

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

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

**D-11-5: Development of a screening  
questionnaire**

Contractual Date of Delivery:	M43 + 45 days
Actual Date of Submission:	19 May 2008
Editors:	J. Koopman, J. Lyzenga , A. Goedegebure
Sub-Project/Work-Package:	SP 5 – WP 11
Version:	1.0
Total number of pages:	68

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

## Deliverable D-11-5

<b>VERSION DETAILS</b>
Version: 1.0
Date: 16 May 2008 / 22-02-2010 (public)
Status: final

<b>CONTRIBUTOR(S) to DELIVERABLE</b>	
<b>Partner</b>	<b>Name</b>
EMC	J. Koopman
EMC	A. Goedegebure
VU	J. Lyzenga

<b>DOCUMENT HISTORY</b>			
<b>Version</b>	<b>Date</b>	<b>Responsible</b>	<b>Description</b>
0.1	3-2008	EMC	Draft to be reviewed
1.0	5-2008	EMC	After review
	22-02-2010	M.Vlaming	Made public

<b>DELIVERABLE REVIEW</b>			
<b>Version</b>	<b>Date</b>	<b>Reviewed by</b>	<b>Conclusion*</b>
0.1	22-4-2008	Gaby Nordbrock	Update then accept
0.1	22-4-2008	Sion Morris	Update then accept

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

## Table of Contents

Pre-Amble.....	7
1 Executive Summary.....	8
2 Introduction .....	9
3 Usability tests .....	13
3.1 Introduction .....	13
3.2 Methods .....	13
3.2.1 The Questionnaire .....	13
3.2.2 Set-up of usability tests .....	14
3.3 Results – Expert Trials.....	16
3.3.1 Introduction .....	16
3.3.2 Style of the questionnaire .....	16
3.3.3 Specific questions .....	18
3.4 Results – User Trials .....	30
3.4.1 Introduction .....	30
3.4.2 Style of the questionnaire .....	30
3.4.3 Specific questions .....	30
4 Validation by end users.....	34
4.1 Introduction .....	34
4.2 Methods .....	34
4.2.1 Development of the questionnaire .....	35
4.3 Results .....	37
4.3.1 Descriptive statistics.....	37
4.3.2 Intercorrelations .....	40
4.3.2.1 Descriptive statistics .....	40
4.3.2.2 Linear regression .....	41

4.3.2.3	Factor Analysis .....	42
4.3.3	Screening.....	46
4.3.3.1	Sensitivity and Specificity .....	46
4.3.3.2	Sensitivity and specificity - different weightings per item	48
5	Discussion .....	53
5.1	Outcome relative to the SSQ .....	53
5.2	Outcome compared to the Kramer survey .....	55
5.3	Questionnaire specificity and sensitivity .....	56
5.3.1	The effect of cut-of hearing loss .....	57
5.3.2	The effect of the width of insufficient hearing.....	59
6	Dissemination and Exploitation .....	62
6.1	Ethical issues .....	62
7	Conclusions .....	63
8	References .....	65
	Appendix – A : Psychometric curves for each item .....	67

## List of Figures

Figure 1:	Scree plot of Eigenvalues as a function of the component number.....	43
Figure 2:	Factorplot after Varimax rotation.....	44
Figure 3:	Scatterplots of the mean of the items in different domains as a function of the regression coefficients of the items relative to the different factors (1-column 1, 2-column 2, 3-column 3, and 4- column 4).....	45
Figure 4:	ROCs based on hearing averaged over 1, 2, 4 kHz of the best ear (BEA; left) and the weighted average over the best ear and worse ear (Handicap; right). .....	47
Figure 5:	Optimisation curves for the different competence domains. Specificity and sensitivity were weighted equally. ....	47

Figure 6: ROC based on all items. Gray curves are based on the averaged hearing loss (1,2,4 kHz) of the best ear. Black curves are based on a 4:1 weight of best and worse ear. Solid curves are based on the direct item scores, dashed curves on the corrected item scores. ....	49
Figure 7: Optimisation curves for the four different situations. ....	50
Figure 8: ROC after optimisation of the area underneath the ROC by removing some of the items. ....	51
Figure 9: Scatterplot between scores of the Kramer questionnaire and the Hearcom questionnaire. Scores were determined by taking the mean of all items in that particular competence domain.....	55

## List of Tables

Table 1: User statistics that participated to the trials.....	15
Table 2: Descriptive statistics for the different auditory tasks studied in the questionnaire. <i>P</i> -value < 0.05: *; <i>P</i> -value < 0.01: **; <i>P</i> -value < 0.001: *** based on Spearman Rank correlation coefficients. ....	39
Table 3: Spearman rank correlation coefficients between different auditory domains. (***) <i>p</i> -value<0.001). ....	41
Table 4: Results from the linear fit of the items score in a given competence domain (first column) as a function from the hearing loss in the better ear (BEA) and the worse ear (WEA). (***) <i>P</i> -value<0.001; ** <i>P</i> -value< 0.01. <i>R</i> provides Pearson's correlation coefficient. ....	42
Table 5: Results from factor analysis. 'Before' and 'after' relate to before or after Varimax rotation.....	44
Table 6: Cut-off value, sensitivity, and specificity for optimal sensitivity and specificity combination.....	50
Table 7: Cut-off value, sensitivity, and specificity for optimal sensitivity and specificity combination after removing some items (see last column). ....	52
Table 8: Rank correlation coefficients between the item-score and the hearing loss in the best ear (BEA) and the weighted average (Handicap) for the SSQ-questionnaire and the Hearcom questionnaire (Hc-q).....	54

Table 9: Correlation coefficients between scores of the Kramer questionnaire and the Hearcom questionnaire. (***) $p < 0.001$ ; ** $p < 0.01$ ; * $p < 0.05$ ).	56
Table 10: Effect of cut-off hearing loss in the best ear between good and insufficient hearing. Column orange provides the number of observations in the category insufficient hearing. Three different sensitivity and specificity combinations were given 1) BEA – based on the average of all items. 2) BEA(org; lim) – based on the average of all items corrected for the items score at 35 dB, and 3) BEA (fin) based on the average of the different item scores corrected for the score at 35 dB leaving out specific items (see section 4.3.3.2).	58
Table 11: Similar to Table 10 but now based on the weighted average between best and worse ear (1:4).	59
Table 12: Difference in averaged hearing loss of the best ear (1, 2, 4 kHz) between the result good and poor. Three different sensitivity and specificity combinations were given 1) BEA – based on the average of all items. 2) BEA(org; lim) – based on the average of all items corrected for the items score at 35 dB, and 3) BEA (fin) based on the average of the different item scores corrected for the score at 35 dB leaving out specific items (see section 4.3.3.2).	60
Table 13: Similar to Table 12 but now for the weighted hearing loss between best an worse ear.	61

## **Acknowledgement**

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

## Pre-Amble

The HearCom project aims at achieving full participation of people with a reduced hearing ability in the modern communication society by reducing limitations in auditory communication. Within the HearCom project, Subproject 5 (SP 5) looks at how the Internet can be used to provide new services that will forward this aim, and will provide a route for the exploitation of many of the outcomes of the HearCom project. Within this subproject a web portal will be developed for the dissemination of project outcomes (Work package 10) and frameworks will be developed for some of these outcomes to fit into (Work package 11 and Work package 12). The structure will focus on the assessment of hearing loss using self-screening tests (WP11) and e-rehabilitation (WP12).

In the early stages of hearing loss, people are often unaware of their limited hearing. Therefore, they generally wait a long time before deciding to seek an audiological assessment. The aim of WP11 is to provide Internet access to self-screening tests for hearing impairment that have been developed within SP1 and SP2. The ability to conduct such self-screening tests is expected to increase awareness of hearing-impairment – which will likely lead to an increasing number of people looking for medical advice. The results of D11-1 showed that 14% of the respondents indicated solely to conduct screening by using a questionnaire. Partly because some people expected that their system would not be accurate enough to provide these data.

This work fits closely with WP12, which builds the infrastructure of the HearCom web site. This work also relates to WP10 (the development and implementation into the portal). This deliverable reports on the development, user trials, and validation of an Internet screening test on the HearCom website. Usability trials were conducted in the United Kingdom and in the Netherlands. The validation of the questionnaire was carried out by NL-VuMC and NL-EMC.

# 1 Executive Summary

The HearCom screening questionnaire was designed to measure a range of hearing disabilities across several domains; speech intelligibility, sound localisation, audibility of sounds and discrimination of sound. The user evaluations did not expose any serious issues that could not be solved.

In the evaluation, 129 participants were asked to report their competence in performing everyday tasks represented by 25 questionnaire items. Answers were scored on a four-interval likert-scale. A factor analysis based on the response of the respondents confirmed the four competence domains after Varimax rotation. Three areas are identified based on the hearing loss averaged over 1, 2, and 4 kHz;

1. the green area, intended to cover respondents with a hearing loss lower than 29 dB.
2. the amber area, indicating insufficient hearing covering respondents with a hearing loss between 29 and 41 dB.
3. the red area, indicating a poor hearing for respondents having a hearing loss more than 41 dB.

Based on the responses, items were removed in order to increase the area under a receiving operating characteristic (ROC) optimizing the sensitivity and specificity. In this way, in order to predict the hearing loss based on the item responses for the best ear, 10 items are sufficient in order to provide a high sensitivity and high specificity (0.88 and 0.85). In order to obtain a high accuracy for the weighted hearing loss in both ears, 18 items are required (0.85 and 0.81).

## 2 Introduction

Hearing loss, with a reported prevalence of about 16% for people above the age of 17, has been identified as one of the most frequent chronic conditions affecting older populations in developing countries (Davis, 1989; Wilson, 1999). The prevalence increases with age; whereas only 5-10 % of the people below 40 have a hearing loss (hearing loss > 25 dB averaged over 0.5, 1, 2, and 4 kHz), 5-6 out of 10 people (50-60 %) suffer from reduced hearing when they are between 60 and 70 years of age (Davis, 1989; Wilson, 1999). Despite this being common knowledge, people with reduced hearing do not always seek medical help. Prevalence of hearing aid use amongst those with a hearing loss above 25 dB HL (averaged over 0.5, 1, 2, and 4 kHz) is limited (Popelka et al., 1998; Davis, 1989). Applying the prevalence rates from Davis(1989) to the situation in the Netherlands, approximately 15-40 % of the people suffering from hearing loss use hearing aids (data obtained from CBS, 2006).

There may be a number of factors that contribute to the relatively low prevalence of hearing aid use. Basically they could either indicate that people are not able to detect a reduced ability to hear (see also Smits et al., 2004; Koopman, 2008), or that they are not willing to accept their reduced hearing. The 25-item Hearing Handicap Inventory for the Elderly (HHIE) questionnaire (Ventry and Weinstein, 1982), therefore focuses on emotional aspects (acceptance) and on competence in certain auditory tasks (ability to detect). From these 25-items, a subset of 10 items were selected to provide the possibility to screen on hearing loss (Weinstein, 1986), resulting in the HHIE-S questionnaire. The questionnaire showed a sensitivity of 0.35 (35 percent of the participants that were hearing impaired were correctly identified) and a specificity of 0.92 (92 percent of the participants without impaired hearing were correctly classified). Other

questionnaires intended to screen on hearing loss show a higher sensitivity and specificity (Gomez et al, 2001; sensitivity 0.77/specificity 0.82; Gates et al. 2003-0.35/0.94; van Schaik 1998 0.76/0.76). Acceptance is most likely a major contributor to the low prevalence of hearing aid use. However, in order to start accepting a hearing aid, people first need to know that their hearing ability has been diminished. Therefore, emotional aspects may be less suitable for hearing screening. The research conducted on the success of screening on hearing loss using competence in different auditory tasks is rather unambiguous. Davis (1989) mentioned that using a single question such as: Do you think you find it difficult to follow a conversation if there is a background noise (e.g. TV, radio, children playing)? "provides an efficient screen for mild/moderate hearing screening [p. 916]". Unfortunately, no specificity of sensitivity has been given and can be determined based on the data provided. Wilson (1999) asked people if they have troubles hearing what people say to them in a quiet room when they speak on a given level to them (loudly, normally, whisper). Results indicated that there was a considerable misclassification on self-reported hearing loss. Clark et al. (1991) and Sindhusake et al. (2000) asked respondents if they felt they had a hearing loss. They reported acceptable sensitivity (respectively, 56% and 78%) and a specificity (respectively 82% and 67%) using a single question. Schaik et al. (1997) showed a reasonable specificity and sensitivity by combining two single questions adapted from the HHIE-S questionnaire ("Can you carry on a one-to-one conversation?"; specificity=1, sensitivity=0.4 and "Can you follow the conversation of a small group of people?"; specificity=0.8, sensitivity=0.75), leading to a specificity of 1 and a sensitivity of 0.75. These results suggest that a higher sensitivity and specificity can be obtained by combining several items within the same questionnaire.

