FP6–004171 HEARCOM

Hearing in the Communication Society

INTEGRATED PROJECT

Information Society Technologies

D-2-5: Optimized, final set of impairment tests included in the auditory profile

Contractual Date of Delivery:	M45: 31, May, 2008 (+45 days)
Actual Date of Submission:	28 July 2008
Editor:	Wouter Dreschler Thamar van Esch
Sub-Project/Work-Package:	SP1, WP2
Version:	V2.0 (modified after review)
Total number of pages:	30

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

Deliverable D-2-5

VERSION DETAILS	
Version:	2.0
Date:	July 22, 2008
Status:	Final version

CONTRIBUTOR(S) to DELIVERABLE

Partner	Name
NL-AMC	Wouter A. Dreschler
NL-AMC	Thamar van Esch
NL-VUMC	Johannes Lyzenga
DE-HZO	Matthias Vorman
DE-HZO	Kirsten Wagener
SE-LINK	Birgitta Larsby
SE-LINK	Matthias Hällgren
UK-ISVR	Sheetal Athalye

DOCUMENT HISTORY

Version	Date	Responsible	Description
0.1	April 28, 2008	Wouter Dreschler	First draft version
0.2	May 7th, 2008	Thamar van Esch	Final results of MCS-1 added
0.3	May 27th, 2008	Wouter Dreschler and Mark Lutman	Discussion parts added (chapters 4 to 7)
1.0	June 26, 2008	Wouter Dreschler Thamar van Esch	Suggestions from different colleagues. Version for review.
2.0	July 22, 2008	Wouter Dreschler Thamar van Esch	Modified after review

DELIVERABLE REVIEW

Version	Date	Reviewed by	Conclusion*
1.0	July 12, 2008	Dr. Andrew Faulkner	Accept with some modifications
1.0	July 12, 2008	Prof.dr. Tammo Houtgast	Accept with some modifications

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

Table of contents

List of Figures

List of Tables

Acknowledgement

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

Pre-Amble

This deliverable concerns the description of the optimized, final set of impairment tests included in the auditory profile.

The choices have been based on the final results of the first multi-centre study incorporating 103 subjects, on extra reference tests in order to support the interpretation, and on the outcomes of a consensus meeting about the reduction of the tests included in the preliminary version of the Auditory Profile. Also, we decided to incorporate a discussion about the relevance of the Auditory Profile for auditory rehabilitation.

The protocol for this multi-centre study has been described in deliverable D-2-2 and preliminary results have been presented in deliverable D-2-3. The auditory profile was designed in order to enable consistent characterization of an individual’s auditory impairment across Europe. The auditory profile has the potential to develop to and to be accepted as an advanced clinical standard that can be used to determine the individual hearing deficiencies in communication and can help to determine the benefit of assistive devices. The auditory profile should include a range of necessary measures to describe the details of, and differences between, different hearing impairments.

The auditory profile is relevant for the work in SP2 (Adverse Conditions), because its outcome values will define the auditory demands for the acoustical conditions required in case of hearing impairment. The work is relevant for SP3 (Rehabilitation) and SP4 (Assistive Technology), because the auditory profile indicates the deficits that need to be compensated, either by signal processing (SP3) or by alternative strategies (SP4). Finally, the implementation of these tests in OMA (Oldenburg Measurement Applications) will have a great impact on the dissemination of the test procedures and will stimulate a broad clinical acceptance of this innovative approach to auditory testing. The plan for dissemination and exploitation is described in D-2-4.

1 Executive Summary

The HEARCOM (Hearing in the Communication Society) project aims at full participation in the modern communication society by reducing the limitations in auditory communication. Two of the focus areas of HEARCOM are on the identification and characterization of auditory communication limitations and on the development of standardized testing and evaluation procedures for hearing-impaired persons. In this context, a preliminary auditory profile has been defined. This preliminary auditory profile has been validated in an international multi-centre study. The first results have been presented in deliverable D-2-3. The results of further analyses will be presented in the current report.

The aim of the auditory profile is that it should be used as a diagnostic tool in a broad population of subjects with complaints about their performance in (auditory) communication tasks. Our ambition is that the Auditory Profile will become a diagnostic standard that can be assessed in a (specialized) hearing centre or clinic or in audiological research. The end user of the auditory profile is the professional interested in the characteristics of the hearing of a particular client/patient.

The focus of the multi-centre field trial was to investigate the clinical applicability, the relevance of the test results to characterize communication ability, and the derivation of the parameters to be used in further testing. All tests have been implemented in OMA (Oldenburg Measurement Applications) software. For applications in research, the software can be left more open in order to be flexible towards diverging applications. However, for clinical use the test procedures should be described in detail and the software may need a CE-approval in due time.

