FP6–004171 HEARCOM

Hearing in the Communication Society

INTEGRATED PROJECT

Information Society Technologies

D-1-9: Report on an optimized inventory of Speech-based auditory screening & impairment tests for six languages

Contractual Date of Delivery:	28 february 2009 (+45 days)
Actual Date of Submission:	9 April 2009
Editor:	K. Wagener
Sub-Project/Work-Package:	WP1
Version:	1
Total number of pages:	20

Dissemination Level		
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)

Deliverable D-1-9

VERSION DETAILS	
Version:	1.0
Date:	8 April 2009
Status:	Final

CONTRIBUTOR(S) to DELIVERABLE	
Partner	**Name**
DE-UOL	Melanie Zokoll
DE-HTCH	Daniel Berg
DE-HZO	Kirsten Wagener
BE-LEU	Sophie Jansen
NL-VUMC	Johannes Lyzenga
PL-IAM	Edward Ozimek
SE-LINK	Birgitta Larsby
UK-ISVR	Mark Lutman

DOCUMENT HISTORY			
Version	**Date**	**Responsible**	**Description**
0.1	3 Mar 09	DE-HZO	Draft for review
1.0	8 April 09	DE-HZO	Revision according reviews

DELIVERABLE REVIEW			
Version	**Date**	**Reviewed by**	**Conclusion***
0.1	2 April 09	Norbert Dillier	Accept
0.1	3 April 09	Tammo Houtgast	Accept

∗ e.g. Accept, Develop, Modify, Rework, Update

Table of Contents

List of Figures

List of Tables

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.

Pre-Amble

The European project HearCom (Hearing in the communication society, FP6–004171) followed the main aim in international audiology that is to obtain similar results across sites when measuring in similar conditions also on an international scale. This aim is rather challenging with regard to speech intelligibility tests since different speakers and languages of course highly influence comparability.

Therefore HearCom established minimum quality requirements for speech intelligibility tests in order to reach highest comparability across European countries.

The available speech intelligibility tests that fulfil these requirements were implemented on a common measurement platform. For this purpose, the originally German Oldenburg measurements applications (OMA) were extended to international use. The outcome of HearCom with regard to speech intelligibility tests available in OMA (i.e. the non-report part of D-1-9) are as following: 18 different speech intelligibility tests are now implemented in OMA. These tests cover 6 languages and include 9 tests with speech material already available before the HearCom project and 9 tests that were also developed/evaluated within HearCom.

This deliverable report presents the inter-language cross-validation analysis of reference data for the speech intelligibility tests that are implemented in OMA.

These data can be used in order to refer speech intelligibility results of hearing-impaired listeners to these normal-hearing reference data in order to minimize differences across languages. This is also a main aspect for the multi-centre study II performed within WP2 of HearCom. Therefore, there are strong interdependencies with deliverable D-2-6 (2009).

1 Executive Summary

This report compares reference speech intelligibility functions of three different types of speech intelligibility tests implemented within the European project HearCom. These data were either collected within HearCom or already available in literature. Comparison was done in order to test the perceptual similarities of these tests across languages. If possible, the same model function was employed to describe the speech intelligibility functions for all tests.

The three types of speech intelligibility tests selected were a digit triplets test, using spoken numbers in a background of noise (screening test of speech recognition under adverse conditions that was also implemented for telephone and internet use and therefore also included in SP5/WP11 eDiagnostics) and two types of sentence intelligibility test: 1) sentence tests with meaningful everyday sentences (so called Plomp type sentences) and 2) unpredictable, non-meaningful sentences (so called Matrix sentences). The sentence intelligibility tests were selected for more precise measurements.

2 Introduction

Speech allows us to communicate with the complex world surrounding us. As long as our auditory system is healthy, we are able to understand our interlocutor even within a noisy environment (for example a crowd of people talking at a high level). For hearing-impaired people, however, such a situation is more difficult. Therefore, a hearing impairment is often primarily observed during communication with interfering noise (the so-called ‘cocktail-party situation’). To gain a better understanding of the effect of hearing impairment on speech perception in everyday listening situations, we have to understand the mechanisms and factors influencing speech recognition in noise. The same is true for the limited effect of modern hearing instruments in improving the situation for hearing-impaired listeners.

As a standard measurement during diagnosis and rehabilitation of hearing diseases, pure-tone audiograms are determined. To complete diagnosis, speech intelligibility tests preferably in noise should also be employed. This is because understanding speech within interfering noise is much more complex than detecting pure tones and thus, represents more adequately the difficult everyday situations of the listener. Speech intelligibility tests in noise help diagnosing above-threshold hearing impairment, judging the individual communication capacity in difficult listening situations, and defining the benefit of hearing aids.

