International cross-validation of sentence intelligibility tests

Kirsten Carola Wagener[1], Thomas Brand[2], Birger Kollmeier[2]

[1]Hörzentrum Oldenburg GmbH
[2]Medizinische Physik, Carl von Ossietzky Universität Oldenburg

Introduction

In audiology it is desirable to obtain similar results across sites when measuring in similar conditions also on an international scale. This aim is quite challenging with regard to speech intelligibility tests since different languages may highly influence comparability.

The European project HearCom (Hearing in the communication society, FP6–004171) tries to establish minimum quality requirements for speech intelligibility tests in order to reach highest comparability across European countries.

Within a multi-centre study, sentence intelligibility tests were applied to normal-hearing and hearing-impaired listeners in four different countries to determine cross-validation data that is necessary for international comparability of these test procedures.

Multi-centre study

The sentence intelligibility measurements were performed as a part of a multi-centre study applying the so-called 'auditory profile' to normal-hearing and hearing-impaired listeners (i.e. a set of extensive audiological, psychoacoustical measurements, and questionnaires to characterize the individual hearing). Five partner sites from four different European countries participated in the measurements. Netherlands: Academic Center Amsterdam and VU University Medical Center Amsterdam, Sweden: Linköping University Dept of Audiology, United Kingdom: University of Southampton Institute of Sound and Vibration Research, Germany: Hörzentrum Oldenburg.

Measurements

In the present study sentence intelligibility was determined in different conditions: So-called Plomp type sentences (short meaningful sentences, HearCom D-1-2, 2005) were used to determine the binaural SRT in quiet (SRT: speech reception threshold, i.e. speech presentation level or signal-to-noise ratio that yields 50% intelligibility), monaural SRT in non-modulated speech shaped icra1 noise (Dreschler et al, 2001) and in modulated speech shaped icra5-250 noise (modulations simulate one interfering talker, Wagener et al, 2006). The noise was either male or female frequency shaped regarding the speaker's gender of the applied sentence test. So-called Matrix sentences (syntactically fixed but semantically non predictable sentences, HearCom D-1-2, 2005) were used to determine binaural aspects of speech intelligibility like intelligibility level difference (ILD=benefit between SRTs of signal and noise presentation from same direction S_0N_0 and signal and noise presentation from different directions S_0N_{90}). Also, the binaural intelligibility level difference was determined (BILD= benefit between listening with only the contralateral ear to the noise source in S_0N_{90} and listening with both ears in this situation).

All measurements were performed via free-field equalized Sennheiser HDA200 headphones. The binaural measurements were performed with virtual acoustics.

The sentence intelligibility measurements in noise were performed at a fixed noise presentation level of 65 dB SPL for normal-hearing listeners. For hearing-impaired listeners, an individual loudness level was chosen (according to a prior individual loudness scaling measurement included in the auditory profile: level yielding a loudness rating of 20 categorical units, i.e. between "soft" and "medium").

Results

Plomp type sentences: Normal-hearing data

Fig. 1 shows the mean SRT results and the respective standard deviations of normal-hearing listeners who performed Plomp type sentence intelligibility tests. The binaural SRT data in quiet are shown in the left part of the figure (given in dB SPL), the monaural SRT data in non-modulated Icra noise are shown in the middle, and the monaural SRT data in modulated Icra noise are shown in the right part of the figure (both given in dB SNR). The country-specific data are indicated as follows: German: dark blue, Dutch: light blue, Swedish: yellow, British: red.

The different results across countries can partly be explained by the procedure differences across countries in applying Plomp type sentences. One difference is the scoring method: Both the Dutch and the Swedish test apply sentence scoring, the German test applies word scoring, and the British test applies key word scoring. Also the adaptive procedure of the Dutch test is different from the other tests: In the German, Swedish, and British tests, an adaptive procedure with decreasing step size was used that is described in Brand & Kollmeier 2002 by procedure A1.

The Dutch test uses a 1up-1down adaptive procedure with fixed step size 2 dB. As a consequence of the different languages, the speakers differ across tests (Dutch and Swedish: female speaker, German and British: male speaker).

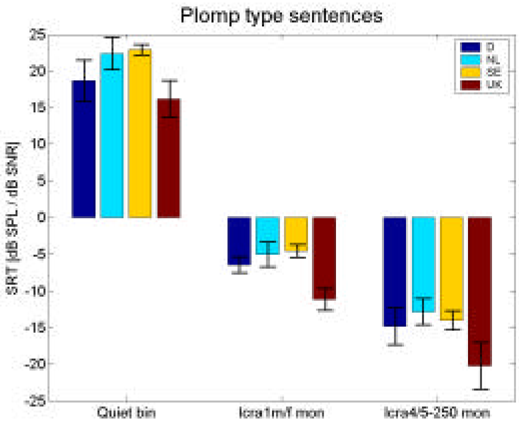


Fig. 1: Mean country-specific normal-hearing SRT data and standard deviations of Plomp type sentences (German: dark blue, Dutch: light blue, Swedish: yellow, British: red). Three different conditions (binaural SRT in quiet, monaural SRT in non-modulated Icra noise, and monaural SRT in modulated Icra noise).