Still, the specificity and sensitivity of such screening questionnaires are much lower than those obtained by the speech in noise test as described by Smits et al. (2004). When adopting the SRT in noise (sentences) following the procedure of Plomp and Mimpen (1979) as the golden standard, the SRT obtained using three digits in noise showed a sensitivity and specificity of 0.9. The test can easily be conducted by phone (using the number pad as a response box) and via the Internet. A major benefit of the screening test as described in Smits et al. (2004) or a self-screening questionnaire compared to other screening methods (e.g. hearing tones and speech, Cienkowski, 2003; an audioscope, Lichtenstein *et al.*, 1988; whispered voice test, Eekhof et al. 1996) is that it can be assessed without help or assistance of qualified people and users can easily conduct the test in the comfort of their own homes. The need for such an ability to assess hearing loss in the privacy of the users own environment is shown by the popularity of the service (Smits, 2005). Most people now experiencing the onset of hearing loss (people in their fifties; often referred to as baby-boomers) are generally proficient at using computers and the Internet. As such, the Internet is widely regarded as an acceptable means of providing such a service (Koopman et al. 2008). However, this survey also indicated that some respondents tend to distrust the outcome of the test, based on poor acoustics or the presence of environmental noises, or the inability to manipulate these environments. Others respondents indicated that they did not have access to appropriate equipment (headphones).

Partly for these reasons, some respondents indicated that a questionnaire would be a good tool. The potential use of Internet, the telephone, and a questionnaire were compared to each other. Results indicated that 78.6% of all respondents would use the test via the Internet, whereas 86.6% of them would use a questionnaire when screening on hearing loss. In addition, 9.4% would use solely the Internet and 13.9% would use solely a questionnaire to screen for hearing loss. This indicates that a

questionnaire to screen on hearing loss would be a great asset to such a service, provided that the questionnaire has a high level of sensitivity and specificity. Such a high sensitivity and specificity may be difficult to establish. Smits et al. (2004) shows, despite their excellent sensitivity and specificity when compared to the SRT of sentences, a reduced but still very acceptable sensitivity (0.79) and specificity (1) when compared to the hearing loss averaged over 0.5, 1, 2, and 4 kHz. Although the choice of the SRT of sentences as a golden standard is very acceptable (assessing functional hearing loss), it illustrates the difficulty and importance of using the correct golden standard.

This Deliverable presents the development of a questionnaire that intends to allow people to screen their hearing accuracy. In line with other questionnaires such as the Auditory Profile of Hearing Aid Benefit (APHAB; Cox 1995), the Speech, Spatial and Qualities of Hearing Scale (SSQ) (Gatehouse and Noble, 2004), the Amsterdam Inventory for Auditory Disability and Handicap (Kramer, 1995), the questions will assess the competence of the respondent in particular situations that often require substantial listening effort from hearing impaired persons.

The evaluation of the questionnaire is divided in two stages. The usability of the questionnaire has been evaluated by a number of experts and novice users. The aim of this evaluation is to identify unclear or inconsistent phrasing of the questions and instructions. The results of this evaluation will be described in chapter 3. Based on the results of the usability tests, a slightly renewed version of the questionnaire (improved formulation of questions, when needed) has been validated by collecting the responses of a large number of end users. The analysis of this validation procedure will be described in chapter 4. The aim of this analysis is to obtain the reliability of the questionnaire by comparing the responses to the questions with the amount of hearing loss of the users.

## 3 Usability tests

### 3.1 Introduction

A Questionnaire has been composed for Hearcom, mainly based on available validated questionnaires. The exact procedure and choices made to obtain this questionnaire are described in detail in chapter 4.2.1. This chapter describes the evaluation of the usability of the questionnaire as a screening tool. For this aim a number of experts and novice users have gone thoroughly through all the questions using a simplified thinking aloud protocol (see <http://jthom.best.vwh.net/usability/thnkalod.htm>).

### 3.2 Methods

#### 3.2.1 The Questionnaire

The questionnaire can be accessed via the Internet by making the appropriate selection ('questionnaire') from the 'testarea' page of the HearCom portal, which hosts all screening tests.

The page opens with the following simple instructions for the user:

*This questionnaire lists a number of everyday activities that can often require a lot of listening effort if you have a hearing loss. It will be your task to indicate how much difficulty you experience when conducting these activities **without the use** of a hearing aid or other listening devices.*

*If a particular activity does not apply to you, please respond by selecting "not applicable".*

In total 25 items are divided over 3 pages and focus on four aspects of hearing (speech intelligibility, sound localisation, sound recognition, and sound detection). Items are provided on a four-interval likert scale with an additional option to indicate that a certain item is not applicable to the respondent. A four item scale was preferred relative to an odd-item scale

(3 or 5), since an even (4) item scale “forces” the user to indicate whether a problem exists and so provides more information.

### 3.2.2 Set-up of usability tests

The simplified thinking aloud protocol was used in the tests. A detailed description of this procedure is given in deliverable D11.4, or check the Internet page: '<http://jthom.best.vwh.net/usability/thnkalod.htm>.' This procedure requires a limited number of users (5 to 10) to identify most of the problems in usability of the questionnaire. The potential users are asked to conduct the tests in the presence of an observer. This method was chosen in order to involve users with the products developed within HearCom and see if they were able to conduct the tests themselves without any help of the observers, as if they were at home.

Two different trials have been performed:

1. Expert trials (5 experts<sup>1</sup>)
2. User trials (8 novice users<sup>2</sup>).

Five experts participated in the expert trials. Since the questionnaire itself is straightforward, we decided to focus on the questions and the instructions themselves. Results from the Dutch (n=3) and from United Kingdom (n=2) were collected and analysed. If an issue raised in a particular language version also applied to the other language version, the solution was applied to both language versions of the questionnaire.

User trials have been conducted in order to involve users with the development of HearCom products. The main target group of the self-screening tests is people suffering from the onset of hearing loss who are familiar with the Internet at work or at home. Therefore, the study focused on recruiting people around the age of 50. The level of experience

---

<sup>1</sup> Users with a lot of experience using the Internet, and having experience with audiology and the intended target group.

<sup>2</sup> People of the intended target group (baby boomers, less experience with the Internet).

with the Internet was no reason to include or exclude users. A summary of the characteristics of the users who participated in the trials is given in Table 1. The column NH/HI indicates whether they were Normal Hearing or Hearing-Impaired, based on their Fletcher Index being smaller or larger than 35 dB HL.

User ID	Age	NH/HI
UNL-1	53	HI
UNL-2	57	HI
UNL-3	53	HI
UNL-4	49	HI
UUK-1	25	NH
UUK-2	58	HI
UUK-3	25	NH
UUK-4	26	NH
Average	43	5 HI/3 NH

**Table 1: User statistics that participated to the trials**

The usability tests first sought to see if users understood what was expected of them. After that, users were asked to comment on their understanding of the questions. The usability tests always started at the page showing the questionnaire. The usability test was conducted without time constraints. The observers followed the actions of the users taking notes of unexpected behaviour. Before each trial started, it was explained to the user that the Internet test itself was being tested, and not their performance (which will be subject of sections **Error! Reference source not found.** and onwards).

## 3.3 Results – Expert Trials

### 3.3.1 Introduction

*This questionnaire lists a number of everyday activities that can often require a lot of listening effort if you have a hearing loss. It will be your task to indicate how much difficulty you experience when conducting these activities **without** the use of a hearing aid or other listening devices.*

*If a particular activity does not apply to you, please respond by selecting "not applicable"*

The following issues were raised:

- Avoid phrases like 'how much difficulty you experience' as this assumes that the subject actually has difficulties.
  - Replaced by "how much difficulty you experience (if any)"
- Mention the intention of the questionnaire.
  - This questionnaire will allow you to indicate whether you experience difficulty in specific areas of hearing, and will provide you with an estimate of whether your hearing has degraded.

### 3.3.2 Style of the questionnaire

- "It is an issue of style, but I always start with positives, work down to negatives and have 'not applicable' at the end. This questionnaire lays out responses in the opposite manner."
  - Option 'not applicable' will be shifted to the end of the line.

- The categories of answer are very unclear. What do the categories actually mean and how might the respondents understand them? For example what is the difference between 'occasionally' and 'frequently', and why are there no categories of 'never' or 'always'?
  - It might be a matter of translation. This scale was used in a validated questionnaire (Kramer, 1995). The categories were translated by the authors. 'Never' belongs to the same category as 'almost never' and 'always' belongs to 'almost always'.
- In general the questions are very ambiguous, the situations are very variable and therefore they will generate very subjective and very diverse answers. I suggest that the author considers exactly what he/she requires from each question and then edits them accordingly.
  - Most questions have proven themselves in daily practice. The fear of not understanding what is been said, will be considered after evaluation. Once there is no clear relation between the competence assessed and the pure tone hearing loss (or asymmetrical hearing for sound localisation), questions may be misunderstood.
- Some questions start with the word 'imagine'. You should never do this as it immediately means that you have a hypothetical question and these are well known for giving unreliable answers.
  - Solved by rewording these items to begin with "Think of a time when/where ..."
- Do you really need to ask 25 questions? It appears that many of the questions here ask similar things so I would advise the author to focus on the key questions/issues and reduce the size of the questionnaire accordingly.

- This will be evaluated and become an outcome of the eventual analysis. The most valid items will be used to provide a good image of the users' hearing.

### 3.3.3 Specific questions

***Q1. When you are watching a film in a cinema or a play in a theatre, are you able to understand what is being said if people around you are whispering or rustling paper?***

**Intention:** This question intends to cover speech intelligibility in the presence of soft environmental noises.

**Comment:** Ambiguous - are you meant to understand the film or the people whispering?

**Reply:** The film – question will be reworded to:

When you are watching a film in a cinema or a play in a theatre, are you able to understand what is being said (either on the screen or on stage) if people around you are rustling paper or whispering?

---

***Q2. Can you hear household sounds like those made by a fan, fridge, oven, dish washer, or a running tap?***

**Intention:** This question intends to focus on the detection of soft environmental sounds that are familiar to the user.

**Comment:**

1. Household products listed produce varied types, frequencies and volumes of sound.
2. Modern devices are often much quieter than older ones – also depends on habituation

**Reply:**

The question will be evaluated which will indicate if the question is appropriate. Noises described in this question are mainly low-frequency sounds of moderate to low sound levels. Low correlation coefficients with other questions will indicate that this question is inappropriate.

---

**Q3. Do you have to concentrate really hard when you are listening to a voice on the radio?**

**Intention:** This question focuses on the intelligibility of speech without the ability to lip-read.

**Comment:** What does 'concentrate really hard' actually mean? And surely 'a voice on the radio' will vary very much according to radio reception, audio quality of the radio and the actual program being listened to?

Maybe avoid the problem by turning the radio loud

**Reply:**

- 1) The Dutch version of the question omits the 'need to concentrate hard'. This should be implemented in the UK-question.
  - 2) The question concerns a situation in which the user cannot see the speakers' face – added to which is the enhanced listening challenge posed by the restricted bandwidth of the radio output. I see no further problems in this question. During the evaluation, comments will be reconsidered if necessary.
  - 3) Add "at normal volume".
- 

**Q4. Are you able to hear the birds singing outside?**

**Intention:** This question focuses on the ability to hear high pitched sounds.

**Comment:**

- 1) The interpretation of this question depends upon many variables so it is likely to be of little actual use as a question. It depends on whether there are birds outside, what type/species they are, what kind of glazing one has, etc.
- 2) How do you know that there are “birds singing outside” if you cannot hear them?

**Reply:** The question has proven itself in the Kramer survey. When people think of birds singing, they generally think of high-pitched sounds at a well defined level.

---

***Q5. Are you able to tell who it is you're talking to on the telephone just by listening to the person's voice?***

**Intention:** This question focuses on the **recognition** of someone's voice

**Comment:** Dependent upon audio quality of the phones, quality of reception and the frequency of the voice you are listening to. Also how would you answer this question if you can hear high pitched voices really well but find lower pitched voices hard to understand?

**Reply:** Telephones do not transmit frequencies beyond 3 kHz. Hence, the frequency of the voice is unlikely to influence the users' response.

---

***Q6. Can you hear the difference in sound between a trumpet and a saxophone?***

**Intention:** This question focuses on the recognition of two instruments that are alike in sound.

**Comment:**

- 1) Why a trumpet and a saxophone? This could confuse a respondent because it relies on them having a good knowledge of how these instruments both sound, but there are also different types of saxophone and trumpet so this could lead to further problems. If you wish to investigate if someone can differentiate between instruments at different frequencies then why not use instruments such as a violin, flute, bass, trombone, etc
- 2) Very dependent on if they play at the same time or not.

**Reply:**

- 1) The intention was to identify if people were able to distinguish between two familiar instruments that are alike in sound. If people indeed experience difficulties distinguishing between these instruments based on unfamiliarity this question will not correlate with auditory data and the question will be removed from the questionnaire.
- 2) The question mentions differentiate, so both cases apply.

---

***Q7. Can you tell how somebody is feeling from the sound of their voice?***

**Intention:** This question focuses on the recognition of inflection (by frequency modulations)

**Comment:** I don't see what this has to do with having a hearing loss. Surely this has more to do with your familiarity with a person and how their voice changes according to their moods?

