

**FP6-004171 HEARCOM**  
**Hearing in the Communication Society**

**INTEGRATED PROJECT**  
**Information Society Technologies**

**D-11-6: Validation of the screening localisation test and an inter comparison of all HearCom screening tests**

Contractual Date of Delivery:	31, Aug, 2008 (+45 days)
Actual Date of Submission:	14, Oct, 2008
Editor:	Johannes Lyzenga
Sub-Project/Work-Package:	SP5, WP11
Version:	V2.0
Total number of pages:	50

<b>Dissemination Level</b>		
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) This information is confidential and may be used only for information purposes by Community Institutions to whom the Commission has supplied it		

## Deliverable D-11-6

<b>VERSION DETAILS</b>
Version: 2.0
Date: 14 Oct 2008 / 22-02-2010(Public)
Status: final

<b>CONTRIBUTOR(S) to DELIVERABLE</b>	
<b>Partner</b>	<b>Name</b>
VUMC	Johannes Lyzenga
EMC	Jan Koopman

<b>DOCUMENT HISTORY</b>			
<b>Version</b>	<b>Date</b>	<b>Responsible</b>	<b>Description</b>
1.0	16 Sept 2008	Johannes Lyzenga	First draft
2.0	14 Oct 2008	Johannes Lyzenga	Final version
	22-02-2010	M.Vlaming	Final version made public

<b>DELIVERABLE REVIEW</b>			
<b>Version</b>	<b>Date</b>	<b>Reviewed by</b>	<b>Conclusion*</b>
1.0	18 Sept 2008	M. Buschermöhle	Accept after modification
1.0	06 Oct 2008	W. Dreschler	Accept after modification

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

## Table of Contents

Pre-Amble.....	7
1 Executive Summary.....	8
2 Introduction .....	9
3 Methods.....	10
3.1 Description of the screening tests .....	10
3.1.1 The minimum-audible angle (MAA) test .....	11
3.1.2 The auditory virtual-environment test (AVE).....	12
3.1.3 The digit-triplet test .....	12
3.1.4 The HearCom questionnaire.....	12
3.2 Description of the comparison tests .....	13
3.2.1 The 8-speaker test .....	13
3.2.2 The Amsterdam Inventory (Kramer list).....	14
3.3 Subjects and procedures .....	14
3.3.1 VUmc, Amsterdam .....	15
3.3.2 EMC, Rotterdam.....	15
4 Validation of the MAA test .....	17
4.1 Subjects and methods.....	17
4.2 Results .....	18
4.3 Discussion .....	19
4.4 Conclusions.....	20
5 Inter comparison of screening tests.....	20
5.1 Subjects and methods.....	20
5.2 Results .....	20
5.2.1 Digit-triplet test .....	21
5.2.2 The AVE and 8-speaker tests .....	23

5.2.3	HearCom questionnaire and Amsterdam Inventory .....	26
5.3	Inter comparison of the tests .....	29
5.4	Effects of locations and instructions .....	32
5.5	Discussion .....	34
5.5.1	The tests.....	34
5.5.1.1	The minimum-audible angle test (MAA).....	34
5.5.1.2	The digit-triplet test (SRT).....	35
5.5.1.3	The auditory virtual-environment test (AVE) .....	36
5.5.1.4	The 8-speaker test (8Sp).....	37
5.5.1.5	HearCom questionnaire (HQ) .....	37
5.5.1.6	The Amsterdam Inventory (AI).....	38
5.5.2	Inter comparison.....	39
5.5.2.1	Correlations for all participants.....	39
5.5.2.2	Correlations for the normally-hearing participants .....	40
5.5.2.3	Correlations for the hearing-impaired participants .....	41
5.5.3	Locations and instructions .....	41
5.6	Conclusions.....	42
6	General discussion .....	43
6.1	The MAA test .....	43
6.2	The screening test battery .....	43
7	Dissemination and Exploitation .....	45
7.1	Ethical issues .....	45
8	General conclusions.....	46
9	Literature.....	47
10	Appendix A. Text of the HQ .....	49

## List of Figures

Figure 3-1: <i>Speaker set-up for the Internet MAA test</i> .....	11
Figure 3-2: <i>Speaker set-up for the 8-speaker test</i> .....	14
Figure 3-3: <i>Average audiograms for the VUmc participants</i> .....	15
Figure 3-4: <i>Average audiograms for the EMC participants</i> .....	16
Figure 5-1: <i>MAA scores versus audiogram (PTA3)</i> .....	34
Figure 5-2: <i>Triple digit scores versus audiogram (PTA3)</i> .....	35
Figure 5-3: <i>Transformed AVE scores versus audiogram (PTA3)</i> .....	36
Figure 5-4: <i>Transformed 8-speaker scores versus audiogram (PTA3)</i> ....	37
Figure 5-5: <i>Overall average HQ scores versus audiogram (BEA)</i> .....	38
Figure 5-6: <i>Overall average AI scores versus audiogram (BEA)</i> .....	39

## List of Tables

Table 4-1: <i>NH data for the MAA test</i> .....	17
Table 4-2: <i>HI data for the MAA test</i> .....	18
Table 5-1: <i>NH data for the digit-triplet test</i> .....	21
Table 5-2: <i>MHI and AHI data for the digit-triplet test</i> .....	22
Table 5-3: <i>HI data for the digit-triplet test along with the averages for the MHI, AHI, and HI groups.</i> .....	23
Table 5-4: <i>NH data for the localisation tests</i> .....	24
Table 5-5: <i>MHI, AHI, and HI data for the localisation tests</i> .....	25
Table 5-6: <i>NH data for the questionnaires</i> .....	27
Table 5-7: <i>MHI, AHI, and HI data for the questionnaires</i> .....	28
Table 5-8: <i>Extra questionnaire data via the Internet (VUmc)</i> .....	29
Table 5-9: <i>Spearman correlations for all groups of participants</i> .....	30
Table 5-10: <i>Partial correlations for all groups corrected for age</i> .....	30
Table 5-11: <i>Pearson correlations for the NH group</i> .....	31

Table 5-12: <i>Spearman correlations for the MHI, AHI, and HI groups</i> .....	31
Table 5-13: <i>Comparison of MAA results with those scored in the quiet office room and at home via the Internet</i> .....	32
Table 5-14: <i>Comparison of SRT laboratory results with those scored in the quiet office room and at home via the Internet</i> .....	32
Table 5-15: <i>Comparison of the laboratory AVE results with those scored in the quiet office room and at home via the Internet</i> .....	33
Table 5-16: <i>Comparison of the laboratory HQ results with those scored at home via the Internet</i> .....	33
Table 6-1: <i>Criteria and PTA3 or BEA thresholds for the screening tests</i> ..	43
Table 6-2: <i>Sensitivity and specificity of the tests</i> .....	44

## **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

An important aim of the HearCom project is to enable the general public, of as many European language groups as possible, easy access to screening tests with which they can check their hearing skills in relation to full participation in the communication society. To create as low a threshold as possible for people to go and try these tests, it should be possible for them to perform them in the safe and well-know location of their own homes. At the outset of the HearCom project, a hearing screening test was already in use in The Netherlands in the shape of a digit-triplet test via telephone and Internet. It was decided to translate that test to a number of languages (versions have been developed for UK English, German, Swedish, French, Polish, while versions for US English, Greek, Turkish, and Russian are under development), and to make them available via the Internet and/or telephone. Next, novel screening tests were to be developed within the HearCom project concerning localisation skills and the subjective experience of hearing problems.

So far, three screening tests have been implemented and evaluated: a digit-triplet speech-in-noise test (see Smits et al., 2004; Smits and Houtgast, 2005; D1-1; D1-4b; D1-6a; D1-6), a screening questionnaire (see D11-5) evaluating the subjective experience of hearing problems, and a localisation test in a virtual environment (see D3-7; D3-8). The present deliverable presents the validation of a two-speaker minimum-audible angle (MAA) localisation test (see D1-5), along with an inter comparison of all screening tests and a well established localisation test and a ditto questionnaire. An MAA test was also used in the Auditory Profile, developed within HearCom (see D2-2; D2-3; D2-5), but the screening version differs in its measurement set-up. While the MAA version investigated in the Auditory Profile uses generic head-related transfer functions (HRTFs) for stimulus presentation over headphones, the screening version uses HRTFs in combination with a cross-talk cancellation technique to enable presentation via two simple computer speakers. This is expected to increase the accessibility of the test, as many more members of the general public will have a simple set of computer speakers rather than a set of high-quality diffuse-field headphones.

In order to establish the value of the screening tests for the users, we have added a section to this deliverable in which the results of all HearCom screening tests are compared amongst each other as well as to two well established measures, a localisation test and a questionnaire. This comparison was executed for both normally-hearing and hearing-impaired listeners. This additional part of the present study is intended to confirm the value of the screening tests and to enhance their credibility and stimulate their Europe-wide dissemination.

# 1 Executive Summary

A screening MAA localisation test was developed and implemented via the Internet (see D1-5). The test uses HRTFs and a cross-talk cancellation technique to enable stimulation via two simple speakers. Therefore, it only requires a computer, an Internet connection, and two such simple speakers to enable people to execute the test at home. The present deliverable presents the validation of this test. In addition, results of this test are compared to those for the other screening tests developed and/or implemented in the HearCom project, a digit-triplet speech-in-noise test, a localisation test in a virtual environment and a screening questionnaire, as well as to the results of a well established localisation test and ditto questionnaire. Both normally-hearing and hearing-impaired listeners participate in the study. The results provide a threshold criterion for the MAA test, differentiating the results into one of the categories “good,” “insufficient,” and “poor.” The results of the inter comparison of the screening tests and the well established hearing measures indicate that all the screening tests can indeed detect hearing problems reliably, in that the results of the screening tests all correlate to some extent with the pure-tone average hearing loss (PTA) or a subjective audibility measure. In addition, they all measure different aspects of hearing, in that they only correlate moderately with each other, if at all.



## 2 Introduction

To give the general public an opportunity to screen themselves for hearing problems that might hinder communication, it was decided to develop and provide a set of screening tests. The need for, and interest in, screening tests was investigated in deliverable D11-1 with a survey study among professionals and end users performed in the UK, The Netherlands and Germany. The results of this study indicated a large public interest in quick and reliable screening tests.

Since some screening tests and all implementations would be language specific, the aim was to make as many language versions of selected hearing tests as possible (see D1-2 for an early overview of tests, D1-3 for a protocol to generate multi-language version of speech tests, and D1-7 for normalization data regarding multi-language versions of speech tests). To create screening tests with very low participation thresholds, it should be possible for users to perform those tests in the safe and well-know location of their own homes. Since users then have to adjust the presentation level of the auditory screening tests to a comfortable level themselves, an important selection criterion for these tests is a large degree of presentation-level independence. Therefore, tests with a large degree of level independence and good applicability for telephone and/or Internet were selected and implemented.

A screening test for hearing was already implemented in The Netherlands in the shape of a digit-triplet test by telephone (since 2004) and Internet (since 2006). In the HearCom project, that test has been constructed in a large number of other languages. They are made available via the Internet and/or telephone. Since their introduction, nearly a million people have performed digit-triplet screening test by telephone in the UK and the Netherlands. German telephone and Internet versions have recently been made available to the public as well.

Next to the aforementioned speech-in-noise test, two localisation tests have been developed in the HearCom project. The first screening localisation test is a minimum-audible angle (MAA) localisation test. It literally measures the minimum angle between two sound sources for which listeners are able to discriminate between those two sources. This test has been constructed in such a way that the stimuli can be presented using two simple speakers, rather than via an extended array of speakers or a virtual system that depends strongly on headphone quality and characteristics.

The second localisation test uses a virtual environment (AVE) for stimulus presentation, and its success may well depend on the set of head-phones used. In addition it requires a lot of Internet-server power. Therefore, this test is not intended for free access via the Internet. Instead it will be accessible via login after invitation, giving greater control to the operator

and the opportunity to ensure that a proper set of high-quality, diffuse-field, headphones is employed by the user.

The fourth and final screening test is a questionnaire, developed to evaluate subjective speech-perception, localisation, sound-recognition, and detection abilities. It originally comprised 25 questions, and its validation has been described in D11-5. From the results of this validation it was decided to remove 7 questions for greater sensitivity and specificity of the list. These questions have still been answered by the participants of the current study, but they will not be taken into consideration in the data analysis, only the results of the remaining 18 questions are analysed.

The usability of the HearCom screening-test implementations was evaluated in a series of usability trials; see D11-2 and D11-4 for trials concerning two localisation tests, the MAA test and the virtual-environment AVE test; and see D11-3 and D11-3a for the digit-triplet test.

To evaluate the value of the screening tests for the end users, we have asked a large number of people to perform as many of the tests as was feasible for them, when possible at home, and when convenient at one of two audiological centres in the Netherlands. To be able to perform the tests at home, participants needed a fast Internet connection and a set of computer speakers and/or a set of headphones. Both normally-hearing and hearing-impaired listeners were invited to participate. The hearing-impaired listeners were selected from the patients visiting either the EMC in Rotterdam, or the VUmc in Amsterdam. Normally-hearing listeners were drafted via the hearing-impaired listeners and by flyers. From the inter comparison of the screening tests and two well established measures we found that all the screening tests can indeed reliably detect hearing problems, and that they all measure different aspects of hearing.

## **3 Methods**

### **3.1 Description of the screening tests**

In this section, only the Dutch versions of the four selected screening tests are described. For the localisation tests, the versions in other languages only differ in the texts used in the implementation. The digit-triplet test and the HearCom questionnaire need to be entirely translated and revalidated for implementations in other languages. This has been done for the digit-triplet test in the following languages: UK English, German, Swedish, French, and Polish, while this is currently being done for US English, Greek, Turkish, and Russian. The questionnaire has been translated into English, but that version has not yet been validated.

### 3.1.1 The minimum-audible angle (MAA) test

The hearing screening test to be validated in the current study is the MAA test. This localisation test was introduced by Mills (1958), and was newly developed for use via the Internet. The test measures the minimum angle of two sound sources for which listeners are able to discriminate between sounds moving in either direction between those two sources. The stimuli consist of two consecutive noise bursts presented from either side of the straight-ahead direction. The user's task is to indicate the direction of the sound: from left to right or right to left. The positions (angles) on the left and the right, from which the noise bursts are presented, are varied adaptively using an adaptive staircase procedure, starting at  $\pm 32$  degrees (64 degrees in total). After one incorrect response, the angle between the presentation positions is increased, after two correct responses it is decreased. The step size for the increments and decrements starts at four degrees per position, and it is halved to two degrees after four turning points, and to one degree after four further turning points. The threshold value for the minimum-audible angle is estimated from eight turning points for the final 1-degree step size, and it is expressed in degrees.

For screening via the Internet, an MAA test was constructed for stimulus presentation via two simple loudspeakers, positioned at  $\pm 15$  degrees from the straight-ahead direction (see Figure 3-1). By means of generic HRTFs and cross-talk cancellation, the virtual position of the 2 speakers can be varied between angles of  $\pm 32$  degrees in 1 degree steps (see Tavan et al., 2005). For this technique to work well, participants need to set up their speakers rather precisely. For this, ample instructions are presented on the Internet pages introducing the test. For extended descriptions of this test and its Internet version, see D1-5 and D11-4, respectively.

