



**FP6–004171 HEARCOM**

**Hearing in the Communication Society**

**INTEGRATED PROJECT**

**Information Society Technologies**

**D-1-7: Report on normalization data and cross-language comparison for sentence tests**

Contractual Date of Delivery:	01-01-2008 (+45 days)
Actual Date of Submission:	14 February 2008
Editor:	Kirsten Wagener, DE-HZO
Sub-Project/Work-Package:	SP1/WP1
Version:	Final 1.0
Total number of pages:	12

<b>Dissemination Level</b>		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	
Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006) This information is confidential and may be used only for information purposes by Community Institutions to whom the Commission has supplied it		

## Deliverable D-1-7

<b>VERSION DETAILS</b>	
Version:	1.0
Date:	14 February 2008
Status:	Revised

<b>CONTRIBUTOR(S) to DELIVERABLE</b>	
<b>Partner</b>	<b>Name</b>
DE-HZO	Kirsten Wagener / Matthias Vormann / Birger Kollmeier
NL-VUMC	Johannes Lyzenga / Tammo Houtgast
NL-EMC	Jan Koopman
NL-AMC	Thamar van Esch / Wouter Dreschler
SE-LINK	Birgitta Larsby / Mathias Hällgren
UK-ISVR	Mark Lutman / Sheetal Athalye

<b>DOCUMENT HISTORY</b>			
<b>Version</b>	<b>Date</b>	<b>Responsible</b>	<b>Description</b>
0.1		M. Vormann	1 <sup>st</sup> Draft
0.2		B. Kollmeier	revision
1.0		K. Wagener	Revision due to internal review

<b>DELIVERABLE REVIEW</b>			
<b>Version</b>	<b>Date</b>	<b>Reviewed by</b>	<b>Conclusion*</b>
0.1	2008-01-20	Torben Poulsen	Accept after revision
0.1	2008-01-30	Tammo Houtgast	Accept after revision

\* e.g. Accept, Develop, Modify, Rework, Update

## Table of Contents

Pre-Amble .....	4
1 Executive Summary .....	4
2 Introduction .....	5
3 Multi-centre study .....	5
4 Measurements .....	6
5 Results .....	6
5.1 Plomp type sentences: Normal-hearing data .....	6
5.2 Matrix sentences: Normal-hearing data .....	8
6 Usage of cross-validation data – Results with hearing-impaired listeners .....	9
7 Conclusions and outlook .....	11
8 Literature .....	11
9 Dissemination and Exploitation .....	12
9.1 Dissemination .....	12
9.2 Ethical issues .....	12

## Acknowledgement

Supported by grants from the European Union FP6, Project 004171 HEARCOM. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

The contributions of the project partners VU Amsterdam (NL), AMC Amsterdam (NL), Linköping University (SE), and University of Southampton (UK) are kindly acknowledged.

## Pre-Amble

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.

This deliverable presents the inter-language cross-validation analysis of reference data for the particular sentence tests that were used in a multi-centre study within WP2 of HearCom. The multi-centre study was performed in order to evaluate the so-called ‘auditory profile’ (an audiological test battery to characterize the functional hearing ability of an individual). The sentence tests used in this multi-centre study were also measured with normal-hearing listeners in order to determine the international comparability of these tests. Knowledge thereof is needed in order to align the multi-centre results of hearing-impaired listeners across countries. In this deliverable, the inter-language cross-validation of this data is presented as well as the inter-language analysis of the multi-centre study results.

Therefore, there are strong interdependencies with deliverable D-2-3 (2007) and D-2-5 (2008).

Basis of this deliverable is the contribution of Wagener et al (2007) at the EFAS congress in Heidelberg 2007.

## 1 Executive Summary

Within an international multi-centre study sentence intelligibility was determined in different conditions in order to ensure the comparability of the normative speech intelligibility data across four languages (i.e., German, Swedish, British English and Dutch): So-called everyday sentences (or Plomp type sentences) were used to determine 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 and in modulated speech shaped ICRA5-250 noise (modulations simulate one interfering talker). Syntactically fixed, but semantically non predictable sentences (“Olsa/Matrix test sentences”) were used to determine binaural aspects of speech intelligibility like intelligibility level difference (ILD) and binaural intelligibility level difference (BILD). The measurements were performed both with normal-hearing and with hearing-impaired subjects in four

different countries (Germany, Netherlands, Sweden, and UK) under comparable conditions.

The data was evaluated with regard to inter-language cross-validation analysis of reference data for the particular sentence tests as well as to inter-language analysis of the multi-centre study results.