**Reply:** Most hearing-impaired people experience a poor spectral resolution. This causes difficulties in the detection of the frequency modulations typical to determine someone's mood.

---

**Q8. When you hear people or cars behind you, are you surprised to find that they are closer than you thought when you look around?**

**Intention:** This question focuses on the estimation of the distance of a known sound source.

**Comment:**

1) Human hearing is most acute in a forwards direction so we all have less effective hearing behind ourselves. As a result almost everyone will give a similar answer to this question and it may therefore be of little help.

2) Most questions require a positive answer in order to indicate if they do not experience difficulties. This question is reversed.

**Reply:** Similar questions are used in the SSQ hearing questionnaire and results show otherwise.

Question will need to be turned around

---

**Q9. Can you tell from its sound how far away a bus or lorry is?**

**Intention:** This question focuses on the estimation of the distance of a known low-frequency sound source.

**Comment:** Again a very ambiguous question with many variables

---

**Reply:** Similar questions are used in the SSQ hearing questionnaire and results are reliable.

---

**Q10. Can you tell from its sound whether a bus or lorry is moving towards you or moving away from you?**

**Intention:** This question focuses on the ability to deduce the direction of movement of a known low-frequency sound source.

**Comment:** no comments.

---

**Q11. When you hear a voice or footsteps behind you, can you tell in which direction (from left to right or right to left) that person is moving without looking round?**

**Intention:** This question focuses on the ability to deduce the direction of movement of a known sound source.

**Comment:** Voices and footsteps are very different sounds and as our hearing is focused towards the front most people will have problems with this situation.

**Reply:** This will be part of the evaluation. It will not be a major problem if certain items are considered to be difficult, since the ability to tell something about the onset of hearing loss may require these difficult items.

---

**Q12. Do you immediately turn your head in the correct direction when somebody calls you?**

**Intention:** This question focuses on sound localization.

**Comment:** no comments

---

**Q13. *Can you follow what somebody is saying when there is much reverberation, such as in a church or railway terminus building? Can you follow what the other person is saying?***

**Intention:** This question focuses on speech intelligibility in reverberant rooms.

**Comment:** What is the relevance of 'reverberation' and do you really think that respondents will understand what you actually mean by this word? Are you possibly more concerned with loud, low frequency noises?

**Reply:** Low frequency noises are represented. It is expected that the term 'reverberation' will be understood through the use of these examples. Another option may be to insert "echoic environment".

---

**Q14. *Can you recognize family members by their voices alone?***

**Intention:** This question focuses on the recognition of people with well known voices.

**Comment:** My understanding is that everyone would answer Yes, or all the time, to this question unless they were profoundly deaf. If they were profoundly deaf then I suggest that they would be aware of this and would have no need for this questionnaire.

**Reply:** The question was obtained from the Kramer survey. This questionnaire showed that this fear is unfounded. Results indicate that not

everyone does not experience problems recognizing people based on their voice alone.

---

***Q15. Can you easily have a quiet conversation with somebody you know?***

**Intention:** This question focuses on the ability to have a conversation in a quiet environment.

**Comment:** In what environment?

**Reply:** This is a matter of translation and will be reworded to:  
Can you easily have a quiet conversation with somebody you know in a quiet environment?

---

***Q16. When you are in a quiet room, can you follow a conversation on the telephone?***

**Intention:** This question focuses on the ability to have a conversation in a quiet environment without the ability to see the person you are talking to.

**Comment:**

- 1) Follow? Do you mean 'easily understand' maybe?
- 2) Depends on the speaker and quality of phone line.

**Reply:**

- 1) This is a matter of translation.
  - 2) This question is intended to refer to an averaged speaker and responses should not be affected by the quality of the telephone line.
-

Reworded to:

When you are in a quiet room, can you easily conduct a conversation on the telephone?

---

***Q17. Can you understand what is said without having to ask people to repeat themselves?***

**Intention:** This question focuses on the ability to have a conversation in general.

**Comment:** Again in what environment?

**Reply:** Validation will indicate if this is true. We expect that the people who have to ask others to repeat themselves will know this.

---

***Q18. Can you understand somebody who talks to you at a busy party or reception?***

**Intention:** This question focuses on the ability to have a conversation in a busy environment – testing for the so-called “cocktail party effect”

**Comment:** no comments

---

***Q19. When you are in a car on the motorway are you able to follow the news on the radio?***

**Intention:** This question focuses on the ability to have a conversation in the presence a low frequency environmental sound.

**Comment:** Dependant upon many variables including whether you are driving, radio reception, audio quality, the car itself, the presenter's voice, etc.

**Reply:** The noise from the car itself is almost constant at a high speed. It emits a relatively defined level of noise and is therefore likely to provide reliable answers.

---

**Q20. *Imagine that you are listening to someone on the telephone and that somebody next to you begins to talk. Are you able to follow what both people are saying?***

**Intention:** This question focuses on the ability to follow two conversations, which provides an indication of the effort a subject has to pay in order to complete a task.

**Comment:** Hypothetical and confusing questions. If the respondent does understand them, then the answers will be very unreliable.

**Reply:** Item to be re-worded to: "Think of a time when you are listening to someone on the telephone and that somebody next to you begins to talk. Are you able to follow what both people are saying?"

---

**Q21. *Imagine that you are listening to someone who is standing next to you and the news reader on the radio at the same time. Can you decide which conversation you want to follow?***

**Intention:** This question focuses on the subject's ability to choose whom he/she will be listening to.

**Comment:** Hypothetical and confusing questions. If the respondent does understand them, then the answers will be very unreliable.

**Reply:** This item will be reworded to

“Think of a time that you are listening to someone who is standing next to you and the news reader on the radio at the same time. Can you decide which conversation you want to follow?”

---

***Q22. In a busy shop or supermarket, can you understand the checkout assistant when they are talking to you?***

**Intention:** This question focuses on the ability to understand a high-pitched voice in the presence of a fluctuating background.

**Comment:** no comments

---

***Q23. Imagine that you and four other people are in a busy restaurant and that you cannot see everybody. Are you able to follow the conversation?***

**Intention:** This question focuses on the ability to have a conversation in a busy restaurant without the ability to lip-read.

**Comment:** Surely most people will have problems in understanding conversations with this number of people and in this environment. As such you are likely to get negative responses to this.

**Reply:** This will be part of the evaluation. It will not be a major problem if certain items are considered to be difficult, since the ability to tell something about the onset of hearing loss may require such difficult items.

The question will be reworded to:

Think of a time that you and four other people are in a busy restaurant and that you cannot see everybody. Are you able to follow the conversation?

---

***Q24. Imagine that you are sitting at a table with four other people in a quiet room and that you can see everybody. Can you follow the conversation?***

**Intention:** This question focuses on the ability to have a conversation in a busy restaurant with the ability to lip-read.

**Comment:** no comments

---

*Q25. When you are travelling in a bus or a car, are you able to talk to the person sitting next to you without difficulty?*

**Intention:** This question focuses on the ability to have a conversation in the presence of a low-frequency environmental noise.

**Comment:** Should refer to hearing rather than talking?

**Reply:** Question will be reworded to:

When you are travelling in a bus or a car, are you able to have a conversation with the person sitting next to you without difficulty?

## 3.4 Results – User Trials

### 3.4.1 Introduction

*This questionnaire lists a number of everyday activities that can often require a lot of listening effort if you have a hearing loss. It will be your task to indicate how much difficulty you experience when conducting these activities **without** the use of a hearing aid or other listening devices.*

*If a particular activity does not apply to you, please respond by selecting "not applicable"*

- 1) Despite the instruction provided, users UUK-1 ;UNL-1 and UNL-2 completed the questionnaire under the assumption that they **could** use hearing aids in the situation described.
  - Although the questionnaire intends to serve people who have concerns about their hearing (i.e. generally people who do not use hearing aids), it may help to clarify why the questionnaire intends to describe the situation without the use of hearing aids.

### 3.4.2 Style of the questionnaire

No comments were given about the style of the questionnaire.

### 3.4.3 Specific questions

***Q1. When you are watching a film in a cinema or a play in a theatre, are you able to understand what is being said if people around you are whispering or rustling paper?***

**Comment:** UUK4: "Should I be able to understand the people that whisper, or the movie/play?"

**Reply:** Question is reworded to “When you are watching a film in a cinema or a play in a theatre, are you able to understand what is being said (either on the screen or on the stage) if people around you are rustling paper or whispering?”

---

**Q3. Do you have to concentrate really hard when you are listening to a voice on the radio?**

**Comment:** UUK4 never listens to the radio.

**Reply:** option ‘not applicable’ is available.

---

**Q6. Can you hear the difference in sound between a trumpet and a saxophone?**

**Comments:**

User UUK1 needed clarification as to what a saxophone was – and how this sound differed from the sound of a trumpet.

User UUK2 and UNL3 stated that they would not have been able to answer this question, as they did not consider themselves able to know the difference between the sound of the two instruments.

**Reply:** Question is reworded by

“Can you distinguish between a trumpet and a saxophone purely based on the sound they make?”

---

**Q7. Can you tell how somebody is feeling from the sound of their voice?**

**Comments:** UUK3 and UUK4 were slightly confused by the purpose of this question.

**Reply:** The question is easy to understand and answer there probably should not be a need for further explanation - the respondent just has to trust that the question has a specific purpose!

---

**Q11. *When you hear a voice or footsteps behind you, can you tell in which direction (from left to right or right to left) that person is moving without looking round?***

**Comments:** UUK3 and UNL1 found it difficult to imagine being in this scenario. Although it happens many times each day – it isn't something they actively think about while it's happening.

**Reply:** 1. if they had experienced problems in such a situation, they probably would have recognized the situation. 2. the option 'not applicable' can be selected in such a case.

---

**Q13. *Can you follow what somebody is saying when there is much reverberation, such as in a church or railway terminus building. Can you follow what the other person is saying?***

**Comment:** UUK4: What is reverberation?

**Reply:** Term might need some more explanation (link?)

---

**Q18. *Can you understand somebody who talks to you at a busy party or reception?***

**Comments:** To user UUK2 the question was unclear, since it did not describe how noisy the party was. UNL-2 was unsure since he was not familiar with such a situation.

**Reply:** Providing more precise statements will probably cause more confusion. Evaluation will indicate if question is appropriate for use.

---

**Q19. *When you are in a car on the motorway are you able to follow the news on the radio?***

**Comment:** User UUK3 indicated that this depended on the volume of the radio.

**Reply:** Add 'normal level'.

---

## **4 Validation by end users**

### **4.1 Introduction**

This chapter describes the user validation of the Hearcom questionnaire. For this aim the responses of a large number of end users have been collected (129). The responses to the questions are compared to the amount of hearing loss of the individual users. In this way we can investigate whether the questionnaire is a reliable screening tool for hearing impairment. The statistically defined measures “sensitivity” and “specificity” are used for this aim. In addition, we have tried to identify clusters of questions that predict different aspects of auditory performance by means of a factor analysis.

### **4.2 Methods**

The authors acknowledge the problem that patients with identical audiometric characteristics may cope quite differently with daily-life situations. A golden standard is required to assess the success of a tool for screening purposes. Although it has its limitations in assessing functional hearing loss (i.e. what can someone do with this hearing loss), the golden standard chosen is the pure tone audiogram. Hearing loss for each ear is averaged over the frequencies 1, 2, and 4 kHz. The side with the lowest average hearing loss is referred to as the better ear average (BEA), the ear with the highest average hearing loss is referred to as the worse ear average (WEA). The binaural impact (handicap) of the hearing loss is assessed according the UK rules in which the better ear is weighted relative to the worse ear as 4:1 (Lutman et al. 1987). Air conduction and bone-conduction thresholds were measured in 5-dB steps using clinical audiometry. Conductive hearing loss was defined as an air bone gap larger than 7.5 dB averaged over 1, 2, and 4 kHz.

### 4.2.1 Development of the questionnaire

The content of the questionnaire was designed to cover an extensive but realistic range of situations in which respondents could experience limitations as a result of their reduced hearing. The twenty-five items are distributed over four categories as distinguished by Kramer (1995), in order to cover different conditions of assumed varying difficulty. The first category, on self assessed speech intelligibility (12 items) was designed to cover a range of different conditions of competing sounds; conversations in crowded (restaurant, shop) and quiet environments (cinema, quiet room); conversation partner being visible or not visible (in regard to lip-reading); conversations in continuous noise (bus, car) or fluctuating noise (restaurant, radio). The second category, on self assessed sound localisation (also often referred to as spatial hearing; 5 items) covers standard-items on directional hearing (someone calling), on distance judgements (footstep, distance; bus), and the discrimination of movement as introduced by the SSQ (bus, footstep). The third category contains items (4) addressed to the self-assessed ability to distinguish sounds, and cover the ability to recognize someone over the phone (familiar and less familiar) and the ability to tell someone's mood based on their voice. The last category contains items (2) on the ability to hear sounds (singing birds or sounds in the house). The choice of the topics was based on a number of existing validated questionnaires: the Amsterdam Inventory (Kramer et al., 1995); ease of listening features in the Abbreviated Profile of Hearing Aid Benefit (Cox & Alexander, 1995); the Speech, Spatial and Qualities of Hearing Scale (Gatehouse and Noble, 2004).