Since the results of such tests depend on how the test is constructed and performed, an internationally comparable quality standard of speech intelligibility tests should be established. This also enables several countries to contribute to international multi-center studies.

In order to achieve the highest possible comparability across European countries the European project HearCom established qualitative and quantitative criteria for speech intelligibility tests (HearCom deliverable D-1-3, 2006). These criteria were already fulfilled by some speech tests in different European languages. The HearCom project extended the arsenal of speech intelligibility tests for European languages to six languages. The following partner sites contributed to this test arsenal (within WP1 and in collaboration with WP7 to increase the number of languages): Belgium: University of Leuven, Lab. Exp. ORL, Germany: Hörzentrum Oldenburg and Center of excellence HörTech, Greece: Institute for Language and Speech Processing, The Netherlands: Academic Medical Centre Amsterdam and VU University Medical Centre Amsterdam, Poland: A. Mickiewicz University, Institute of Acoustics, Sweden: Linköping University, Dept. of Audiology, United Kingdom: University of Southampton, Institute for Sound and Vibration Research.

Within HearCom three different types of speech intelligibility tests were selected (HearCom deliverable D-1-2, 2005) and compared across languages: the digit triplets test (Smits et al, 2004) for screening purposes and two types of sentence tests for extended measurements. The first type of sentence tests consists of everyday sentences and is called Plomp type sentences in the following (e.g., Plomp and Mimpen, 1979). The second type of sentence tests consists of short unpredictable sentences, so-called Matrix sentences (Hagerman, 1982). The two types of sentence tests were chosen for more precise measurement. They assess phonetic, lexical and, in case of the Plomp type sentence tests, also semantic information and thus incorporate the whole language system.

The digit triplets tests were developed as fast screening tests for end user usage via telephone and internet, while the sentence tests were developed for professional usage. In order to disseminate the speech intelligibility tests, they were implemented as separate measurement modules in a common measurement platform within HearCom: the Oldenburg Measurement Applications (OMA, HearCom deliverable D-1-1 and D-1-4). All tests have the same look and the data is stored and can be analysed within one single database. At the moment OMA and the different modules are available as a research version. Parts of the test arsenal (the German sentence tests) are also already available as CE certified versions.

This report summarizes the reference data of this speech intelligibility test arsenal determined with normal-hearing listeners. This data is used to refer the results of hearing-impaired listeners to normal-hearing data and to quantify the test (and language) specific differences.

3 OMA

The Oldenburg Measurement Applications (OMA) were developed with the aim to offer audiologists a convenient instrument to perform audiological measurements using a flexible and modular system, independent of whether their workplace is a clinic, a research facility, or a hearing aid acousticians office (or hearing aid dispenser). OMA is a modular software package which consists of one basic setup containing all shared components (e.g. database, hardware modules) and various test procedures which can be added individually. Due to its modular structure, new measurement procedures can relatively easy be implemented in the system.

A modern personal computer equipped with a high-quality sound card is a prerequisite for OMA. The sound signal is output from the PC and connected to the external input of the audiometer. The audiometer is software controlled via a serial PC interface. Customizations for a number of audiometers on the market currently exist. For research purposes also an integrated virtual audiometer can be used.

The use of one common software platform for all tests offers several advantages for multi-centre purposes: all tests can be easily operated because they have the same look and feel. All patient data are stored in the same database which allows for easy evaluation and analysis of test results. All hardware components as well as calibration data are shared across the different modules.

4 Comparison across languages

All speech tests in this report focus on measuring speech intelligibility in noise, i.e. the speech reception threshold (SRT: signal-to-noise ratio that yields 50% intelligibility). Different tests can be compared by, for example comparing the reference speech intelligibility functions determined with normal-hearing listeners. Ideally, these functions are obtained in applying the same model function and therefore the parameters of the model functions can be compared as standardized reference values. For most speech intelligibility tests in this report the logistic model function was employed (Equation 1):

$$p\,(L, SRT, s) = \frac{1}{1 + e^{4s(SRT - L)}} \qquad (1)$$

p denotes the mean probability that the words of a sentence are correctly repeated by the subject, if the sentence was presented with a signal-to-noise ratio L. The speech reception threshold SRT and the slope s of the intelligibility function at the SRT describe the entire function. SRT and s therefore act as reference values of these speech intelligibility tests and will be compared for the three types of tests under consideration.

In order to compare the sentence intelligibility tests with each other that differ in several details, re-analyses of the original data were made whenever possible/or necessary (e.g., Hagerman's original data of the matrix-type sentence test in order to compare the data with other matrix-type sentence tests) for comparison.