It seems that the scoring method mostly influences the results: When analyzing the German data according to sentence scoring (by applying the j factor concept by Boothroyd & Nittrouer 1988), the results are similar to the Dutch results.

Matrix sentences: Normal-hearing data

Fig. **2** (left panel) shows the mean monaural SRT results and the respective standard deviations of normal-hearing listeners who performed Matrix sentence intelligibility tests. The monaural SRT data in quiet are shown in the left part of the figure (given in dB SPL), the monaural SRT data in non-modulated Icra noise are shown in the middle, and the monaural SRT data in modulated Icra noise are shown in the right part of the figure (both given in dB SNR). The country-specific data are indicated as follows: German: dark blue, Dutch: green, Swedish: red.

Fig. **2** (right panel) shows the mean binaural SRT results and the respective standard deviations of normal-hearing listeners who performed Matrix sentence intelligibility tests. The SRT data for S_0N_0 presentation are shown in the left part of the figure (given in dB SNR), the ILD data are shown in the middle, and the BILD data are shown in the right part of the figure. The country-specific data are indicated as follows: German: dark blue, Dutch: light blue, Swedish: yellow, British: red.

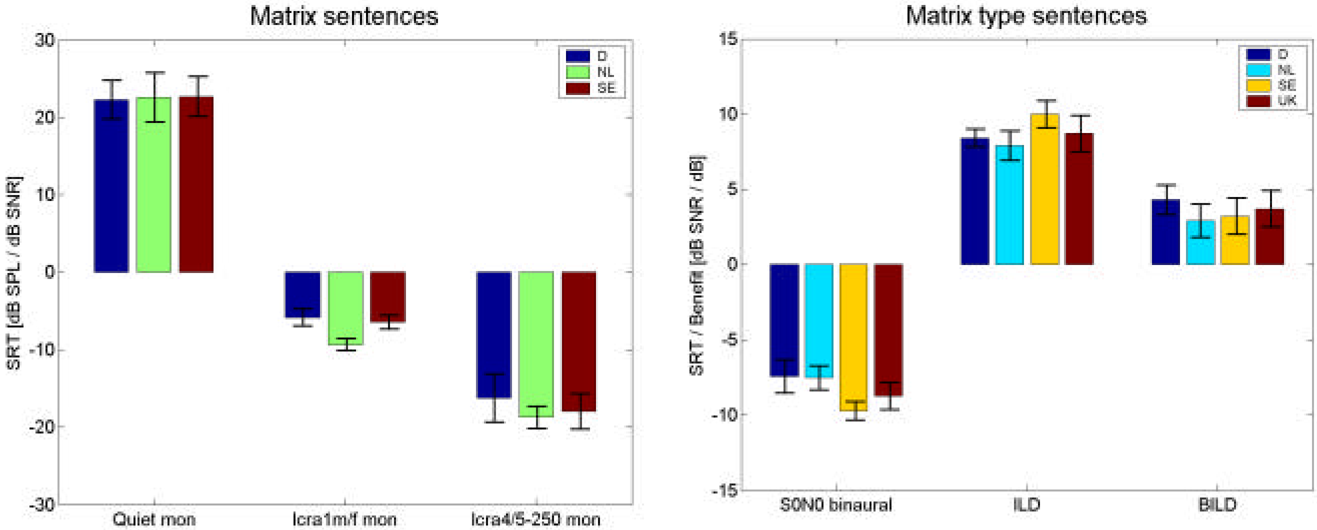


Fig. 2: Left panel: Mean country-specific monaural normal-hearing SRT data ***and standard deviations of*** *Matrix sentences (German: dark blue, Dutch: green, Swedish: red). Three different conditions (SRT in quiet, SRT in non-modulated Icra noise, and SRT in modulated Icra noise).*
Right panel: Mean country-specific binaural normal-hearing SRT data and standard deviations of Matrix sentences (German: dark blue, Dutch: light blue, Swedish: yellow, British: red). Three different conditions (SRT in S_0N_0, ILD, and BILD).