The wording of the items was chosen carefully, in order to cancel out the factor of audibility when possible. Patients did not indicate any problems with the wording of the questions when they filled in the questionnaires prior to their hearing assessment at the clinic. Items were selected and translated by the authors into the Dutch language, except for questionnaires already available in English and Dutch. Two other native

speakers translated the questionnaire into English. After that, they were re-translated by two Dutch clinicians into Dutch to assure that the English and Dutch translations contained correctly corresponding questions.

Before the questionnaire started it was explicitly mentioned that questionnaires pointed to situations in which the respondent would not wear a hearing device (if applicable). Questions were asked in the affirmative way (*i.e.* can you still listen ...), instead of randomising this by applying reversal scaling, in order to reduce the possibility of confusion. The four response categories provided ranged from “rarely”, “sometimes”, “often”, to “almost always”, a fifth response category was provided to indicate non-familiarity to the situations.

No reward was given for completing the questionnaire and no reminders were sent. The responses were analysed anonymously.

### **Statistical analysis**

First of all, a descriptive statistics will be conducted to provide a full overview of the 129-member focus group and their results. Hearing loss is determined as the average hearing loss at 1, 2, and 4 kHz. Furthermore, a weighted average based on the hearing loss of both ears is referred to as the handicap.

The questionnaire intends to describe the auditory performance of people in different domains (as suggested by Kramer, 1995). Linear regression will be used determine how the best ear and the worse ear contribute to the performance in one of these domains. A difference in hearing loss between both ears is likely to affect localisation performance. In order to compare the outcome of this questionnaire to that of a validated questionnaire, a factor analysis will be conducted. In addition, a principal components analysis (PCA) was conducted. The multi-dimensional space, that follows from this PCA, can be compared to the space as determined

by Kramer (1995). Finally, the reliability of the questionnaire (in terms of sensitivity and specificity) will be determined and optimised.

Scores were determined as the proportion of the maximal score. For each question, the score the different intervals are given by 1 (rarely), sometimes (0.6667), often (0.3333), almost always (0). The average disability score according to the questionnaire is determined by the sum of the items falling within the competence domain.

Unless otherwise stated, the percentages were calculated using the number of respondents giving a valid response to a particular question. A  $p$ -value of 0.05 was considered statistically significant. The statistical analysis was carried out using SPSS version 11.5. Curves were fitted using least-square methods using a logarithmic function.

## 4.3 Results

### 4.3.1 Descriptive statistics

In total 129 people completed the questionnaire (64 male; 65 female), on average they were 51 years old (s.d. 15 years). The better ear average (BEA) averaged over 1, 2, and 4 kHz equalled 29.5 dB (s.d. 19.9 dB). The worse ear average (WEA) averaged over 1, 2, and 4 kHz equalled 51.8 dB (s.d. 28.4 dB). The averaged asymmetry between the best and worse ear averaged equalled 22.2 dB (s.d. 24.1 dB; quartiles 5-34 dB). Table 2 provides the averaged scores and the Spearman rank correlation coefficients for all items together and for each category separately. Table 2 clearly indicates that asymmetrical hearing loss mainly contributes to problems experienced with sound localization (both for mixed and sensorineural hearing losses) and for speech intelligibility in noise to some extent.

All items independently had a high significant correlation with either the hearing loss in the best ear (BEA), in the worst ear (WEA), or with the asymmetry between both ears. On average, each item had 1.7% missing values. There is only one clear exception, Q6 (distinction between a saxophone and trumpet) which resulted in 17% missing values (including not applicable). This indicates that almost all questions are correctly understood.

	All items	Speech Intelligibility		Localisation	Recognition	Detection
		Quiet	Noise			
Mean (Sd)	2.47 (0.68)	2.79 (0.74)	2.02 (0.75)	2.40 (0.85)	3.11 (0.76)	2.67 (1)
<i>Mixed (37)</i>	2.89 (0.69)	3.17 (0.64)	2.47 (0.85)	2.71 (0.89)	3.53 (0.53)	3.2 (1.0)
<i>Sensorin.(96)</i>	2.31 (0.61)	2.64 (0.72)	1.84 (0.63)	2.27 (0.80)	2.95 (0.78)	2.46 (0.92)
$\rho$ score BEA	-0.60 <sup>***</sup>	-0.56 <sup>***</sup>	-0.54 <sup>***</sup>	-0.39 <sup>***</sup>	-0.52 <sup>***</sup>	-0.60 <sup>***</sup>
<i>Mixed (37)</i>	-0.48 <sup>**</sup>	-0.53 <sup>**</sup>	-0.51 <sup>**</sup>	-0.34	-0.34	-0.51 <sup>**</sup>
<i>Sensorin.(96)</i>	-0.52 <sup>***</sup>	-0.47 <sup>***</sup>	-0.44 <sup>***</sup>	-0.32 <sup>**</sup>	-0.42 <sup>***</sup>	-0.53 <sup>***</sup>
$\rho$ score WEA	-0.47 <sup>***</sup>	-0.38 <sup>***</sup>	-0.43 <sup>***</sup>	-0.53 <sup>***</sup>	-0.18 <sup>*</sup>	-0.35 <sup>***</sup>
<i>Mixed (37)</i>	-0.49 <sup>**</sup>	-0.45 <sup>**</sup>	-0.53 <sup>**</sup>	-0.52 <sup>**</sup>	0.05	-0.10
<i>Sensorin.(96)</i>	-0.47 <sup>***</sup>	-0.31 <sup>**</sup>	-0.35 <sup>**</sup>	-0.51 <sup>**</sup>	-0.16	-0.36 <sup>**</sup>
$\rho$ score Asym	-0.14	-0.06	-0.15	-0.35 <sup>***</sup>	0.14	-0.01
<i>Mixed (37)</i>	-0.35 <sup>*</sup>	-0.3	-0.40 <sup>*</sup>	-0.42 <sup>*</sup>	0.17	0.07
<i>Sensorin.(96)</i>	-0.15	-0.06	-0.14	-0.37 <sup>***</sup>	0.09	-0.07
$\rho$ score (diss.) 0.2(4BEA+WEA)	-0.59 <sup>***</sup>	-0.54 <sup>***</sup>	-0.54 <sup>***</sup>	-0.47 <sup>***</sup>	-0.47 <sup>***</sup>	-0.58 <sup>***</sup>
<i>Mixed (37)</i>	-0.59 <sup>***</sup>	-0.62 <sup>***</sup>	-0.62 <sup>***</sup>	-0.47 <sup>**</sup>	-0.25	-0.47 <sup>**</sup>
<i>Sensorin.(96)</i>	-0.52 <sup>***</sup>	-0.45 <sup>**</sup>	-0.43 <sup>***</sup>	-0.40 <sup>***</sup>	-0.37 <sup>***</sup>	-0.51 <sup>***</sup>

**Table 2: Descriptive statistics for the different auditory tasks studied in the questionnaire.**

***P*-value < 0.05: \*; *P*-value < 0.01: \*\*; *P*-value < 0.001: \*\*\* based on Spearman Rank correlation coefficients.**

## 4.3.2 Intercorrelations

### 4.3.2.1 Descriptive statistics

A standard approach to study inter-item-relationships is by means of a factor analysis. This paper described 25 items (independent variables) using 129 observations. It is generally thought that a stable solution can be obtained by using a 5-fold to 10-fold (Gatehouse and Noble, 2004). Given that these requirements are fulfilled, but only on the lower edge, care should be taken by interpreting the results of this analysis. Therefore, systematic observations will be described prior to the results of the factor analysis.

All independent items are positively intercorrelated<sup>3</sup>. The items within each domain (detection, discrimination, speech in quiet, speech in noise, sound localisation) correlate significantly with the independent items within this domain (Spearman's  $\rho > 0.62$ ;  $P$ -value  $< 0.0001$ ). The only exception was found for the domain sound detection (Spearman's  $\rho > 0.44$ ;  $P$ -value  $< 0.0001$ ). In this domain, the separate items ( $n=4$ ) also show high mutual correlation coefficients (Spearman's  $\rho > 0.54$ ;  $P$ -value  $< 0.0001$ ) when excluding item Q6. Table 3 shows the intercorrelations between the different domains. In general, there is a high coherence between the different domains, except for sound localisation.

---

<sup>3</sup> After rewording Q8, in order to be in line with the other items.

	All	Speech (Q)	Speech (N)	LOC	DISCR	DETECT
All	1	0.90 <sup>***</sup>	0.92 <sup>***</sup>	0.81 <sup>***</sup>	0.73 <sup>***</sup>	0.82 <sup>***</sup>
Speech(Q)	0.90 <sup>***</sup>	1	0.80 <sup>***</sup>	0.62 <sup>***</sup>	0.62 <sup>***</sup>	0.75 <sup>***</sup>
Speech(N)	0.92 <sup>***</sup>	0.80 <sup>***</sup>	1	0.69 <sup>***</sup>	0.73 <sup>***</sup>	0.73 <sup>***</sup>
LOC	0.81 <sup>***</sup>	0.62 <sup>***</sup>	0.69 <sup>***</sup>	1	0.48 <sup>***</sup>	0.58 <sup>***</sup>
DISCR	0.73 <sup>***</sup>	0.62 <sup>***</sup>	0.73 <sup>***</sup>	0.48 <sup>***</sup>	1	0.58 <sup>***</sup>
DETECT	0.82 <sup>***</sup>	0.75 <sup>***</sup>	0.73 <sup>***</sup>	0.58 <sup>***</sup>	0.58 <sup>***</sup>	1

**Table 3: Spearman rank correlation coefficients between different auditory domains. (\*\*\*)  $p$ -value<0.001).**

#### 4.3.2.2 Linear regression

Linear regression was used to predict the outcome of the questionnaire based on the hearing loss in the best ear and worse ear. Table 4 shows the results of a linear fit of the mean item score for each of the competence domains as a function of the hearing loss in the better and worse ear (averaged over 1, 2, and 4 kHz).

Domain	R	Equation
All items	0.63	$-0.017*BEA^{***} - 0.004*WEA^* + 3.18$
Speech (quiet)	0.59	$-0.020*BEA^{***} + 3.46$
Speech (noise)	0.56	$-0.014*BEA^{**} - 0.006*WEA^{**} + 2.70$
Sound localisation	0.54	$-0.006* BEA - 0.013* WEA^{***} + 3.24$
Audibility	0.61	$-0.031*BEA^{***} + 3.57$
Sound discrimination	0.58	$-0.025*BEA^{***} + 3.63$

**Table 4: Results from the linear fit of the items score in a given competence domain (first column) as a function from the hearing loss in the better ear (BEA) and the worse ear (WEA). \*\*\* P-value<0.001; \*\* P-value< 0.01. R provides Pearson's correlation coefficient.**

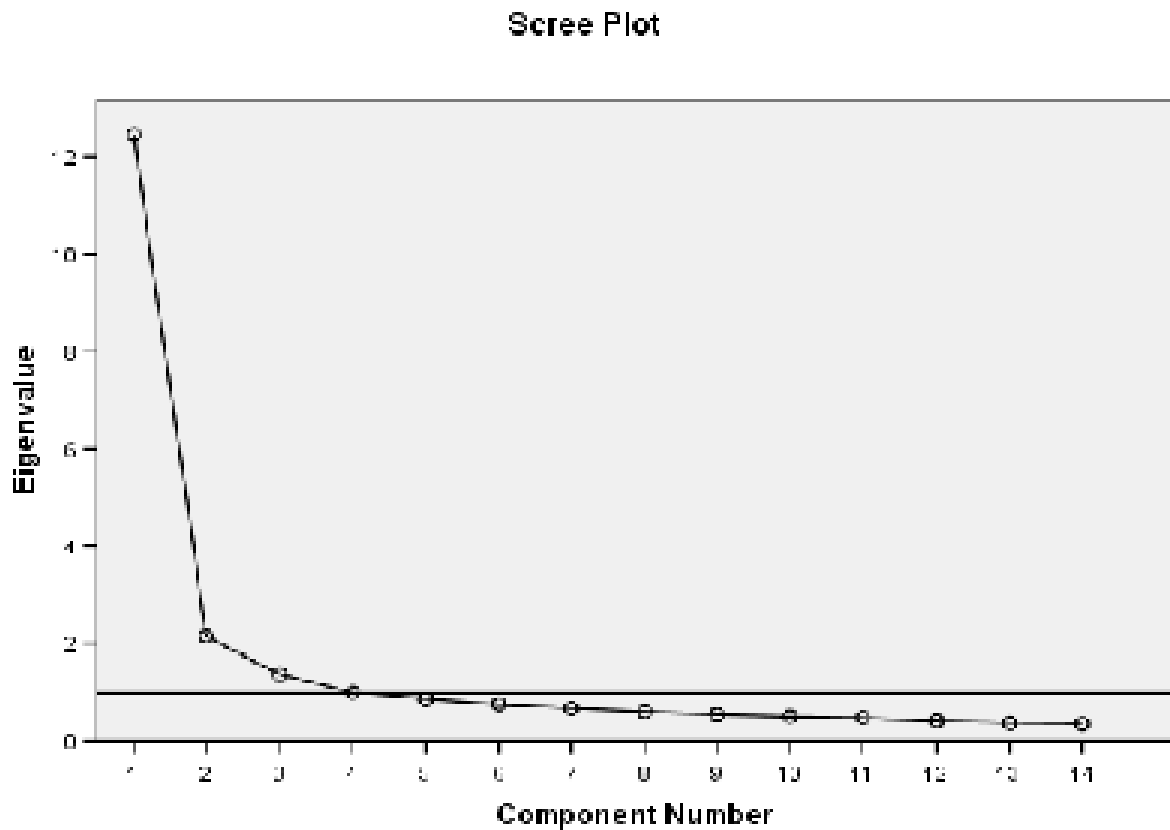
The results indicate that the hearing loss of the worse ear contributes to the perceived problems for speech in noise and sound localization. Moreover, for sound localization, the worse ear seems to contribute more to the perceived difficulties locating sound sources than the hearing loss in the better ear.