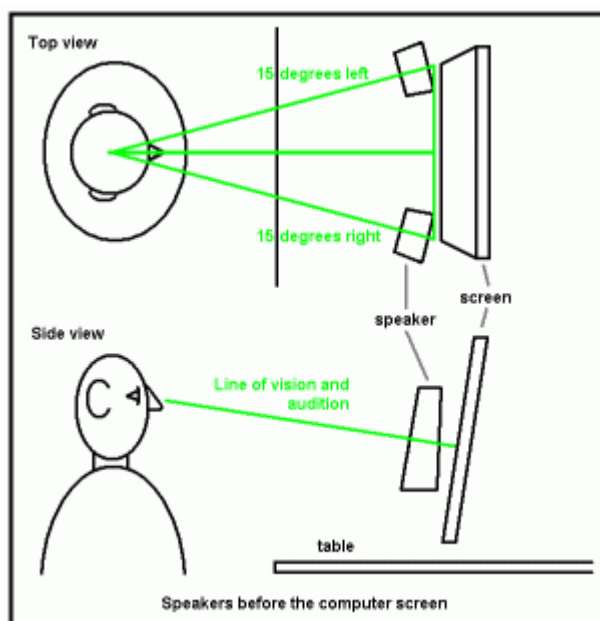


Figure 3-1: *Speaker set-up for the Internet MAA test*

### 3.1.2 The auditory virtual-environment test (AVE)

The second localisation test uses a complex virtual environment (AVE) for stimulus presentation via headphones. The virtual environment comprises a living room with 6 telephones, in which the participant can change his or her head position using a slider. Both the sound field in the living room, as perceived via the headphones, and the picture of the living room on the computer screen change according to the indicated head movement. In the test, one of the telephones is ringing, while the radio plays a song. The task of the participant is to identify the ringing telephone, in which they can make use of head movements. The ringing telephone can be selected by making a mouse click on its icon in the picture of the living room. During the test, all six telephones ring two times in a randomised order, resulting in 12 presentations and responses. The results are scored in relation to two response criteria classifying the scores in one of the three categories "good," "intermediate," and "poor." These criteria were established in the validation study in D3-8. For extended descriptions of this test and its Internet version, see D3-7 and D11-2 plus D11-4, respectively.

### 3.1.3 The digit-triplet test

In the current study, only the Dutch Internet version of the digit-triplet test was used. Before the test, participants chose a comfortable presentation level for the test. In the test, 23 digit triplets (e.g., "2 5 3") are presented in varying levels of speech-shaped noise. The participant responds by keying in the perceived numbers on the computer keyboard, or by selecting the corresponding buttons in a Java applet using a mouse. The levels of the triplets are adaptively varied to estimate the threshold as the SNR at which 50-percent of the triplets are repeated correctly (named SRT in the rest of this report). The digit-triplet test for telephone and Internet was initially developed for the Dutch language (Smits et al., 2004; Smits and Houtgast, 2005) and its results have been shown to correlate strongly (0.87) with results of the standard Dutch clinical speech-in-noise test (Plomp and Mimpen, 1979). For the Dutch version of the test, the average score for normally-hearing listeners is -6.9 dB. To detect hearing problems, the criterion of the Dutch screening test has been set to larger than -4.1 dB for an "insufficient" result, and to larger than -1.4 dB for a "poor" result. For extended descriptions of this test and its Internet version, see D1-4b and D1-6 plus D11-3, respectively.

### 3.1.4 The HearCom questionnaire

The HearCom questionnaire was designed for Internet participation. It comprises 18 questions (18 Qs), and its validation was described in D11-5. The list was developed to subjectively measure speech-perception in quiet (4 Qs) and in noise (5 Qs), localisation (4 Qs), sound-recognition (3 Qs), and detection (2 Qs) abilities. The response categories are: 'niet van

toepassing' (not applicable) 'bijna nooit' (almost never), 'soms' (occasionally), 'vaak' (frequently), and 'bijna altijd' (almost always). In D11-5, this questionnaire was found to have a sensitivity of 0.75 and a specificity of 0.86 for corrected handicap (see D11-5, page 52). The exact questions of the Dutch questionnaire are given in Appendix A.

## 3.2 Description of the comparison tests

We used two well established tests to compare the results of the screening tests with. The digit-triplet test has already been successfully compared to standard clinical speech tests by Smits et al, 2004. To set a benchmark for the localisation tests, we chose the 8-speaker test as developed at the VUmc by Goverts during his thesis work (Goverts, 2004). The results of the HearCom questionnaire are compared with those of the Amsterdam Inventory, developed by Kramer during her thesis work (Kramer, 1998). These two measures have been linked in a study comparing several measures of auditory disability with self-reported hearing disability (Kramer et al., 1996), showing a strong relationship between the chosen localisation test and scores of the Amsterdam Inventory.

### 3.2.1 The 8-speaker test

The 8-speaker test that was used in the experiments is described in detail in the thesis of Theo Goverts (Goverts, 2004). In short, eight speakers are positioned at 45-degree intervals around a listener, who is sitting in a chair and faces forward towards the front speaker (see Figure 3-2). The speakers are positioned at ear height, at a distance of approximately 1.5 meter from the head of the listener. From all speakers, distracting sounds with an average length of 2.9 s. are presented in a random fashion, with a maximum overlap of 70% between the sounds, and with a maximum of three distracters playing at the same time. The distracters were: the chime of the Big Ben, a signing bird, water flowing from a bottle, a laughing baby, a fragment of acoustic-guitar playing, a barking dog, and a police siren. The target sound was an old-fashioned telephone ringing for a duration of 3.3 s. All sounds were presented at 60 dB SPL for normally-hearing listeners, and at levels up to 85 dB SPL for hearing-impaired listeners, depending on their audiograms and preferences. The task of the listeners was to ignore all distracters and point to the speaker from which the target sound was presented. Twelve targets were presented from random positions with pauses of 4 to 10 seconds between them. The reverberation time of the room in which the test was executed was approximately 0.5 s. at the VUmc, while it may have been slightly shorter at the EMC, as a smaller room was used there. With such a test set-up, Goverts reported an average correct score of 86% for fifteen normally-hearing listeners. In contrast, he found an average score of 47-percent correct for a group of 39 hearing-impaired listeners.

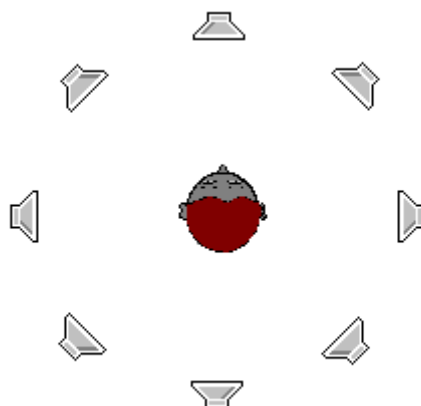


Figure 3-2: *Speaker set-up for the 8-speaker test*

### 3.2.2 The Amsterdam Inventory (Kramer list)

For the current study, the Amsterdam Inventory was adapted for Internet participation. Paper versions were also available, which were used by 11 subjects. It comprises 28 questions (28 Qs) which are illustrated with cartoons depicting the listening situations. The response categories are: 'bijna nooit' (almost never), 'soms' (occasionally), 'vaak' (frequently), 'bijna altijd' (almost always), and 'niet van toepassing, ik kom nooit in deze situatie' (not applicable). The validation of the questionnaire was described in the Thesis of Sophia Kramer (Kramer, 1998) and in Kramer et al., 1995, 1998. The list was developed to subjectively measure speech-perception in quiet (5 Qs) and in noise (5 Qs), as well as localisation (5 Qs), sound-distinction (8 Qs), and detection (5 Qs) abilities, while intolerance to loud sounds was evaluated using a single question.

## 3.3 Subjects and procedures

The data for the present study were collected at two audiological centres. Both normally-hearing listeners and hearing impaired listeners were asked to participate in the tests. Many of them were recruited from the patient data bases of the hospitals. Many participants did some of the tests at home and some at either audiological centre, creating a degree of overlap (details are given below). The aim was to collect data for the validation of the MAA test and to mutually compare the tests results of various tests. So, participants were asked to perform as many tests of the battery as was feasible for them, and not all participants performed all the tests. All HI listeners were asked to perform the measurements without the assistance of any hearing aids, and they were asked to fill in the questionnaires for unaided situations. Next to severely hearing-impaired listeners, we also included listeners with developing hearing impairments for more resilient data analyses, as well as a few with asymmetric losses.

### 3.3.1 VUmc, Amsterdam

At the VUmc, 12 hearing-impaired listeners were asked to participate in all the tests. We asked them to first try the tests at home and subsequently to come and visit the audiological centre to perform them there once more under supervision (they were paid for that visit). In addition, 9 normally-hearing listeners were recruited. They followed the same procedure: first trying the tests at home and then coming over to the audiological centre to perform them in the laboratory. By comparing the data collected from their homes and the data collected at the laboratory, the effect of unsupervised versus supervised test execution can be studied. This is particularly interesting for the MAA test, for which the users need to set up the speakers according to the rather precise specifications that are explained on the Internet pages introducing the test (see Figure 3-1). The average audiograms of all VUmc participants are shown for right and left ears separately in Figure 3-3.

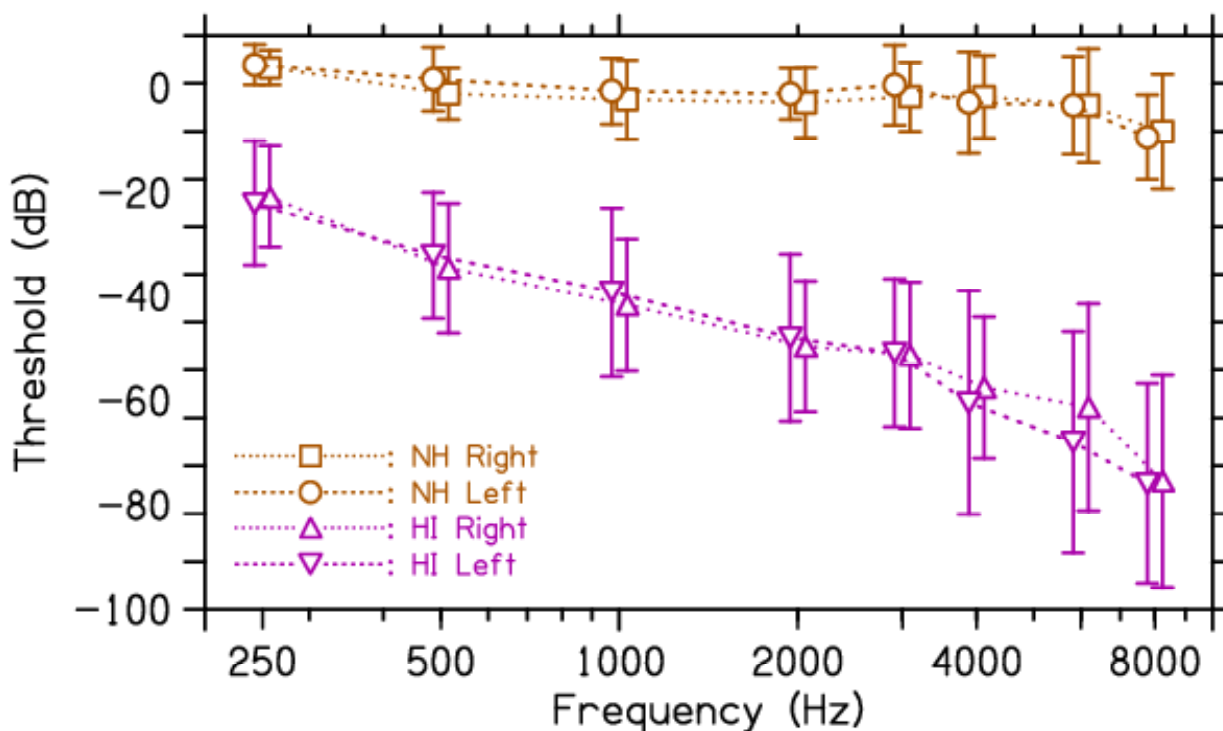


Figure 3-3: Average audiograms for the VUmc participants

### 3.3.2 EMC, Rotterdam

At the EMC, both normally-hearing listeners and hearing impaired listeners were asked to participate for those tests that could be conveniently executed by them, depending on time and possibilities. One group of participants was selected from the patient data base, and they were asked to perform the tests from their homes via the Internet. The selection criteria for this group were: (1) participants needed to be younger than 70 years, (2) have a total air-bone threshold gap of less than 15 dB, and (3) the availability of a recent audiogram (less than 1 year old). Another

group was asked to participate briefly, executing one or more tests, during an already scheduled visit to the hospital. This group included both hearing-impaired and normally-hearing listeners. They were asked to perform the tests from their homes later on, where many of them performed several of the tests. So, many participants executed tests for a subset of the measurement conditions. The EMC subjects received no payment for their participation. On the basis of their audiograms, they were classified into a normally-hearing (20 NH) group, a moderately hearing-impaired (12 MHI) group, a hearing-impaired group (24 HI), and a group with asymmetrical hearing losses (8 AHI). The normally-hearing group had no losses larger than 20 dB at the audiometric frequencies of 0.25, 0.5, 1, 2, and 4 kHz, and no losses larger than 30 dB at 8 kHz. The hearing-impaired group had an average loss of more than 20 dB over all those frequencies. The group with moderate losses fell between the two already described groups. Finally, the group of eight listeners with asymmetrical losses had a difference of more than 12 dB between the PTA3 values (averaged over 0.5, 1 kHz, and 2 kHz) of their two ears (their average PTA3 values are given in the tables below). Six of them would otherwise have fallen into the HI group, and the remaining two would have been part of the MHI group. The average audiograms of all EMC participants are shown for right and left ears separately in Figure 3-4. Taken together with the subjects from the VUmc, the NH, MHI, AHI, and HI groups now contain 29, 12, 8, and 36 listeners, respectively.

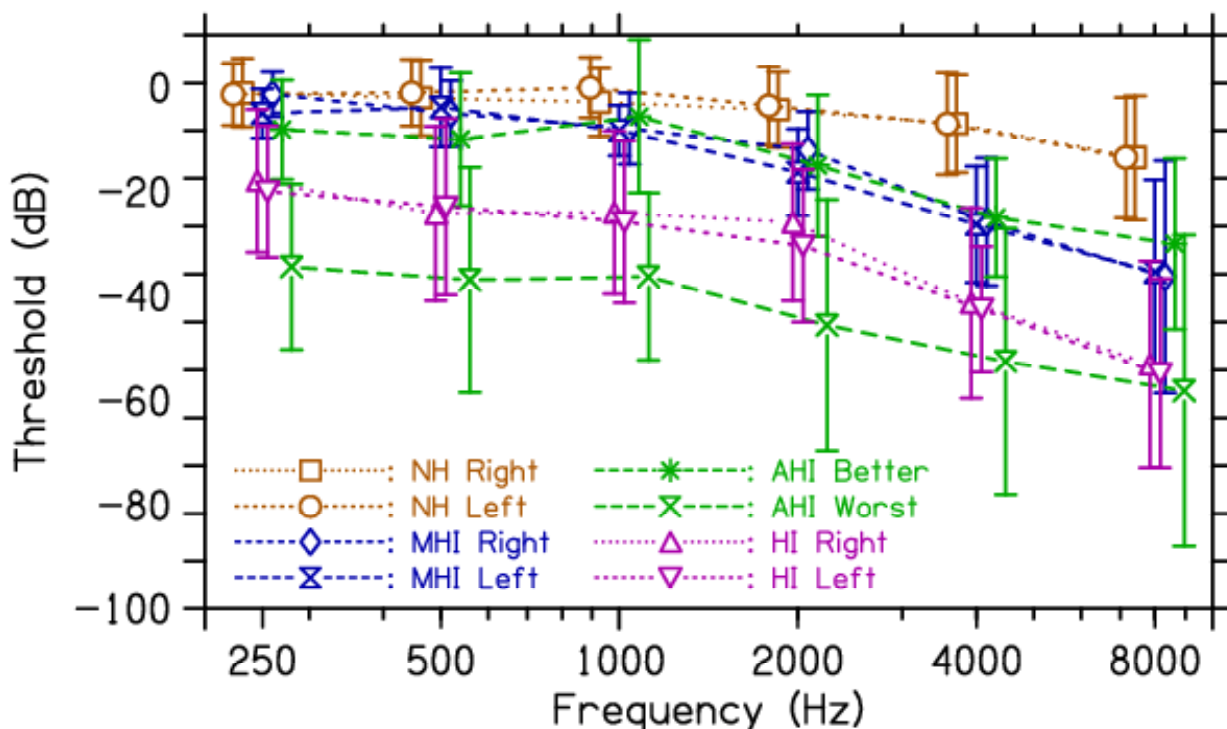


Figure 3-4: Average audiograms for the EMC participants



## 4 Validation of the MAA test

The first objective of the study was to validate the newly developed Internet version of the MAA test as an Internet screening test. For this, two criteria need to be established, with which the results of the participants can be classified as “good,” “intermediate,” and “poor.” In Section 5, the results of the MAA will be compared to those of the other screening tests, and to those of two well-established tests.

