Prediction of Speech Intelligibility in Fluctuating Noise for Listeners with Normal and Impaired Hearing

Ralf Meyer[1], Thomas Brand[2], Birger Kollmeier[3]

[1] *Carl von Ossietzky Universität Oldenburg, Medizinsche Physik, Email: ralf.m.meyer@uni-oldenburg.de*

[2,3] *Carl von Ossietzky Universität Oldenburg, Medizinsche Physik*

Introduction

Speech intelligibility prediction based exclusively on the pure-tone audiogram (e.g. Speech Intelligibility Index (SII)-based models) shows some limitations. For listeners with similar pure-tone audiograms the same speech reception thresholds (SRT) are predicted even though the observed SRT values may differ. For SRT values measured in fluctuating noises, these differences are even larger than in stationary noises. One hypothesis is that cognitive parameters significantly influence the observed SRT in fluctuating noise.

Predicting SRT values measured for listeners with a hearing aid is very useful for assessing the maximum achievable benefit of a hearing aid. With the Master Hearing Aid (MHA) [5] it is possible to control and access the complete path of signal processing for different hearing aid algorithms. This feature can be used to integrate the MHA into a framework for predicting speech intelligibility. This framework consists of a short term preprocessing and the SII similar to [9].

In this study the speech reception thresholds for stationary and different fluctuating noises are measured for 9 listeners with impaired hearing. Three SRT values for different noises are measured using the MHA, to evaluate the capabilities of a short term SII model in predicting aided thresholds. Additionally a pure-tone audiogram and a categorical loudness scaling are measured. For the assessment of cognitive parameters the Text Reception Threshold test (TRT,[12]) and the lexical decision test (e.g. [2]) are used.

Two hypotheses are investigated. First, short term SII models are able to predict the SRT in fluctuating noise even when hearing aids are used. Second, the results of the cognitive tests should be useful to explain parts of the remaining deviations between the audiogram based predictions of the short term SII model and the measured values.

Method

Listeners

The measurements were done using 9 listeners with impaired hearing (age: 60-78 years, mean: 62 years) and a control group of 5 listeners with normal hearing (age: 21-29 years, mean: 25 years) for the non-auditive tests. The hearing loss of the hearing impaired group was mostly broadband between 30 dB HL and 70 dB HL, only two audiograms showed a significant decrease to higher frequencies.

Auditive measurements

For all listeners a pure-tone audiogram was measured. The speech reception threshold (SNR value at 50% intelligibility) measurements were done using the Oldenburg sentence test [10]. The Oldenburg sentence test is a matrix test. All sentences have the same structure (name - verb - number - adjective - noun). The threshold was measured with an adaptive procedure with lists of 20 sentences. The noise level was constant and the speech level was adaptively varied according to [3].

In order to perform measurements using the Master Hearing Aid (MHA) a special measurement setup is used. This special setup is shown in figure 1. For the measurements two PCs were used. The first PC provided the Oldenburg Measurement Applications (OMA,[8]), which controlled the measurement parameters (noise type, noise level, sentence list). This PC was connected via an optical digital audio connection to the second PC, which provided the MHA. The audio signals (speech and noise) were processed by the MHA according to the listeners hearing loss, then free field equalized, and presented to the listeners monaurally via Sennheiser HDA200 headphones. In this study, two different hearing aid algorithms were investigated. First the linear 'camfit linear' algorithm and second the 'NAL NL1' algorithm, which applies compression.

The SRT measurements were done for 3 different conditions, which differed in Master Hearing Aid usage:

- Noise level 65 dB, no MHA,
- Noise level 65 dB, MHA with camfit linear algorithm,
- Noise level 65 dB, MHA with NAL NL1 algorithm.