Based on the fit on all items, it is important to note that the best ear versus the worse ear follows the 4:1 rule known from other research and often applied to determine the handicap.

#### 4.3.2.3 Factor Analysis

As mentioned in Section 4.3.2.1, the ideal way to describe intercorrelations is by means of a factor analysis. The Eigenvalue as a function of the number of components is called a Scree plot (see Figure 1). Eigenvalues larger than 1 explain more than a single component in

the original domain and is therefore considered to be significant. Figure 1 shows that four components significantly contribute to the explained variance.



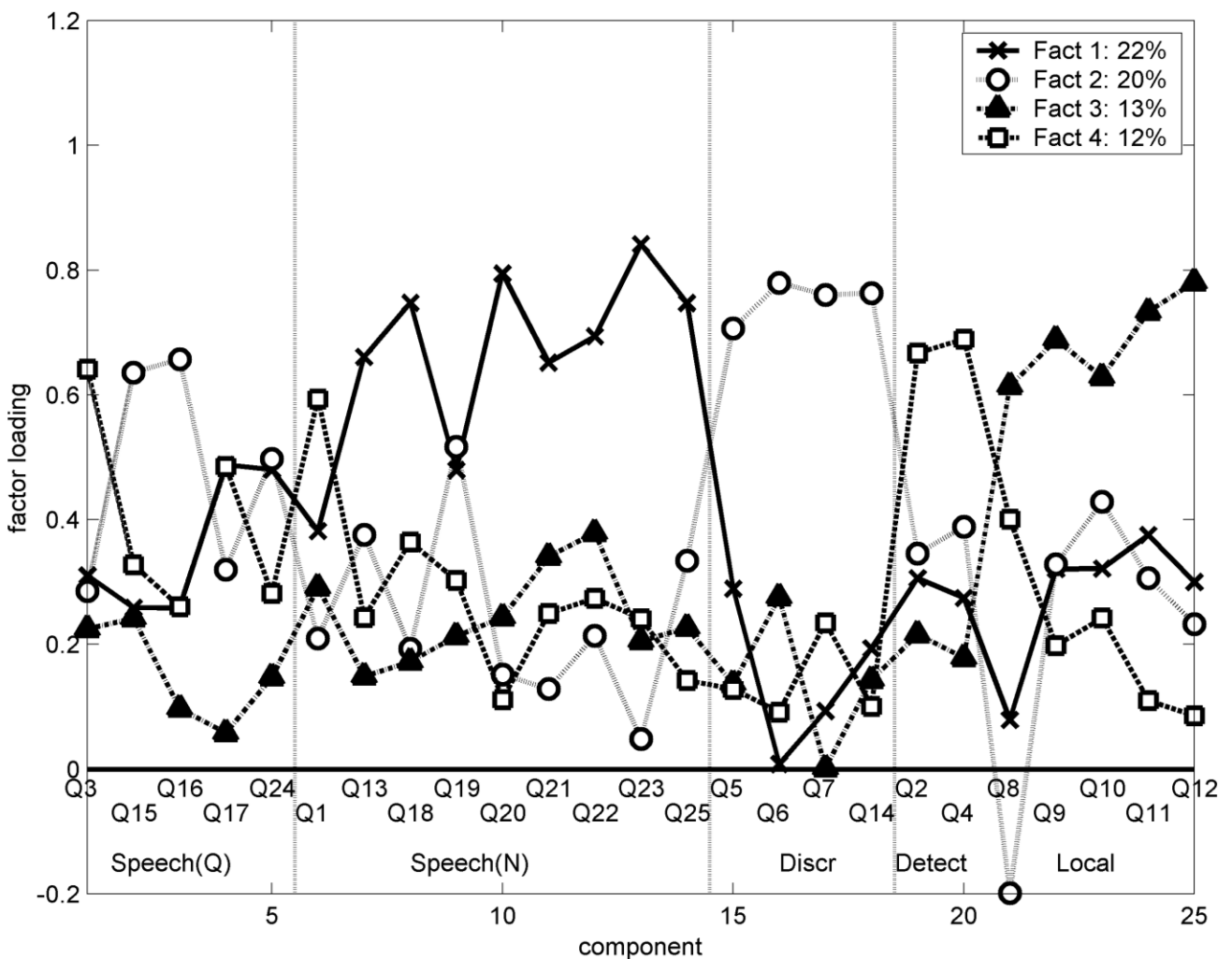
**Figure 1: Scree plot of Eigenvalues as a function of the component number.**

Table 5 shows how the four significant components explain an overall total variance of 68%. Table 5 shows the results before and after Varimax rotation. Before Varimax rotation, there is one main component that accounts for most of the variance (49.8%). After Varimax rotation, the variance is better divided over the different components.

Component	Eigenvalue		Variance expl. (%)	
	before	after	before	After
1	12.5	5.6	49.8	22.5
2	2.14	5.0	8.6	19.9
3	1.38	3.3	5.5	13.3
4	1	3.1	4	12.2

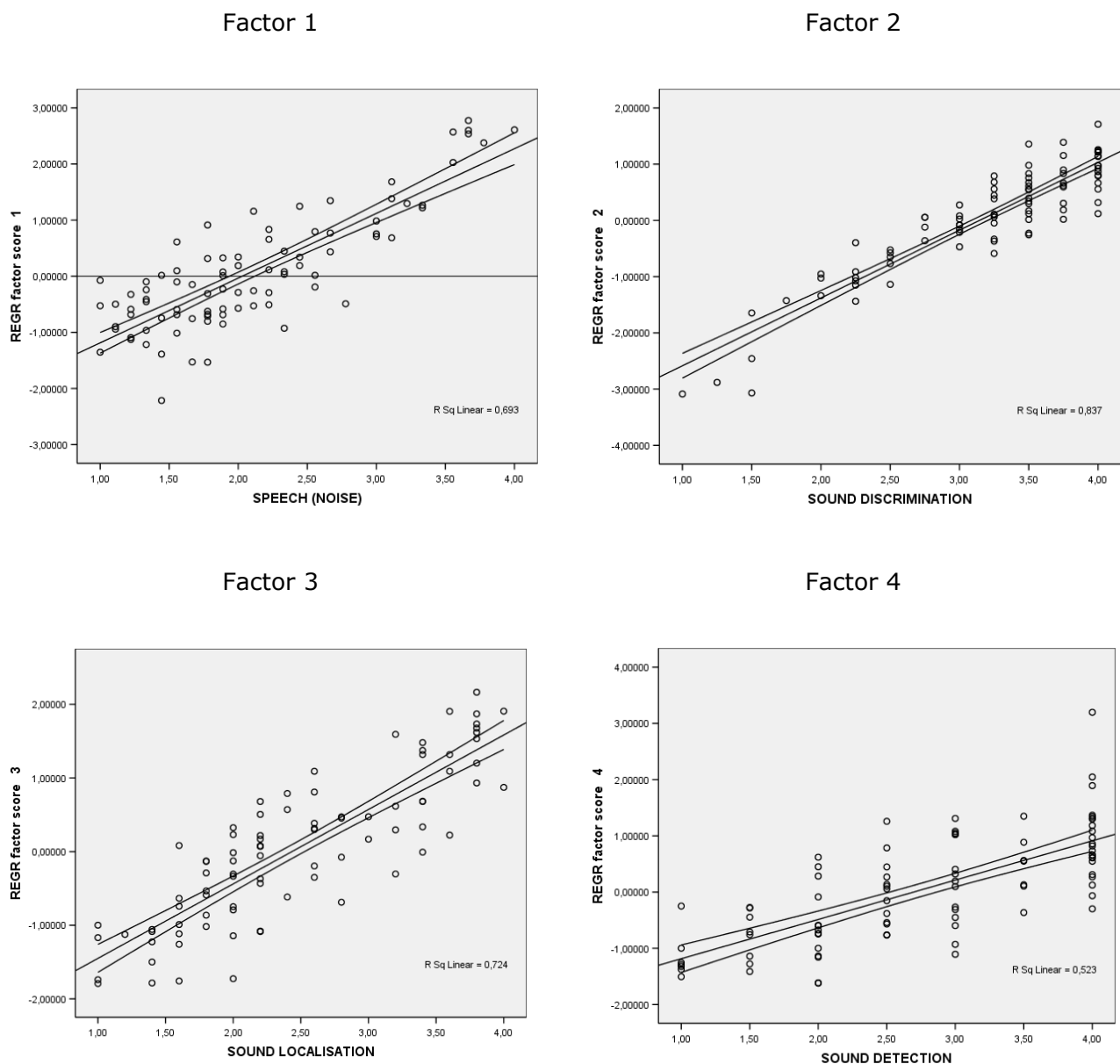
**Table 5: Results from factor analysis. 'Before' and 'after' relate to before or after Varimax rotation**

The factorplot after Varimax rotation is given in Figure 2.



**Figure 2: Factorplot after Varimax rotation.**

The results indicate that the first factor mainly loads on the items that focus on speech intelligibility in noise (Speech(N)). The second factor, mainly loads on items focussing on speech in quiet (Speech (Q)) and Discrimination of sounds (Discr). The third factor loads on the items focussing on sound localisation (Local). The fourth factor largely focuses on items describing audibility (Detect).



**Figure 3: Scatterplots of the mean of the items in different domains as a function of the regression coefficients of the items relative to the different factors (1-column 1, 2-column 2, 3-column 3, and 4- column 4).**

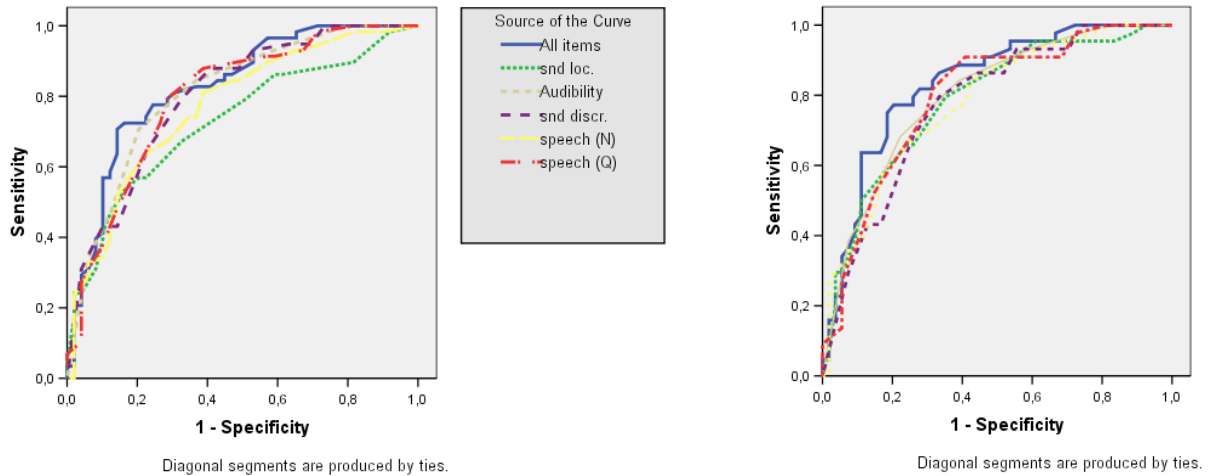
To illustrate the difference between the different factors (see Figure 2) and the original auditory competence domains (speech intelligibility in noise, speech intelligibility in quiet, sound localisation, audibility, and sound discrimination), a scatterplot of the regression coefficients relative to the original factors and the auditory competence domain is given (see Figure 3). These figures indicate that there is indeed good agreement with the original domain, although there is some scatter.

Finally, the internal consistency reliability values denoted by Cronbach's  $\alpha$  is fairly high (0.96). It should be noted that, because the items were chosen based on item-total correlations for this particular sample of subjects, it would not be surprising if results from a different sample reveal some decrease in the internal reliability statistics. For each item,  $\alpha$  is determined if this item were deleted. All  $\alpha$ 's were larger than 0.95, indicating a high reliability for the different items.