### 4.1 Subjects and methods

At the VUmc, 9 NH and 12 HI listeners participated. At the EMC, data were collected for 12 NH, 6 MHI, 4 AHI, and 13 HI listeners (some EMC data were lost due to technical problems at partner DE-FIT). To evaluate the scoring criteria, the data of the total group of 21 NH listeners will be compared to the data for the remaining 35 listeners with some degree of hearing problem. This analysis will concentrate on the scores collected in the laboratories, where a fixed set-up of computer and speakers was used for the measurements, set up according to the instructions. The laboratory scores will later be compared to those collected, either at an earlier or a later instant, via the Internet (see Section 5).

	Sex	Age	MAA 1	MAA 2	Room	Home	PTA3
NH01	F	40	0.5	0.75		0.5	-6.67
NH02	M	22	0.875	1.5		0.75	-4.17
NH03	M	32	3.125	2.75		0.875	5.00
NH04	F	55	1	1.125			5.00
NH05	F	26	1.625	1.625		2.875	-2.50
NH06	F	26	0.5	0.625		0.625	0.00
NH07	F	32	1	0.5		0.75	-3.33
NH08	F	22	1.625	1.375		1.375	0.83
NH09	F	30	0.5	0.875		2	-3.33
NH10	F	51	2.625				11.67
NH16	M	35	1.375				6.67
NH17	F	45	5.625				0.00
NH18	F	23	1.125		3.875		0.83
NH19	F	22	0.625				-4.17
NH21	M	25	4		2		5.83
NH22	M	36	1.625				-5.00
NH24	F	51	5.75				2.50
NH25	F	49	5.25				0.83
NH27	F	50	2				3.33
NH28	F	35	1.875				1.67
NH29	M	28	1.375				9.17
Avg.		35.0	2.10	1.24	2.94	1.22	1.15
SD		11.1	1.69	0.69	1.33	0.83	4.88
SE		2.4	0.37	0.23	0.94	0.29	1.06
Nr		21	21	9	2	8	21

Table 4-1: NH data for the MAA test

## 4.2 Results

Table 4-1 shows the NH results for the laboratory version of the MAA test. The 4<sup>th</sup> column shows the results for the sound-proof booth, the 5<sup>th</sup> column shows the retest results for the first nine NH listeners. Columns 6 and 7 show the results for the NH listeners who performed the test in a quiet office room and/or at home, respectively. Columns 2, 3, and 8 show the listeners’ Sex, Age, and PTA3, respectively. The four bottom rows show average (Avg.), standard deviation (SD), standard error (SE), and number of entries (Nr) of the columns above.

	Sex	Age	Box 1	Box 2	Room	Home	PTA3
MHI01	M	54	1.5		1.25		-3.33
MHI02	F	37	3.875				8.33
MHI03	F	43	2.5		1.5	2.125	6.67
MHI08	M	59	6.25		7		5.00
MHI09	M	46	0.625				4.17
MHI11	F	48	4.75				12.50
AHI01	M	41	3			4.125	43.33
AHI05	F	50	2.625				19.17
AHI06	F	70	31.125				22.50
AHI07	F	21	1.625				21.67
HI01	F	74	2.125	4.375			43.33
HI02	F	48	1	2.5			30.83
HI03	F	62	0.625	0.875			9.17
HI04	F	57	3.5	1.875			45.00
HI05	F	49	2.25	3.375			37.50
HI06	M	49	1.75	1.625			21.67
HI07	F	65	3.375	0.875			29.17
HI08	M	58	2.375	2.875			32.50
HI09	F	72	1.875	0.625			32.50
HI10	F	72	4.5	2.125			25.83
HI11	F	63	5	3.5			36.67
HI12	M	63	1.125	2.25			36.67
HI13	F	64	2.25				16.67
HI14	M	61	1.125		2.375		10.83
HI15	M	60	3.125				10.00
HI16	F	64	13		9.875		58.33
HI17	M	53	2.875		8.25	1.5	20.00
HI18	F	52	8.375				53.33
HI21	M	46	3.125				22.50
HI29	M	65	5.5				25.83
HI30	M	61	10.125				31.67
HI32	M	37	1.625		1.5		17.50
HI34	F	30	3.375				25.00
HI35	M	71	5.75		1		11.67
HI36	F	70	2.125				9.17
Avg.		55.3	4.3	2.2	4.1	2.6	23.8
SD		12.7	5.4	1.2	3.6	1.4	14.7
SE		2.1	0.9	0.3	1.3	0.8	2.5
Nr		35	35	12	8	3	35

Table 4-2: HI data for the MAA test

Table 4-2 shows the data for the HI listeners: the subsequent sections show the data for the MHI, AHI, and HI groups, respectively. As indicated in the 1<sup>st</sup> column, the results for the MHI, AHI, and HI groups are shown from top to bottom, respectively. The structure of this table is identical to that of Table 4-1.

The average age of the NH group is 35.0 (SD = 11.1) years, while that of the collected hearing-impaired group is 55.3 (SD = 12.7). This seems to be in agreement with the relatively high degree of presbycusis in the hearing-impaired groups. The average PTA3 of the NH group is 1.2 (SD = 4.9) dB, while that of the collected hearing-impaired group is 23.8 (SD = 14.7) dB, which is considerably higher, though some very low values occur in column 8 of Table 4-2. These are listeners with significant losses in the higher frequencies that are not taken into account in the PTA3.

The average MAA test score of the NH group is 2.10 (SD = 1.69) degrees, while that of the collected hearing-impaired group is 4.28 (SD = 5.38). So, the degree of spread in these data is rather large, signifying the occurrence of outliers. This will be discussed in the next section. The test (Avg. = 1.19) – retest (Avg. = 1.24) reliability recorded from the first nine NH listeners is 0.26 degrees. For the first 12 HI listeners the test (Avg. = 2.46) – retest (Avg. = 2.24) reliability was found to be 1.13 degrees.

### 4.3 Discussion

The average MAA score of the NH group was found to be 2.10 (SD = 1.69) degrees. However, the results for NH17, NH24, and NH25, are very close to the average plus two times the SD (5.48 degrees), so these listeners will be excluded from further analysis. For the remaining 18 NH listeners, the average MAA score now is 1.52 (SD = 0.95). This would set the criterion for “insufficient” at Avg. + 2·SD to 3.4 degrees, and the criterion for “poor” at Avg. + 4·SD to 5.3 degrees. Such a set of criteria would score all NH listeners as “good.” This is in good agreement with the earlier results from the study in which the MP3 quality of the stimuli was determined (D1-5). In that study, an average score of 1.60 (SD = 0.78) was found for 12 NH listeners, leading to the preliminary criteria of 3.5 and 5.0 degrees for the categories “insufficient” and “poor” (see D1-5, Section 4.3).

The average MAA score of the collected hearing-impaired group was 4.28 (SD = 5.38). For one AHI listener (AHI06), the result falls outside the range of Avg. ± 2·SD (from 0 to 15 degrees). For the remaining 34 collected hearing-impaired listeners, the average MAA score now is 3.49 (SD = 2.70). For listener AHI06, with a strongly asymmetrical hearing loss, the test was completely impossible to perform. The remaining three AHI listeners scored between 1.6 and 3 degrees, bringing them in the range of “good” scores. Three of the moderately hearing-impaired listeners (MHI) are classified in this range as well, while two are scored as

“insufficient,” and one is scored as “poor.” For the HI group, 17 listeners score “good,” 3 score “insufficient,” and 5 score “poor.” So, overall the hearing-impaired listeners score rather well at the present localisation task. This is in agreement with the results of Lorenzi et al. (1999b), who showed relatively good localisation scores for HI listeners in quiet, using clearly audible stimuli without the use of any hearing aids. The localisation skills of their hearing-impaired listeners appeared to suffer more from the presence of background noise than was the case for normally-hearing listeners (Lorenzi et al., 1999a). This observation seems to make the MAA test somewhat less suitable as a detector of absolute threshold problems. How the test functions as a detector of general localisation problems will be discussed in the Section 5.

## 4.4 Conclusions

From the present data set, the MAA criterion for “insufficient” can be set at 3.4 degrees, and the criterion for “poor” at 5.3 degrees. The test-retest reliability for nine NH listeners was found to be less than 0.3 degrees, for twelve HI listeners it was found to be close to 1 degree.

# 5 Inter comparison of screening tests

The purpose of this section is to study the inter relationship between the HearCom screening tests, and to compare their results to two well established measures: an 8-speaker localisation test (Goverts, 2004) and a questionnaire (Amsterdam Inventory, Kramer, 1998), the results of which were shown to be strongly correlated (Kramer et al., 1996). In addition, we will investigate the effects of locations and instructions on the results of the screening tests by comparing the results of tests executed at home and at the audiological centres.

## 5.1 Subjects and methods

At the VUmc, the same 9 NH and 12 HI listeners participated as in the MAA validation. At the EMC, data were collected for 20 NH, 12 MHI, 8 AHI, and 24 HI listeners. Taken together over the two institutes, the NH, MHI, AHI, and HI groups contained 29, 12, 8, and 36 listeners, respectively. However, not all listeners participated for all tests (see below for details). The listeners that performed tests at home did so before coming in to the VUmc, or after visiting the EMC.

## 5.2 Results

In this section the results are presented per test, except for the MAA test, for which the results have already been presented in Section 4.2.

### 5.2.1 Digit-triplet test

ID	Sex	Age	Box 1	Box 2	Room HP	Room EP	Home	PTA3
NH01	F	40	-8.2	-8.8			-7	-6.67
NH02	M	22	-6.4	-8.6			-6.4	-4.17
NH03	M	32	-6.4	-7.8			-6.6	5.00
NH04	F	55	-8.4	-7.8			-6.8	5.00
NH05	F	26	-6.8	-7.6			-7.8	-2.50
NH06	F	26	-9.8	-8			-7.6	0.00
NH07	F	32	-6.6	-5.8			-5.4	-3.33
NH08	F	22	-7	-6.4			-5.4	0.83
NH09	F	30	-8.6	-7.6			-7.4	-3.33
NH10	F	51	-5.6					11.67
NH11	M	35	-6.8					1.67
NH12	F	53	-4.6		-7.6			-5.00
NH13	F	34	-6.2				-7	-2.50
NH15	M	21	-6.4					11.67
NH16	M	35	-7.6					6.67
NH17	F	45	-6.4					0.00
NH18	F	23	-9.2	-8.2	-8	-6.4		0.83
NH19	F	22	-9.8					-4.17
NH20	F	27	-7.4		-5.4			-5.00
NH21	M	25	-8.6					5.83
NH22	M	36	-9.4					-5.00
NH24	F	51	-4.8					2.50
NH25	F	49	-7.2					0.83
NH26	F	43	-4.4					0.83
NH27	F	50	-7.2					3.33
NH28	F	35	-8.2					1.67
NH29	M	28	-8.4					9.17
Avg.	35.1	-7.27	-7.66		-7.00	-6.40	-6.74	0.96
SD	10.9	1.50	0.92		1.40		0.83	5.09
SE	2.1	0.29	0.29		0.81		0.26	0.98
Nr	27	27	10		3	1	10	27

Table 5-1: NH data for the digit-triplet test

Table 5-1 shows the digit-triplet SRTs for the normally-hearing listeners. The layout of this table is identical to that of Table 4-1, with the SRTs for the quiet office room measured with headphones (Room HP) or ear-plug speakers (Room EP). In the following, the SRT of NH26 is excluded, because it is approximately Avg. + 2·SD. For NH01 to NH09 (the VUmc NH), the “Box” results were measured with speakers, for the others (the EMC NH) with headphones. The average score of the VUmc group is -7.6 (SE = 0.4) dB, and for the EMC group it is -7.3 (SE = 0.4) dB. Since the difference of the averages is smaller than the SE values, this difference is not significant, and the “Box 1” data from VUmc and EMC are pooled.

Both the average sound-proof booth score of -7.3 (SD = 1.5) dB, and the test (Avg. = -7.7) – (Avg. = -7.7) retest reliability of 0.8 dB for the 10 NH

listeners who did the test twice in the booth, are in good correspondence with previously found values (Smits et al., 2004; Smits and Houtgast, 2005). The current data (excluding NH26) lead to “insufficient” and “poor” criteria of -4.6 and -1.7 dB for the Internet version of the test. These values are very close to the values of -4.1 and -1.4 dB that are already in use for the Dutch telephone version of the test.

ID	Sex	Age	Box 1	Box 2	Room HP	Room EP	Home	PTA3
MHI01	M	54	-6.4		-6	-8.6		-3.33
MHI02	F	37	-6.2					8.33
MHI03	F	43	-5.6		-6.2	-7.8	-2.6	6.67
MHI04	F	40	-7					3.33
MHI05	F	24	-4.8					10.83
MHI06	F	54	-7.6					5.00
MHI07	M	52	-8.4					19.17
MHI08	M	59	-7.6					4.17
MHI09	M	46	-6		-7.2			8.33
MHI10	F	64	-7.8					6.67
MHI11	F	48	-6.4					12.50
MHI12	M	32	-8.6		-8.8			4.17
AHI01	M	41	-4.8		-3.4		-6.8	43.33
AHI02	M	70	-7.8		-9			21.67
AHI03	M	41	-6.8				-5.2	9.17
AHI04	F	38	-5.4		-5.4			29.17
AHI05	F	50	-8.4					19.17
AHI06	F	70	-6.6					22.50
AHI07	F	21	-9.2		-7.2	-7.8		21.67
AHI08	F	63	-6		-7.2			8.33

Table 5-2: MHI and AHI data for the digit-triplet test

The digit-triplet SRTs for the hearing-impaired listeners are shown in the Tables 5-2 (MHI and AHI) and 5-3 (HI). The layouts of these tables are identical to that of Table 5-1, except that the averages for all hearing-impaired listeners (the MHI, AHI, and HI groups) are shown at the bottom of Table 5-3. For one AHI listener (HI13), the result of -0.2 dB falls outside the range of Avg. ± 2·SD (from 0 to 15 dB). This subject will be discarded in the next analysis. For the remaining 52 collected hearing-impaired listeners, the average SRT now is -5.66 (SD = 2.73). Using the criteria of -4.6 and -1.7 dB, as found above for 16 NH listeners, all 12 MHI and 8 AHI are scored as “good.” For the HI group, 5 listeners are scored as “poor,” 8 are scored as “insufficient,” and 20 score “good.” The good results of the MHI listeners indicate that the digit-triplet test does not detect moderate hearing losses very well. This is in agreement with the results of Smits and Houtgast (2005), who also found relatively good results for such listeners. The good results of the AHI group indicate that these listeners used their better ear in this task, which is in line with expectations for asymmetrical hearing losses. A fair number of the HI listeners score rather well in this speech-in-noise task, but 13 of the 33 HI listeners do not reach a “good” score and would be advised to seek help.

ID	Sex	Age	Box 1	Box 2	Room HP	Room EP	Home	PTA3
HI01	F	74	-2.4	-2			-2	43.33
HI02	F	48	-4.6	-4.8				30.83
HI03	F	62	-5	-5.8				9.17
HI04	F	57	-2.8	-3.6				45.00
HI05	F	49	-2.8	-2.4				37.50
HI06	M	49	-5.4	-8.4				21.67
HI07	F	65	-4.8	-8.8				29.17
HI08	M	58	-0.8	-2.6				32.50
HI09	F	72	-3.8	-5.2				32.50
HI10	F	72	-1	-0.2				25.83
HI11	F	63	-7	-5.6				36.67
HI12	M	63	-6.6	-5.8				36.67
HI13	F	64	-0.2				-3.8	16.67
HI14	M	61	-7		-5.8			10.83
HI15	M	60	-6					10.00
HI16	F	64	-0.6		-1	-3.8		58.33
HI17	M	53	-9		-7.4		-6.8	20.00
HI18	F	52	6.2		3			53.33
HI19	M	52	-6.8					22.50
HI21	M	46	-7				-5.4	22.50
HI22	F	60	-7.2		-7.4		-4	19.17
HI23	M	55	-6		-8.6			20.00
HI24	F	55	-8.2		-8	-7.6	-7.2	20.00
HI25	M	53	-8.6					2.50
HI26	F	43	-4.8		-8.4			18.33
HI27	F	57	-2.2					42.50
HI29	M	65	-5.6					25.83
HI30	M	61	-3.6					31.67
HI31	M	29	-8.6		-7.4	-8.2	-6.6	17.50
HI33	F	31	-3.6		-2.8			16.67
HI34	F	30	-8.2					25.00
HI35	M	71	-6.2		-7.4	-4.4		11.67
HI36	F	70	-6.8					9.17
Avg.		53.0	-5.55	-4.60	-6.08	-6.89	-5.04	21.05
SD		13.1	2.80	2.56	2.98	1.94	1.86	13.77
SE		1.8	0.38	0.74	0.67	0.73	0.59	1.89
Nr		53	53	12	20	7	10	53

Table 5-3: HI data for the digit-triplet test along with the averages for the MHI, AHI, and HI groups.

