FP6–004171 HEARCOM

Hearing in the Communication Society

INTEGRATED PROJECT

Information Society Technologies

D-1-2: Report on the proposed set of communication performance tests

Contractual Date of Delivery:	31-10-2005 + 45 days
Actual Date of Delivery:	28-11-2005
Editor:	J. Lyzenga
Sub-Project/Work-Package:	SP1/WP1
Version:	2.0
Total number of pages:	28

Dissemination Level		
PU	Public	PU
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-1-2

VERSION DETAILS	
Version:	2.0
Date:	21 November 2005
Status:	final

CONTRIBUTOR(S) to DELIVERABLE	
Partner	**Name**
VUMC	J. Lyzenga. First draft.
HZO	K. Wagener. Edits.

DOCUMENT HISTORY			
Version	**Date**	**Responsible**	**Description**
0.1	16-09-2005	J. Lyzenga	First draft
1.0	18-10-2005	K. Wagener	Various edits
2.0	21-11-2005	J. Lyzenga, K Wagener	Revised version

DELIVERABLE REVIEW			
Version	**Date**	**Reviewed by**	**Conclusion***
1.0	31-10-05	N. Dillier	Accept with suggestions for update
	3-11-05	B. Larsby	Accept; update for small corrections

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

Table of Contents

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.

1 Executive Summary

This report describes the proposed set of communication performance tests as chosen in WP1/SP1. We have chosen to concentrate on three types of tests: a digit triplets test as a quick screening test of speech recognition under adverse conditions, two types of sentence tests for more precise measurements of that ability, and localization tests to estimate spatial hearing abilities. One type of sentence test uses an open response set, the other a closed response set. The digit triplets and sentence tests use spoken text, either numbers or sentences, presented in a background of noise to estimate recognition thresholds. In the localization tests, listeners are required to identify the direction from which a target sound is presented. All tests will be made available in several languages and, when applicable, through a number of channels (audiological work station, telephone, Internet). The languages involved comprise English, Danish, Dutch, French, German, Polish, and Swedish. The results of this report will be input to SP1/WP2 (measures of the auditory profile) and SP5/WP11 (Internet implementations).

2 Introduction

To establish measures of the communication performance of listeners, a number of tests can be used. One of the objectives of the HearCom project, under Sub Project 1, Work Package 1, is to develop tests that enable rapid or precise estimation of a person’s communication performance. These tests should be, or made to be, suitable for a number of countries or language areas. This report describes the proposed set of communication performance tests as chosen in SP1/WP1. We have chosen to concentrate on three types of tests. First, a digit triplets test, using spoken numbers in a background of noise, as a quick screening test of speech recognition under adverse conditions is described. This test will be made available in a number of languages and for a number of different channels such as the Internet, the telephone, and an audiological work station. Second, two types of sentence tests were included for more precise measurements of speech recognition abilities, one using an open response set and one that can also be used with a closed response set. The first type requires a measurement assistant, the second type can be executed in an automated format as well, enabling a larger number of channels through which it can be executed. Third, localization tests are included to estimate spatial hearing abilities. The precise implementation of the localization tests depends on the channel, either Internet or audiological work station, through which it is presented. The results of this report will be input to SP1/WP2, where the tests to measure the auditory profile are selected and validated, and SP5/WP11, where the Internet implementations of the HearCom Internet portal are built.

3 Speech communication performance tests

One aspect of communication is the ability to understand speech. This is primarily a problem under adverse listening conditions. For example, for sounds that have low levels compared to the absolute auditory threshold, amplification can be applied to ensure audibility, but such amplification will then raise the levels of both target sounds and background noise, thus not improving the signal-to-noise ratio (SNR).

A type of test that has been under consideration for inclusion in the test battery are auditory cluster-identification tests. Such tests measure the ability to identify a particular sound from a group of sounds. They are often set up as identification measurements using consonant-vowel, vowel-consonant, or consonant-vowel-consonant combinations. They provide a measure of basic identification skills and they are suitable for automatic operation. However, as these basic skills are still at a considerable distance from the actual ability to understand speech under adverse conditions, the results may be hard to interpret. Correct interpretation of such results will require a considerable knowledge of auditory perception, implying the involvement of an expert in the evaluation of the outcomes. For that reason, such tests have not been included in the current project where we aim at tests that provide outcomes that are easy to understand for a wide variety of users.