For each situation the SRT value is measured for three different noise types. First, the ICRA1 noise [4] is used. This noise is stationary noise with a speech-like long-term spectrum. Second, the ICRA5-250 noise [11] is used. This noise is a fluctuating noise with speech-like fluctuations and long-term spectrum. Both ICRA noise were generated from english speech and simulate a male speaker. Third, the International Speech Test Signal (ISTS,[7]) is used. This noise is also a fluctuating noise, but in contrast to the ICRA5-250 noise the ISTS is generated in a different way, has longer pauses, simulates a female speaker, and was generated from speech samples from several different languages.

Non-auditive measurements

In order to measure the speech processing skills of the listeners two non-auditive test were used. First, the 'Text Reception Threshold'-Test (TRT-test,[12]) was used. This test measures the capabilities of reading when parts of the text are hidden and was developed as an visual analogue to the SRT measurement. The TRT-test used in this study is slightly different from the test presented in [12]. The task for listeners was to read a sentence from the Oldenburg sentence test displayed on a computer screen. The sentence with black font was covered by a black bar pattern. The width of the bars was varied adaptively according to the listeners answers. The same lists of 20 sentences as for the Oldenburg sentences test were used. The result of the test was the percentage of the area covered with black bars at 50% percent intelligibility. Second, a lexical decision test (e.g. [2]) was used. The used implementation of the test was developed within the Hearcom project [6]. On a computer screen a combination of upper-case letters were presented to the listeners and they had to decide, if the displayed combination of letters is a word or a non-word by pressing one of two keys on the keyboard. The score of correct answers in percent and the response time was measured.

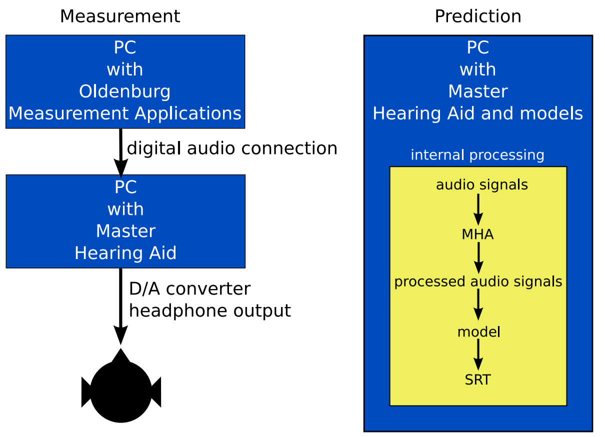

Figure 1: Measurement and prediction setup. The measurements were done using two PCs. The first PC provides the Oldenburg Measurement Applications (OMA) and is connected with the second PC via a digital optical audio connection. The second PC provides the MHA and generates the headphone output for the listener. For the prediction the MHA and the models run together on a third PC. The audio signals are processed by the MHA and then fed into the model.

Model

For the prediction of SRT values a model based on the Extended Speech Intelligibility Index (ESII) by Rhebergen et al. [9] was used. The calculation procedure for the model is shown in figure 2 together with the calculation procedure of the standard Speech Intelligibility Index (SII)[1].

Because the standard SII is unable to predict the release of masking in fluctuating noise for speech reception thresholds, Rhebergen et al. developed a preprocessing to modify the SII concept for the prediction of SRTs in fluctuating noise. This preprocessing consists of three steps. First, the input signals (speech and noise) are filtered into 21 frequency channels using a 'critical-band' filter bank according to the standard SII. Second, in each frequency channel the envelope is calculated using the Hilbert transform of the signal. Third, the intensity in each frequency channel is estimated using frequency dependent time windows. For every time step an SII value is calculated and finally the mean over all SII values is taken. According to [9] also a forward masking function is applied to model forward masking.

In this study the ESII model with forward masking[9] was modified. Rhebergen et al. used a stationary noise with a speech like long-term spectrum instead of the original speech signal. In this way, for a stationary noise the SII-values calculated with the ESII were identical to those calculated with the standard SII. Using a stationary speech signal and fluctuating noise signal results in an large SII value for every time step with a noise pause. In the measurement, not for every time step speech energy is present, because of the fluctuations of the speech signal. In this study, real speech signals were used as input speech signals, in order to reduce this effect. Five sentences of the Oldenburg sentence test[10] were used. It was necessary to use more than one sentence, because of the different fluctuations of the sentences. For each sentence an SRT value was calculated with the ESII and finally the mean over these SRT values is calculated. In this way the averaging of the standard SII (taking the long-term spectra of speech and noise) is 'transformed' to the averaging over the SRT values for the different sentences of the Oldenburg sentence test.