### 5.2.2 The AVE and 8-speaker tests

Table 5-4 shows the data for the AVE and 8-speaker localisation tests for the normally-hearing listeners. The first three columns contain the listener ID and their sex and age. The next four columns contain the results of the first and second AVE test (test and retest) and the AVE tests performed in a quiet office and at home. These scores contain two presentations per location (12 in total), and they are expressed as a correct ratio. The next four columns contain the same data as transformed correct scores. These

scores were calculated according to the results and methods described in deliverable D3-8. In short, in this method the individual scores are compared with the logarithm of the chance distribution of a large group of normally-hearing listeners (“normal” distribution) and the guessing chances, and they are expressed as the log-chance to obtain the recorded distribution by drawing from the “normal” distribution. Larger (more positive) numbers imply better scoring. The next two columns show the data for the 8-speaker test, for six presentations per speaker position. They are expressed as a correct ratio in the first column, and as transformed log-chances in the second column. These latter scores were calculated in a similar way as those for the AVE tests; the average distribution of all NH listeners was used as the “normal” distribution. The last column shows the PTA3. At the bottom of each column, the average, SD, SE, and number of listeners are shown.

ID	Sex	Age	AVE Box1	AVE Box2	AVE Room	AVE Home	AVE Box c1	AVE Box c2	AVE Rm. c	AVE Hm. c	8-Sp Avg.	8-Sp c	PTA3
NH01	F	40	0.750	0.750			1.145	1.145			0.625	1.000	-6.67
NH02	M	22	0.500	0.750			0.819	1.345			0.833	1.516	-4.17
NH03	M	32	0.667	0.583			0.863	0.717			0.792	1.386	5.00
NH04	F	55	0.750	0.583			1.365	0.827			0.688	1.193	5.00
NH05	F	26	0.250	0.667			-0.602	1.037			0.854	1.504	-2.50
NH06	F	26	0.667	0.667			1.224	1.057			0.688	1.106	0.00
NH06	F	32	0.833	0.917			1.299	1.253			0.938	1.690	-3.33
NH08	F	22	0.917	1.000			1.503	1.520			0.875	1.660	0.83
NH09	F	30	0.750	0.667			1.235	0.729			0.833	1.468	-3.33
NH10	F	51	0.250				-1.209						11.67
NH12	F	53									0.542	0.991	-5.00
NH15	M	21									0.521	1.076	11.67
NH16	M	35				0.917				1.303			6.67
NH18	F	23	0.750				0.991				0.542	1.098	0.83
NH19	F	22	0.750				1.158						-4.17
NH20	F	27									0.583	1.200	-5.00
NH22	M	36	0.917				1.503				0.583	1.059	-5.00
NH24	F	51	0.917				1.316						2.50
NH25	F	49	0.917				1.503						0.83
NH26	F	43									0.750	1.368	0.83
NH28	F	35	0.917				1.373				0.583	1.150	1.67
NH29	M	28									0.625	1.380	9.17
Avg.		34.5	0.719	0.731		0.917	0.968	1.070		1.303	0.697	1.285	0.80
SD		11.3	0.217	0.143			0.768	0.278			0.135	0.228	5.48
SE		2.4	0.054	0.048			0.192	0.093			0.033	0.055	1.17
Nr		22	16	9		0 1	16	9		0 1	17	17	22

Table 5-4: NH data for the localisation tests

Table 5-5 shows the data for the AVE and 8-speaker localisation tests for the three groups of hearing-impaired listeners, from top to bottom: MHI, AHI, and HI. The layout of this table is the same as that of Table 5-4. The averages at the bottom of the table are calculated over all three groups.



ID	Sex	Age	AVE Box1	AVE Box2	AVE Room	AVE Home	AVE Box c1	AVE Box c2	AVE Rm. c	AVE Hm. c	8-Sp Avg.	8-Sp c	PTA3
MHI01	M	54	0.583		0.833		0.660		1.278				-3.33
MHI02	F	37	0.833				1.382				0.583	-0.447	8.33
MHI03	F	43	0.333		0.500	0.167	-1.642		0.152	-4.270			6.67
MHI04	F	40	0.833				1.382						3.33
MHI08	M	59	0.917				1.399						5.00
MHI09	M	46				0.750				1.138			4.17
MHI11	F	48	0.583				0.361				0.521	0.220	12.50
AHI01	M	41	0.167		0.333	0.250	-2.756		-1.097	-2.118			43.33
AHI02	M	70	0.167		0.417	0.083	-2.389		-1.833	-2.965			21.67
AHI03	M	41									0.479	-0.410	9.17
AHI05	F	50	0.750				1.235						19.17
AHI07	F	21	0.833				1.125				0.688	-0.494	21.67
AHI08	F	63	0.250				-0.335				0.458	-0.851	8.33
HI01	F	74	0.833	0.917			1.252	1.373			0.354	-1.510	43.33
HI02	F	48	0.167	0.250			-2.694	-1.454			0.521	0.067	30.83
HI03	F	62	0.500	0.583			0.391	0.114			0.354	0.302	9.17
HI04	F	57	0.417	0.500			-1.087	-0.845			0.500	-0.659	45.00
HI05	F	49	0.917	0.583			1.373	0.640			0.271	-2.697	37.50
HI06	M	49	0.167	0.250			-3.619	-3.241			0.313	0.583	21.67
HI07	F	65	0.333	0.583			-1.611	0.626			0.542	-1.088	29.17
HI08	M	58	0.417	0.500			-2.815	0.259			0.458	-0.369	32.50
HI09	F	72	0.250	0.083			-1.865	-2.819			0.271	0.861	32.50
HI10	F	72	0.250	0.333			-0.911	-1.839			0.208	0.081	25.83
HI11	F	63	0.333	0.250			-1.436	0.068			0.458	1.477	36.67
HI12	M	63	0.167	0.167			-2.663	-1.562			0.271	-0.191	36.67
HI13	F	64	0.083			0.750	-3.720			1.345	0.583	-0.095	16.67
HI14	M	61	0.667		0.917		0.948		1.212				10.83
HI15	M	60	0.833				1.252						10.00
HI16	F	64	0.250		0.750		-0.934		0.882				58.33
HI17	M	53	1.000		0.917		1.520		1.373				20.00
HI18	F	52	0.417		0.583	0.750	0.119		0.744	0.985			53.33
HI21	M	46	0.500				0.353				0.646	0.568	22.50
HI22	F	60	0.333				-0.580						19.17
HI23	M	55	0.500		0.333		-1.795		0.091		0.604	0.666	20.00
HI29	M	65				0.667				0.941			25.83
HI30	M	61	0.750				1.047				0.604	1.162	31.67
HI35	M	71	0.000				-7.622						11.67
HI36	F	70	0.417				-0.559						9.17
Avg.		56.0	0.479	0.417	0.620	0.488	-0.721	-0.723	0.311	-0.707	0.461	-0.135	22.37
SD		11.6	0.280	0.236	0.240	0.306	2.021	1.464	1.120	2.344	0.139	0.944	14.79
SE		1.9	0.047	0.068	0.080	0.116	0.342	0.423	0.373	0.886	0.030	0.206	2.40
Nr		38	35	12	9	7	35	12	9	7	21	21	38

Table 5-5: MHI, AHI, and HI data for the localisation tests

When comparing the average AVE results of the NH group with those of the hearing-impaired groups, we can see that the average correct score of the NH is 0.72 (SD = 0.22), while that of the hearing-impaired groups is 0.48 (SD = 0.28). So, these scores are only one NH SD apart. However, for the transformed scores we find an average of 0.97 (SD = 0.77) for the

NH, and  $-0.72$  ( $SD = 2.02$ ) for the hearing-impaired groups. So, using this method, the scores of the hearing-impaired groups are about two NH SDs away from the NH scores. Clearly this method (described in detail in deliverable D3-8) gives the better results for the AVE test. When checking the results for the NH group, the AVE-Box-c1 results of NH05 and NH10 are more than two SDs away from the average and will be discarded in the next analysis. The same is true for the results of HI35 in the group of hearing-impaired listeners, who produced no correct answers at all. These scores will be disregarded in the next analysis as well. Now the average transformed score of the NH is  $1.24$  ( $SD = 0.22$ ), and of the hearing-impaired group it is  $-0.52$  ( $SD = 1.65$ ). This is in good general agreement with the results in D3-8 (page 15).

A similar effect for the average correct and transformed scores is seen for the results of the 8-speaker test that was included in the experiments as a reference test. Here we find average correct scores of  $0.70$  ( $SD = 0.14$ ) for the NH group, and  $0.46$  ( $SD = 0.14$ ) for the hearing-impaired groups. So, for this outcome measure, these average scores are about 1.7 NH SDs apart. The average transformed scores are  $1.29$  ( $SD = 0.23$ ) for the NH group and  $-0.14$  ( $SD = 0.94$ ) for the hearing-impaired groups. So, here the average scores are about 6.2 NH SDs apart. For this reason, the transformed scores will be used in all later analyses. For this measurement, no clear outliers were observed.

### 5.2.3 HearCom questionnaire and Amsterdam Inventory

Table 5-6 shows the average results of the questionnaires for the NH group. The scores vary from 0.00 (no meaningful reception of sounds) to 1.00 (no problems). For both the HearCom questionnaire (HQ) and the Amsterdam Inventory (AI), the results are shown per category: speech perception in noise (Sp-N), speech perception in quiet (Sp-Q), detection of sounds (Det.), recognition of sounds (Rec.), localisation (Loc.), and the overall score (All). In addition, the column "QE" shows the average results for just the localisation questions of the AI, filled in by all participants who performed the AVE test via the Internet. In addition, the first three columns give the ID, sex and age of the participants, and the last column shows their PTA3. At the bottom of each column, the average, SD, SE, and number of participants are shown.

The averages for the NH group in Table 5-6 show that the overall scores for the HearCom questionnaire are slightly lower than those of the Amsterdam inventory. This is mainly caused by the categories recognition (Rec.) and localisation (Loc.), where the AI produces higher scores than the HQ. For all categories and the overall score, the SDs are somewhat larger for the HQ than for the AI. However, the intention of the HQ is that it is a screening tool, while the more extensive AI is intended to be a research tool.

ID	Sex	Age	HearCom questionnaire						Amsterdam Inventory						QE Loc	PTA 3	
			Sp-N	Sp-Q	Det.	Rec.	Loc	All	Sp-N	Sp-Q	Det.	Rec.	Loc	All			
NH01	F	40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97		-6.67
NH02	M	22	0.93	0.92	0.67	0.78	0.75	0.83	0.93	0.80	0.80	0.88	0.73	0.83	0.89		-4.17
NH03	M	32	1.00	0.92	1.00	0.78	0.83	0.89	1.00	0.93	1.00	0.92	0.93	0.94			5.00
NH04	F	55	0.93	0.89	1.00	0.89	0.83	0.88	0.87	1.00	0.93	0.88	0.93	0.89			5.00
NH05	F	26	0.93	1.00	1.00	1.00	1.00	0.96									-2.50
NH06	F	26	0.73	1.00	0.83	0.67	0.75	0.78	0.87	0.87	0.87	0.90	0.80	0.86			0.00
NH06	F	32	1.00	1.00	1.00	1.00	1.00	0.94									-3.33
NH08	F	22	0.93	0.92	1.00	1.00	0.75	0.91	1.00	1.00	0.93	0.96	0.93	0.97			0.83
NH09	F	30							0.93	0.87	1.00	0.92	1.00	0.94			-3.33
NH11	M	35	0.87	0.67	0.83	1.00	0.75	0.82	1.00	1.00	0.87	1.00	0.93	0.94	0.89		1.67
NH12	F	53													0.83		-5.00
NH13	F	34													0.50		-2.50
NH14	F	19													0.61		1.67
NH15	M	21	0.60	0.75	1.00	0.33	0.67	0.59	0.67	0.80	0.93	0.79	0.73	0.79	0.78		11.67
NH17	F	45	0.87	1.00	1.00	0.89	0.83	0.91							1.00		0.00
NH19	F	22	0.87	1.00	1.00	0.89	0.83	0.91	0.93	0.87	1.00	0.96	0.87	0.92	0.94		-4.17
NH20	F	27	0.67	0.75	0.67	0.67	0.58	0.65	0.80	0.80	0.93	0.92	0.80	0.83	0.67		-5.00
NH21	M	25	0.60	0.92	0.83	0.89	0.67	0.72	0.67	0.80	1.00	0.83	0.80	0.80	0.89		5.83
NH22	M	36	1.00	1.00	1.00	1.00	1.00	0.97							1.00		-5.00
NH24	F	51	0.93	1.00	1.00	1.00	0.83	0.90	0.87	1.00	1.00	1.00	1.00	0.97	1.00		2.50
NH25	F	49	1.00	1.00	1.00	0.89	0.83	0.90	0.93	1.00	1.00	1.00	0.80	0.94	0.83		0.83
NH27	F	50	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00	0.94	1.00		3.33
NH28	F	35	0.80	1.00	1.00	0.89	0.92	0.87	0.87	1.00	1.00	1.00	0.80	0.94	0.89		1.67
Avg.		34.2	0.88	0.93	0.94	0.87	0.83	0.86	0.90	0.92	0.95	0.93	0.88	0.91	0.85		-0.07
SD		11.4	0.14	0.10	0.11	0.17	0.13	0.11	0.11	0.09	0.06	0.07	0.10	0.06	0.15		4.45
SE		2.4	0.03	0.02	0.03	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.04		0.93
Nr		23	19	19	19	19	19	19	16	16	16	16	16	16	15		23

Table 5-6: NH data for the questionnaires

Table 5-7 shows the results of the questionnaires for the three groups of hearing-impaired participants, from top to bottom: MHI, AHI, and HI. The layout of this table is the same as that of Table 5-6. The averages at the bottom of the table are calculated over all three groups. As was found for the NH group, the average scores of the hearing-impaired participants tend to be lower for the HQ than for the AI. For this group this is mainly due to the categories detection (Det.), recognition (Rec.), and localisation (Loc.). In comparison with Table 5-6, it can be seen that the average scores of the hearing-impaired participants are much lower than those of the NH group. Over the complete HQ, the average NH score of 0.86 (SD = 0.11) lies about 3 SDs above the average hearing-impaired score of 0.53 (SD = 0.27). For the entire AI, the average NH score of 0.91 (SD = 0.06) lies nearly 5 SDs above the average hearing-impaired score of 0.63 (SD = 0.24). So, when using the average score, both lists are well able to distinguish between normal and impaired hearing. In D11-5, criteria were proposed for the HearCom Questionnaire for response categories named “green,” “amber,” and “red.” How this works out for the overall scores will be treated further in the Discussion of Section 5.5.