3.1 Digit triplets tests

Digit triplets tests are suitable for swift screening of communication abilities. They provide speed over accuracy, producing a threshold estimation in just a few minutes at the expense of measuring only a subgroup of all the abilities necessary for good communication skills. The latter shortcoming is mainly caused by the small number of actual stimuli used in the test, compared to, for example, open-set sentence tests.

Digit triplets tests are speech-recognition-in-noise tests using spoken combinations of three digits, presented in a noise background. By adaptively varying the signal-to-noise ratio, a threshold value can be established for a preset correct score (usually 50 percent correct). This threshold then represents a measure of speech recognition abilities under adverse conditions. Because of the simple instructions, digit triplets tests are extremely suitable for automatic and self-screening operation. Combined with the speed of the test, this makes them particularly suitable as screening tests (Smits et al., 2004, Smits and Houtgast 2005).

A telephone version of the digit triple test has been running in The Netherlands since 2003 and has collected more than 160000 responses. This telephone version provides users with a test outcome in one of three categories: good, insufficient, and poor. The “good” results lie within 2 standard deviations of the average score of young normally-hearing listeners. If the measured SNR falls above this threshold value with up to 2 dB, the outcome “insufficient” is given, and for even worse results the outcome “poor” is provided. For the latter two outcomes, the advice is given to the callers to have their hearing further examined by a professional (Smits et al., 2004).

For further details of the test procedure and implementation, see the deliverables D-2-1 and D-1-1. An overview of the specifications of the Dutch and German versions and the construction of their stimuli is given in the Appendix.

3.1.1 Stand-alone versions

Digit triplets tests can be executed as stand-alone tests using a set of head phones or loudspeakers and a PC or other type of computer. This can be achieved using dedicated measurement programs or by incorporating them in audiological measurement stations. In the present project, such tests have been, or are being, developed in a number of languages. The Dutch test was originally implemented using tokens with coarticulation, while the German implementation did not preserve the coarticulation present in the original recordings. This systematic difference may affect the test results. Therefore, an investigation into its possible consequences is presently being performed at the VU Medical Center in Amsterdam, the Netherlands. The outcome of this comparison is expected by the end of 2005.

3.1.1.1 Dutch

Validated Dutch versions of the digit triplets test (Smits et al., 2004) are available as a PC-program running under Windows and as part of the Oldenburg Measurement Applications (OMA)(for the latter, see deliverable D-1-1).

3.1.1.2 German

A validated German version of the digit triplets test has been constructed and validated (Wagener et al., 2005; Bräcker, 2005), and is available as part of the Oldenburg Measurement Applications (see deliverable D-1-1).

3.1.1.3 English (UK)

An English version of the digit triplets test has been constructed. It is available as a PC-program running under Windows which is currently used for the validation of the test. An implementation in OMA is planned.

3.1.1.4 Swedish

A Swedish version of the digit triplets test is under construction. All stimuli have been recorded, cut, and edited. A computer implementation in OMA is currently under construction so the validation can be executed in the first half of 2006.

3.1.1.5 French

A French version of the digit triplets test is under construction. All stimuli have been recorded and are ready to be cut and edited. A computer implementation in OMA will be constructed so the validation can be started in the first quarter of 2006. The validation will be executed in a cooperative effort of the University of Leuven, Belgium, the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland, and probably the University of Paris, France.

3.1.2 Telephone versions

Thanks to the simple instructions and measurement procedure, digit triplets tests can be executed as stand-alone tests via the telephone. For this, band-limited (300 to 3400 Hz) versions of the digit triplets test have been validated. The responses of the listeners are collected using the numerical pad of the telephone, and a call center can control the automatic execution of the test.

3.1.2.1 Dutch

The Dutch version of the telephone test was implemented in 2003 and has been running successfully since then. The test costs callers 35 eurocent per minute, and takes approximately three minutes to execute. More than 160000 callers have executed the test, leading to a large database of results (Smits et al., 2005).

3.1.2.2 German

A lab version of the German telephone test with experimenter interaction to start the measurement was set up for the validation of the test via telephone. A German version of the fully automatic telephone test has beenimplemented and is presently tested. It is likely to become operational during the course of 2006.

3.1.2.3 English (UK)

An English version of the telephone test is currently being implemented. It is awaiting the validation that is under way, and is expected to become operational at the end of 2005.

