Application of the Binaural Speech Intelligibility Model for prediction of intelligibility in in-car noise

R.-L. Fischer[1], S. Meyer[1], J.Otten[2], C. Meinecke[1]
[1] *Friedrich-Alexander-University of Erlangen-Nuremberg, Germany, Email: extern.rosa-linde.fischer@audi.de*
[2] *AUDI AG, 85045 Ingolstadt, Germany, Email: joern.otten@audi.de*

Introduction

In order to optimize in-car communication and infotainment applications in vehicles, intelligibility measurements in selected conditions could be used to predict speech intelligibility for occupants. Intelligibility is one important factor for communication quality [6, 3] and can be measured reliably. However measurements of intelligibility with subjects are time consuming and expensive, what makes the use of Speech Indices attractive. Their application is well known in typical communication situations with concurrent speech noise but in-car acoustics is special. Therefore, the scope of application has to give proof in this situation.

Comprehensive analysis of speech transmission and noise in vehicles showed a complex situation. Different velocities, cars, seats etc. change the spectrum and level of noise. Speech in cars is transmitted via an ensemble of loudspeakers with different positions around the occupants. Moreover, there are several signal processing channels for entertainment and announcements including delays and different settings for optimizing the entertainment for sitting positions, which results in very different spectra. As can be seen in Figure 1 there is also a spatially distributed signal and noise situation, resulting in channel differences measured with an artificial head. Additionally, energy in a spectrum of in-car noise is concentrated under 100 Hz, while Speech Indices only account for the range above this frequency.

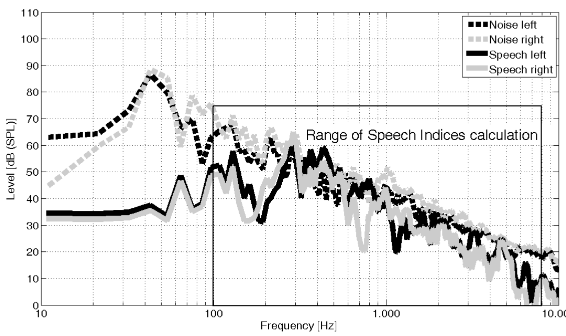

Figure 1: Spectra of typical speech and noise in cars.

The purpose of this paper is to identify the speech prediction method which works best for in-car situations. Investigated are the Speech Intelligibility Index (SII) and the Speech Transmission Index (STI). Moreover, the SII is extended for binaural measurements. The so-called Binaural Speech Intelligibility Model (BSIM, [1, 2]) combines the binaural equalization-cancellation processing [4] with the SII.

Speech intelligibility data were collected in two experiments carried out at the University of Erlangen-Nuremberg. Experiment 1 focused on several noise conditions in vehicles, while the purpose of Experiment 2 was to investigate the influence of different speech transmission ways.

Methods

Recording of Speech and Noise Stimuli

For speech signals the sentences of the Oldenburger Sentence Test (OLSA, [10, 11, 12]) were convoluted with the binaural transfer function to an artificial head, measured via several transmission ways in cars. The noise signals were binaural recordings while driving. All binaural recordings were made with the artificial head HMS II (Head Acoustics).

In Experiment 1 noise conditions differ across cars, seats and velocities. Speech was transmitted via entertainment channel to the respective seat.

Experiment 2 focused on different ways of speech transmission in cars which are realized in several cars. In addition to transmission via entertainment channel speech was transmitted via telephone and navigation announcement channel. Furthermore in-car communication was simulated by a speaking head at driver´s place and speech recordings at right rear seat and at the 3rd row in a Sport Utility Vehicle (SUV). Velocity was kept at 150 km/h in all conditions. Experimental conditions are listed in Table 1.

Empirical Speech Related Tresholds

Eight normal-hearing subjects participated in each of the two experiments. Their age ranged in Experiment 1 from 24 to 31 and in Experiment 2 from 23 to 59. Speech intelligibility measurements were made using the Oldenburg Measurement Applications. Both speech and noise were presented via headphones (Sennheiser HDA 200) and calibrated to 65 dB (SPL) on the left channel measured by the artificial head HMS II. For intelligibility measurements the noise level was kept at 65 dB (SPL), while the speech level was adaptively adjusted to get the Speech Related Threshold (SRT). The SRT represents the difference level of speech related to the noise, needed for 50% intelligibility. The more negative the SRT, the better speech intelligibility is, because speech could be quieter for understanding 50%.