ID	Sex	Age	HearCom questionnaire						Amsterdam Inventory						QE Loc	PTA 3
			Sp-N	Sp-Q	Det.	Rec.	Loc	All	Sp-N	Sp-Q	Det.	Rec.	Loc	All		
MHI02	F	37	0.53	0.58	0.67	0.44	0.58	0.51	0.40	0.67	0.73	0.83	0.67	0.68	0.67	8.33
MHI03	F	43	1.00	0.75	1.00	1.00	1.00	0.92	0.73	0.87	1.00	1.00	1.00	0.92	1.00	6.67
MHI04	F	40	0.93	1.00	1.00	1.00	0.92	0.91	1.00	1.00	1.00	1.00	1.00	0.94		3.33
MHI05	F	24	0.33	0.67	0.17	0.89	0.50	0.49							0.56	10.83
MHI06	M	59	0.80	1.00	1.00	0.78	0.92	0.81	0.60	0.93	1.00	1.00	1.00	0.91	1.00	5.00
MHI10	F	64	0.87	1.00	1.00	0.89	1.00	0.94	0.80	0.87	1.00	0.92	1.00	0.90		6.67
MHI11	F	48	0.27	0.58	0.33	0.56	0.58	0.45	0.47	0.67	0.73	0.71	0.60	0.63	0.61	12.50
MHI12	M	32	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	4.17
MHI13	F	54	1.00	1.00	1.00	0.78	0.83	0.92	1.00	0.87	0.93	0.92	1.00	0.94		0.83
AHI01	M	41	0.53	0.42	0.00	0.78	0.08	0.47	0.53	0.50	0.40	0.88	0.13	0.53	0.11	43.33
AHI02	M	70	0.40	0.58	1.00	1.00	0.33	0.54	0.13	0.67	0.73	0.79	0.27	0.57	0.44	21.67
AHI03	M	41	0.40	0.58	1.00	0.78	0.42	0.54	0.13	0.67	0.67	0.58	0.33	0.50	0.22	9.17
AHI04	F	38													0.17	29.17
AHI05	F	50													0.94	19.17
AHI07	F	21	0.80	0.83	1.00	1.00	0.75	0.86	0.73	0.93	1.00	1.00	0.87	0.92	0.94	21.67
AHI08	F	63							1.00	1.00	1.00	0.96	0.93	0.98	1.00	8.33
HI01	F	74	0.07	0.50	0.17	0.89	0.00	0.27	0.20	0.50	0.40	0.88	0.13	0.46		43.33
HI02	F	48	0.25	0.25	0.17	0.33	0.17	0.19	0.27	0.17	0.13	0.50	0.20	0.29		30.83
HI03	F	62	0.40	0.67	0.67	0.89	0.58	0.51	0.27	0.53	0.67	0.83	0.75	0.62		9.17
HI04	F	57	0.13	0.17	0.00	0.56	0.08	0.15	0.27	0.13	0.13	0.33	0.20	0.26		45.00
HI05	F	49	0.00	0.17	0.00	0.67	0.58	0.42	0.00	0.50	0.20	0.83	0.33	0.47		37.50
HI06	M	49	0.40	0.67	0.33	1.00	1.00	0.64	0.40	0.87	0.80	0.96	0.93	0.82		21.67
HI07	F	65	0.25	0.58	0.17	0.33	0.42	0.36	0.27	0.60	0.73	0.75	0.60	0.59		29.17
HI08	M	58	0.20	0.50	0.33	0.67	0.50	0.38	0.20	0.40	0.47	0.50	0.67	0.48		32.50
HI09	F	72	0.13	0.42	0.00	0.44	0.33	0.28	0.33	0.20	0.47	0.67	0.47	0.43		32.50
HI10	F	72	0.13	0.08	0.00	0.17	0.00	0.06	0.07	0.13	0.13	0.42	0.20	0.23		25.83
HI11	F	63	0.08	0.25	0.17	0.44	0.33	0.24	0.13	0.33	0.13	0.50	0.33	0.33		36.67
HI12	M	63	0.00	0.17	0.17	0.22	0.00	0.09	0.00	0.00	0.47	0.38	0.25	0.25		36.67
HI16	F	64	0.27	0.50	0.17	0.44	0.25	0.31	0.40	0.47	0.60	0.63	0.40	0.51	0.61	58.33
HI17	M	53													1.00	20.00
HI18	F	52	0.20	0.33	0.33	0.44	0.17	0.24	0.27	0.00	0.40	0.42	0.33	0.30	0.61	53.33
HI19	M	52	0.27	0.42	0.67	0.83	0.25	0.36	0.07	0.27	0.33	0.46	0.20	0.32	0.22	22.50
HI20	F	59	0.53	0.58	0.67	0.89	0.92	0.67	0.47	0.40	0.80	0.79	0.93	0.70		22.50
HI21	M	46	0.67	0.92	1.00	1.00	0.89	0.79	0.67	0.93	0.93	1.00	0.73	0.85	0.89	22.50
HI22	F	60	0.33	0.58	0.50	0.89	0.42	0.42	0.27	0.47	0.67	0.71	0.53	0.54	0.39	19.17
HI23	M	55	0.60	0.83	0.67	0.67	0.58	0.65	0.67	0.67	0.87	0.92	0.80	0.80	1.00	20.00
HI26	F	43	0.13	0.50	0.33	0.44	0.17	0.28	0.13	0.47	0.40	0.54	0.40	0.44	0.39	18.33
HI27	F	57	0.20	0.50	0.17	0.89	0.17	0.32	0.20	0.40	0.13	0.63	0.40	0.41	0.33	42.50
HI28	M	50	0.87	0.92	1.00	0.89	0.92	0.87	0.87	0.87	1.00	0.96	1.00	0.91		25.00
HI29	M	65	0.67	0.75	0.83	0.89	0.75	0.69	0.53	0.73	0.87	0.92	0.73	0.79	0.67	25.83
HI30	M	61	0.40	0.67	0.83	0.78	0.75	0.58	0.33	0.67	0.67	0.71	0.60	0.61	0.83	31.67
HI31	M	29	0.93	1.00	0.83	0.89	0.92	0.91	0.93	1.00	0.93	1.00	0.87	0.94		17.50
HI35	M	71													0.94	11.67
HI36	F	70	0.33	0.42	0.67	0.89	0.50	0.45							0.56	9.17
Avg.		53.3	0.44	0.60	0.54	0.73	0.53	0.53	0.44	0.59	0.65	0.76	0.60	0.63	0.66	22.99
SD		13.1	0.31	0.26	0.38	0.24	0.33	0.27	0.31	0.30	0.30	0.21	0.30	0.24	0.30	14.01
SE		2.0	0.05	0.04	0.06	0.04	0.05	0.04	0.05	0.05	0.05	0.03	0.05	0.04	0.06	2.11
Nr		44	39	39	39	39	39	39	38	38	38	38	38	38	26	44

Table 5-7: MHI, AHI, and HI data for the questionnaires

At the VUmc, a number of participants also filled in the questionnaires at home via the Internet, next to filling them in the laboratory. Table 5-8 shows these results. They will be used in Section 5.4 (Effects of locations and instructions) to compare the laboratory results with the home results. At the EMC, participants either filled in the questionnaires at home or in the lab, without any overlap. For the inter comparison of all the tests, the results of the laboratory questionnaires will be used, supplemented with the Internet results where applicable.

ID	Sex	Age	HearCom questionnaire						Amsterdam Inventory						QE	PTA
			Sp-N	Sp-Q	Det.	Rec.	Loc	All	Sp-N	Sp-Q	Det.	Rec.	Loc	All	Loc	3
NH01	F	40	1.00	1.00	1.00	1.00	1.00	1.00							1.00	-6.67
NH02	M	22	0.67	0.83	0.67	0.67	0.75	0.77	0.73	0.87	0.80	0.83	0.73	0.81	0.89	-4.17
NH03	M	32	0.93	0.92	1.00	0.78	0.92	0.85							0.94	5.00
NH04	F	55	0.87	0.92	1.00	1.00	1.00	0.88							0.94	5.00
NH05	F	26	0.87	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00	0.97	1.00	-2.50
NH06	F	26	0.80	0.92	1.00	1.00	0.83	0.82	0.93	0.87	0.93	0.95	1.00	0.93		0.00
NH06	F	32	1.00	1.00	1.00	1.00	1.00	0.97	1.00	0.93	1.00	0.96	1.00	0.98	1.00	-3.33
NH08	F	22	0.87	0.92	1.00	0.89	0.75	0.85	1.00	1.00	1.00	1.00	0.87	0.97	1.00	0.83
NH09	F	30	1.00	1.00	1.00	0.83	1.00	0.93							1.00	-3.33
HI01	F	74	0.07	0.33	0.00	0.78	0.17	0.24							0.22	43.33
HI02	F	48													0.44	30.83
HI11	F	63	0.17	0.17	0.00	0.44	0.25	0.23								36.67
HI12	M	63	0.00	0.25	1.00	0.11	0.50	0.26								36.67

Table 5-8: *Extra questionnaire data via the Internet (VUmc)*

### 5.3 Inter comparison of the tests

For the inter comparison of all the tests we have used only the test data, i.e. the data of the first executed attempt for a certain test. Using average data may shed more light on the correlations between the tests, but those correlations might then not be representative for users who do a test only once. The outliers identified in the previous sections were included in this analysis, as they produced scores for more tests than just the outlying one, and there may still be a relation between those different tests.

The correlations between the test results were calculated for the largest group of participants for which the appropriate data was available. So, for many tests different numbers of participants were involved. Table 5-9 shows correlations between the combined test results of all groups of participants. Since the distribution of the variables was not necessarily normally distributed over all groups of participants, Spearman correlations (rhos) were calculated. In each cell, the number at the top shows the correlation, while the number in brackets at the bottom shows the number of participants for which data could be included. The significance of the correlations is shown by the asterisks: \*\* indicates  $p < 0.01$ , and \* indicates  $p < 0.05$ . This table shows many significant correlations. Because both the NH and the hearing-impaired groups were included, nearly all results correlate with each other. Exceptions are the MAA test

that does not correlate with AVE, 8Sp, HQ, and AI, and the AVE test that does not correlate with the 8Sp. However, since age correlates with all results including the PTA3, it is more interesting to examine the correlations corrected for the factor Age. These partial correlations are shown in Table 5-10. The structure of this table is the same as that of Table 5-9.

Spear.	SRT	MAA	AVE	8Sp	HQ	AI	Age	PTA3
SRT	.	0.421 ** (55)	-0.504 ** (51)	-0.484 ** (38)	-0.662 ** (56)	-0.587 ** (53)	0.442 ** (63)	0.570 ** (63)
MAA		.	-0.105 (46)	-0.348 (31)	-0.332 (41)	-0.301 (38)	0.436 ** (46)	0.514 ** (46)
AVE			.	0.282 (32)	0.557 ** (41)	0.555 ** (40)	-0.428 ** (45)	-0.533 ** (45)
8Sp				.	0.638 ** (32)	0.516 ** (32)	-0.516 ** (34)	-0.610 ** (34)
HQ					.	0.905 ** (55)	-0.521 ** (59)	-0.802 ** (59)
AI						.	-0.571 ** (55)	-0.809 ** (55)
Age							.	0.569 ** (63)

Table 5-9: Spearman correlations for all groups of participants

Part.	SRT	MAA	AVE	8Sp	HQ	AI	PTA3
SRT	.	0.064 (52)	-0.143 (48)	-0.258 (35)	-0.522 ** (53)	-0.488 ** (50)	0.588 ** (60)
MAA		.	0.071 (43)	0.105 (28)	-0.004 (38)	0.005 (35)	0.377 ** (43)
AVE			.	-0.080 (29)	0.413 ** (38)	0.375 * (37)	-0.175 (42)
8Sp				.	0.416 * (29)	0.277 (29)	-0.477 ** (31)
HQ					.	0.959 ** (52)	-0.681 ** (56)
AI						.	-0.672 ** (52)

Table 5-10: Partial correlations for all groups corrected for age

Table 5-10 shows correlations at the 1% level between the PTA3 on the one side and the SRT, MAA, 8Sp, HQ, and AI on the other side. So, all tests except the AVE show a good correlation with the audiogram. However, the AVE correlates at the 1% level with the HQ and at the 5% level with the AI. In addition, the SRT correlates at the 1% level with the HQ and AI. The 8Sp correlates at the 5% level with the HQ, but not with

the AI. The MAA results do not correlate with those of the questionnaires. As before, the HQ and the AI correlate very strongly with each other.

Pears.	SRT	MAA	AVE	8Sp	HQ	AI	Age	PTA3
SRT	.	0.562 ** (21)	-0.399 (16)	0.262 (17)	-0.059 (20)	0.201 (18)	0.328 (22)	0.120 (22)
MAA		.	-0.100 (16)	0.149 (13)	-0.167 (17)	0.038 (15)	0.466 (17)	0.494 * (17)
AVE			.	-0.231 (12)	-0.153 (14)	-0.057 (13)	0.344 (14)	0.147 (14)
8Sp				.	0.290 (13)	0.493 (12)	-0.469 (14)	-0.051 (14)
HQ					.	0.877 ** (18)	0.432 (20)	-0.485 * (20)
AI						.	0.403 (18)	-0.356 (18)
Age							.	0.024 (22)

Table 5-11: Pearson correlations for the NH group

Pearson correlations for the NH group are given in Table 5-11. The layout of this table is the same as that of Table 5-9. This analysis shows correlations at the 1% level between the SRT and MAA, and between the HQ and AI. In addition, correlations at the 5% level are found between the PTA3 on the one side, and the MAA and HQ on the other side.

Spear.	SRT	MAA	AVE	8Sp	HQ	AI	Age	PTA3
SRT	.	0.143 (34)	-0.357 * (35)	-0.201 (21)	-0.643 ** (36)	-0.597 ** (35)	0.273 (41)	0.626 ** (41)
MAA		.	-0.199 (30)	0.181 (18)	0.027 (24)	0.048 (23)	0.056 (30)	0.200 (30)
AVE			.	-0.302 (20)	0.295 (27)	0.346 (27)	-0.247 (31)	-0.296 (31)
8Sp				.	0.192 (19)	0.023 (20)	0.040 (20)	-0.105 (20)
HQ					.	0.956 ** (37)	-0.400 ** (39)	-0.705 ** (39)
AI						.	-0.414 ** (37)	-0.718 ** (37)
Age							.	0.300 (41)

Table 5-12: Spearman correlations for the MHI, AHI, and HI groups

Spearman correlations for the hearing-impaired groups MHI, AHI, and HI are shown in Table 5-12. The layout of this table is the same as that of



Table 5-9. For the SRT, this table shows correlations at the 1% level with the HQ, AI, and PTA3, and at the 5% level with the AVE. In addition, the HQ and AI correlate very strongly with each other, and they both correlate with the PTA3 and the Age ( $p < 0.01$ ).

## 5.4 Effects of locations and instructions

In this section we compare the effects of listening condition and instructions by comparing laboratory results for different rooms, and by comparing laboratory results with those scored at home via the Internet. These comparisons are made using the data of all subject groups. The upper half of Table 5-13 shows the average MAA results of those EMC participants who performed the test in the quiet office room and in the sound-proof booth. From left to right, the table shows condition, average, SD, SE and number of participants for which the data could be calculated. The averages are much less than 1 SE apart, so we observe no significant differences. The lower half of the table shows the same results for 8 VUmc and 3 EMC participants who performed the MAA test in the laboratory and at home via the Internet. Here too, the averages are much less than 1 SE apart and we observe no significant differences.

Cond.	Avg.	SD	SE	Nr.
Box1	3.98	3.67	1.16	10
Room	3.86	3.28	1.04	10
Box1	1.65	1.06	0.32	11
Home	1.59	1.12	0.34	11

Table 5-13: *Comparison of MAA results with those scored in the quiet office room and at home via the Internet*

The condition-comparison data for the digit-triplet test is shown in Table 5-14. It has the same structure as Table 5-13. The top section shows the comparison of the head-phone results for the quiet office room with the speaker results of the sound-proof booth (23 EMC participants). The mid section shows the same comparison while using ear-plug speakers in the quiet office room (8 EMC participants). In both comparisons, no significant differences are observed (t-test, pair-wise comparisons). The bottom section shows the comparison of the laboratory results with the home results for 9 VUmc and 10 EMC participants, here a small but significant difference is found (t-test, pair-wise comparisons:  $t = -3.4$ ,  $p < 0.005$ ).

Cond.	Avg.	SD	SE	Nr.
Box1	-5.91	3.33	0.70	23
Room HP	-6.20	2.82	0.59	23
Box1	-6.75	2.86	1.01	8
Room EP	-6.83	1.80	0.64	8
Box1	-7.05	1.69	0.39	19
Home	-6.00	1.62	0.37	19

Table 5-14: *Comparison of SRT laboratory results with those scored in the quiet office room and at home via the Internet*



The top section of Table 5-15 shows the comparison of the AVE results scored in the quiet office room and in the sound-proof booth (9 hearing-impaired EMC participants). The difference in the average scores was found to be significant (t-test, pair-wise comparisons:  $t = -3.9$ ,  $p = 0.005$ ). Scores in the quiet room are better than those in the sound-proof booth. This may well be related to the sequence of these tests, those in the quiet room were performed after those in the sound-proof booth. So, rather than the quiet room being a better listening environment than the sound-proof booth, these differences may well be the result of practice. The bottom section of table 5-15 shows the average AVE results scored in the laboratory (EMC) using the preferred AKG K240DF (diffuse field) and those scored at home via the Internet using the head-phones that were available. Here, no significant AVE differences were observed (t-test, pair-wise comparisons:  $t = -0.53$ ,  $p = 0.6$ ), but only 5 participants contributed to this comparison.