It can be concluded that the inter-language differences are smallest in Matrix tests due to smallest procedure differences. Differences in hearing-impaired data across countries minimize when using the inter-language differences based on normal-hearing data as correction terms in the hearing-impaired data.

## 2 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.

In order to reach highest comparability across European countries the European project HearCom (Hearing in the communication society, FP6–004171) tries to establish qualitative and quantitative criteria for speech intelligibility tests. These criteria are fulfilled by a predefined set of speech tests in different European languages, i.e., both tests taken from the literature and tests that have been developed within the HearCom project. Once these criteria are established (e.g., by a standardization body or by a de-facto standard) it is assumed that similar tests in more languages can be added that both fulfill these criteria and reach the same level of comparability as the existing tests.

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.

## 3 Multi-centre study

The sentence intelligibility measurements were performed as a part of a multi-centre study in WP2 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.

## 4 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). For ILD and BILD measures, also the ICRA1 noise was used.

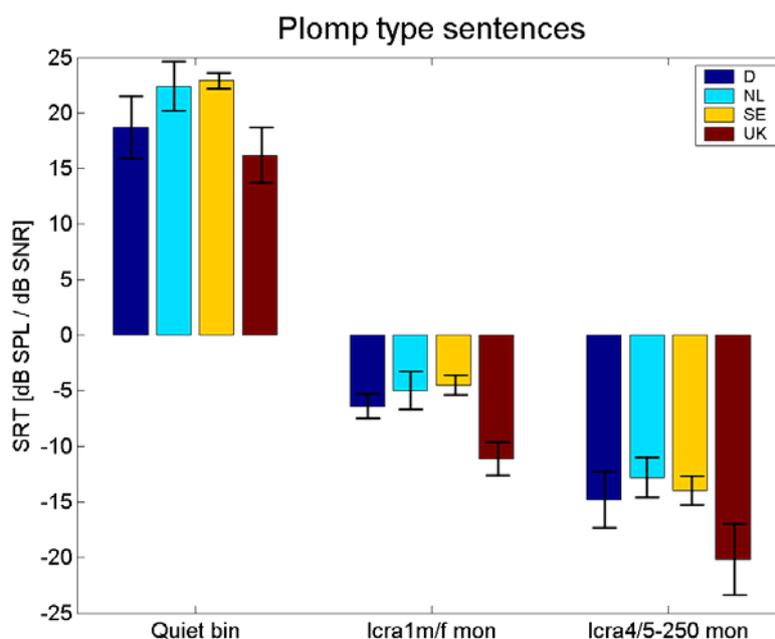
All measurements were performed via free-field equalized Sennheiser HDA200 headphones. The binaural measurements were performed with virtual acoustics.

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").

## 5 Results

### 5.1 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.



**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).**

Slightly different numbers of subjects participated in the measurements. In the German tests, 20 subjects participated in the quiet condition and 13 subjects participated in both noise conditions. In the Dutch and Swedish tests, 20 subjects participated in the noise conditions; 17 (NL) and 6 (SE) participated in the quiet condition. In the English test, 15 subjects participated in all conditions.

The differences in SRT results across languages are statistically significant within each condition due to a one way ANOVA (1% error probability).

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. That means that in the British test not all words per sentence are scored but particular key words per sentence are scored. All tests measure 50% correct score as basis for SRT, i.e. 50% of sentences are correct in case of sentence scoring, 50% of words are correct in case of word scoring, and 50% of key words are correct in case of key word scoring. There is one additional difference between the Dutch and all other tests: in all tests except the Dutch all presented words/sentences are used to calculate the SRT. In the Dutch test, only the last 8 sentences per list were used to calculate SRT. 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