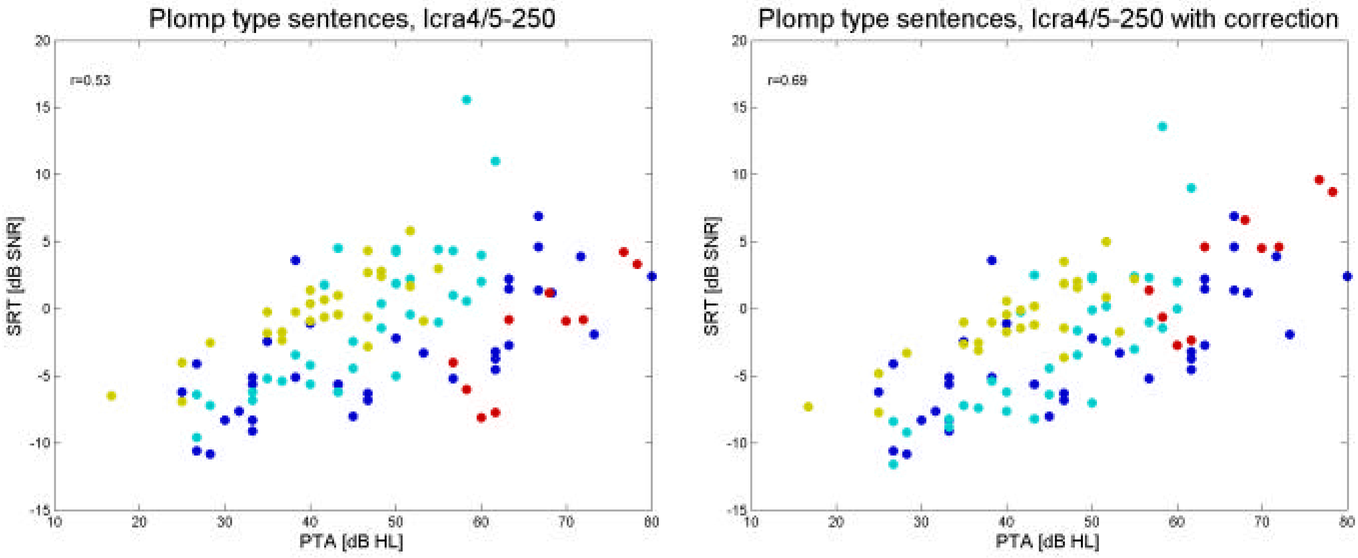
As shown in the figures, the differences across countries are smaller compared to the Plomp type sentences data. This can be explained by the fact that for the Matrix sentence tests the same measurement procedure was used in all countries and the only difference apart from the language itself was the speaker of the test.

Usage of cross-validation data – Results with hearing-impaired listeners

Since the general aim is to achieve comparable results of audiological tests across countries also for speech tests, the country-specific normal-hearing cross-validation data can be used to equalize country-specific differences.

As a first approach the normal-hearing differences are applied as linear correction terms to the hearing-impaired sentence intelligibility data obtained in the multi-centre study.

Fig. 3 shows the speech intelligibility data obtained with Plomp type sentences with hearing-impaired listeners within the multi-centre study of HearCom. The left panel shows the raw individual data, the right panel shows the same data but with the normal-hearing cross-validation data applied as linear correction. The individual SRTs are given for the different pure tone averages per subject (PTA, average across 1, 2, and 4 kHz). Dark blue points: German, light blue: Dutch, yellow: Swedish, and red: British data.



.Fig. 3: Individual hearing-impaired SRT data with Plomp type sentences. Left panel: raw data. Right panel: Cross-validation data applied as correction. Data are given for different PTA (average across 1, 2, 4 kHz). Dark blue: German, light blue: Dutch, yellow: Swedish, red: British data.

The correlation of PTA and SRT data could be raised from r=0.53 to r=0.69 by applying the normal-hearing cross-validation data as simple linear correction.

Conclusions and outlook

The analysis of the normal-hearing cross-validation data with two types of sentence intelligibility tests (Plomp type and Matrix sentences) has shown that some of the country-specific differences can be explained by procedure differences like word scoring versus sentence scoring. Since the procedure differences are less in the Matrix sentences, also the country-specific differences are smaller in these sentences.

It was shown that the normal-hearing cross-validation data can be used to improve international comparability of speech intelligibility results of hearing-impaired listeners. In this study, only a simple linear correction was applied here. It may be that more sophisticated approaches may even enlarge comparability (e.g. based on intelligibility predictions).

Acknowledgements

The study was supported by the European community: HearCom (FP6–004171). The contributions of the project partners VU Amsterdam (NL), AMC Amsterdam (NL), Linköping University (SE), and University of Southampton (UK) are kindly acknowledged.

Literature

Boothroyd A, Nittrouer S (1988) Mathematical treatment of context effects in phoneme and word recognition. J. Acoust. Soc. Am. 84(1), 101-114.

Brand T, Kollmeier B (2002) Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. J. Acoust. Soc. Am. 111(6), 2801-2810.

Dreschler WA, Verschuure H, Ludvigsen C, Westermann S (2001) 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.

HearCom Deliverable D-1-2 (2005) Report on the proposed set of communication performance tests. www.hearcom.org

Wagener KC, Brand T, Kollmeier B (2006) The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. Intern J Audiol 45(1), 26-33.