In the current deliverable we start with an extra reference study that was required for a correct interpretation of the results across different countries. This report further describes the final analyses of the results of the first multi-centre study and the consensus that was reached with respect to a reduction of the tests included in the preliminary version of the Auditory Profile. Finally, we added a discussion about the relevance of the Auditory Profile for auditory rehabilitation. This in agreement with the requests from the last review, in which was also decided to devote extra experimental work to this topic (WP2, task 4). Those results will be reported in a separate deliverable (D-2-8: The Auditory Profile in auditory rehabilitation).

2 Reference study ILD, BILD, GP

2.1 Introduction

In the preliminary analyses of the multi-centre study data it was found that there were unforeseen centre/language effects in ILD, BILD and Gothenburg Profile data. Because of these effects, it was decided in Wassenaar to perform an extra reference study on these tests. The purpose of this reference study was to see if the effects were consistent, and if they were, to have reference data to perform language-specific corrections on the multi-centre study data.

2.2 Protocol

Ten normal-hearing subjects in each language were included in the reference study (10 at HZO, ISVR and LINK and 5 at VUMC and AMC). All subjects had air-conduction thresholds better than 20 dB HL and were between 18 and 75 years old. Tests were conducted in exactly the same manner as in the multi-centre study. Besides ILD, BILD, and GP, also an Acalos with broadband noise was conducted to determine measurement levels. Consequently, the protocol included the following tests:

- Audiogram (needed for inclusion and definition of ‘better ear’ based on PTA(1,2,4)).
- Acalos with broadband noise (ICRA1 or ICRA1female, depending on the gender of the speaker) at the better ear. MCL-low derived from this test was the measurement level for the following tests.
- Training matrix test: two 20-items lists, presented binaurally at fixed SNR (0 dB).
- Intelligibility level difference (ILD) measurement, left and right, measured with matrix-type sentences.
- Binaural intelligibility level difference (BILD) measurement, left and right, measured with matrix-type sentences.
- Gothenburg Profile

For protocol details we refer to deliverable D-2-2 in which the protocol for the multi-centre study has been described. The order of the tests was counterbalanced across subjects using a Latin square. All tests except the audiogram and Acalos were repeated for retest during a second run one to three weeks after the first run.

2.3 Results

Before analysing differences between languages, results from VUMC and AMC were compared. Paired t-tests showed that there were no significant differences between the two Dutch centres. As a result, we could pool all Dutch data.

2.3.1 ILD

Figure 1 shows ILD results (averaged over test and retest) of the reference data. The vertical axis represents the release of masking obtained by subtracting the SNR in the S0N0 condition from the SNR in the S0N90 condition. Clearly, the same pattern of differences between languages as seen in the preliminary multi-centre study analyses appears. For ILDleft (noise from left side) Dutch results differ significantly from Swedish and German data (ANOVA with Bonferroni correction) whereas for ILDright Dutch data are significantly different from all other languages. Differences between languages are typically 1.5 to 2.0 dB, which is relevant compared to the ILD effect size.