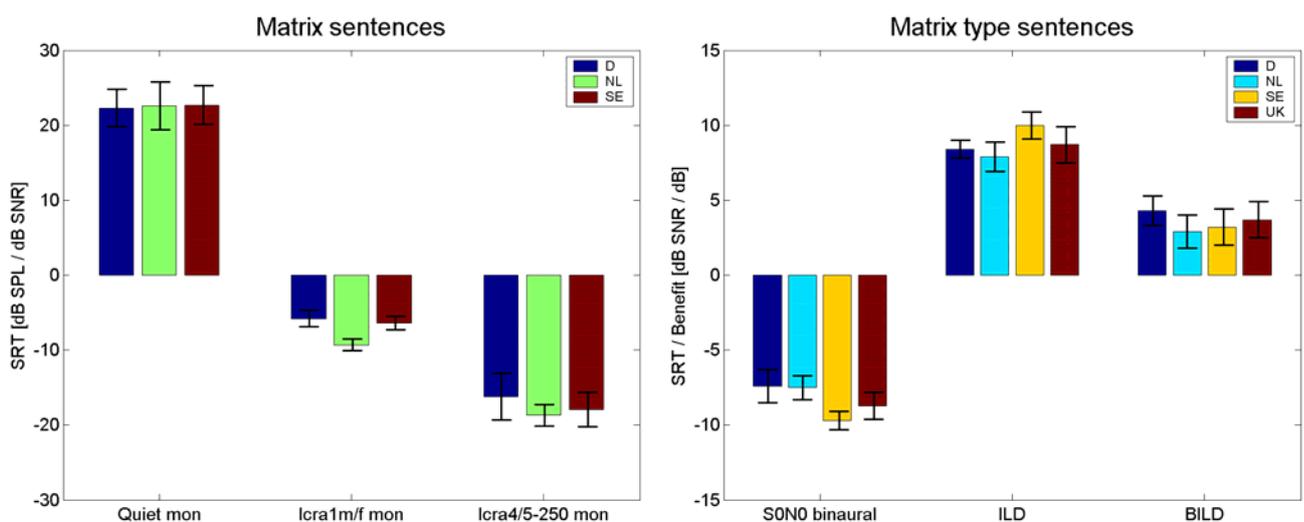
by procedure A1. The Dutch test uses a 1up-1down adaptive procedure with fixed step size 2 dB including the first sentence per list to be repeated as level off sentence. As a consequence of the different languages, the speakers differ across tests (Dutch and Swedish: female speaker, German and British: male speaker).

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 & Nitttrouer 1988), the results are similar to the Dutch results.

## 5.2 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).**

Different numbers of subjects participated in the measurements. In the German tests, 8 subjects participated in the quiet and ICRA1 condition and 7 subjects participated in the ICRA5-250 condition. In the Dutch and Swedish tests, 20 subjects participated in all conditions except for the Dutch quiet condition: 13 subjects participated in these measurements.

The differences in SRT results across languages are statistically significant in condition ICRA1, whereas there are no significant differences in condition quiet and ICRA4/5-250 due to a one way ANOVA (1% error probability).

Although the ILD and BILD measures are difference values between two measured SRT results (and thus any language effects are excluded) there are statistically significant differences across languages both for ILD and BILD measures.