For prediction the MHA was installed on one PC, together with the model. The same signal processing as for the measurements was applied by the MHA, accept for the free field equalization. The processed audio signals were then used as input variables for the prediction model.

Results

Auditive measurements

In figure 3 the measured SRT values for the 9 listeners with impaired hearing are shown. For each listener the SRT values for the different noise types are denoted with different colors (blue: ICRA1, green: ICRA5-250, orange: ISTS). The control measurements at 65 dB noise level are shown in the left panel. For most of the listeners and noise types the thresholds are positive, that means at the SRT the speech level was higher than the noise level. The next two panels show the results for the two MHA algorithms. The middle panel shows the results for the 'camfit linear' algorithm. Compared to the unaided situation all listeners have a benefit using the MHA. In this case a much lower speech level is needed to achieve 50% intelligibility. Using the compressive MHA algorithm (right panel) the benefit is much smaller.

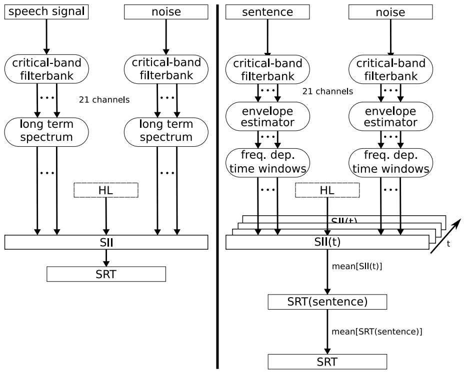

Figure 2: Comparison between the standard SII [1] on the left side and the model used in this study on the right side. For the standard SII, long-term spectra are calculated from the input speech and noise signals and used as input for the SII calculation. The basis for the model used in this study is the ESII from Rhebergen et al. [9]. The signals are passed through three preprocessing steps. First, a 'critical band' filter bank. Second, in each frequency band the signal envelope is estimated using the Hilbert transform. Third, the intensity is estimated in frequency dependent time windows and a forward masking function is applied [9]. For each time step an SII value is calculated. Finally the mean over all SII values from the different time windows is taken. In contrast to the original model the speech is not constant. For 5 random sentences from the Oldenburg sentence test [10] an SRT value is calculated with the ESII, then the mean is taken over the 5 SRT values. This model incorporates the frequency dependent variations of the speech and the noise level. In each model version the hearing-loss is incorporated as an input parameter for the SII (HL)

Predictions

Figure 4 and 5 show the results for the predictions done with the model presented above. For each of the two algorithms the predictions for the SRT values are plotted on the x-axis and the observed SRT values are plotted on the y-axis. The square of Pearson's linear correlation coefficient r^2 is calculated over all noise types. The r^2 value is a measure of the amount of the variance which can be explained by the model. The r^2 value is shown in each plot. The diagonal denotes perfect prediction and the dashed lines above and under the diagonal denote the 95% confidence intervals for the Oldenburg sentence test. The different noise types are denoted by different colors and shapes (black downwards triangle: ICRA1, blue x: ICRA5-250, red circle: ISTS).

The prediction error for the two different algorithms is the range of prediction errors for predictions done without the MHA. For the NAL NL1 algorithm the predictions are closer to the diagonal and the r^2 value shows that 75% of the variance can be explained by the model predictions. For the camfit linear algorithm predictions from the model were to high. The listeners perform better than the model predicts. The r^2 value shows that only 64% of the variance could be explained.