Cond.	Avg.	SD	SE	Nr.
Box1	-0.70	1.56	0.52	9
Room	0.31	1.12	0.37	9
Box1	-2.08	1.44	0.64	5
Home	-1.40	2.47	1.11	5

Table 5-15: Comparison of the laboratory AVE results with those scored in the quiet office room and at home via the Internet

Table 5-16 shows the average results of each specific sub scale (Sp-N, Sp-Q, Det., Rec., Loc., and overall Avg.) of the HQ, for the 11 VUmc participants who filled in the questionnaire both at home and in the laboratory. All averages in the table are much less than 1 SE apart, indicating no significant differences. There were not enough subjects who filled in the AI at home and in the laboratory to make that comparison.

Cond.	Avg.	SD	SE	Nr.
Sp-N Lab.	0.69	0.42	0.13	11
Sp-N Hm.	0.66	0.39	0.12	11
Sp-Q Lab.	0.78	0.32	0.10	11
Sp-Q Hm.	0.75	0.33	0.10	11
Det. Lab.	0.73	0.37	0.11	11
Det. Hm.	0.79	0.40	0.12	11
Rec. Lab.	0.79	0.26	0.08	11
Rec. Hm.	0.79	0.29	0.09	11
Loc. Lab.	0.66	0.38	0.11	11
Loc. Hm.	0.74	0.31	0.09	11
All. Lab.	0.71	0.34	0.10	11
All. Hm.	0.71	0.31	0.09	11

Table 5-16: Comparison of the laboratory HQ results with those scored at home via the Internet

## 5.5 Discussion

### 5.5.1 The tests

#### 5.5.1.1 The minimum-audible angle test (MAA)

The results and the validation of the MAA test were described in Section 4. The analysis of the results of 18 NH listeners and 34 hearing-impaired listeners (6 MHI, 3 AHI, 25 HI) lead to criteria for “insufficient” and “poor” of 3.4 and 5.3 degrees, respectively (with results smaller than 3.4 degrees classified as “good”). Figure 5-1 shows the test results for NH (yellow squares) and MHI, AHI, and HI (magenta, blue, and green squares) listeners (except AHI6 who scored 32 degrees) plotted as function of the audiogram expressed in the PTA3.

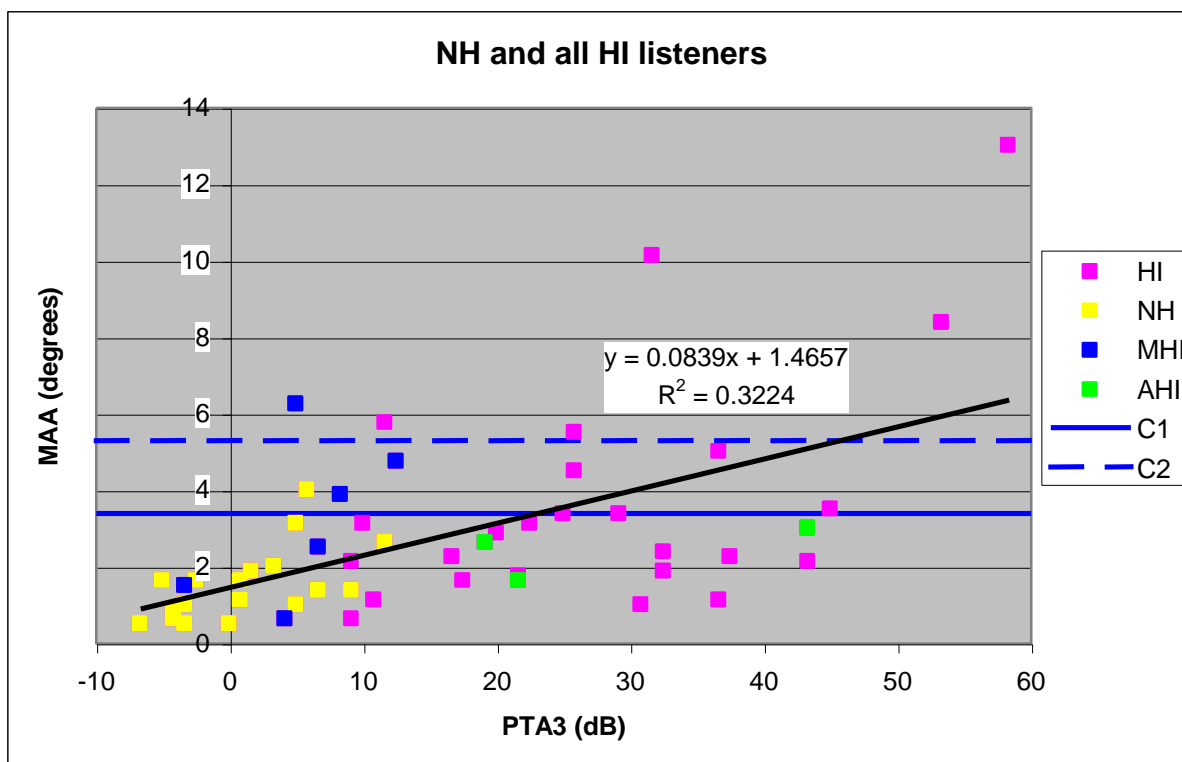


Figure 5-1: MAA scores versus audiogram (PTA3)

Figure 5-1 shows that six hearing-impaired listeners were scored as “poor” while five were scored as “insufficient.” The rest scored “good.” Of the NH listeners one participant was scored as “insufficient” while the rest scored “good.” To have some sort of means to compare the criteria of the different screening tests among each other, a regression line can be drawn through the thresholds, so they can be related to hearing losses. The regression line ( $r^2 = 0.32$ ) in Fig. 5-1 shows that the criteria of 3.4 and 5.3 degrees would roughly correspond to PTA3 values of about 23 and 46 dB.

### 5.5.1.2 The digit-triplet test (SRT)

The results of the digit-triplet test were described in Section 5.2.1. Figure 5-2 shows the test results for the NH (26) and the MHI (12), AHI (8), and HI (32) participant groups.

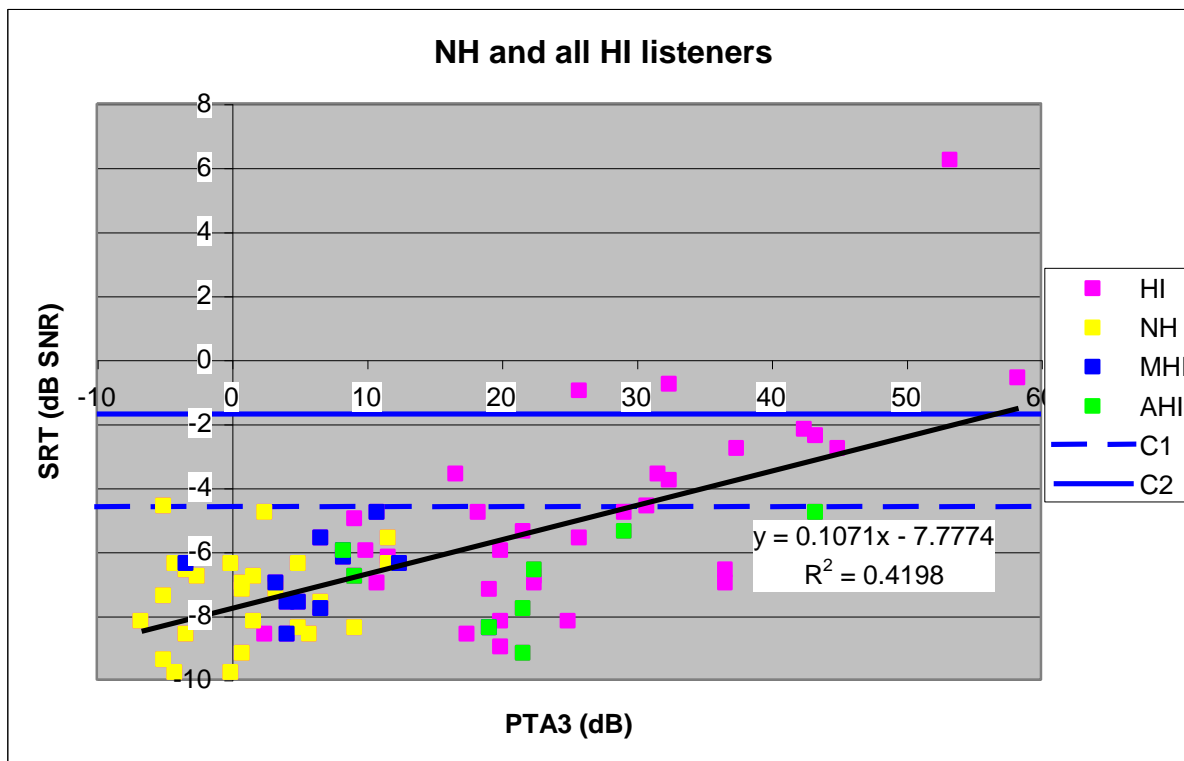


Figure 5-2: Triple digit scores versus audiogram (PTA3)

The criteria for “insufficient and “poor” results, as derived from the present study for the Internet version of the test, lie at -4.6 and -1.7 dB, respectively, with scores lower than -4.6 scoring as “good.” Figure 5-2 shows that four hearing-impaired listeners are scored as “poor” while eight are scored as “insufficient.” The remaining 40 hearing-impaired listeners score “good.” The regression line ( $r^2 = 0.42$ ) shows that the criteria of -4.6 and -1.7 dB correspond roughly to PTA3 values of about 31 and 57 dB. However, in the study of Smits et al. (2004), PTA3 values of about 20 and 42 dB were found to correspond to their two criteria. However, that study included relatively more severely hearing-impaired listeners than the hearing-impaired listeners who participated in the present study. So, the listeners in the study of Smits et al. had larger average losses, and may have suffered more from audibility problems than the present listeners (for severely hearing-impaired listeners, high frequency information is often lost when self chosen presentation levels are used). They often scored positive SRT values, while there is only one such score in the present study. This may have skewed the present outcomes to more favourable thresholds, and so produce a shallow regression line.

The criteria for the telephone and Internet versions of the test lie very close together at -4.1 plus -1.4 dB, and -4.6 plus -1.7 dB, respectively. So, the advantage of the binaural stimulus presentation of the Internet version appears to be cancelled out by the headphone quality at home or the room acoustics that play a role when two computer speakers are used.

5.5.1.3 The auditory virtual-environment test (AVE)

The results of the AVE localisation test were described in Section 5.2.2. Figure 5-3 shows the results of 14 NH listeners and 34 hearing-impaired listeners (6 MHI, 5 AHI, 23 HI). The shown criteria for “good,” “insufficient,” and “poor” were derived from the NH data of VUmc and RUB (see D3-8) using the current “NL-VUmc” implementation to calculate the transformed scores (this implementation gives slightly different numbers than the “DE-RUB” implementation because the zeros in the confusion matrices are dealt with slightly differently). The criteria used in Figure 5-3 lie at 2-SD and 4-SD from the NH average at 0.38 and -0.40, respectively.

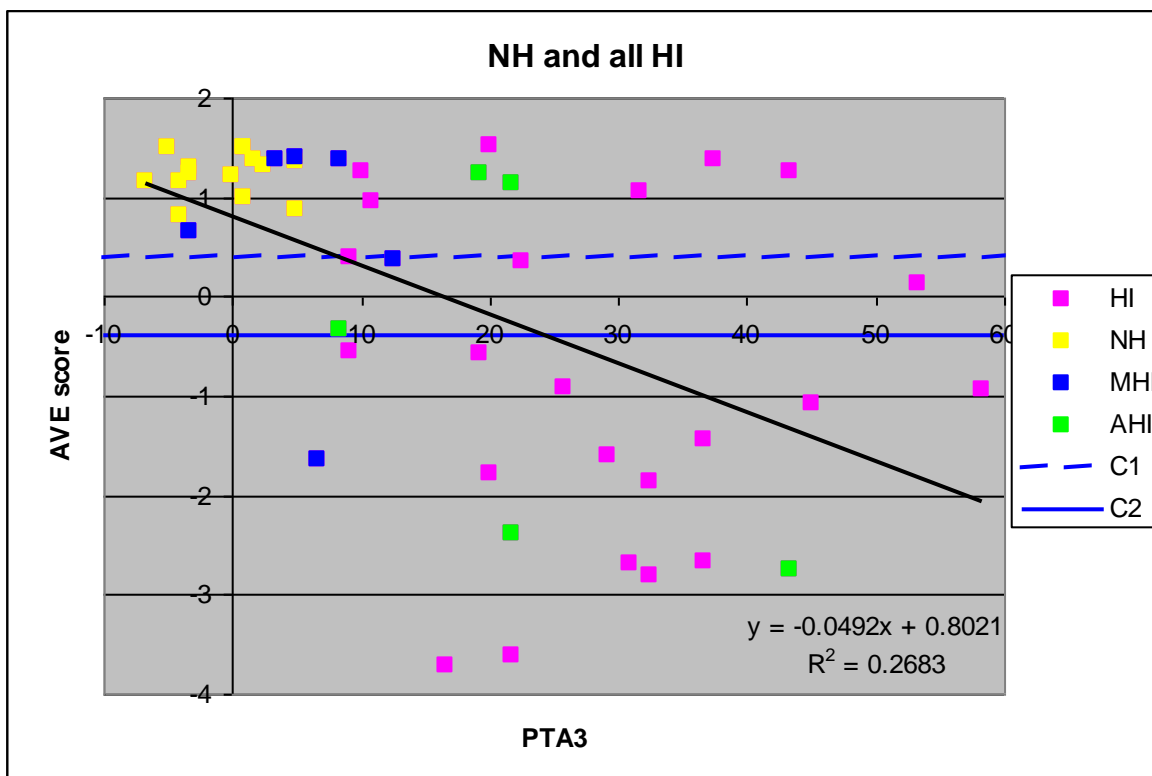


Figure 5-3: Transformed AVE scores versus audiogram (PTA3)

The regression line ( $r^2 = 0.27$ ) indicates that the criteria of 0.38 and -0.40 would correspond to PTA3 values of roughly 9 and 24 dB. The criteria in Figure 5-3 classify all NH participants as “good”. For the hearing-impaired groups (MHI, AHI, and HI) 18 participants are scored as “poor,” four are scored as “insufficient,” and 11 score “good.” Of the AHI listeners, two score “poor,” one scores “insufficient,” and two score “good.”

### 5.5.1.4 The 8-speaker test (8Sp)

The results of the 8-speaker localisation test were described in Section 5.2.2. Figure 5-4 shows the results of 17 NH listeners and 21 hearing-impaired listeners (2 MHI, 3 AHI, 16 HI). The shown criteria for "good," "insufficient," and "poor" were derived from the NH data of VUmc and EMC (see D3-8) using the current "VUmc" implementation to calculate the transformed scores. The criteria used in Figure 5-3 were chosen at 2-SD and 4-SD below the average of 1.29 (SD = 0.23) at 0.83 and 0.37, respectively.