Table 1: Experimental conditions

Experiment 1		
Condition	**Signal**	**Noise**
1	Entertainment	Estate car, co-driver, 150 km/h
2	Entertainment	Estate car, rear seat, 150 km/h
3	Entertainment	Estate car, co-driver, 50 km/h
4	Entertainment	Cabrio, co-driver, open, 150 km/h
5	Entertainment	Cabrio, co-driver, closed, 150 km/h
6	Entertainment	Coupé, co-driver, 150 km/h
7	Entertainment	SUV, co-driver, 150 km/h

Experiment 2		
Condition	**Signal**	**Noise**
1	Entertainment, rear seat	Estate car 1
2	In-Car Communication, rear seat	Estate car 1
3	Entertainment, co-driver	Estate car 2
4	Navigation, co-driver	Estate car 2
5	Telephone, co-driver	Estate car 2
6	In-Car Communication, rear seat	SUV
7	In-Car Communication, 3rd row	SUV

Prediction of SRTs and Speech Indices

For binaural SII predictions a MATLAB [9] implementation by Beutelmann (2008) was used, which combines a gammatone filterbank, an equalization-cancellation (EC) process in each frequency band, a gammatone resynthesis and the SII (see Figure 2). The One-third octaveband SII procedure (ANSI S3.5-1997) with the band importance function SPIN was implemented. The SRT is related to 50% intelligibility, calculated by SII procedure.

For monaural SII calculations the binaural part of the model was bypassed.

For the STI calculation a NTI Acoustilyzer AL1 was used, which has implemented the STI-PA procedure (DIN EN 60268-16).

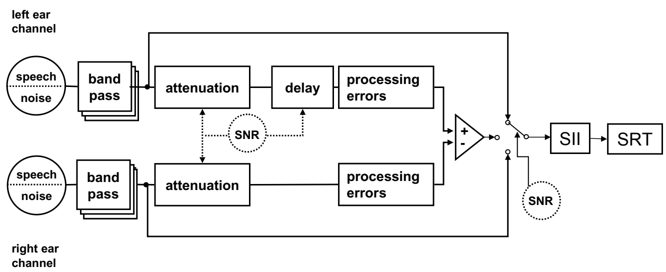

Figure 2: Binaural Speech Intelligibility Model (BSIM) [1].

Results

Figure 3 shows the results of Experiment 1, which analyzed several noise conditions combined with a speech transmission via entertainment channel in vehicles. Speech intelligibility achieved by subjects is significantly underestimated by monaural and binaural SII (t-Tests, $p \leq .001$), but the prediction of the binaural model fits better than the monaural one (ANOVA, $p \leq .05$), except in the closed Cabrio (Condition 4) and the Coupé (Condition 6).

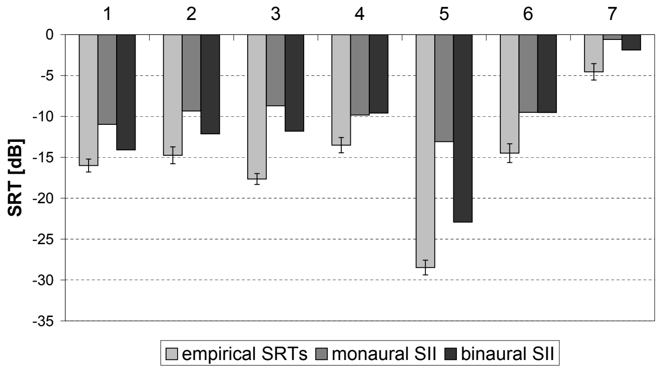

Figure 3: Mean SRTs for several noise conditions in cars (see Table 1, Exp. 1). Error bars are the mean standard deviations averaged over subjects.

The scatter plot for empirical SRT data as a function of monaural SII, binaural SII and STI for all noise conditions are presented in Figure 4. The correlation between Speech Indices and empirical SRTs shows, that prediction by the monaural SII (r=-.88) is improved by use of the binaural model to r=-.98. The correlation for the STI is r=-.76.