Non-auditive measures

For the two non-auditive measures the following results were measured: The TRT-test assesses the percentage of

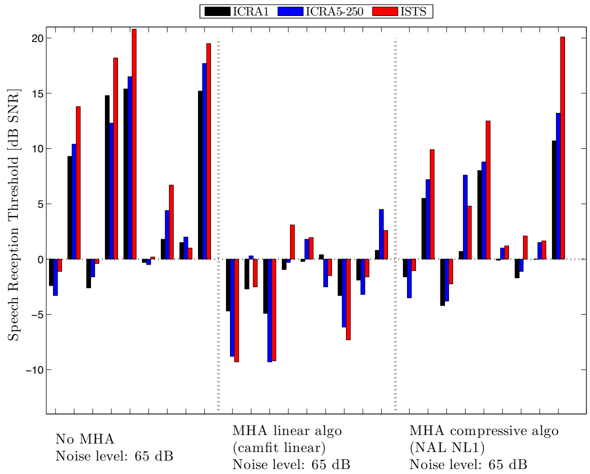

Figure 3: Observed Speech Reception Threshold (SRT) values for the three conditions measured in this study. For each condition MHA algorithm and noise level are denoted below the panels. For each listener the SRT values for the different noise types are denoted with different colors (blue: ICRA1, green: ICRA5-250, orange: ISTS). In the left panel the SRT values for 65 dB noise level and no MHA are shown. In the middle and right panel the SRT values for the two MHA algorithms ('camfit linear' and 'NAL NL1') are shown.

the text-field, which is covered by the black bar pattern at 50% intelligibility. For the TRT-test the measured threshold was between 55% and 70% with a mean of 62%. So over 50% of the visible area could be covered at 50% intelligibility. For the lexical decision test the mean measured score was 96.5% with only small variances. The mean response time was 802ms. Compared with the control group of listeners with normal hearing, there was no significant difference between the groups.

In order to evaluate if the cognitive parameters can explain parts of the predictions errors of the model described above, the correlation between the predictions errors and the individual results of the non-auditive test were calculated. These correlations were not significant.

Conclusion & Discussion

The low performance of the hearing impaired listeners for the situation with a noise level of 65 dB and no MHA can be explained by the hearing loss of the listeners. For most of the listeners a noise level of 65 dB is far below medium loudness. For some of them it is close to a measurement in silence. Some listeners even reported that they did not hear the noise. The situation with 65 dB noise level and no hearing aid serves as a baseline for the possible benefit due to the MHA processing.

The prediction of the measured SRT values using the ESII with forward masking and for fluctuating speech signals shows that up 74% of the variance can be explained by the model. The first hypotheses of this study was confirmed: Current models for the prediction of speech intelligibility thresholds are able to predict the SRT values of aided measurement with a accuracy compared to normal-hearing and unaided measurement. The second hypothesis, however, was not confirmed so

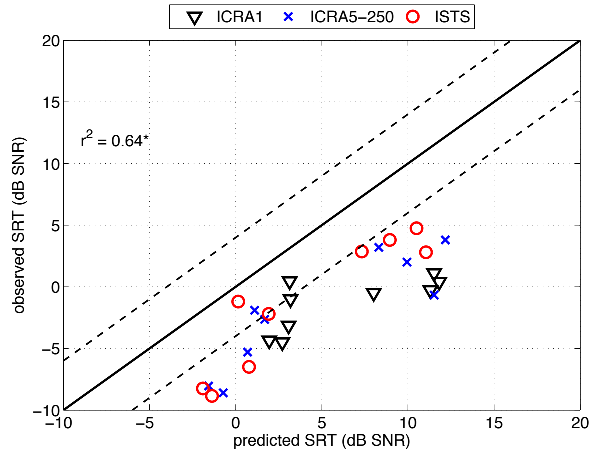

Figure 4: Observed SRT values over predicted values for the 'camfit linear' algorithm of the MHA in dB SNR. The observed values are shown on the y-axis and the predicted values on the x-axis. The diagonal denotes perfect prediction. The dashed lines above and under the diagonal denote the 95% confidence intervals for the Oldenburg sentence test. The different noise types are denoted by different colors and shapes (black downwards triangle: ICRA1, blue letter x: ICRA5-250, red circle: ISTS). The r^2 values over all noise types is shown in the upper left corner.