3.1.2.4 Swedish

A Swedish version of the telephone test will be implemented once the test validation has been performed. It is likely to become operational during the course of 2006.

3.1.2.5 French

A French version of the telephone test will be implemented once the test validation has been performed. The validation will be executed in a cooperative effort of the University of Leuven, Belgium, the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland, and the probably University of Paris, France. It is possibly becoming operational during the course of 2006.

Table 1. *Digit triplet tests per language*

Language	Version	Material	Validation	Status
Dutch	Stand alone	Available	Ready	Running
	Telephone	Available	Ready	Running
	Internet	Available	Ready	Under construction for 2006
German	Stand alone	Available	Ready	Running
	Telephone	Available	Ready	Running, presently tested
	Internet	Available	Ready	Under construction for 2006
English	Stand alone	Available	In progress	Planned for 2005
	Telephone	Available	In progress	Under construction for 2005
	Internet	Available		Under consideration
Swedish	Stand alone	Available	Planned for 2006	Under construction for 2006
	Telephone	Available	Planned for 2006	Under construction for 2006
	Internet	Available		Under consideration
French	Stand alone	Under construction for 2006	Planned for 2006	Planned for 2006
	Telephone	Planned for 2006	Planned for 2006	Planned for 2006
	Internet			Under consideration

3.1.3 Internet versions

As the digit triplets test is extremely suitable for self screening, it is one of the obvious candidates for application via the Internet. These Internet implementations will be realized in cooperation with Sub Project 5 and the first ones are scheduled for month 31 of the HearCom project. These first implementations will be realized in German and Dutch. English, French, and Swedish versions are under consideration.

3.2 Closed-set sentence tests

Closed-set sentence tests (often called Hagerman or OLSA tests) are suitable for rapid estimation of communication abilities. They provide accurate results, but as they use a limited set of word material of 50 words, there is an upper limit to how complete they estimate all the abilities necessary for good communication skills, as compared to, for example, open-set sentence tests. On the other hand, the word set in this type of test is much larger than is the case for the digit triplets test, putting this test somewhere between digit triplets tests and open-set sentence tests. For example, the phoneme distributions of the test materials represent the mean phoneme distribution of the respective language, which is virtually impossible to achieve with digit triplets tests. A major advantage of closed-set sentence tests is that they offer possibility of prolonged testing. Since the sentences used in this test do not have a meaning that can be easily memorized, the available sets of sentence lists can be used again and again without the risk of listeners learning the sentences by heart. This is in sharp contrast with open-set sentence tests, where a long period (many months) is required before the same list can be presented to the same listener to prevent running the risk of memory effects confounding the estimated thresholds.

Closed-set sentence tests are speech-recognition-in-noise tests using spoken five-word sentences (name verb numeral adjective object), presented in background noise. Each word is taken from a set of ten alternatives. By adaptively varying the signal-to-noise ratio, a threshold value can be established for a preset correct score. This threshold gives a measure of speech recognition abilities under adverse conditions. Because of the relatively simple instructions and the possibility to use a closed response set, these sentence tests are suitable for automatic operation and may even be considered for self-screening operation, though the latter application might be less suited for elderly listeners. Often, these sentence tests are not used in closed mode (i.e. the response alternatives are not visually presented to the listener). However, the Oldenburg Measurement Applications provides both options, with and without visual alternatives. For further details of the test procedure and implementation, see the deliverables D-2-1 and D-1-1.

3.2.1 Stand-alone versions

Closed-set sentences tests can be executed as stand-alone tests using a set of head phones or loudspeakers and a PC or other type of computer. This can be achieved using dedicated software or by incorporating them in audiological measurement stations such as OMA. In the present project, such tests have been, or are being, developed in a number of languages.

3.2.1.1 German

A validated German version of the closed-set sentence test is the Oldenburg sentence test (OLSA), constructed and validated by Wagener et al. (1999a, b, c), and it is available as part of the Oldenburg Measurement Applications (see deliverable D-1-1).

3.2.1.2 Swedish

A validated Swedish version of the closed-set sentence test is the Hagerman test, constructed and validated by Hagerman (1982), and Hagerman and Kinnefors (1995). It will be made available via the Oldenburg Measurement Applications by the end of 2005.