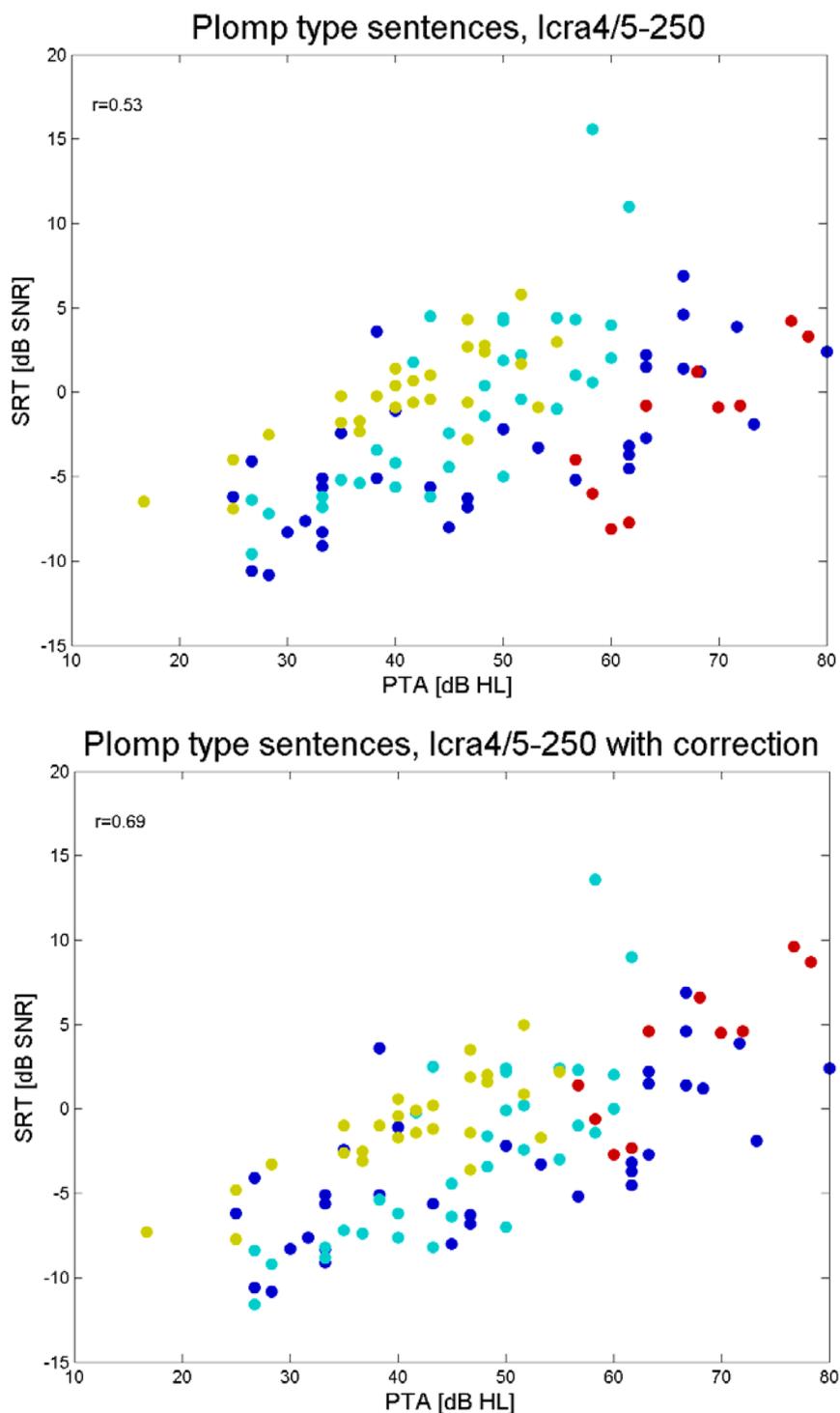
The differences across countries are smaller compared to the Plomp type sentences data. This can be explained by the fact that for the Matrix sentence tests the same measurement procedure was used in all countries and the only difference apart from the language itself was the speaker of the test.

## **6 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.

Fig3 shows the speech intelligibility data obtained with Plomp type sentences with hearing-impaired listeners within the multi-centre study of HearCom. The upper panel shows the raw individual data, the lower 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. Upper panel: raw data. Lower 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 significantly raised from  $r=0.53$  to  $r=0.69$  by applying the normal-hearing cross-validation data as simple linear correction (error probability 5%, assuming independent samples which is not really given in this case).

## 7 Discussion and conclusions

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).

## 8 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](http://www.hearcom.org)

HearCom Deliverable D-2-3 (2007) *Report about the results of the multi-centre evaluation of the Auditory Profile*. [www.hearcom.org](http://www.hearcom.org)

HearCom Deliverable D-2-5 (2008) *Optimized, final set of impairment tests included in the Auditory Profile*. In preparation.

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.

Wagener KC, Brand T, Kollmeier B (2007) *International cross-validation of sentence intelligibility tests*. Proceedings of EFAS conference 2007, in press.

## 9 Dissemination and Exploitation

### 9.1 Dissemination

Parts of the data of this deliverable were already disseminated to the audiological scientific community on the EFAS congress in June 2007 by oral presentation and contribution to the conference proceedings.

The results of the complete cross-language comparison study will be published in a scientific paper in an international journal. Further, the dissemination of the cross-language validation results should be part of HEARCOM eServices, i.e., the webpage should include the paper manuscript for download and should provide links to the appropriate tests in the different languages. The cross-language validation is needed for all attempts to harmonize speech intelligibility procedures across different countries since this data is needed to align the speech intelligibility results of hearing-impaired listeners in case of international comparisons. Hence, the results will provide input to appropriate standardization bodies.

### 9.2 Ethical issues

The data presented in this deliverable were determined as part of the multi-centre study described in D-2-3 (2007). The measurements for the multi-centre study have been approved by the medical ethical committee of all participating centres:

- AMC-NL: MEC 05/127 # 05.17.0934, dated August 3<sup>rd</sup> 2005
- HZO-DE: "Klinische Tests zur Bestimmung individueller Hördefizite und Kommunikationsfähigkeiten", dated November 15<sup>th</sup> 2006
- ISVR-UK: 791, dated February 13<sup>th</sup> 2007
- LINK-SE: M83-06
- VUMC-NL: MEC05/12 - 2006/171, dated November 2<sup>nd</sup> 2006