### 4.3.3 Screening

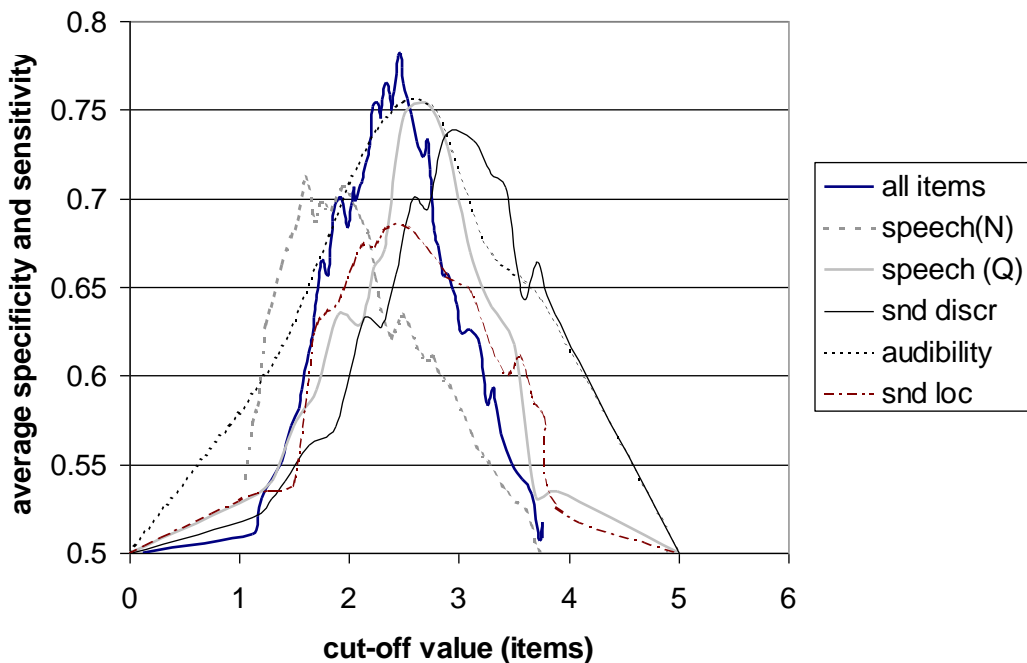
#### 4.3.3.1 Sensitivity and Specificity

The quality of screening is often expressed by the receiving operating characteristics (ROC) of the test. On the abscissa, "*1-specificity*" is expressed providing a direct measure of the proportion of false positives. The ordinate displays the sensitivity of the test, the proportion of true negatives. Since the test should have a high sensitivity and a low specificity, the cut off will be determined by maximizing the results in the left upper corner. Figure 4 displays receiver operating curves for the BEA (left) and the weighted hearing loss over two ears (right).



**Figure 4: ROCs based on hearing averaged over 1, 2, 4 kHz of the best ear (BEA; left) and the weighted average over the best ear and worse ear (Handicap; right).**

The average value of the sensitivity and specificity is plotted as a function of the cut-off score averaged over the items in a specific competence measure (see Figure 5). The optimum of the curve provides the highest accuracy.

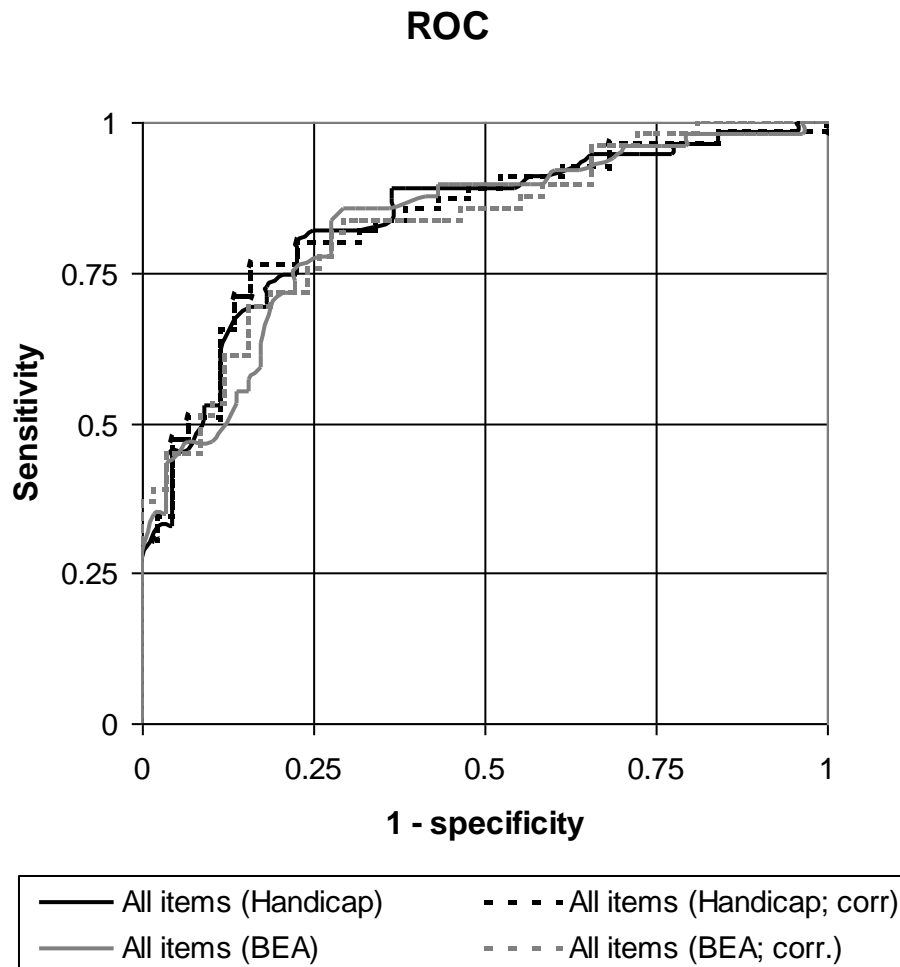


**Figure 5: Optimisation curves for the different competence domains. Specificity and sensitivity were weighted equally.**

Figure 5 indicates that the different competence measures have a different optimal cut-off value. Speech in noise (dashed curve) has a clear lower cut-off (more to the left) than speech in quiet (grey curve). These results also imply that a better performance of the questionnaire can be obtained by a different weighting for the different items/competence domains.

#### 4.3.3.2 Sensitivity and specificity - different weightings per item

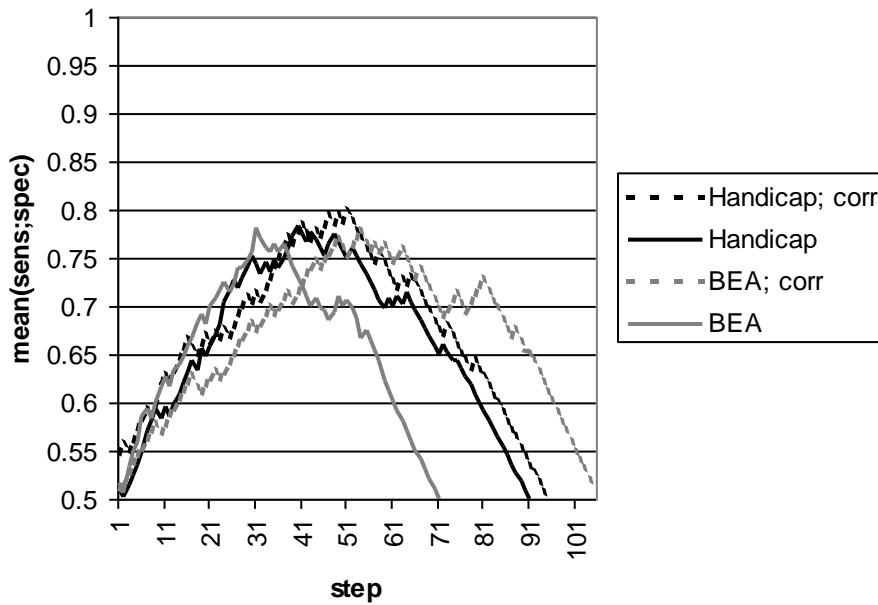
Psychometric curves based on the individual items are given in Appendix A. For each item-score, the averaged hearing loss of the best ear was determined. Based on these four data points, psychometric curves were determined for each item. The estimated percentage of respondents reporting difficulties with a particular item was determined at an average hearing loss of 30 dB (BEA). This was subtracted from the actual item score and item scores were limited between 0 and 1.



**Figure 6: ROC based on all items. Gray curves are based on the averaged hearing loss (1,2,4 kHz) of the best ear. Black curves are based on a 4:1 weight of best and worse ear. Solid curves are based on the direct item scores, dashed curves on the corrected item scores.**

The receiver operating characteristics of these two methods can be seen in Figure 6. The optimisation curves determined by the average of the sensitivity and specificity are given in Figure 7. The cut-off value for the optimal combination of sensitivity and specificity were determined and are given in Table 6.

**Optimalisation curves**



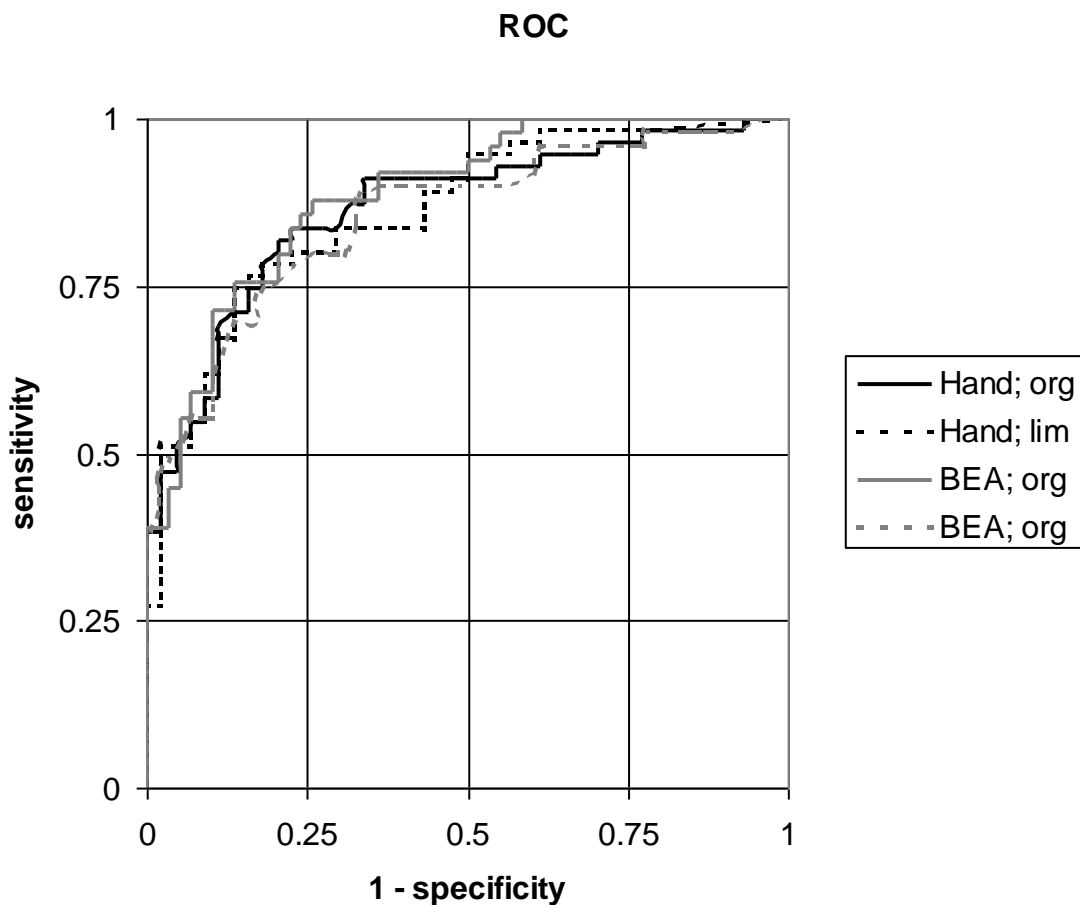
**Figure 7: Optimalisation curves for the four different situations.**

Condition	Cut-off	Sensitivity	Specificity
Handicap	0.381	0.818	0.75
Hand.; corr	0.090	0.745	0.841
BEA	0.129	0.816	0.723
BEA; corr.	0.220	0.800	0.759

**Table 6: Cut-off value, sensitivity, and specificity for optimal sensitivity and specificity combination.**

The above described data were determined based on all items. In the remainder of this section, we will discuss if a better performance can be obtained by removing some of the items. As a criterion to remove items, the maximum area underneath the ROC curve was used. In a first

approximation, the individual items were removed with the smallest area underneath the ROC up to the point that the area of the mean of the remaining items did not decrease any further. In this way, items with a relatively low sensitivity and specificity were removed from the analyses. After that, the effect of removing additional items on the area underneath the ROC was studied. The resulting ROC curve is given in Figure 8.