4.1 Digit triplets tests

The digit triplets tests determine speech intelligibility in a speech shaped noise using combinations of three digits per stimulus. It is a relatively quick and easy test for screening purposes with high sensitivity and specificity. The test can be repeated with the same subject because there is a low risk of remembering the triplets. Additionally, it is suitable for a large number of hearing impaired subjects and also non native listeners (Warzybok et al, 2007). The test is performed in a closed response format, i.e. the listener has to mark the digits that were understood out of given response alternatives. These alternatives can be given in form of a telephone pad and since no instructor is needed to administer the test, it can be done by telephone or via internet (e.g., Smits et al., 2004). At the moment the digit triplets tests are available in six languages [Dutch: Dutch 3-digit test (VU, Smits et al., 2004), English: English 3-digit test

(Phipps, 2007), French: French 3-digit test (Jansen, 2008), German: Digit triplets test (Wagener et al., 2005, 2006), Polish: Polish Digit Triplets Test (PDTT, Ozimek et al. 2009, HearCom deliverable D-1-11, 2008), and Swedish: Swedish 3-digit test, (Hällgren and Larsby, 2009, HearCom deliverable D-1-8, 2009)]. The number of syllables varies across languages since in general the tests exclude multisyllabic digits (e.g. 'sieben' or 'seven'). Only the Polish test includes multisyllabic digits because excluding these digits would mean to exclude more than half of the material.

All tests are already implemented in OMA. The English, French, Polish and Swedish digit triplets tests were developed and evaluated during the HearCom project.

There is one additional test that was also developed and evaluated within HearCom with regard to usage via headphones and as Internet application, i.e. a Greek digit triplets test (HearCom deliverable D-1-6b, 2009). Although this test is currently not available as OMA application or telephone test the reference data will be reported in the cross-language comparison. This test only consists of disyllabic since only '4' consists of only one syllable in Greek and therefore was excluded from the material.

4.1.1 Digit triplets tests: reference values via headphones

Table 4.1 presents the reference values defining the mean list-specific intelligibility functions for the digit triplets tests: mean list-specific slope s and mean Speech Reception Threshold SRT (i.e. signal-to-noise ratio that yields 50% intelligibility). The scoring method was triplet scoring (i.e., all digits of the triplet had to be correct and in correct order to be counted as correct answer). On the left hand side the data determined during headphone measurements are given, on the right hand side the data via telephone are given. All measurements were performed monaurally. As Ozimek et al. (2009) pointed out, it should be emphasized that due to certain differences in linguistic structure and measurement methods used in different laboratories, only a general comparison is possible across languages.

The steepest slope of the speech intelligibility functions can be observed for the French and Swedish digit triplets test (27.1, and 24.2 %/dB, respectively), whereas the shallowest slope is observed for the Dutch and English digit triplets tests (16.0, and 17.0 %/dB, respectively). The mean SRTs for the different digit triplets tests are also quite consistent across languages with a significantly higher SRT of the Greek test (-7.7 dB SNR). Functions characterizing the German and Polish tests are almost identical.

Table 4.1: Reference values of digit triplets tests via headphones and telephone. Given are: mean list-specific slopes *s* and mean SRTs.

language	speaker gender	via headphones s [%/dB]	via headphones SRT [dB SNR]	via telephone s [%/dB]	via telephone SRT [dB SNR]	references
Dutch	f	16,0	-11,2	20,0	-7,1	Smits et al. (2004)
British English	f	17,0	-11,1	/	-5,5	Phipps (2007); Lutman (2009)
French	f	27,1	-10,5	17,1	-6,4	Jansen (2008)
German	f	19,6	-9,3	17,9	-6,5	Wagener et al. (2006)
Greek	f	16,8	-7,7			HearCom D-1-6b
Polish	m	19,4	-9,4	/	-7,4	Ozimek et al. (2009a); HearCom D-1-11
Swedish	f	24,2	-6,9	24,3	-4,3	Hällgren and Larsby (2009), HearCom D-1-8

The speech intelligibility functions (speech intelligibility performance depending on signal-to-noise ratio, Fig. 4.1) show a steep slope for all languages, hence enabling accurate determination of the speech reception threshold.