3.2.1.3 English (UK)

An English version of the closed-set sentence test is currently under construction; at the moment the validation is being performed. The validation is expected to be completed before the end of 2005, after which it can be made available via the Oldenburg Measurement Applications during the course of 2006.

3.2.1.4 Dutch

A Dutch version of the closed-set sentence test is currently under construction. The measurement material is expected to be finalized in the first quarter of 2006, after which the validation can be started. The validation will be executed in a cooperative effort of the Erasmus Medical Center of Rotterdam and the Amsterdam Medical Center in the Netherlands, and the University of Leuven in Belgium.

3.2.1.5 French

A French version of the closed-set sentence test is currently under construction. The raw material has been recorded, and the cutting and editing will be performed next. The validation of the test is expected to be performed during the first half of 2006.

3.2.1.6 Danish

A validated Danish version of the closed-set sentence test is currently available (Wagener et al., 2003). It will be made available via the

Oldenburg Measurement Applications during the course of 2005 and 2006. As the number of languages in which the test was planned to be implemented will be more than met, this implementation has low priority. The license to implement this material in OMA has to be checked but is not expected to be problematic.

3.2.1.7 Polish

A Polish version of the closed-set sentence test is currently being planned. The construction and validation of this test will be executed in cooperation with Work Package 7.

Table 2. *Closed-set sentence tests per language*

Language	Version	Material	Validation	Status
German	Stand alone	Available	Ready	Running
	Internet	Available		Under consideration
Swedish	Stand alone	Available	Ready	Under construction for 2005
	Internet	Available		Under consideration
English	Stand alone	Available	In progress (2005)	Planned for 2006
	Internet			Under consideration
Dutch	Stand alone	Under construction for 2006	Planned for 2006	Planned
	Internet			
French	Stand alone	Under construction for 2006	Planned for 2006	Planned
	Internet			
Danish	Stand alone	Available	Ready	Planned for 2006
	Internet	Available		Under consideration
Polish	Stand alone	Planned for 2006	Planned	Planned
	Internet			

3.2.2 Internet versions

Closed-set sentence tests are suitable for self screening, so, they are potential candidates for application via the Internet. Such Internet implementations would then be realized in cooperation with Sub Project 5.

Potential implementations may be realized for those versions that are, or will be, available via the Oldenburg Measurement Applications. This would then involve tests in languages as German, Danish, English, and Swedish. Such Internet versions are being considered.

3.3 Open-set sentence tests

Open-set sentence tests (often called Plomp or HINT tests) are suitable for rapid estimation of communication abilities. They provide accurate results, and, as they use an open set of response options, these are the tests within the presently proposed set that make the more complete estimation of all the abilities necessary for good communication skills. Nevertheless, as these tests use short, meaningful, sentences, while in everyday situations listeners need to cope with running speech, these tests still approximate the total set of skills involved in listeners' communication abilities. In addition, various binaural skills will need to be estimated as well (see Section 2.4) to obtain a more complete investigation. The price of providing the most complete communication-abilities estimate of the set is that this test is not suitable for prolonged testing. To prevent running the risk of memory effects confounding the estimated thresholds, a relatively long period (several months) is required before the same sentence list can be presented to the same listener. So, extended testing is limited by the number of available sentence lists. This contrasts with closed-set sentence tests, where memory problems have a much smaller effect.

Open-set sentence tests are speech-recognition-in-noise tests using short, meaningful, sentences, presented in background noise. By adaptively varying the signal-to-noise ratio, a threshold value can be established for a preset correct score. This threshold gives a measure of speech recognition abilities under adverse conditions. Because of the open response set, these sentence tests usually require a measurement assistant for the response evaluation, and are, therefore, less suitable for automatic operation. For further details of the test procedure and implementation, see the deliverables D-2-1 and D-1-1.

3.3.1 Stand-alone versions

Open-set sentences tests can be executed as stand-alone tests using a set of head phones or loudspeakers and a PC or other type of computer. This can be achieved using dedicated software or by incorporating them in audiological measurement stations such as OMA. In the present project, such tests have been, or are being, developed in a number of languages.

3.3.1.1 Dutch

Validated Dutch versions of open-set sentence tests have been developed and validated by Plomp and Mimpen (1979) and Versfeld et al. (2000).

Both tests are available as part of the Oldenburg Measurement Applications (see deliverable D-1-1).

