending on the hearing loss of the listener. For normal-hearing listeners the noise level was 65 dB SPL.

Results & Discussion

Fig. 2 shows scatter plots of the results for all model versions used. On the abscissa the predicted SRT values are shown. On the ordinate the observed SRT values are shown. The diagonal line displayed in all figures represents perfect predictions of the measured data. The dashed lines around the diagonal line show a deviation of 4dB from perfect prediction. Furthermore for each plot the resulting correlation between observed an predicted SRT is displayed

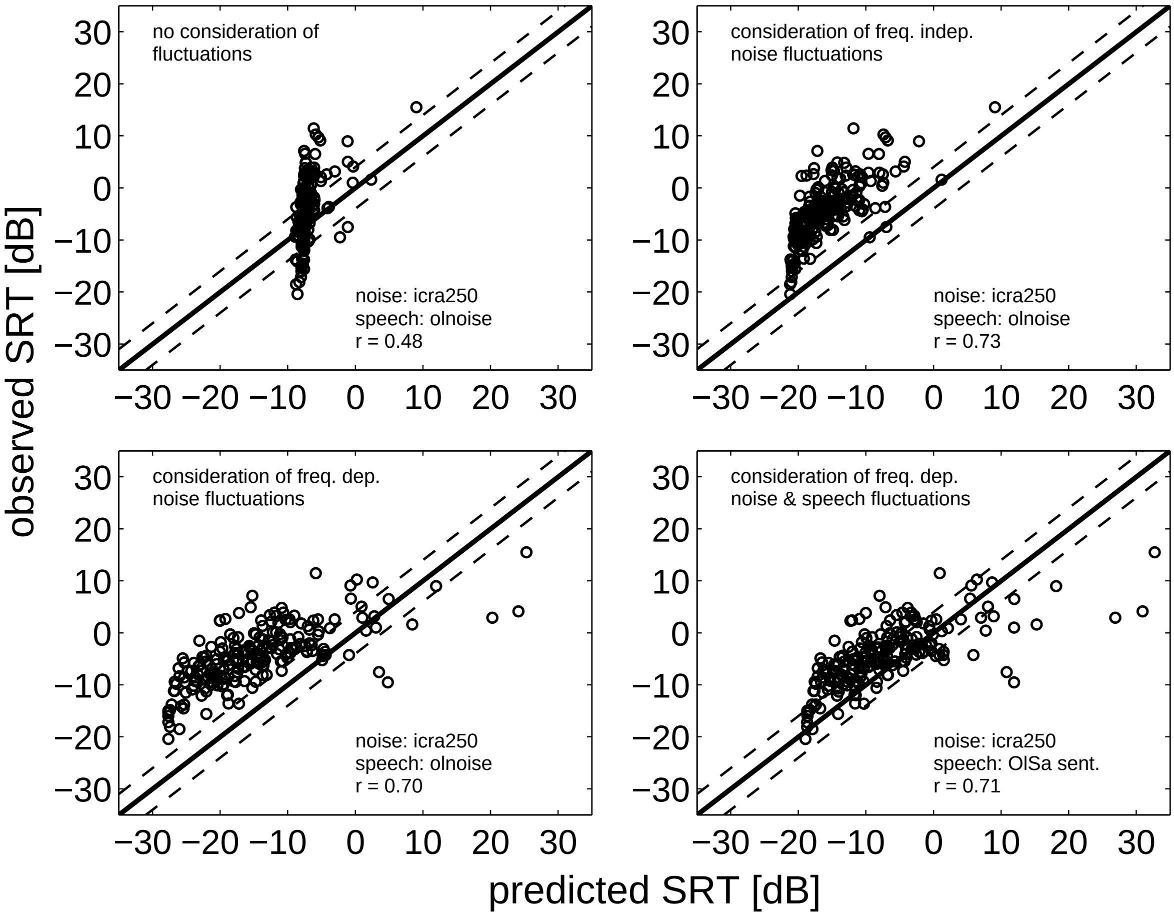

Figure 2: Observed over predicted SRTs of 113 normal-hearing and hearing-impaired subjects. Predictions using: SII (upper left), short-term SII version according to Brand et al. (2002) (upper right), short-term SII version according to Rhebergen et al. (2005) (lower left), short-term SII version introduced in this study (lower right).

The correlation for the standard SII is r=0.48. For the three other model versions the correlation is about r=0.7. This means that the consideration of some temporal information in terms of the frequency independent fluctuations of the noise (Brand et al. 2002) results in a higher correlation between the predicted and the measured SRTs. However, considering further temporal information in terms of the frequency dependent fluctuations of the noise (Rhebergen et al. 2005) and the frequency dependent fluctuations of the noise and the speech (extension presented in this study) does not result in a significantly higher correlation.

On the other hand, the consideration of more temporal information yields a closer alignment between predictions and observations even though the correlation does not improve. Although the SII was not designed to predict SRTs in fluctuating noise, it yields good results for some subjects, i.e. about 50% of the predictions are within the 4dB interval. If we consider the frequency independent temporal information (model version based on Brand et al. (2002)) the correlation is higher as for the SII (SII: r=0.48, Brand: r= 0.73), but there are less points close to the diagonal, only about 4% of the predictions are within the 4dB interval. If we also take the frequency dependent fluctuations of the noise into account (model version based on Rhebergen et al.

(2005)) the correlation is slighty smaller than for the version from Brand et al (2002) (Brand: r=0.73, Rhebergen: r=0.70). However, more predictions are close to the diagonal, i.e. about 8% of the predictions are within the 4dB interval. If we now also consider the fluctuations of the speech signal (extension introduced in this study), the correlation is between the version from Brand et al. (2002) and Rhebergen et al. (2005) (new extension: r=0.71). Now, the predictions for a lot of subjects were close to the data (about 50% of the predictions are within the 4dB interval), but for some subjects the predictions show a very large deviation from the observed SRTs. This results in a lower correlation than for the version of Brand et al. (2002). The consideration of more temporal information results in a prediction closer to the measurement for some of the subjects (4% and 8% vs. 50%), however for other subjects large deviations occur.

This improvement of predictions for the versions which consider more temporal information is achieved with an increase of computational complexity. The complexity and calculation time in the last two versions is much higher than in the versions from Brand et al. (2002).

Acknowledgement

This study was supported by the 'Audiologie Initiative Niedersachsen'.

References

ANSI (1997). „Methods for Calculation of the Speech Intelligibility Index“, American National Standard S3.5-1997

Brand, T. (2002). „Vorhersage der Sprachverständlichkeit in Ruhe und im Störgeräusch aufgrund des Reintonaudiogramms“, DGA 2002

Rhebergen, K. & Versfeld, N. (2005). „A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners.“ J Acoust Soc Am, 117, 2181-92

Wagener K., Brand T., Kühnel V. und Kollmeier B. (1999). „Entwicklung und Evaluation eines Satztestes für die deutsche Sprache I-III: Design, Optimierung und Evaluation des Oldenburger Satztestes“, Zeitschrift für Audiologie 38(1-3)

Wagener K., Brand T., Kollmeier B. (2006). „The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners“, Int J Audiol, 45, 26-3