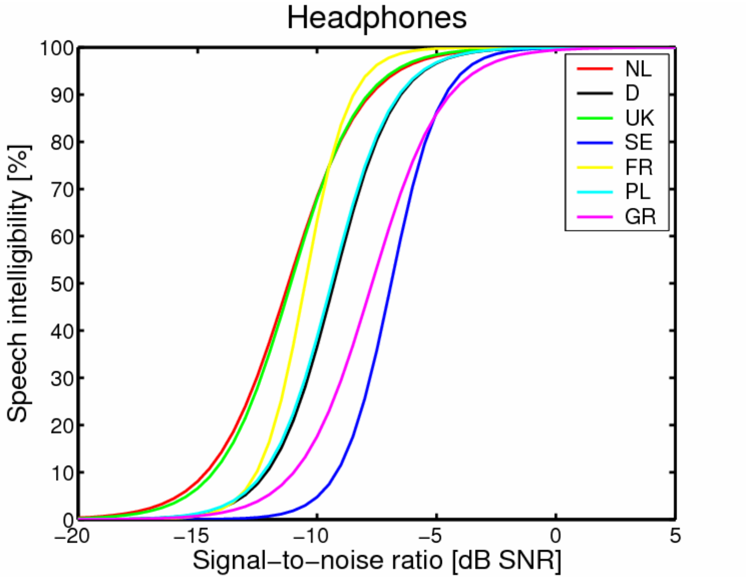

Fig. 4.1 Intelligibility functions of Dutch, German, English, Swedish, French, Polish, and Greek digit triplets tests via headphones.

4.1.2 Digit triplets tests: reference values via telephone

Table 3.1 also presents the reference values defining the mean list-specific intelligibility functions for the digit triplets tests via telephone (right hand side). The scoring method was also triplet scoring. Similar results as for headphone usage can be seen. The steepest slope of the speech intelligibility functions was observed for the Swedish digit triplets test (24.3 %/dB) with nearly the same value as observed for the headphone presentation. There is no slope available for the Polish test via telephone. The mean SRTs for the different digit triplets tests are also quite consistent across languages for telephone usage with a significantly higher SRT of the Swedish test (-4.3 dB SNR). For telephone usage the intelligibility functions of the German and French tests are almost identical.

In the moment, no normal-hearing data of the English digit triplets test via telephone are available. Due to Lutman (2009), the normal-hearing data via telephone are also about 4 dB higher compared to the presentation via headphones.

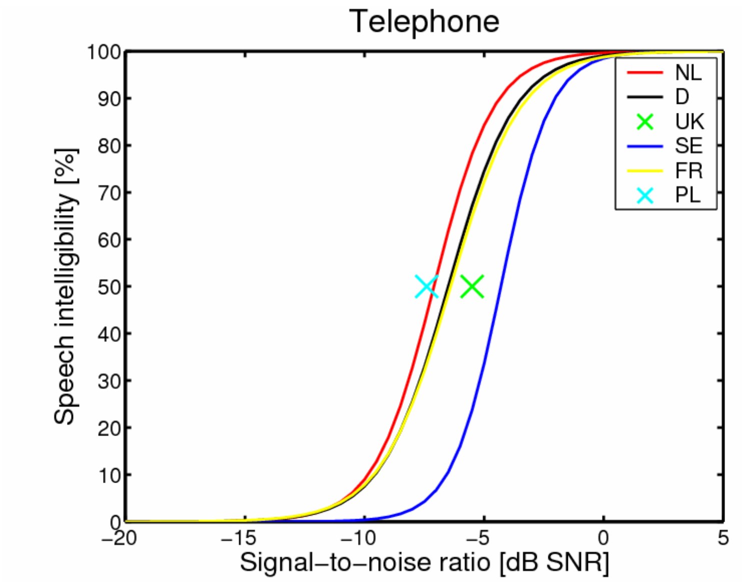

Fig. 4.2 Intelligibility functions of Dutch, German, Swedish, and French digit triplets tests via telephone. The crosses denote the SRTs of the English and Polish digit triplets tests via telephone, where no slope is available.

As can be seen in Figure 4.2 the speech intelligibility functions for the Dutch, French, and German digit triplets tests are almost identical. The speech intelligibility function of the Swedish digit triplets test has a similar slope to the other languages, the higher SRT, however, results in a shift of the intelligibility function compared to the other languages.

4.1.3 Influence of headphone/telephone presentation

The digit triplets tests show an increase in SRT of 2 up to 4 dB SNR when conducted via telephone compared to headphone presentation. The slopes of the different languages stayed about the same for the Swedish digit triplets test, became higher for the Dutch test and shallower for the remaining languages.

Generally, speech intelligibility functions become more consistent over languages when determined via telephone, indicating that language specific influences especially on the SRT are reduced.