**Figure 8: ROC after optimisation of the area underneath the ROC by removing some of the items.**

Again four curves are displayed, providing the sensitivity specificity for the best ear averaged over 1, 2, and 4 kHz (BEA) and the weighted average (4:1) of the better and worse ear. Both were given for the average of the different items (org) and by score corrected by the fitted item-score at 30 dB HL. Table 7 summarizes the results of the optimisation. The

sensitivity and specificity were highest for the original analysis for the weighted analysis between the better and worse ear. By removing 10 items, the sensitivity and specificity increases.

Condition	Cut-off	Sensitivity	Specificity	Removed items
Handicap (org)	0.54	0.82	0.80	(n=10) Q5, Q8, Q12, Q14, Q15, Q16, Q19, Q20, Q21, Q22;
Hand.; corr	0.088	0.75	0.86	(n = 7) Q5, Q8, Q16, Q18, Q20, Q21, Q23
BEA	0.136	0.735	0.828	(n = 9) Q11, Q12, Q13, Q15, Q16, Q19, Q20, Q22, Q23
BEA; corr.	0.143	0.878	0.742	(n = 15) Q2, Q5, Q9, Q8, Q11, Q12, Q13, Q15, Q16, Q19, Q20, Q21, Q22, Q23, Q24

**Table 7: Cut-off value, sensitivity, and specificity for optimal sensitivity and specificity combination after removing some items (see last column).**

By removing these 10 items, the original intention of the questionnaire to focus on four categories remains intact. Instead of 12 items, the first category on self assessed speech intelligibility covers 8 items. The second category, on self assessed sound localisation, covers 3 instead of 5 items. The third category with focus on distinguishing sounds covers 2 items instead of 4.

## 5 Discussion

### 5.1 Outcome relative to the SSQ

The outcome of the questionnaire can be compared to established questionnaires such as the SSQ and the Kramer questionnaire. Table 8 provides insight in the rank correlation coefficients based on the hearing loss (1, 2, 4 kHz) in the best ear (BEA) and the weighted average over ears (Hand) on specific items. The items were identical or in some cases very comparable to the items discussed in the SSQ. Before both questionnaires are compared, it is important to note the difference in response categories. The Hearcom questionnaire uses a four-interval lickert scale, while the SSQ uses a sliding scale between 0 and 10. Considering these differences, the correlation coefficients are alike, a paired *t*-test indicates no significant differences between both observations (*p*-values; BEA: 0.39/ Hand: 0.08). More specifically, the items focussing on speech intelligibility (*p*-values; BEA: 0.67/ Hand: 0.05) or spatial hearing (*p*-values; BEA: 0.77/ Hand: 0.64) also indicate no significant differences between the two questionnaires.

Noble & Gatehouse (2004) mention that:

*"the division of the original sample into those with interaurally symmetrical and asymmetrical hearing loss reveals substantial differences in rated abilities, and substantial differences in the way in which those disabilities drive the experience of the handicap. Members of the asymmetry group rate their ability more poorly across all the domains addressed in the SSQ. This is especially over the three components (direction, distance and movement) of the spatial domain."*

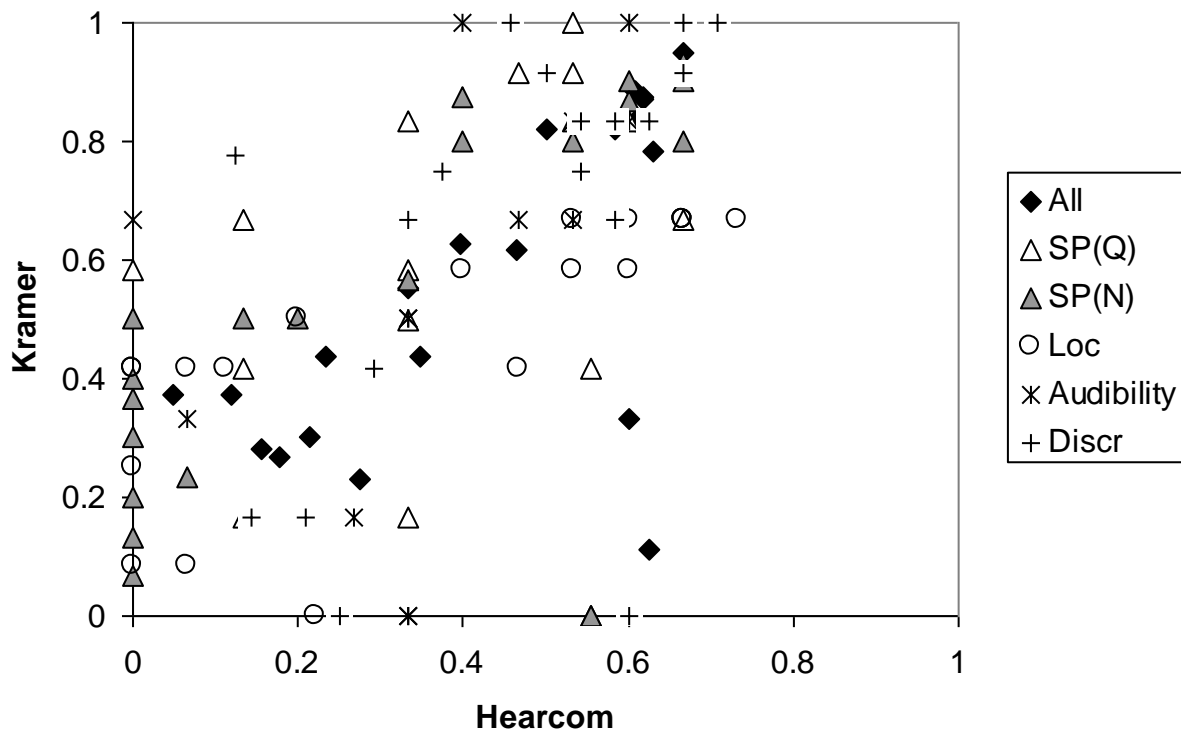
This effect is also shown in the effect of the predictive power in the worse ear in the spatial domain (see section 4.3.2.2).

<i>Item</i>	<i>BEA (SSQ)</i>	<i>Hand (SSQ)</i>	<i>BEA (HC-q)</i>	<i>Hand (HC-q)</i>
20) Follow 1 person speaking and telephone at same time.	-0.41	-0.64	-0.33	-0.36
24) Having conversation with 5 people in noise with vision.	-0.21	-0.52	-0.44	-0.44
23) Having conversation with 5 people in noise no vision.	-0.20	-0.49	-0.36	-0.36
15) Talking with 1 person in quiet room.	-0.49	-0.43	-0.35	-0.34
18) Follow one conversation with many people talking	-0.35	-0.41	-0.47	-0.45
16) Have conversation on telephone	-0.09	-0.41	-0.36	-0.32
25) Talking with one person in continuous background	-0.52	-0.37	-0.39	-0.40
13) Having conversation in echoic environment	-0.48	-0.32	-0.41	-0.40
10) Identify whether vehicle is approaching or receding	-0.47	-0.58	-0.47	-0.51
11) Identify lateral movement from voice or footsteps	-0.48	-0.56	-0.34	-0.40
12) Lateralize a talker to left to right	-0.42	-0.38	-0.26	-0.35
8) Sounds further than expected	-0.09	-0.32	-0.19	-0.23
9) Sounds closer than expected	-0.03	-0.23	-0.37	-0.43
14) Identify different people by voice	-0.30	-0.53	-0.36	-0.33

**Table 8: Rank correlation coefficients between the item-score and the hearing loss in the best ear (BEA) and the weighted average (Handicap) for the SSQ-questionnaire and the Hearcom questionnaire (Hc-q).**

## 5.2 Outcome compared to the Kramer survey

Twenty one subjects were asked to fill in the Kramer questionnaire after filling in the Hearcom questionnaire. In this way, the items in the five different competence areas, can be compared. The score of each competence domain was averaged over items. Missing items were left out of the analysis.



**Figure 9: Scatterplot between scores of the Kramer questionnaire and the Hearcom questionnaire. Scores were determined by taking the mean of all items in that particular competence domain.**

Figure 9 shows the scatterplot of the different subjects. Although there appears to be rather much scatter, the correlation coefficients between both questionnaires indicate a significant relation in all competence domains (see Table 9). The lowest correlation coefficient (speech in quiet) reported is probably relatively low since both questionnaires differ in their intention. Whereas the Kramer questionnaire intends to expose the

functionality, the Hearcom questionnaire focuses on hearing screening, where speech in quiet is probably not troublesome for those with onset hearing loss.

Competence domain	Correlation coeff.
All	0.66 <sup>***</sup>
Speech(q)	0.43 <sup>*</sup>
Speech (n)	0.73 <sup>***</sup>
Localisation	0.78 <sup>***</sup>
Audibility	0.66 <sup>***</sup>
Identification	0.60 <sup>**</sup>

**Table 9: Correlation coefficients between scores of the Kramer questionnaire and the Hearcom questionnaire. (\*\*\*)  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ ).**

### 5.3 Questionnaire specificity and sensitivity

The authors recognize that hearing thresholds derived from pure tone audiometry do not always adequately predict the need for and benefit of rehabilitation (Newman et al 1990; Taylor 1993). The usefulness and appropriate use of different outcome measurements have been under debate, and there are some indications that some of the measurements used are not sufficiently sensitive (Bentler & Kramer, 2000; Gatehouse, 2000). All respondents included in this analysis were patients visiting our audiological clinic. The fact that these people chose to visit the clinic indicates that they must feel that they have some difficulties with their hearing. This will be reflected in the normal population, and therefore it can be expected that the actual specificity and sensitivity will be higher

than reported here. On the other hand, it can be argued that the actual sensitivity and specificity will be higher than reported here, since the questionnaire needs to identify normal hearing respondents that actually experience problems with their hearing (such as tinnitus). The intention of this questionnaire is not to advise people that they have no significant problems with their hearing, but rather that there is no need for audiological rehabilitation (such as the use of a hearing aid, for example). The criteria to define good, insufficient, and poor hearing are arguable. In addition, sensitivity and specificity are conflicting properties. In the following sections, the effect of this area between good and poor hearing on sensitivity and specificity will be discussed.

### 5.3.1 The effect of cut-of hearing loss

The hearing threshold at which someone is said to be hearing-impaired is arguable. In the above mentioned analysis, 30 dB HL is considered to be a valid figure. All respondents with a hearing loss averaged over 1, 2, and 4 kHz were considered to be good for an averaged value below 25 dB HL, insufficient for hearing losses between 25 dB HL and 35 dB HL and poor for hearing losses above 35 dB HL. In the Netherlands, people with a hearing loss smaller than 35 dB HL will not receive a co-payment from their insurance, which may also provide a valid criterion for the good–insufficient hearing (green-orange) differentiation.

Cut-off (dB HL)	Orange (N)	BEA			BEA (org; lim)			BEA(fin)		
		sens	spec	sum	sens	spec	sum	sens	spec	sum
30	26	0.71	0.86	1.57	0.76	0.80	1.56	0.86	0.76	1.62
31	30	0.71	0.89	1.60	0.77	0.84	1.61	0.74	0.93	1.67
32	31	0.70	0.90	1.60	0.84	0.80	1.64	0.89	0.80	1.69
33	31	0.70	0.90	1.60	0.80	0.83	1.63	0.89	0.80	1.69
34	32	0.80	0.81	1.61	0.83	0.84	1.67	0.88	0.81	1.69
35	28	0.79	0.81	1.60	0.82	0.84	1.66	0.87	0.84	1.71
36	32	0.79	0.82	1.61	0.85	0.79	1.64	0.87	0.85	1.72
37	28	0.78	0.84	1.62	0.81	0.87	1.68	0.87	0.87	1.74
38	31	0.78	0.84	1.62	0.81	0.87	1.68	0.86	0.87	1.73
39	23	0.76	0.87	1.63	0.81	0.87	1.68	0.83	0.90	1.73
40	19	0.77	0.82	1.59	0.82	0.87	1.69	0.82	0.90	1.72

**Table 10: Effect of cut-off hearing loss in the best ear between good and insufficient hearing. Column orange provides the number of observations in the category insufficient hearing. Three different sensitivity and specificity combinations were given 1) BEA – based on the average of all items. 2) BEA(org; lim) – based on the average of all items corrected for the items score at 35 dB, and 3) BEA (fin) based on the average of the different item scores corrected for the score at 35 dB leaving out specific items (see section 4.3.3.2).**

Table 10 and Table 11 illustrate the effect of the cut-off hearing loss, at which is determined if people are normal hearing or have insufficient hearing, on the specificity and the sensitivity. The area of insufficient hearing was chosen to be equal to 10 dB (for instance 30 dB – 40 dB). Results in Table 10 are based on the hearing loss in the best ear, while Table 11 is based on the weighted hearing loss of both ears. The performance of the questionnaire seems to increase slightly with increasing cut-off hearing loss. Choosing a higher cut-off hearing loss (in

order to follow the Dutch guidelines for co-payment of hearing aids, for instance), offers a slightly better sensitivity and specificity.