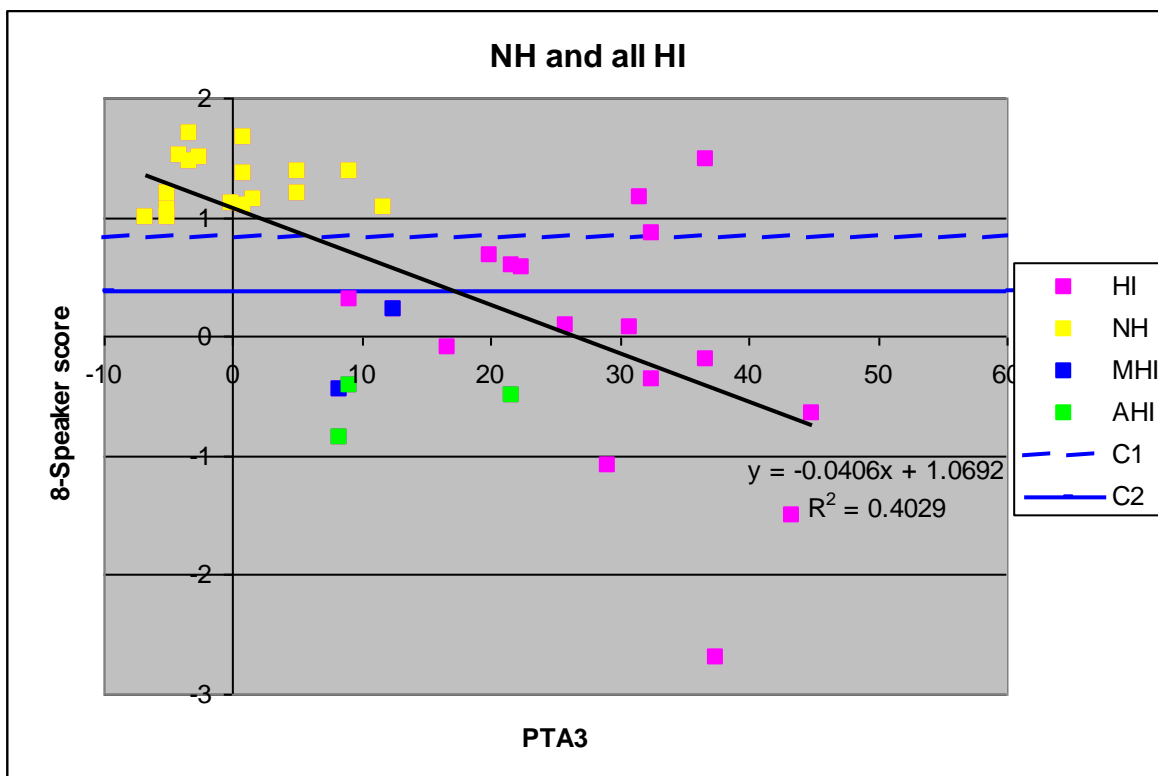


Figure 5-4: *Transformed 8-speaker scores versus audiogram (PTA3)*

The regression line ( $r^2 = 0.40$ ) indicates that the criteria of 0.83 and 0.37 would correspond to PTA3 values of roughly 6 and 17 dB. Figure 5-4 shows that using these criteria, all NH participants would be scored as "good." Of the hearing-impaired participants, 14 would be scored as "poor," three would be scored as "insufficient," while three would score "good." All three AHI listeners were scored as "poor."

### 5.5.1.5 HearCom questionnaire (HQ)

The results of the HearCom questionnaire were described in Section 5.2.3. The overall average scores of the HQ are shown in Figure 5-5 for 19 NH listeners and for 39 hearing-impaired listeners (9 MHI, 4 AHI, 26 HI). For the HearCom questionnaire, criteria for "green," "amber," and "red" response categories of the HQ (corresponding to "good," "insufficient,"

and “poor”) were proposed in D11-5, for hearing losses expressed as the best-ear average over 1, 2, and 4 kHz (BEA), at the thresholds of 29 and 41 dB between those categories. The criteria shown are derived from those proposed in D11-5 and the current data using the regression line ( $r^2 = 0.63$ ) in the figure. They lie at 0.54 and 0.39 for the categories “amber” and “red,” respectively.

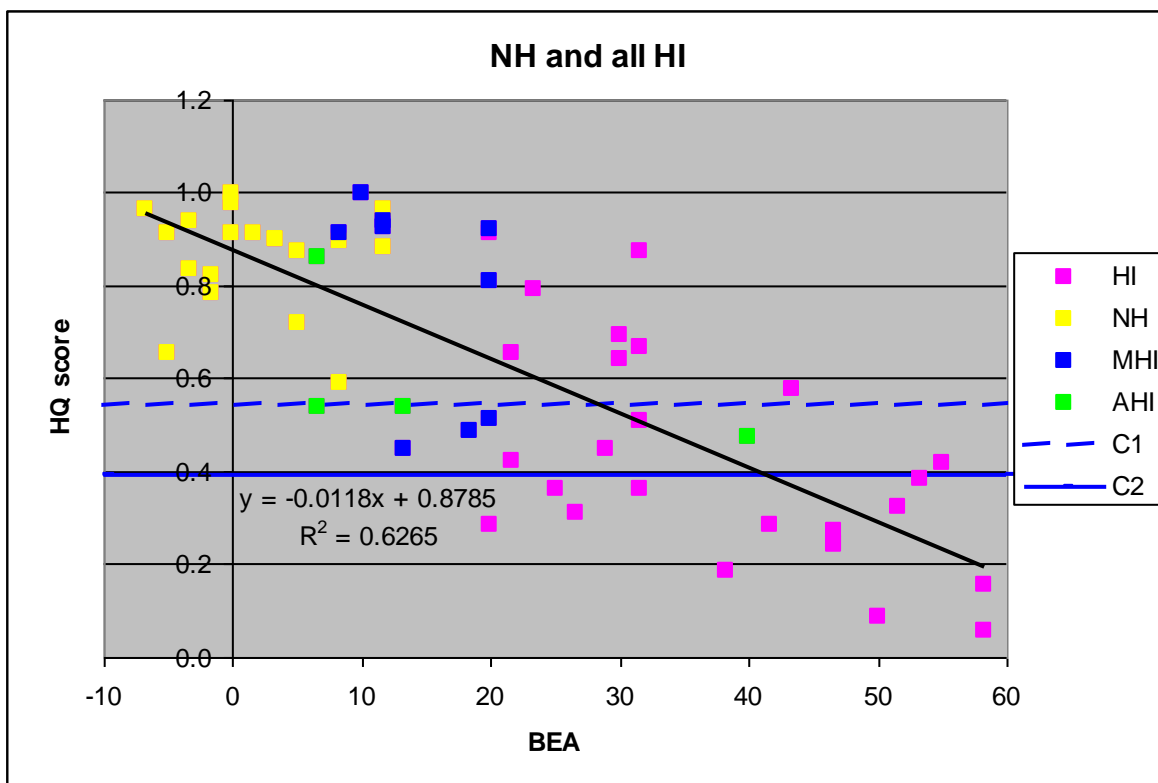


Figure 5-5: Overall average HQ scores versus audiogram (BEA)

Figure 5-5 shows that all NH listeners are scored as “good.” Of the hearing-impaired listeners, 15 are scored as “poor,” 10 as “insufficient,” while the remaining 14 score “good.”

5.5.1.6 The Amsterdam Inventory (AI)

The results of the Amsterdam Inventory were described in Section 5.2.3. The overall average scores of the HQ are shown in Figure 5-6 for 16 NH listeners and for 38 hearing-impaired listeners (8 MHI, 5 AHI, 25 HI). The shown criteria for “good,” “insufficient,” and “poor” were derived from the NH data of VUmc and EMC. These criteria were chosen at 2·SD and 4·SD below the NH average of 0.91 (SD = 0.06) at 0.78 and 0.66, respectively.

The criteria in Figure 5-6 would translate to BEA values of roughly 15 and 26 dB. Figure 5-6 shows that using these criteria, all NH participants would be scored as “good.” Of the hearing-impaired participants, 22 would be scored as “poor,” two would be scored as “insufficient,” while the remaining 14 would score “good.”

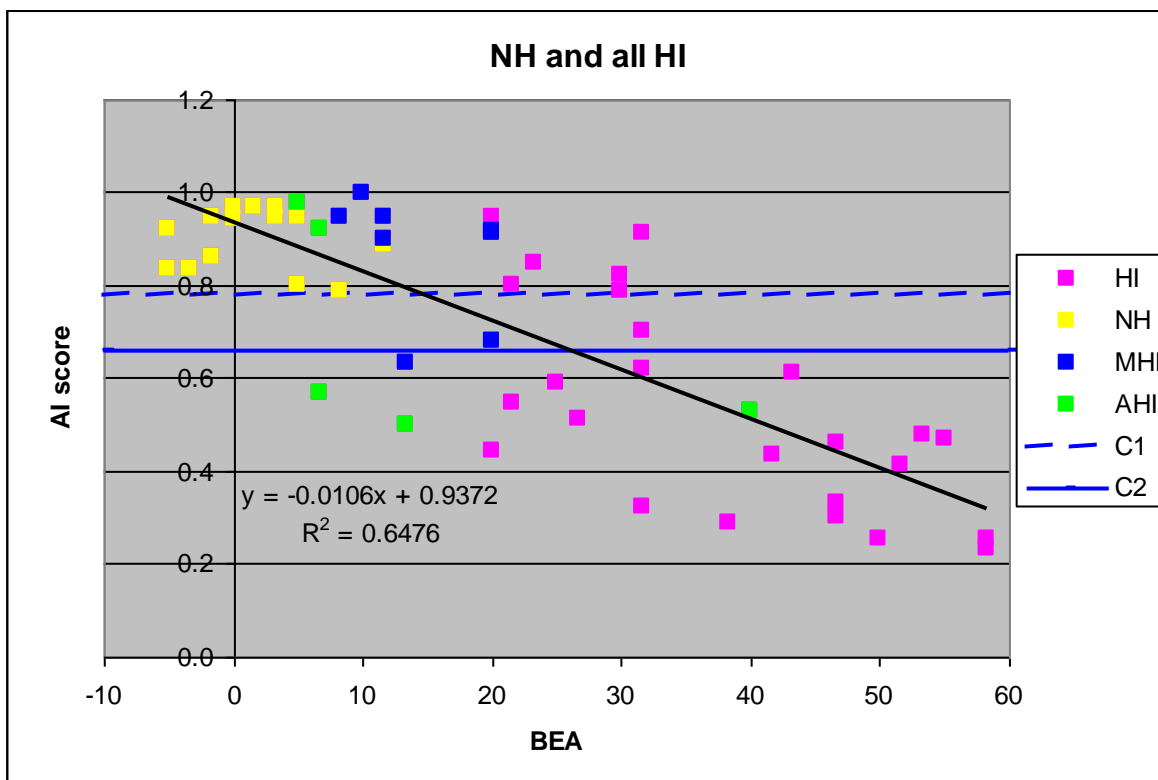


Figure 5-6: Overall average AI scores versus audiogram (BEA)

### 5.5.2 Inter comparison

#### 5.5.2.1 Correlations for all participants

When examining the correlations for all participants analysed together in one group, we can test the effectiveness of the included tests as detectors of hearing problems. In addition, by comparing them to the results of the questionnaires, we can gain insight in what their results mean in daily-life situations.

The correlations between the test scores and audiogram, corrected for the factor Age are shown in Table 5-10. These results show that the PTA3 correlated significantly with all test scores except the AVE. This indicates that the SRT, MAA, and HQ screening tests are good indicators of hearing problems. However, the AVE did correlate with the average results of the HQ and the AI. When examined in more detail, the AVE scores correlate significantly with the Sp-N (0.36\*), Sp-Q (0.44\*\*), Det. (0.45\*\*), and Loc. (0.33\*) scores of the HQ (N = 38). So, somewhat surprisingly, the AVE scores correlate somewhat more strongly with the HQ results for speech in noise and sound detection than with the HQ localisation scores.

It was found in Table 5-10 that the digit-triplet SRTs correlated at the 1% level with the overall HQ and AI scores. The same is true for all sub scores of the HQ, here the correlations range from -0.38 to -0.55, with -0.47 and -0.50 for Sp-N and Sp-Q, respectively (N = 53). So, the SRT scores

correlate well with the audiogram (Table 5-10: 0.59\*\*) and the results of the questionnaires.

The results of the MAA test correlated significantly with the audiogram (0.38\*\*), but they did not correlate significantly with those of the HQ and AI (correlation were smaller than 0.01!). The same is true for all the sub scores of the questionnaires ( $p \geq 0.5$  in all cases). So, the MAA does register hearing problems, but its scores do not correlate well with subjective 360-degree localisation. Nor do they correlate well with the results of the AVE and 8-speaker localisation tests. So, the MAA measures an aspect of localisation that does not relate very strongly to the more "realistic" localisation measures of the HQ and the 8-speaker test. The MAA test estimates the threshold of the perceivable angle between two sound sources, while all other discussed localisation measures are related to the accuracy with which sound-source locations can be identified for inter-source angles that are usually much larger than the threshold value recorded using the MAA test.

The overall average of the HQ correlated very strongly with that of the AI (0.96\*\*) and with the PTA3 (-0.68\*\*). The same is true for the sub scores of the HQ: their correlations with the corresponding sub scores of the AI are all between 0.75 and 0.93, and they all correlate at about -0.65 with the PTA3, apart from the sound recognition (Rec.) score for which that correlation is -0.42. So, next to the overall average, all sub scores of the HQ are good indicators of hearing problems.

#### 5.5.2.2 Correlations for the normally-hearing participants

When examining the correlations for only the group of normally-hearing participants, we can study the inter relationships of the tests in the absence of hearing problems.

The correlations for the NH group in Table 5-11 showed a correlation at the 1% level between the SRT and MAA (0.56\*\*). This is rather surprising, as they measure different aspects of hearing. However, as these are the only tests that use an adaptive procedure, this might reflect a general effect of adaptive procedures. Such procedures require a lot of guessing when no clear percept is present, and familiarity with them can help to not give up and keeping on guessing and guessing.

In addition, the average scores of the HQ and AI correlated strongly (0.88\*\*), showing good correspondence between the two questionnaires. Finally, correlations at the 5% level were found between the PTA3 on the one side, and the MAA (0.49) and overall average of the HQ (-0.49) on the other side. However, such a relation is not found for the sub scores of the HQ. All in all, this implies that small variations in audiogram may already affect the scores of the MAA test and the overall score of the HQ.



### 5.5.2.3 Correlations for the hearing-impaired participants

When examining the correlations for all the groups of hearing-impaired participants, we can study the inter relationships of the tests in the presence of hearing problems.

The Spearman correlations for the hearing-impaired groups MHI, AHI, and HI were shown in Table 5-12. As before, the HQ and AI correlate very strongly with each other ( $0.96^{**}$ ). Therefore, in the following analysis only the HQ will be considered. For the SRT, this table showed correlations with the HQ ( $-0.64^{**}$ , and true for all sub scales) and the PTA3 ( $-0.63^{**}$ ), as found before for the whole group of participants. In addition, the SRT was found to correlate to some degree with the AVE ( $-0.37^*$ ). Next to the SRT, the HQ also correlated with the PTA ( $-0.71^{**}$ , and true for all sub scales) and the Age ( $-0.40^{**}$ ). This implies that the SRT and the HQ are able to make distinctions between the different hearing-impaired listeners, which shows they have a high sensitivity.

### 5.5.3 Locations and instructions

In Tables 5-13 to 5-16, no significant effects were seen in most of the comparisons. A small but significant difference was found between the SRTs measured at the institutes and those measured at home, where the laboratory results were better by 1 dB, which corresponds to the test – retest reliability of the digit triplet test. However, 9 of the 13 listeners in this comparison first executed the test at home before coming into the laboratory to do it there once more. So, this difference probably includes some degree of learning in the execution of the Internet digit triplet test. In addition, a significant difference was found in comparison of the AVE data where results from the quiet office room were better than those from the sound-proof booth. However, it seems very unlikely that the quiet room was a better listening environment than the sound-proof booth, and this difference probably reflects a learning effect. So, overall we observe no important significant effects of listening environment (laboratory versus office room and versus at home) and instructions (laboratory versus at home).

For the MAA test, this implies that the participants were well able to set up their speakers in such a way that the test functioned properly. In addition, it suggests that the speaker quality did not have a large influence on the results. For the SRT test, the listening environment was not important at all, as long as it is a quiet environment. The AVE results were not better in the sound-proof booth than in a quiet room, leading to the same conclusion as for the SRT test. For the HQ questionnaire it appeared to be completely irrelevant whether participants filled them in at home or in the clinic. That is an important result for the HearCom questionnaire.

## 5.6 Conclusions

All tests could be constructed so as to classify the scores into one of the categories “good,” “insufficient,” or “poor” (named “green,” “amber,” and “red” in D11-5). For this, criteria were derived here (MAA, SRT, 8-speaker test, AI) or before (AVE, HQ), leading to estimations of the boundary values expressed in PTA3 or BEA.

For the MAA test, these boundaries fell at PTA3s of 23 dB and 46 dB.

For the SRT test, these boundaries were found at PTA3s of 20 dB and 42 dB by Smits et al. (2004). In the preset study we found them to lie at the rather surprising PTA3s of 31 and 57 dB.

For the AVE test, we found the boundaries at PTA3s of 9 dB and 24 dB.