4.2 Plomp type sentences

Plomp type sentence tests consist of short meaningful sentences. The measurements are normally performed in speech-shaped interfering noise, but are partly also suitable to be presented in quiet. The advantage of speech intelligibility tests with Plomp type sentences is that they need no training. The meaningful sentences result in the disadvantage, that the test lists usually cannot be used twice with the same subject within a certain time interval (i.e., shorter than half a year). The meaningful sentences can easily be memorized or particular words can be guessed from the context, which would affect the SRT result. However, this ability of guessing from the context is also used in normal conversations and therefore influences speech intelligibility in daily life. As the amount of test lists is limited, these sentence tests are not suitable when many speech intelligibility measurements have to be performed with the same listener, e.g., during hearing instrument fitting or in research. Plomp type sentences are available in six European languages on OMA [Dutch: Plomp sentences (Plomp and Mimpen, 1979) and Versfeld sentences (Versfeld et al, 2000), French: French Intelligibility Sentence Test (Luts et al., 2008), English: Bamford-Kowal-Bench (BKB) test (e.g., Cattermole, 2003), German: Göttingen sentence test (Kollmeier and Wesselkamp, 1997), Polish: Polish sentence test (Ozimek et al., 2006), Swedish: hearing in noise test (HINT, Hällgren et al., 2006)]. The Polish sentence test was developed and evaluated during the HearCom project.

4.2.1 Plomp sentences: reference values

Table 4.2 presents the reference values defining the mean list-specific intelligibility functions for the Plomp type sentences: mean list-specific slope s and mean Speech Reception Threshold SRT (i.e. signal-to-noise ratio that yields 50% intelligibility). In general, all speech intelligibility functions were obtained in monaural measurements via headphones in speech-shaped stationary noise. Only the Swedish test was measured in free field via a loudspeaker from the front in 1m distance and listening with both ears. The frequency spectrum of the noise was always shaped according to the long-term spectrum of the male or female speaker respectively of the sentence test material. The scoring method was either sentence scoring (left hand side of table) or word scoring (right hand side of table). For sentence scoring, only the whole sentence repeated correctly is considered as a correct answer. For word scoring, each word repeated correctly is counted.

The results of the different Plomp type sentence tests are rather similar. For sentence scoring, the slope of the speech intelligibility functions range from 15.2 %/dB (Dutch Versfeld sentences) up to 25.5 %/dB (Polish sentence test). The slope of the Polish test is very high in comparison to the slopes of the other European languages. The SRTs of the different Plomp type sentences are more consistent. The highest SRT was found for the Swedish test (-3.0 dB SNR, for sentence scoring) and the lowest SRT for the Polish sentence test (-7.5 dB SNR, word scoring).

Table 4.2: Reference values of Plomp type sentences. Given are: mean list-specific slopes *s* and mean SRTs. The values are given for sentence scoring (left hand side) and word scoring (right hand side).

		sentence scoring		word scoring		
	speaker	s	SRT	s	SRT	
language	**gender**	**[%/dB]**	**[dB SNR]**	**[%/dB]**	**[dB SNR]**	**references**
Dutch	f	15,9	-5,9			Plomp and Mimpen (1979)
Dutch	f	16,6	-4,1			Versfeld et al. (2000)
Dutch	m	15,2	-4,0			Versfeld et al. (2000)
British English	m			/	-6,9	Cattermole (2003)*
French	f	20,2	-7,4			Luts et al. (2008)
German	m			19,2	-6,2	Kollmeier and Wesselkamp (1997)
Polish	m	25,5	-6,1	20,8	-7,5	Ozimek et al. (2009b)
Swedish	f	17,9	-3,0	15,4	-5,0	Hällgren et al. (2006)

*key word scoring

As expected, word scoring results in lower SRTs and lower slopes of the speech intelligibility functions (compare data for sentence and word scoring methods for the Swedish and Polish Plomp type sentences).

Figure 4.3 shows the intelligibility functions of the different tests and conditions. Here, it can be seen, that the different results across countries can partly be explained by the procedure differences. One difference is the scoring method: Most of the tests applied only sentence scoring, the German test applied only word scoring, and the British test applied key word scoring. For the Polish and Swedish Plomp type sentences, both methods can be compared and thus, show how much of the variance between tests can have resulted from the scoring method. 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 an 1up-1down adaptive procedure with a fixed step size of 2 dB. As a consequence of the different languages, the speakers differ across tests, which might have caused much of the variance between the different tests. Please note that the Dutch Versfeld sentences (2000) suggest that speaker gender and identity only marginally contribute to the variance between tests. But actually the Versfeld sentences were optimized in a way that only sentences with similar SRT values of both speakers were included in the final test material.