Cut-off (dB HL)	Orange (N)	Handicap			Hand. (org; lim)			Handicap (fin)		
		sens	spec	sum	sens	spec	sum	sens	spec	Sum
30	34	0.80	0.77	1.57	0.82	0.80	1.62	0.75	0.86	1.61
31	32	0.79	0.73	1.52	0.77	0.77	1.54	0.75	0.79	1.54
32	35	0.86	0.79	1.65	0.88	0.73	1.61	0.84	0.78	1.62
33	37	0.87	0.72	1.59	0.85	0.78	1.63	0.83	0.82	1.65
34	36	0.89	0.72	1.61	0.91	0.74	1.65	0.84	0.81	1.65
35	37	0.90	0.71	1.61	0.85	0.80	1.65	0.85	0.80	1.65
36	34	0.90	0.69	1.59	0.83	0.81	1.64	0.85	0.79	1.64
37	29	0.90	0.71	1.61	0.82	0.82	1.64	0.85	0.80	1.65
38	29	0.89	0.71	1.60	0.84	0.80	1.64	0.84	0.80	1.64
39	26	0.92	0.69	1.61	0.92	0.73	1.65	0.86	0.79	1.65
40	22	0.81	0.77	1.58	0.86	0.76	1.62	0.89	0.75	1.64

**Table 11: Similar to Table 10 but now based on the weighted average between best and worse ear (1:4).**

### 5.3.2 The effect of the width of insufficient hearing

Next to indicating a clearly degraded hearing (red area) or clearly good hearing (green area), screening tests often have an intermediate area that indicates that someone's hearing has only been degraded to some extent (amber area). This area needs to provide a different message than the red area - in which the respondent is alerted to the potential hearing loss - and to encourage the user to seek further testing.

From a statistically view point, the number of respondents falling in this amber area are left out of further analysis (and coded as missing values).

Hence, the size of the amber area will impact on the sensitivity and specificity. The effect of the size (in dB hearing loss symmetrically around 35 dB) of this amber area is given in Table 12 for the best ear and in Table 13 for the weighted difference between both ears.

width	BEA			BEA (org; lim)			BEA (fin)		
	sens	spec	sum	sens	spec	sum	sens	spec	Sum
0	0.71	0.82	1.53	0.78	0.78	1.56	0.82	0.75	1.57
2	0.71	0.82	1.53	0.71	0.84	1.55	0.79	0.80	1.59
4	0.74	0.83	1.57	0.81	0.80	1.61	0.84	0.83	1.67
6	0.76	0.80	1.56	0.81	0.80	1.61	0.84	0.83	1.67
8	0.79	0.81	1.60	0.82	0.84	1.66	0.87	0.84	1.71
10	0.79	0.81	1.60	0.82	0.84	1.66	0.87	0.84	1.71
12	0.81	0.80	1.61	0.83	0.82	1.65	0.88	0.85	1.73
14	0.82	0.84	1.66	0.87	0.84	1.71	0.85	0.87	1.72
16	0.82	0.84	1.66	0.84	0.87	1.71	0.89	0.87	1.76
18	0.83	0.87	1.70	0.88	0.87	1.75	0.88	0.90	1.78
20	0.83	0.87	1.70	0.88	0.87	1.75	0.90	0.90	1.8

**Table 12: Difference in averaged hearing loss of the best ear (1, 2, 4 kHz) between the result good and poor. Three different sensitivity and specificity combinations were given 1) BEA – based on the average of all items. 2) BEA(org; lim) – based on the average of all items corrected for the items score at 35 dB, and 3) BEA (fin) based on the average of the different item scores corrected for the score at 35 dB leaving out specific items (see section 4.3.3.2).**

Without amber area, the questionnaire manages a sensitivity and specificity of respectively 0.75 and 0.82 (best ear) or 0.74 and 0.75 (weighted difference). These numbers increase by increasing the amber area. Allowing an amber area from 25 to 45-dB losses will even provide a sensitivity and specificity of 0.9 (best ear) and 0.89 and 0.86 (weighted difference in ears). This 20-dB area may be adequate to provide clinically

interesting outcomes. After all, people on the lower edge of the amber area (25 dB – 35 dB) experience problems hearing. Not all of these problems are necessarily restricted to audibility (pure tone audiogram), but may also refer to other properties of auditory functioning. Gatehouse and Noble also indicate that the pure tone audiogram provides a first approximation of the experience restrictions by the subject. There are other effects that restrict auditory functioning in the daily life of a subject – e.g., sound localization, speech intelligibility, and sound discrimination. It may be good to alert people to this and to draw these people to auditory clinics, since there may be solutions available that provide additional help for these respondents. Reducing the width of the orange area to 12 dB (29-41 dB) still provides an adequate sensitivity and specificity (0.88 and 0.85 – BEA, or 0.85 and 0.81-weighted difference).

width	Handicap			Handicap			Handicap		
	Sens	spec	sum	Sens	spec	sum	sens	spec	Sum
0	0.72	0.75	1.47	0.81	0.69	1.50	0.74	0.75	1.49
2	0.74	0.76	1.50	0.83	0.71	1.54	0.74	0.79	1.53
4	0.80	0.77	1.57	0.80	0.79	1.57	0.80	0.80	1.60
6	0.80	0.77	1.57	0.80	0.80	1.60	0.80	0.80	1.60
8	0.82	0.76	1.58	0.82	0.79	1.61	0.84	0.79	1.63
10	0.90	0.71	1.61	0.85	0.80	1.65	0.85	0.80	1.65
12	0.88	0.74	1.62	0.83	0.83	1.66	0.85	0.81	1.66
14	0.90	0.72	1.62	0.85	0.82	1.67	0.85	0.82	1.67
16	0.89	0.71	1.60	0.84	0.82	1.66	0.84	0.82	1.66
18	0.92	0.71	1.63	0.84	0.83	1.67	0.86	0.81	1.67
20	0.92	0.75	1.67	0.86	0.86	1.72	0.89	0.86	1.75

**Table 13: Similar to Table 12 but now for the weighted hearing loss between best an worse ear.**

The reported sensitivity and specificity is comparable to the reported sensitivity and specificity for the averaged hearing loss over 0.5, 1, and 2 kHz for the triplet digit test (0.75 and 0.91; Smits et al. 2004). It will be of much interest to test both screening tests in the same subjects and to see if one test provides additional information above the other. Such a study will form part of D11-6.

## **6 Dissemination and Exploitation**

The dissemination of this information is restricted to the consortium. It would be of no benefit to the user to receive any information concerning the development of the questionnaire.

The questionnaire assesses an important part of hearing without the need of additional hardware at the testing site. The subjective evaluation of one's hearing adds to the objective methods provided by the localisation tests (MAA and DE-RUB localisation test) or the assessment of speech in noise (triplet digit test). Users are given the ability to obtain additional information regarding their abilities to use auditory cues.

### **6.1 Ethical issues**

When developing a hearing test that can be conducted using the Internet without the involvement of a professional, there are various ethical issues to be considered. First, the users must know what is expected from them. Not understanding the instructions can lead to unexpected results and therefore the advice given at the end of the test will not be suitable. The focus of the usability study was mainly on the instructions and the understanding of the different items.

Secondly, there is the question of what should be done with the data once people have conducted the test. Obviously, once the test has been implemented in the HearCom site, the name and address of the user will not be asked for. The questionnaire should be freely accessible and should be considered as an additional tool to assess one's hearing. This should also be made clear in the instructions. Once D11-6 (validation and comparison of all screening tests) is finished, a choice should be made about the most reliable tools for screening that will be placed on the web portal.

Thirdly, the outcome of the test should be clear. In other words, the test should have a high sensitivity and specificity. The discussion focused on this issue by assessing the effect of the size and location of the insufficient (amber) area. Error messages should be very clear, in order to provide the correct meaning.

## **7 Conclusions**

The HearCom screening questionnaire was designed to measure a range of hearing impairments across several domains; speech intelligibility, sound localisation, audibility of sounds and discrimination of sound. The usability tests did not expose any serious issues that could not be solved. Competence was asked on 25 items for 129 participants. Answers were scored on a four-interval likert scale. A factor analysis identified the four competence domains after Varimax rotation. This is in agreement with the Kramer survey. Three areas are identified based on the hearing loss averaged over 1, 2, and 4 kHz;

1. the green area, intended to cover respondents with a hearing loss lower than 29 dB.
2. the amber area, indicating insufficient hearing covering respondents with a hearing loss between 29 and 41 dB.
3. the red area, indicating a poor hearing for respondents having a hearing loss more than 41 dB.

Based on these outcomes, some items should be removed in order to optimize the sensitivity and specificity by increasing the area under a receiving operating characteristic (ROC). In this way, for the best ear, 10 items are sufficient in order to provide a high sensitivity and high specificity (0.88 and 0.85). In order to obtain such a high accuracy for the weighted hearing loss in both ears, 18 items are required (0.85 and 0.81). By selecting the advised number of items we can create a useful clinical tool to screen on hearing loss.

## 8 References

Bentler RA, Kramer SE. (2000) Guidelines for choosing a self-report outcome measure. *Ear Hear.* 21(4 Suppl):37S-49S.

CBS (2006) - <http://www.cbs.nl/nl-NL/menu/themas/gezondheid-welzijn/methoden/dataverzameling/korte-onderzoeksbeschrijvingen/permanent-onderzoek-leefsituatie-pols-basisvragenlijst1.html>

Cienkowski KM, McHugh MS, McHugo GJ, Musiek FE, Cox RM, Baird JC. (2006) A computer method for assessing satisfaction with hearing aids. *Int J Audiol.* 45:393-9.

Clark K, Sowers M, Wallace RB, Anderson C. (1991) The accuracy of self-reported hearing loss in women aged 60-85 years. *Am J Epidemiol.* 134:704-8.

Cox RM, Alexander GC. (1995) The abbreviated profile of hearing aid benefit. *Ear Hear.* 16:176-86.

Davis AC. (1989) The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. *Int J Epidemiol.* 18:911-7.

Eekhof JA, de Bock GH, de Laat JA, Dap R, Schaapveld K, Springer MP. The whispered voice: the best test for screening for hearing impairment in general practice? *Br J Gen Pract.* 46:473-4.

Gatehouse S, Noble W. (2004) The Speech, Spatial and Qualities of Hearing Scale (SSQ). *Int J Audiol.* 43(2):85-99.

Gates GA, Murphy M, Rees TS, Fraher A. (2003) Screening for handicapping hearing loss in the elderly. *J Fam Pract.* 52:56-62.

Gomez MI, Hwang SA, Sobotova L, Stark AD, May JJ. (2001) A comparison of self-reported hearing loss and audiometry in a cohort of New York farmers. *J Speech Lang Hear Res.* 44:1201-8.

Koopman, J, Davey E., Wittkopf T., Verschuure, J. (2008) How should hearing screening tests be offered? (2008) accepted for publication in *Int. J. Audiol.*

Kramer SE, Kapteyn TS, Festen JM, Tobi H. (1995) Factors in subjective hearing disability. *Audiology.* 34(6):311-20.

Lichtenstein MJ, Bess FH, Logan SA. (1988) Diagnostic performance of the hearing handicap inventory for the elderly (screening version) against differing definitions of hearing loss. *Ear Hear.*9(4):208-11.

Lutman ME, Brown EJ, Coles RR. (1987) Self-reported disability and handicap in the population in relation to pure-tone threshold, age, sex and type of hearing loss. *Br J Audiol.* 21:45-58.

Newman CW, Weinstein BE, Jacobson GP, Hug GA. (1990) The Hearing Handicap Inventory for Adults: psychometric adequacy and audiometric correlates. *Ear Hear.* 11:430-3.

Noble W, Gatehouse S. (2004) Interaural asymmetry of hearing loss, Speech, Spatial and Qualities of Hearing Scale (SSQ) disabilities, and handicap. *Int. J. Audiol.* 43:100-14.

Noble W, Gatehouse S. (2006) Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial, and Qualities of Hearing Scale (SSQ). *Int. J. Audiol.* 45:172-81.

Plomp R, Mimpen AM. (1979) Improving the reliability of testing the speech reception threshold for sentences. *Audiology* 18:43-52.

Popelka MM, Cruickshanks KJ, Wiley TL, Tweed TS, Klein BE, Klein R. (1998) Low prevalence of hearing aid use among older adults with hearing loss: the Epidemiology of Hearing Loss Study. *J Am Geriatr Soc.* 46:1075-8.

Schaik van, V.H.M. (1998) "Hearing loss in the elderly – hearing questionnaires in general practice", PhD-dissertation – Utrecht.

Sindhusake D, Mitchell P, Smith W, Golding M, Newall P, Hartley D, Rubin G. (2001) Validation of self-reported hearing loss. The Blue Mountains Hearing Study. *Int J Epidemiol.*30:1371-8.

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

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

Taylor KS. (1993) Self-perceived and audiometric evaluations of hearing aid benefit in the elderly. *Ear Hear.*14:390-4.

Wilson DH, Walsh PG, Sanchez L, Davis AC, Taylor AW, Tucker G, Meagher I. (1999) The epidemiology of hearing impairment in an Australian adult population. *Int J Epidemiol.* 28(2):247-52.

# Appendix – A : Psychometric curves for each item