For the 8-speaker test, we found PTA3 boundaries of 6 dB and 17 dB.

For the HQ, we found that the BEA boundaries of 29 dB and 41 dB, proposed in D11-5, led to scoring criteria of 0.54 and 0.39.

For the AI, we found criteria of 0.78 and 0.66, leading to BEA boundaries at 15 dB and 26 dB.

The MAA was found to be a screening test that is well able to detect audibility problems but that does not relate well to subjective localisation experiences. It was found to be very robust with regards to the accuracy with which people set it up and to the speakers they used.

The SRT was found to be a very sensitive screening test with a good relation to the audiogram and subjective experiences. It was found to be very robust with regards to the listening location (as long as it is a quiet location).

The AVE was found to be a screening test that does not detect audibility problems very well, but that does relate well to subjective experiences. It was found to be robust with regards to the listening location (as long as it is a quiet location).

The HQ was found to be a very sensitive screening test with a good relation to the audiogram. Its results were found to correlate very well with those of the well-established AI. The same is true for all its sub scales. The HQ results were found to be very robust to the place where participants filled in the questionnaire.

## 6 General discussion

### 6.1 The MAA test

The MAA test performed well with respect to its ability to detect audibility problems and with respect to its robustness for speaker quality and the accuracy of the speaker set up. Apparently the participants generally had speakers that were good enough and they were well able to properly set the test up. It is, however, somewhat disappointing that it did not correlate with the results of the 8-speaker test or with subjective localisation. The main reason why the MAA test was originally chosen was its large insensitivity to presentation levels (D1-2). It clearly measures another localisation skill than the 360-degree localisation measured by the 8-speaker test and that forms the basis of the localisation questions in the questionnaires. Nevertheless it can be a useful screening test in that it does efficiently measure a localisation skill and compares the produced results to a benchmark set by normally-hearing listeners.

### 6.2 The screening test battery

The proposed screening test battery consists of the MAA test, the digit-triplet SRT test, and the HearCom questionnaire (HQ), supplemented with the AVE virtual-environment localisation test. The MAA scores correlate well with the audiogram, though they do not relate to subjective localisation scores. The SRT scores correlate very well with the audiogram and the results of the questionnaires, and they show a good sensitivity, making the digit-triplet test an excellent screening test. The HQ scores correlate well with the audiogram and they show a very good sensitivity, making the HQ a very good screening test. The AVE scores correlate well with subjective scores, though they do not relate very strongly to the audiogram. No important effects of location and instructions were found for any of these screening tests. In this section we will consider how they can act together as one screening-test battery.

For the present battery of screening tests, the estimations of the PTA3 (SRT, MAA, AVE) and BEA (for the HQ) thresholds, corresponding to the chosen criteria, are shown in Table 6-1.

	C1	C2	Thr1	Thr2
SRT	-4.6	-1.7	31	57
MAA	3.4	5.3	23	46
HQ	0.54	0.39	29	41
AVE	0.38	-0.40	9	24

Table 6-1: *Criteria and PTA3 or BEA thresholds for the screening tests*

For the SRT and the MAA, the PTA3 thresholds are taken from this study, for the AVE from the present interpretation of the results of D3-8, while for the HQ the BEA threshold are taken from D11-5. Table 6-1 shows that

the estimated PTA3 and BEA thresholds of the digit-triplet SRT, MAA, and HQ lie reasonably close together. That means that the proposed freely-available Internet set of digit-triplet test, MAA test, and HQ all classify users into practically the same categories. This implies that they will cooperate well when put together into one battery. The criteria of the AVE classify users into a different range, but this test is not intended for general access via the Internet. The three tests of the proposed Internet screening-test battery (SRT, MAA, HQ) all have a certain chance of "missing" users with a hearing impairment. This can be for two reasons, (1) a lack of sensitivity of the test, or (2) the user not having any problems for the topic examined by the test. The digit-triplet test classified 13 of the original group of 53 hearing-impaired listeners as not scoring "good." For the MAA test this was 12 of 35, and for the HQ it was 25 of 39. When expressing this as a ratio we can estimate the sensitivity of the test for the present groups of HI subjects. For the NH listeners (including outliers), it turned out that 1 listener out of 27 was not classified as "good" for the digit-triplet test. For the MAA test this was 4 out of 21, and for the HQ it was 0 out of 19. These numbers lead to estimations of the specificity of the tests. These sensitivity and specificity estimates are shown in Table 6-2. For comparison, this table also shows these numbers for the AVE test, the 8-speaker test and the AI.

Test	Not "good"	HI	"good"	NH	Sensitivity	Specificity
SRT	13	55	26	27	0.24	0.96
MAA	12	35	17	21	0.34	0.81
HQ	25	39	19	19	0.64	1.00
AVE	21	34	14	16	0.62	0.88
8-Sp	18	21	17	17	0.86	1.00
AI	24	38	16	16	0.63	1.00

Table 6-2: Sensitivity and specificity of the tests

Table 6-2 shows that the 8-speaker test is the overall winner with the highest sensitivity and specificity, followed in sensitivity by the AVE, the HQ, and the AI. The specificity of all tests is very good, with the MAA scoring worst at 0.81. The SRT and the MAA tests score the lowest sensitivity for the present group of HI listeners. It needs to be remarked here that our HI listeners generally scored unusually well on the SRT test, the sensitivity found in Smits et al. (2004) was much higher at approximately 0.75 (their specificity was 0.91). So, the present low value may well be caused by the inclusion criteria of the present study. Whether the same is true for the MAA test is unknown.

The sensitivity of the screening test battery can be improved by performing all three tests, as all listeners then need to score "good" at all tests to be classified as such. So, for the same reason, this may affect the specificity. From the results of the 24 hearing-impaired listeners (4 MHI, 2 AHI, and 18 HI) who performed all three screening tests in the laboratory, the total sensitivity can be estimated to lie at 0.83. This value is very

close to the very good performance of the professional 8-speaker test. For the data of the 17 NH listeners who performed all three screening tests, the specificity can be estimated to be 0.76. This value is limited by the MAA test, which has the lowest specificity of the set. From the point of optimum sensitivity, it is probably a good idea to advise users to execute all three screening tests, and to only conclude that their hearing is fine when a “good” score is reached on all those tests. However, from the point of optimum specificity, it may be a good idea to not include the MAA test in the battery, or to relax its criteria. Not including it would lead to a combined specificity of 1.0, and a combined sensitivity of 0.75. However, relaxing its criterion to 4-SD would give a combined specificity and sensitivity of 0.88, and 0.83, respectively.

## 7 Dissemination and Exploitation

The entire HearCom screening-test battery will be disseminated via the HearCom website in a number of languages. At present, pages for providing the tests and explaining what they measure, how they should be executed, and what the results mean are available in UK English, German, Dutch, Swedish, and Polish. More languages will presently follow, such as French and Greek. The digit-triplet and the MAA test are, or will soon be made generally available in UK English, German, Dutch, Swedish, and Polish. The HearCom questionnaire will soon become available in UK English and Dutch. At the moment, many of those tests are only available after password login. This is also true for the AVE test, but that test will not be made generally available due to its high server load and high equipment requirements.

The availability of the tests needs to be guaranteed after the official end of the HearCom project. A decision will be needed whether the HearCom site will be maintained as it is, or whether it will be broken up into different country and language specific versions. This will also depend on the exploitation possibilities aspired by the different partners in the different countries.

### 7.1 Ethical issues

Several ethical issues are important here. For example, the users must know what is expected from them when they try to do the tests. Not understanding the instructions can lead to unexpected results and to an inappropriate advice at the end of the test. They should also be explained that the tests screen for hearing problems of sensorineural origin, and that they cannot detect conductive losses. In addition, the outcome of the test should be very clear. So, the tests need to have a high sensitivity and specificity, and the results need to be explained properly to the users.

The outcomes of the screening tests are given in three categories: “good,” “insufficient,” or “poor” (named “green,” “amber,” and “red” for the HQ in

D11-5). The corresponding criteria have been chosen carefully, so as to minimise the chance of falsely alarming normally-hearing users who happen to perform rather poorly, while not missing too many hearing-impaired listeners for whom it could be very beneficial to seek professional help.

The texts provided after finishing a test always include directions on how to find further help in the case of enduring doubts, or of poor results. They also explain that this is just the result of a single screening test, and that more professional hearing tests may shed a clearer light on the state of the users' hearing.

The tests are believed to be safe to perform by any adult user, but the safe execution of the tests will have to remain the responsibility of the participants, as they perform them at home and unaccompanied.

The data from the tests are collected anonymously, so they can be studied later. It is important to keep checking the results, with regards to usage and scoring statistics, to check the effectiveness of the test battery. In addition, a large body of data can be collected to better understand the occurrence of hearing problems in many European countries. The processing of these data will need to continue after the official end of the HearCom project.

## 8 General conclusions

The MAA scores correlate well with the audiogram, though they do not relate to subjective localisation scores. The MAA test can function well as a screening test.

The triple digit SRTs correlate very well with the audiogram and the results of the questionnaires, and they show a good sensitivity, making the digit-triplet test a very good screening test.

The average HQ scores correlate well with the audiogram and they show a very good sensitivity, making the HQ a very good screening test.

The AVE scores correlate well with subjective scores, though they do not relate very strongly to the audiogram. It can function well as a screening test for those invited users who have the proper equipment.

The sensitivity of the screening can be improved by performing the whole battery of three screening tests. Excellent combined specificities and sensitivities of 0.88, and 0.83 can then be reached when the criterion of the MAA test is relaxed to  $4 \cdot SD$ .

## 9 Literature

Goverts, T.S. (2004). "Assessment of spatial and binaural hearing in hearing impaired listeners," Ph.D. Thesis, VU University, Amsterdam, The Netherlands.

HearCom Deliverable D1-1. (2005). Communication self-screening test for two languages operating on a PC-based system.

HearCom Deliverable D1-2. (2005). Report on the proposed set of communication performance tests. [www.hearcom.eu](http://www.hearcom.eu)

HearCom Deliverable D1-3. (2006). Protocol for implementation of communication tests in different languages. [www.hearcom.eu](http://www.hearcom.eu)

HearCom Deliverable D1-4b. (2006). First version on Internet screening tests in three languages. [www.hearcom.eu](http://www.hearcom.eu)

HearCom Deliverable D1-5. (2007). First version of Internet screening tests for localisation discrimination (Demonstrator). [www.hearcom.eu](http://www.hearcom.eu)

HearCom Deliverable D1-6. (2007). Multi-language version of the Digit-triplet test for self-screening (Demonstrator).

HearCom Deliverable D1-6a. (2007). Development of Polish triple-digit test.

HearCom Deliverable D1-7. (2008). Report on normalization data and cross language comparison for sentence tests.

HearCom Deliverable D2-2. (2006). Procedures for the tests included in the auditory profile in four languages.

HearCom Deliverable D2-3. (2007). Report about the results of the multi centre evaluation of the Auditory Profile. [www.hearcom.eu](http://www.hearcom.eu)

HearCom Deliverable D2-5. (2008). Optimized, final set of impairment tests included in the Auditory Profile.

HearCom Deliverable D3-7. (2007). Improved Internet-based AVE for demonstration and self-screening purposes. [www.hearcom.eu](http://www.hearcom.eu)

HearCom Deliverable D3-8. (2008). Implementation of the automatic scoring procedure for the AVE localisation test

HearCom Deliverable D11-1. (2005). Requirements specifications for screening and diagnostic tests.

HearCom Deliverable D11-2. (2006). Usability trials of the screening localisation test.

HearCom Deliverable D11-3. (2007). Usability evaluation of the Internet Triplet digit test.

HearCom Deliverable D11-3a. (2008). Update on D11.3 Usability evaluation of the internet triplet digit test.

HearCom Deliverable D11-4. (2007). Evaluating the usability of two tests to screen for localisation performance.

HearCom Deliverable D11-5. (2008). Development of a screening questionnaire.

Kramer, S.E. (1998). "Assessment of hearing disability and handicap," Ph.D. Thesis, VU University, Amsterdam, The Netherlands.

Kramer, S.E., Kapteyn, T.S., Festen, J.M., and Tobi, H. (1995). "Factors in subjective hearing disability," *Audiol.* 34, pp. 311-320.

Kramer, S.E., Kapteyn, T.S., Festen, J.M., and Tobi, H. (1996). "The relationships between hearing disability and measures of auditory handicap," *Audiol.* 35, pp. 277-287.

Kramer, S.E., Kapteyn, T.S., and Festen, J.M. (1998). "The self-reported handicapping effect of hearing disabilities," *Audiol.* 37, pp. 302-312.

Lorenzi, C., Gatehouse, S., and Lever, C. (1999). "Sound localization in noise in normal-hearing listeners," *J. Acoust. Soc. Am.* 105, pp. 1810-1820.

Lorenzi, C., Gatehouse, S., and Lever, C. (1999). "Sound localization in noise in hearing-impaired listeners," *J. Acoust. Soc. Am.* 105, pp. 3454-3463.

Mills, A.W. (1958). "On the minimum audible angle," *J. Acoust. Soc. Am.* 30, pp. 237-246.

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

Smits, C., Kapteyn, T.S., and Houtgast, T. (2004) "Development and validation of an automatic speech-in-noise screening test by telephone," *Int. J. Audiol.* 43, pp. 15-28.

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

Tavan, D., Mannerheim, P. and Nelson, P.A. (2005). "An investigation of the design and implementation of cross talk cancellation filters for virtual acoustic imaging," ISVR Technical Memorandum 951.



## 10 Appendix A. Text of the HQ

Dutch HearCom questionnaire. Version 2

**Q1. U kijkt in een theater naar een film of een toneelstuk, er is geritsel of er wordt gefluisterd. Kunt u nog steeds verstaan wat er gezegd wordt?**

Target: Speech reception in noise (1).

**Q2. Hoort u de geluiden in het huishouden (ventilator, koelkast, oven, vaatwasser, stromende kraan)?**

Target: Detection of sounds (1).

**Q3. Kunt u, zonder u erg hard te concentreren, naar iemand op de radio luisteren?**

Target: Speech reception in quiet (1).

**Q4. Hoort u buiten de zangvogeltjes zingen?**

Target: Detection of sounds (2).

**Q5. Kunt u het verschil in geluid tussen een saxofoon en een trompet goed horen?**

Target: Recognition of sounds (1).

**Q6. Kunt u gemakkelijk iemands humeur afleiden uit de klank van de stem?**

Target: Recognition of sounds (2).

**Q7. Kunt u op basis van het geluid afleiden hoe ver een bus of vrachtwagen van u verwijderd is?**

Target: Localisation (1).

**Q8. Kunt u op basis van het geluid zeggen of een bus of vrachtwagen naar u toe, of van u weg rijdt?**

Target: Localisation (2).

**Q9. Als u iemands stem of voetstappen achter u hoort, kunt u dan zonder omkijken zeggen in welke richting die persoon zich**

**verplaatst? Bijvoorbeeld van links naar rechts of van rechts naar links?**

Target: Localisation (3).

**Q10. Draait u uw hoofd gelijk de goede kant op wanneer iemand u roept?**

Target: Localisation (4).

**Q11. U voert een gesprek met iemand in een sterk galmende ruimte (een kerk, of de hal van een treinstation). Kunt u volgen wat de andere persoon zegt?**

Target: Speech reception in noise (2).

**Q12. Kunt u familieleden aan hun stem alleen herkennen?**

Target: Recognition of sounds (3).

**Q13. Kunt u zonder problemen een rustig gesprek voeren met een bekende?**

Target: Speech reception in quiet (2).

**Q14. Kunt u verstaan wat er wordt gezegd zonder dat u mensen om herhaling hoeft te vragen?**

Target: Speech reception in quiet (3).

**Q15. Kunt u het nieuws op de autoradio volgen, terwijl u op de snelweg rijdt?**

Target: Speech reception in noise (3).

**Q16. Kunt u de caissière verstaan wanneer zij tegen u praat in een drukke winkel of supermarkt?**

Target: Speech reception in noise (4).

**Q17. U zit met een groepje van ongeveer vijf personen rond een tafel in een rustige ruimte. U kunt iedereen van de groep zien. Kunt u het gesprek volgen?**

Target: Speech reception in quiet (4).

**Q18. Kunt u iemand verstaan zonder al te veel inspanning, wanneer zij naast u in een rijdende bus of auto zit?**

Target: Speech reception in noise (5).