3.3.1.2 German

A validated German version of the open-set sentence test is available in the Göttingen sentence test. It was constructed and validated by Kollmeier and Wesselkamp (1997), and is available as part of the Oldenburg Measurement Applications (see deliverable D-1-1).

3.3.1.3 English (UK)

A validated English version of the open-set sentence test is available in the BKB sentence test, constructed and validated by Bench et al. (1979). It will be made available as part of the Oldenburg Measurement Applications in the beginning of 2006.

3.3.1.4 Swedish

A validated Swedish version of the HINT-type open-set sentence test has recently been developed at the University of Linköping (Sweden) as part of the HearCom project. It is presently being made available as part of the Oldenburg Measurement Applications in the beginning of 2006.

3.3.1.5 French

A French version of the open-set sentence test is based on a set of Audivox sentences, as described by Wable (unpublished). It is currently being validated at the University of Leuven, Belgium, as part of Work Package 7. It is still under consideration whether it will be included in the Oldenburg Measurement Applications.

Table 3. *Open-set sentence tests per language*

Language	**Version**	**Material**	**Validation**	**Status**
Dutch	Stand alone	Available	Ready	Running
German	Stand alone	Available	Ready	Running
English	Stand alone	Available	Ready	Planned for 2006
Swedish	Stand alone	Available	Ready	Available
French	Stand alone	Available	In progress (2006)	Under consideration

4 Spatial communication performance tests

One aspect of communication is the ability to localize sound sources. This can be important in quickly establishing the direction from which dangers approach, or, less dramatically, from which information is arriving so that the attention can be directed at correct location. Such skills can be measured using localization tests that directly measure the listener's localization capabilities. Another aspect of spatial sound perception is the binaural advantage listeners may have in the free field, which arises from mental processes that combine the information arriving at both ears. Such skills are measured using difference scores of conditions with and without the presence of a binaural advantage. Because of their direct connection to personal safety in everyday situations, we have decided to concentrate on the first category of spatial hearing test, the localization tests. Tests of binaural advantages are investigated further in Work Package 2, in relation to the auditory profile.

To measure a listener's localizations abilities, sounds need to be presented to a listener from different directions. By collecting responses of the perceived directions, correct scores can be obtained. In the horizontal plane, two main categories of measurement setups are in use with presentation directions covering a full circle or covering half a circle centered around the frontal direction of the listener. Using a full circle, front-back confusions can be measured separately from localization errors, while using a half circle only localization errors can be recorded. However, for a full circle more presentation directions are required, demanding a more complex and expensive setup. On the other hand, using head-related transfer functions (HRTFs), virtual setups can be constructed, either for head-phone use or for two-speaker use. In general, the head-phone approach is found to be the more accurate option, but by including a two-speaker setup, a much wider user group may be reached. These virtual approaches reduce the hardware requirements, but are likely to lead to reduced measurement accuracy. In the present project, full and half circle setups are included in both free-field and virtual options. So, users are free to choose the setup that meets their requirements the closest. For more extensive and detailed information on spatial tests, see deliverable D-2-1.

4.1 Localization tests

Localization scores can be recorded as minimal audible angles (MAAs) and as correct scores, either presented in percentages or in confusion matrices. For measuring MAAs, a very small speaker separation is needed. To record correct scores larger speaker separations can be used, enabling coverage of a full or a half circle.

Depending on the setup and the measurement procedures used, localization tests may be suitable for automatic execution and self-screening. For application via the Internet virtual setups are particularly interesting as they do not require any special hardware beyond a good soundcard combined with a good pair of head phones or speakers.

4.1.1 Stand-alone versions

Localization tests can be executed as stand-alone tests using a set of speakers positioned around a listening position and a PC or other type of computer. Alternatively, a virtual 2D environment can be used in combination with a computer. Such setups can be achieved using dedicated software or by incorporating them in audiological measurement stations such as OMA.

4.1.1.1 5 or 8-speaker setups

Stand-alone versions of a 5 speaker setup are available via the software provided within the HearCom project by the Institute of Sound and Vibration Research (ISVR), Southampton, United Kingdom. Running implementations are available there, at the VU Medical Center in Amsterdam, the Netherlands, and at Hörzentrum Oldenburg, Germany. At both latter institutes an 8 speaker setup is also available.