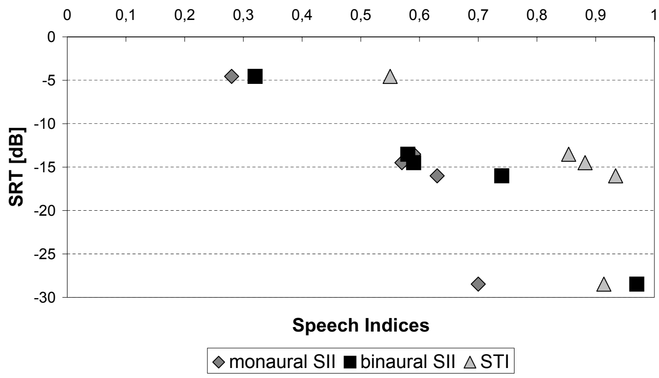

Figure 4: Correlations between Speech Indices and empirical SRTs in several noise conditions in cars.

Experiment 2 examined the influence of different ways of speech transmission on the prediction accuracy. Results are shown in Figure 5. In most conditions the monaural and binaural model again underestimate the empirical speech intelligibility (t-Tests, $p \leq .01$). Only for the transmission of speech via telephone announcement (Condition 5) both models calculate more negative SRTs than the subjects achieved. Again, the fit of the binaural model to empirical SRTs is better than this of the monaural model (ANOVA, $p \leq .05$), except in conditions 3 and 4 (Estate car 2, Entertainment and Navigation), where the differences are not significant.

There are merely gradual differences between the values of the correlations between the empirical SRTs and the Speech Indices values (Figure 6). The correlations are r=-.86 for the monaural SII, r=-.89 for the binaural SII and r=-.84 for the STI.

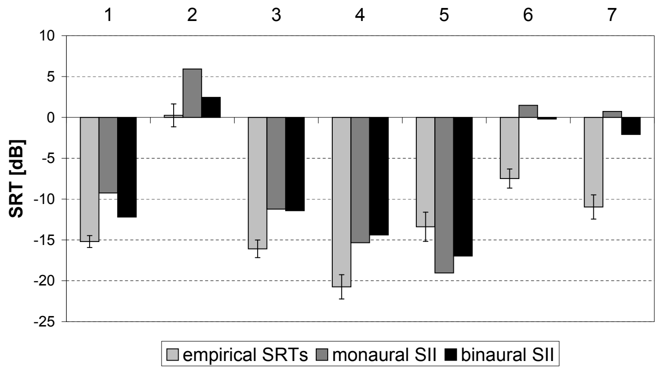

Figure 5: Mean SRTs for several ways of speech transmission in cars (see Table 1, Exp. 2). Error bars are the mean standard deviations averaged over subjects.

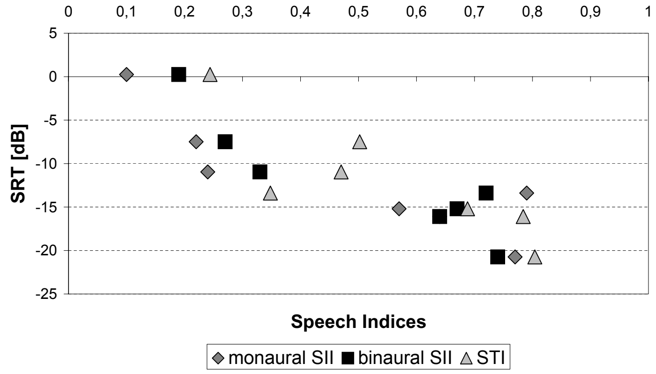

Figure 6: Correlations between Speech Indices and empirical SRTs in several speech transmission ways in cars.

Discussion

The high overall correlation between binaural SII and empirical SRTs in several noise conditions implies that the binaural SII is the best predictor for a general in-car speech intelligibility.

Although the binaural SRT calculations fitted better to empirical results, both SII models showed a general underestimation of the empirical speech intelligibility. It can be supposed, that a major amount of this deviation is caused by a miscalculation of noise energy. The models calibrate the input without accounting for energy under 100 Hz, due to this noise gets a higher level and SRTs are therefore underestimated.