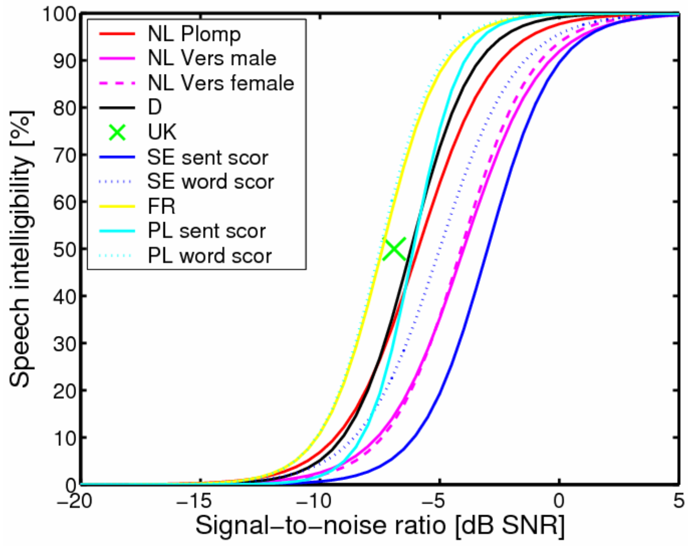

Fig. 4.3 Intelligibility functions of Dutch, German, Swedish, French, and Polish Plomp type sentences. The cross indicates the SRT of the English test, for which no slope is available.

4.3 Matrix-tests

Because of the constraints of Plomp type sentences regarding repeated use, a second type of sentence intelligibility tests has been included in the test arsenal, the so-called Matrix sentences. These tests consist of syntactically fixed, but semantically unpredictable (nonsense) sentences, i.e., sentences with a fixed grammatical structure but using words that do not necessarily make sense in their respective combinations. The word material consists of 10 different word alternatives within the word groups 'name', 'verb', 'numeral', 'adjective', 'object' (i.e. 50 words altogether). The sentences are normally presented in speech-shaped interfering noise, which is either male or female frequency shaped regarding the speaker's gender of the applied sentence test. This type of test may also be presented in quiet. The advantage of this type is that the test lists can be used repeatedly. The disadvantage is that the listener first has to get familiar with the structure of the test and the limited word material. This causes a learning effect and has to be overcome by appropriate training. Meanwhile Matrix sentences are available in OMA in five languages [Dutch: English: English Matrixtest, Hewitt (2007), French: French Matrixtest, (Jansen 2009), German: Oldenburg sentence test (Wagener et al., 1999a-c), Polish: Polish Matrix test (HearCom deliverable D-1-10, 2008), Swedish: Hagerman test (Hagerman, 1982)]. The Matrix tests in English, French and Polish language were developed and evaluated during the HearCom project. The Matrix test for Dutch language is already available in OMA but the evaluation is not yet completed. The Polish test is already prepared for OMA use but due to open license questions of the material it is currently not available in OMA.

4.3.1 Matrix sentences: reference values

Since Hagerman (Hagerman, 1982) originally used a different model function to describe the intelligibility function of the Swedish Matrix test data, comparison with those data was difficult. Wagener (2004) reanalysed the raw data of Hagerman with the logistic model function (Equation 1). Thus, when comparing the results for the Swedish Matrix test, these values (Wagener 2004) were used.

The experimentally determined mean list-specific slopes s and the mean SRTs for these tests are given in Table 4.3 for each of the Matrix sentence tests, separately. All experiments were conducted monaurally via headphones.

Table 4.3 Reference values of Matrix sentences. Given are: mean list-specific slopes *s* and mean SRTs. The values are given for sentence scoring (left hand side) and word scoring (right hand side).

language	speaker gender	sentence scoring s [%/dB]	SRT [dB SNR]	word scoring s [%/dB]	SRT [dB SNR]	references
Dutch	f	17,0	-7,8			HearCom, preliminary data
British English	f			12,8	-8,0	Hewitt (2007)
French	f			14,0	-6,0	HearCom
German	m			17,1	-7,1	Wagener et al. (1999c)
Polish	f	21,8	-8,0	17,1	-9,6	HearCom D 1-10
Swedish	f,m			16,0	-8,1	Hagerman (1982), data reanalyzed by Wagener (2004)

The mean SRTs of the different Matrix tests are very similar, although obtained in different laboratories. The French Matrix test shows the highest SRT (-6.0 dB SNR), the Polish Matrix test shows the lowest SRT (-9.6 dB SNR). Generally, the speech intelligibility functions of the Matrix tests in different languages are very similar (Fig. 4.4).