4.1.1.2 Virtual versions

At the moment two lines of investigation are followed to evaluate the type of test that will be made available as a screening test. The first line involves an investigation into the applicability of a two-speaker setup. The results of such a setup will be compared to those obtained using more customary head-phone versions. This line of investigation is currently being executed at the Institute of Sound and Vibration Research in Southampton, United Kingdom. The second line of investigation involves the comparison of MAA measurements obtained with a virtual setup as compared to a dedicated hardware setup. The necessary HRTFs will be supplied by the Ruhr University in Bochum, Germany, and the actual measurements will be performed at the Amsterdam Medical Center in the Netherlands. Any necessary additional measurements using an 8-speaker setup can be performed at the VU Medical Center in Amsterdam. These evaluations are expected to come to a result in the next six months.

4.1.2 Internet versions

Localization tests can be made suitable for self screening, so, such versions would be good candidates for application via the Internet. These Internet implementations will be realized in cooperation with Sub Project 5 and the first implementation is scheduled for month 30 of the project.

4.1.2.1 5 or 8-speaker setups

These options are not under consideration anymore, as practical problems involved in such implementations (complex hardware and lack of Internet implementation alternatives) make it far less suitable solutions than the virtual setups mentioned in the next section.

4.1.2.2 Virtual versions

Making localization tests available for Internet application will only be feasible via the Oldenburg Measurement Applications. To limit the amount of necessary hardware and the complexity of the required setup, virtual solutions are considered to be the better candidates. The exact versions that will be chosen depend on the outcome of the currently executed evaluations mentioned in Section 3.1.1.

4.2 Other spatial tests

There are a few other good candidates for spatial tests that can be applied in the HearCom project. For example, in Work Package 2, measurements of sound detection differences using different spatial masking conditions (binaural masking level differences or BMLDs) and speech intelligibility differences between different spatial masking conditions (intelligibility level differences, or ILD and binaural ILDs, or BILDs) are being considered in relation to the auditory profile. BMLDs are often determined as the difference between the detection thresholds for a condition where target and masker are both presented diotically and a condition where one of the stimuli is presented dichotically by reversing its phase in one ear. ILDs are often measured as the difference in intelligibility scores between a condition where speech and noise are presented from the same direction and one where they are presented from different directions. BILDs are usually measured as differences between intelligibility scores recorded using both ears and scores obtained with one ear plugged. The presentation directions of the speech and the noise can then be varied to create various measurement conditions.

Inclusion of the described tests in the auditory profile will lead, or in some cases has already led, to them being implemented on the Oldenburg Measurement Applications. This opens the possibility to provide such tests over the Internet. However, some caution needs to be taken here as (a) not all tests produce easily understandable results for non-experts, and (b) some tests require an expert in the execution, for example for the correct plugging of an ear in BILD measurements. So, if the inclusion of such tests in the Internet test battery is considered, only tests like ILD measurements may be suitable.

5 Dissemination and Exploitation

This report describes the communication performance tests that have been selected in Work Package 1. The results of this report will primarily be used be used in Work Package 10 as guideline in the construction of the HearCom portal. In addition, they are of interest to Work Package 2 in relation to the development of the auditory profile and its related tests.

Exploitation of the results of the current report will not occur directly, but will be realized via its effects on the HearCom Internet service (HearCom portal), which has a clear exploitation potential. The way in which the implementation of the selected tests will be exploited via the portal is discussed in detail in Section 6 of deliverable D-13-1.

6 Conclusions

A set of tests to investigate the communication performance of listeners has been selected. These tests may be executed using different channels (via: audiological work stations, the Internet, or by telephone). The tests will be implemented in several languages, including English, Danish, Dutch, French, German, Polish, and Swedish. It was decided to concentrate on three types of measurements.

The first type are a digit triplets tests, used for a rapid screening of speech recognition under adverse conditions. The digit triples test estimates recognition thresholds using spoken number triplets, presented in a noise background. These tests will be made available for use via audiological work stations, Internet, and telephone.

The second type comprise sentence tests for more precise measurements of that ability. One type of sentence test uses an open set of sentences, the other uses a closed set and can also be used as a closed response-set test. These tests estimate recognition thresholds using spoken sentences, presented in a noise background. These tests will be made available for use via audiological work stations and the Internet.