But if we take into account several ways of speech transmission in cars (Experiment 2) the advantage of the SII calculation decreases. So it could be supposed that there are typical degradations of speech signals, which aren´t accounted for by the SII calculation. For instance transmission via telephone comes along with deterioration of signals caused by data compression. One consequence is limitation of bandwidth, which is accounted for by SII as well as the STI concepts. But the signal is also fragmentary transmitted over time by codecs. It can be assumed that modulation in speech signal is flattened, which degrades intelligibility for subjects. This effect isn´t regarded by the SII calculation, but by the STI. This could explain the exceptional overestimation of speech intelligibility in Telephone condition by both SII models. However, the correlation of the binaural SII isn´t worse than that of the STI, so it still has the best overall performance. Precision could be improved by an additional implementation of a Modulation Transfer Function (MTF) as in the STI.

Future work

The adaption of the BSIM for in-car noise is in progress - first results show an increasing fit between empirical data and model predictions by extention of implemented filterbank to lower frequencies. Adaption of further model parameters are considered too.

A second step after optimizing the prediction of speech intelligibility is the calculation of a comfort factor for communication quality in cars. It is known that for higher speech-to-noise ratios intelligibility doesn´t increase anymore and consequently can´t reflect existing changes in communication quality. Hearing difficulty seems to be a good equivalent for communication quality in such situations [5, 7, 8]. Additionally, for occupants in cars it has be shown that hearing difficulty is an important factor for the experience of communication quality which needs to be optimized [3].

For this purpose another study will be done to estimate the comfort factor of communication quality.

References

[1] Beutelmann, R. & Brand, T.: Prediction of speech intelligibility in spatial noise and reverberation for normal hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 120(1)(2006), 331-342

[2] Beutelmann, R.: Modelling binaural speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired subjects. Dissertation, Universität Oldenburg, Physik, Oldenburg, 2008

[3] Fischer, R.-L., Meinecke, C. & Otten, J.: Anforderungen jüngerer und älterer Autofahrer an Insassenkommunikation und die Sprachwiedergabe von Infotainmentsystemen. In Fortschritte der Akustik - DAGA 2008, ed: Jekosch, Hoffmann, DEGA e.V., Berlin, 609-610

[4] vom Hövel, H.: Zur Bedeutung der Übertragungseigenschaften des Außenohrs sowie des binauralen Hörsystems bei gestörter Sprachübertragung. Dissertation, RWTH Aachen, Elektrotechnik, 1984

[5] Morimoto, M., Sato, H. & Kobayashi, M.: Listening difficulty as a subjective measure for evaluating speech transmission performance in public spaces. J. Acoust. Soc. Am. 116(31)(2004), 1607-1613

[6] Preminger, J. E. & van Tassel, D. J.: Quantifying the relation between speech quality and speech intelligibility. Journal of Speech and Hearing Research 38 (1995b), 714-725

[7] Sato, H., Bradley, J. S. & Morimoto, M.: Using listening difficulty ratings of conditions for speech

communication in rooms. J. Acoust. Soc. of Am. 117(3) (2005), 1157-1167

[8] Sust, Ch. A., Steckel, R., Kulka, M., Lazarus, H. & Kurtz, P.: Beurteilung der Sprachkommunikation bei Verkehrsgeräuschen - Gestaltungsgrundsätze. Wirtschaftsverlag NW, Bremerhaven, 2007

[9] The MathWorks. MATLAB 7.6. (2008)

[10] Wagener, K., Brand, T., Kühnel, V. & Kollmeier, B.: Entwicklung und Evaluation eines Satztests für die deutsche Sprache. I. Design des Oldenburger Satztests. Zeitschrift für Audiologie 38(1) (1999a), 4–14.

[11] Wagener, K., Brand, T., Kühnel, V. & Kollmeier, B.: Entwicklung und Evaluation eines Satztests für die deutsche Sprache. II. Optimierung des Oldenburger Satztests. Zeitschrift für Audiologie 38(2) (1999b), 44–56.

[12] Wagener, K., Brand, T., Kühnel, V.& Kollmeier, B.: Entwicklung und Evaluation eines Satztests für die deutsche Sprache. III. Evaluation des Oldenburger Satztests. Zeitschrift für Audiologie 38(3) (1999c), 86–95.