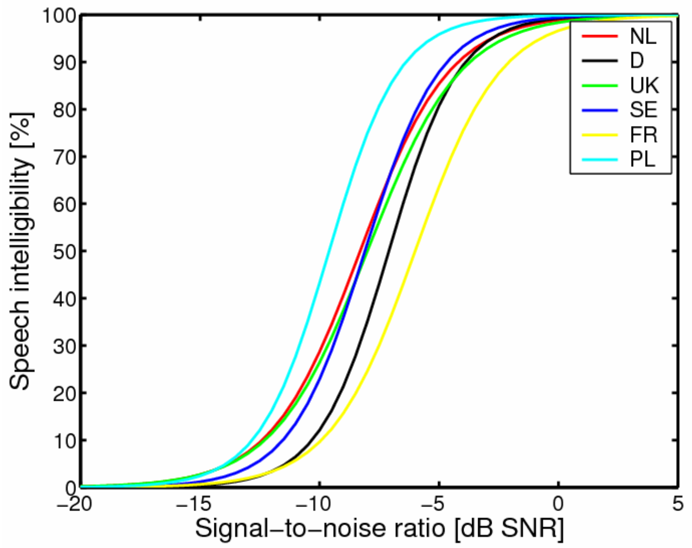

Fig. 4.4 Intelligibility functions of Dutch, German, English, Swedish, French, and Polish Matrix sentences.

4.4 Comparison across tests

Generally, all speech intelligibility tests in the HearCom arsenal show steep slopes, indicating a high accuracy in determining SRTs. The SRTs are always lower in the digit triplets tests compared to sentence tests. This difference between digit triplets tests and sentence tests is due to the limited test word material in the digit triplets test that is composed of only 10 words, which markedly improves speech intelligibility (e.g., Smits et al., 2004).

When comparing the normal hearing reference values across languages, the digit triplets tests via telephone as well as the Matrix sentences show best comparability. This indicates that applying a common test format and performing a more or less equivalent optimization increases comparability of results across languages.

5 Dissemination and Exploitation

The speech materials and the reference values described in this report are mainly of interest to audiologists (researchers and clinicians). The summary of reference values may act as database to refer international multi centre study results with hearing-impaired listeners to the respective normal-hearing listeners and therefore can be used to minimize across language differences that are not due to the hearing impairment. The data base can also act as benchmark for newly developed speech intelligibility tests in other languages. Since all tests presented here fulfil the so-called 'HearCom minimum requirements for speech tests' other tests that will be developed with regard to these requirements should show comparable reference values. A more or less step-by-step guideline for developing speech intelligibility test due to the HearCom requirements is given in HearCom deliverable D-1-3 (2006).

All described speech intelligibility tests are implemented in a common measurement platform (OMA) and are available as a research version by mid of this year (most procedures are already available).

The digit triplets test is used as a self screening test via telephone and via Internet. The telephone version is already available to the public in Dutch, English, and German. The French version will be launched soon. The Swedish and Polish version is still under construction. The Internet version of the digit triplets test is publicly available in five languages (Dutch, German, English, Swedish, and French) on the HearCom website www.hearcom.eu. The Polish and the Greek version are still password protected since the validation of the Internet versions are not yet completed.

6 Conclusions

This report gives an overview on the speech intelligibility tests in noise included in the HearCom arsenal in order to establish comparable tests across languages that fulfil the HearCom quality standard. As the outcome of HearCom within WP1, 18 different speech intelligibility tests are now implemented in OMA. These tests cover 6 languages and include 9 tests with speech material already available before the HearCom project and 9 tests that were also developed/evaluated within HearCom.

The HearCom OMA implementations together with the summary of normal-hearing reference data in this report helps to perform international multi centre studies in a simple harmonized way and helps to analyze hearing-impaired listeners' data and reduce differences across languages that are not due to hearing impairment.

7 References

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.

Cattermole, C, 2003. Comparison Of Speech Recognition In Noise With Sound Field And Earphone Presentation. MSc thesis, University of Southampton.

Hällgren, M., Larsby, B., Arlinger, S., 2006. A Swedish version of the hearing in noise test (HINT) for measurement of speech recognition. Inter. J. Acoust. 45, 227-237.

Hällgren, M., Larsby, 2009. The Swedish digit triplet test; optimization and validation of broadband and telephone versions (in preparation).

Hagerman, B., 1982. Sentences for testing speech intelligibility in noise. Scand. Audiol., 11, 79-87.

HearCom deliverable D-1-1, 2005. Communication self-screening test for two languages operating on a PC-based system.

HearCom deliverable D-1-2, 2005. Report on the proposed set of communication performance tests.