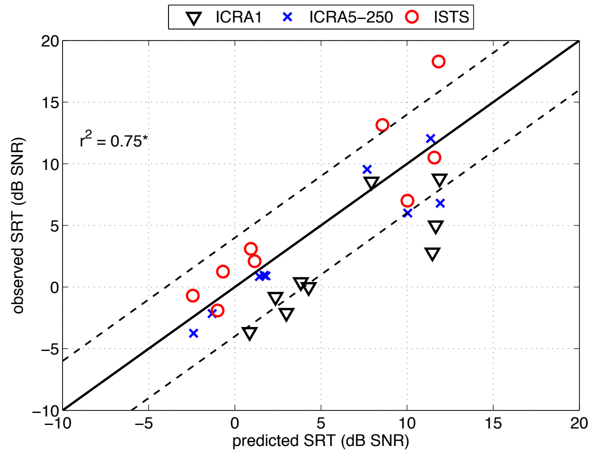

Figure 5: Observed SRT values over predicted values for the 'NAL NL1' algorithm of the MHA in dB SNR. The observed value are shown on the y-axis and the predicted values on the x-axis. The diagonal denotes perfect prediction. The dashed lines above and under the diagonal denote the 95% confidence intervals for the Oldenburg sentence test. The different noise types are denoted by different colors and shapes (black downwards triangle: ICRA1, blue x: ICRA5-250, red circle: ISTS). The r^2 values over all noise types is shown in the upper left corner

far: the remaining variance could until now not be explained by the measured cognitive parameters. For the investigation of this hypotheses more measurements and listeners are needed.

The next steps will be to find a way to integrate other auditive parameters (results of categorical loudness scaling, uncomfortable level) into the model. On the other hand, more subjects are needed to achieve more variance in the results of the non-auditive measures and to understand the role of these parameters in speech processing. Afterwards it might be possible to integrate these parameters into the model.

References

[1] ANSI S3.5-1997. American national standard: Methods for calculation of the speech intelligibility index. American National Standards Institute, New York, June 1997.

[2] C. A. Becker. Semantic context and word frequency effects in visual word recognition. *J Exp Psychol Hum Percept Perform*, 5(2):252–259, May 1979.

[3] T. Brand and B. Kollmeier. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J. Acoust. Soc. Am.*, 111(6):2801–2810, Jun 2002.

[4] W. A. Dreschler, H. Verschuure, C. Ludvigsen, and S. Westermann. ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. International Collegium for Rehabilitative Audiology. *Audiology*, 40(3):148–157, 2001.

[5] G. Grimm, T. Herzke, D. Berg, and V. Hohmann. The master hearing aid: A PC-based platform for algorithm development and evaluation. *Acta acustica united with acustica*, 92(4):618–628, 2006.

[6] Hearcom. http://hearcom.eu.

[7] I. Holube, S. Fredelake, J. Bitzer, and M. Vlaming. Erstellung eines Testsignals mit Sprachcharakteristik. In *Proceedings*. Jahrestagung für Akustik (DAGA), 2007.

[8] Hörtech. Oldenburg measurement applications, http://www.hoertech.de.

[9] K. S. Rhebergen, N. J. Versfeld, and W. A. Dreschler. Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. *J. Acoust. Soc. Am.*, 120(6):3988–3997, Dec 2006.

[10] K. Wagener, V. Kühnel, and B. Kollmeier. Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Zeitschrift für Audiologie/Audiological Acoustics*, 38(1):4–15, 1999.

[11] K. C. Wagener, T. Brand, and B. Kollmeier. The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. *Int. J. Audiol.*, 45(1):26–33, Jan 2006.

[12] A. A. Zekveld, E. L. J. George, S. E. Kramer, S. T. Goverts, and T. Houtgast. The development of the text reception threshold test: a visual analogue of the speech reception threshold test. *J Speech Lang Hear Res*, 50(3):576–584, Jun 2007.