The third type comprise localizations tests to estimate spatial hearing abilities. In the localization tests, listeners are required to identify the direction from which a target sound is presented. Such tests will be made available for use via audiological work stations and the Internet.

The results of this report will be input to SP5/WP11, the construction of the HearCom Internet services (portal). In addition they are of interest to SP1/WP2, in relation to the measures included in the auditory profile.

7 Appendix

7.1 Specifications of Dutch digit triplets test

The Dutch 3-digit test for telephone screening has the following specifications:

- Single-syllable digits are used: 0,1,2,3,4,5,6,8 for Dutch
- Presented in triplets with coarticulation
- Adaptive procedure is used to estimate threshold
- Termination after 23 presented triplets
- Presented triplets are chosen at random from a set of 80 (see below)
- The speech level is varied in the test
- The noise level is fixed at approximately 70 dBA (dependent on the phone)
- The triplets are scored (not each digit)
- First triplet is presented at 0 dB SNR
- Subsequently 2-dB step size in adaptive procedure (+2 for incorrect, -2 for correct)
- The SNRs are limited to a range of -12 to +8 dB
- All triplets are pre-mixed at these SNRs and stored in separate files
- Threshold is the average SNR of the last 20 triplets (19 + SNR after last triplet)

Stimulus construction:

- Each triplet was recorded separately and stored in a wave file
- A female speaker was used
- A 22050-Hz sample rate was used
- Five lists of 23 triplets were recorded (a total of 115 different triplets)
- The stimulus durations, including 0.5-s initial and final pauses, ranged from 1.7 to 2.4 s

- On average, the level of the digits decreased within the triplet

- This decrease was corrected using a level function with a 6-dB linear increase

- The level increase was performed over triplet plus leading and trailing pauses

- On average, the correction raised the last digit by about 3.5 dB relative to the first one

- Speech noise was generated by shaping white noise with the average triplet spectrum

- For all 115 triplets, psychometric functions were recorded using 80 subjects

- In these experiments the noise level was fixed at 62 dBA

- First triplet was presented with increasing SNR (4-dB steps) until a correct response

- For the rest the adaptive procedure was executed as described above

- 80 triplets with steep psychometric functions were selected for further use

- These triplets had SNRs at threshold (50% correct) between -2 and -12 dB

- Averaged over these 80 triplets, the SNR at threshold was -7 dB

- These SNRs were determined over the total duration of the stimulus files

- Level corrections were performed to achieve thresholds of -7 dB for all triplets

- So, all triplets were equally intelligible while they could have different SNRs at threshold

- At threshold, average SNRs of the 1st, 2nd, and 3rd digits were -6.4, -7.0, and -7.6 dB

Experimental hard- and software (development phase):

- Soundcard: Creative Labs Soundblaster 16 Value PNP

- Modem: E-tech, PC336RVP

- Modem software: Katalina Technologies, Voiceguide V2.9, used for answering the telephone and detecting the keys pressed on the number pad

- SRT software: custom made in Delphi (Borland software), used for choosing, mixing, and playing stimuli and controlling SRT protocol

Implementation and execution (exploitation phase):

- The implementation and execution were performed by a call center: SNT Connect Services

- SNT also operates in Belgium, Germany, France, Sweden, Norway, Denmark, and Finland

- All necessary files (about 600) were constructed by the VU Medical Center

- A test takes approximately three minutes including explanations and feedback of results

- Only calls from land lines are accepted, calls from mobile phones are excluded

- The call center has reserved 70 lines for the test and charges are 35 eurocent per minute

- The revenues are split between the telecom operator, SNT, and the VU Medical Center

Advantages:

- Using separately recorded triplets gives a very small dependence on specific verbalizations of the numbers

7.2 Specifications of German digit triplets test

The German 3-digit test for telephone screening has the following specifications:

- Single-syllable digits are used: 0,1,2,3,4,5,6,8,9 for German

- Presented in triplets with co-articulation given by the position in the triplet

- Each triplet starts with the announcement words “Die Ziffern” (“The digits”)

- Noise has same long-term spectrum as speech material (generated by superimposing digit speech material 30 times using randomly chosen fixed pauses between the particular digits)

- Test lists with 27 triplets, each digit occurs three times at each position in the triplet

- Maximum of 243 different test lists

- The digits for different triplet positions are stored in separate files

- The triplets are combined online by using the digit-position files

- The SNRs are mixed online

Stimulus construction:

- The digits were recorded in three different triplet lists, such as each digit was recorded three times at each position in the triplet. These three triplet lists were recorded 5 times (15 recordings for each digit at each position)

- All triplets were recorded with the announcement words ‘Die Ziffern’ (“The digits”)

- The digits were windowed (hanning window, with 5ms flanks) and stored separately per digit and position (without initial and final pauses)

- For each digit at each position in the triplet the two best recordings due to listening quality were chosen as speech material for optimization

- A female speaker was used

- A 44100-Hz sample rate was used

- The digit levels were averaged by the level of the announcement words: The mean level of all announcement words was calculated. The difference between the particular announcement words of a digit and the mean announcement word level was calculated. This difference was balanced by adjusting the level of the digit.

- To build the triplets during the measurements, the digits at desired positions are combined with 200ms pauses in between, resulting in triplet lengths of 1.9-2.2 s. The pause between the indicating words and the triplets equals 250ms

- The indicating words are attenuated by 2 dB relative to the digit material for the measurements

Optimization measurements:

- The digit-specific psychometric functions for all digits at all positions were determined with 12 normal-hearing subjects in the lab by presenting the digits in triplets and using digit-scoring

- The triplets were measured at different fixed SNRs, the noise presentation level was fixed at 65 dB SPL

Optimization:

- For each digit at each position the recording was chosen whose SRT better corresponds with the mean SRT of the whole material

- The level of the digits were adjusted to equalize intelligibility

Differences to and advantages compared to Dutch version:

- Less speech material means that the speech material is more homogeneous, the optimization is less expensive and more different test lists are possible.

8 Literature

Bench, J., Kowal, A., and Bamford, J. (1979) The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br. J. Audiol. 13, 108-112.

Bräcker, T. (2005) Evaluation des deutschen Ziffern-Tripel-Tests (Evaluation of the German digit triplets test). Diploma thesis. Fachhochschule Oldenburg Ostfriesland Wilhelmshaven.

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

Hagerman, B., and Kinnefors, C. (1995) Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. Scand. Audiol. 24, 71-77.

HearCom Deliverable D-1-1 (2005) Communication self-screening test for two languages operating on a PC-based system.

HearCom Deliverable D-2-1 (2005) Implementation of a preliminary test set for auditory impairments.

HearCom Deliverable D-13-1 (2005) Action plan for dissemination and exploitation activities.

Kollmeier, B., and Wesselkamp, M. (1997) Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. J. Acoust. Soc. Am. 102, 2412-2421.

Plomp, R., and Mimpen, A. M. (1979) Improving the reliability of testing the speech reception threshold for sentences. Audiol. 18, 43-52.

Smits, C., and Houtgast, T. (2005) Results from the Dutch speech-in-noise screening test by telephone. Ear & Hearing 26, pp. 89-95.

Smits, C., Kapteyn, T.S., Houtgast, T. (2004) Development and validation of an automatic speech-in-noise screening test by telephone. International Journal of Audiology 43, 15-28.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (2000) Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J. Acoust. Soc. Am. 107, 1671-1684.

Wable, J. (2005) Normalisation d’un test de reconnaissance de la parole dans le bruit chez le sujet deficient auditif. Unpublished.

Wagener, K., Brand, T., and Kollmeier, B. (1999a) Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a German sentence test part I: Design of the Oldenburg sentence test). Z. Audiol. 38, 4-15.

Wagener, K., Brand, T., and Kollmeier, B. (1999b) Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests (Development and evaluation of a German sentence test part II: Optimization of the Oldenburg sentence test). Z. Audiol. 38, 44-56.

Wagener, K., Brand, T., and Kollmeier, B. (1999c) Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a German sentence test part III: Evaluation of the Oldenburg sentence test). Z. Audiol. 38, 86-95.

Wagener, K.C., Eeenboom, F., Brand, T., Kollmeier, B. (2005) Ziffern-Tripel-Test: Sprachverständlichkeitstest über das Telefon. (Digit triplets test: speech intelligibility test via telephone) Tagungs-CD der DGA Jahrestagung 2005

Wagener, K., Josvassen, J. L., and Ardenkjaer, R. (2003) Design, optimization and evaluation of a Danish sentence test in noise. Int. J. Audiol. 42, 10-17.