HearCom deliverable D-1-3, 2006. Protocol for implementation of communication tests in different languages.

HearCom deliverable D-1-4, 2006. First version on Internet screening tests in three languages (final report and demonstrator).

HearCom deliverable D-1-6b, 2009. Development of a digit-triplet test in modern Greek.

HearCom deliverable D-1-8, 2009. User-group specific performance in screening tests in different countries.

HearCom deliverable D-1-10, 2008. Development and evaluation of Polish matrix sentence test.

HearCom deliverable D-1-11, 2008. implementation of Polish Digit Triplet Test.

HearCom deliverable D-2-6, 2009. Report about the validation of the Hearcom auditory profile in four countries. In preparation.

Hewitt, D. R., 2007. Evaluation Of An English Speech-In-Noise Audiometry Test. MSc thesis, University of Southampton.

Jansen, S. (2008). Unpublished Internal Report. (cited from D-1-8)

Jansen 2009 Personal communication

Kollmeier, B., Wesselkamp, M., 1997. Development and evaluation of a sentence test for objective and subjective speech intelligibility assessment. J. Acoust. Soc. Am. 102 (4), 1085–1099.

Lutman, M., 2009. Personal communication.

Luts, H., Boon, E., Wable, J., and Wouters, J. (2008). FIST: A French sentence test for speech intelligibility in noise. Int. J. Audiol. 47: 373-374.

Nelissen M (2000). A Dutch speech-in-noise test. Final thesis Speech- and Language pathology. University of Nijmegen

Ozimek, E., Kutzner, D., Se˛k, A., Wicher, A., Szczepaniak, O., 2006. The Polish sentence test for speech intelligibility evaluations. Arch. Acoust. 31 (4), 431–438.

Ozimek, E., Kutzner, D., Sek, A., Wicher, A., 2009a. Development and evaluation of Polish digit triplet test. Speech Communication (in press)

Ozimek, E., Kutzner, D., Sek, A., Wicher, A., 2009b. Polish sentence tests for measuring the intelligibility of speech in interfering noise. International Journal of Audiology (in press)

Phipps,H.L 2007. Assessment of telephone bandwidth on the English number recognition in noise test. MSc thesis, University of Southampton.

Plomp, R., Mimpen, A.M., 1979. Improving the reliability of testing the speech reception threshold for sentences. Audiology 18, 43–53.

Smits, C., Kapteyn, T., Houtgast, T., 2004. Development and validation of an automatic speech-in-noise screening test by telephone. Internat. J. Audiol. 43, 15–28.

Versfeld, N.J., Daalder, L., Festen, J.M., Houtgast, T., 2000. Method for the selection of sentence material for efficient measurement of the speech reception threshold. J. Acoust. Soc. Amer. 107, 1671–1684.

Wable, J., 2001. Normalisation d'un test de reconnaissance de la parole dans le bruit chez le sujet déficient auditif. Cahiers de l'Audition, 14, 29-38.

Wagener, K.C. (2004) Factors influencing sentence intelligibility in noise. Phd thesis. bis-Verlag Oldenburg. ISBN = 3-8142-0897-8

Wagener, K.C., Bräcker, T., Brand,T., Kollmeier, B., 2006. Evaluation des Ziffern-Tripel-Test über Kopfhörer und Telefon. DGA 9. Jahrestagung Köln 2006, TagungsCD, ISBN = 3-9809869-5-0.

Wagener, K., Brand, T., Kollmeier, B., 1999a. Development and evaluation of a German sentence test I: design of the Oldenburg sentence test (in German). Z. Audiol. 38, 4–15.

Wagener, K., Brand, T., Kollmeier, B., 1999b. Development and evaluation of a German sentence test II: optimisation of the Oldenburg sentence tests (in German). Z. Audiol. 38, 44–56.

Wagener, K., Brand, T., Kollmeier, B., 1999c. Development and evaluation of a German sentence test III: evaluation of the Oldenburg sentence test (in German). Z. Audiol. 38, 86–95.

Wagener, K., Eeenboom, F., Brand, T., Kollmeier, B., 2005. Ziffer- Tripel-Test: Spracherverstandlichkeitstest über das Telefon. Proceedings of 8th congress of the German Society of Audiology.

Wagener, K., Josvassen, J.L., Ardenkjaer, R., 2005b. Design, optimization, and evaluation of a Danish sentence test in noise. Internat. J. Audiol. 42 (1), 10–17.

Warzybok, A., Wagener, K.C., Brand, T. 2007. Intelligibility of German digit triplets test by non-native listeners. Proceedings of 8th EFAS congress joined with 10th congress of the German Society of Audiology.