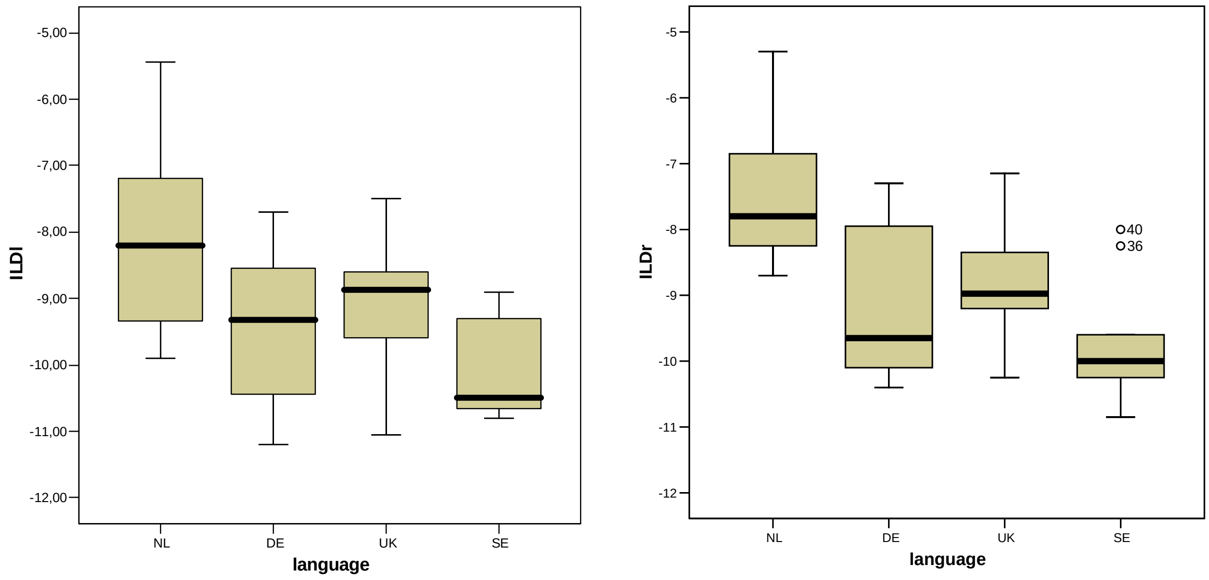

Figure 1: Differences between languages in ILD data of normally-hearing listeners for noise from left and right sides (left and right panels respectively).

2.3.2 BILD

BILD data of the reference study (averaged over test and retest) are presented in figure 2. Plotted values are the release of masking obtained by subtracting the monaural S0N90 condition from the binaural S0N90 condition. As can be seen, results are similar to those presented in D-2-3. Differences between languages are around 1.5 to 2.0 dB, which is quite large compared to the BILD effect size. For BILD$_{\text{left}}$, German data differs significantly from Dutch data, and for BILD$_{\text{right}}$ German data is significantly different from all other languages. This probably reflects the fact that the German speech tests used a male talker with a higher content of low-frequency speech information. Since the effect of binaural unmasking is more pronounced at low frequencies, it can be expected that the male/female difference is larger for the BILD data (i.e., pure binaural processing) than for the ILD data (i.e., combination of monaural and binaural effects).

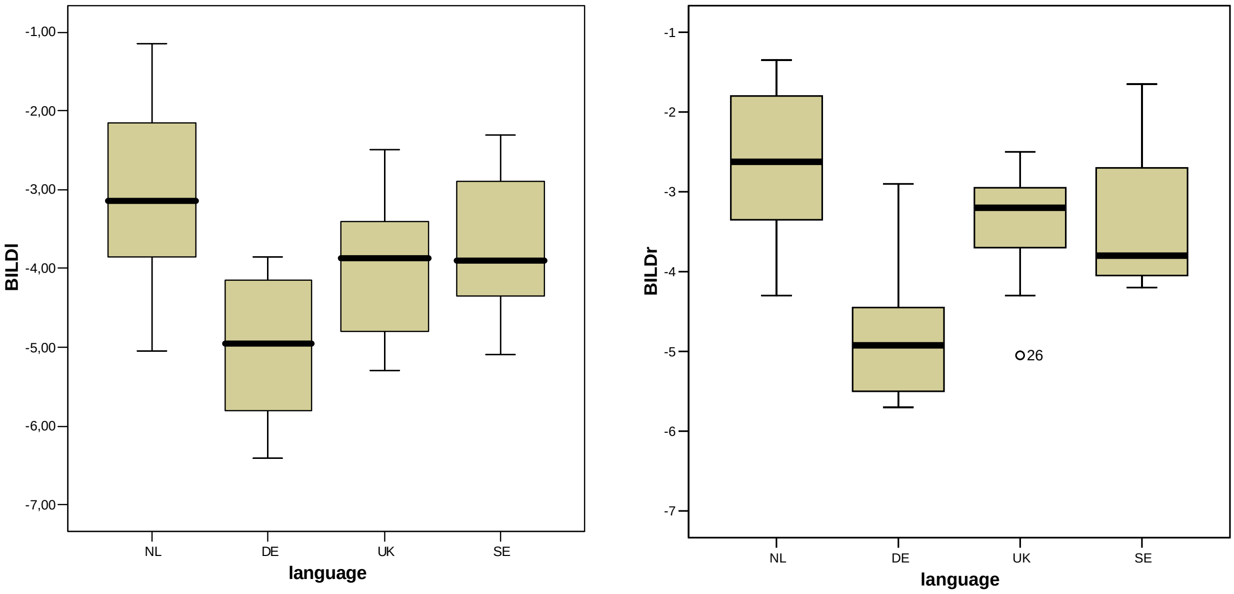

Figure 2: Differences between languages for BILD data of normally-hearing listeners. Left panel shows results for BILD$_{\text{left}}$ (noise from left side), BILD$_{\text{right}}$ data are presented in the right panel.

2.3.3 Gothenburg Profile

Figure 3 shows differences between languages for the Gothenburg Profile, averaged across test and retest. Like in the preliminary analyses of the multi-center study, British results are very deviant from all other languages on the ‘speech perception’ subscale. The average difference between UK data and other countries is 16.8% on a 100% scale. Also on the ‘spatial hearing’ subscale are significant differences between languages: UK data differs significantly from SE and DE data (mean difference: 12.9%), and the difference between NL and DE data is significant too (9%).

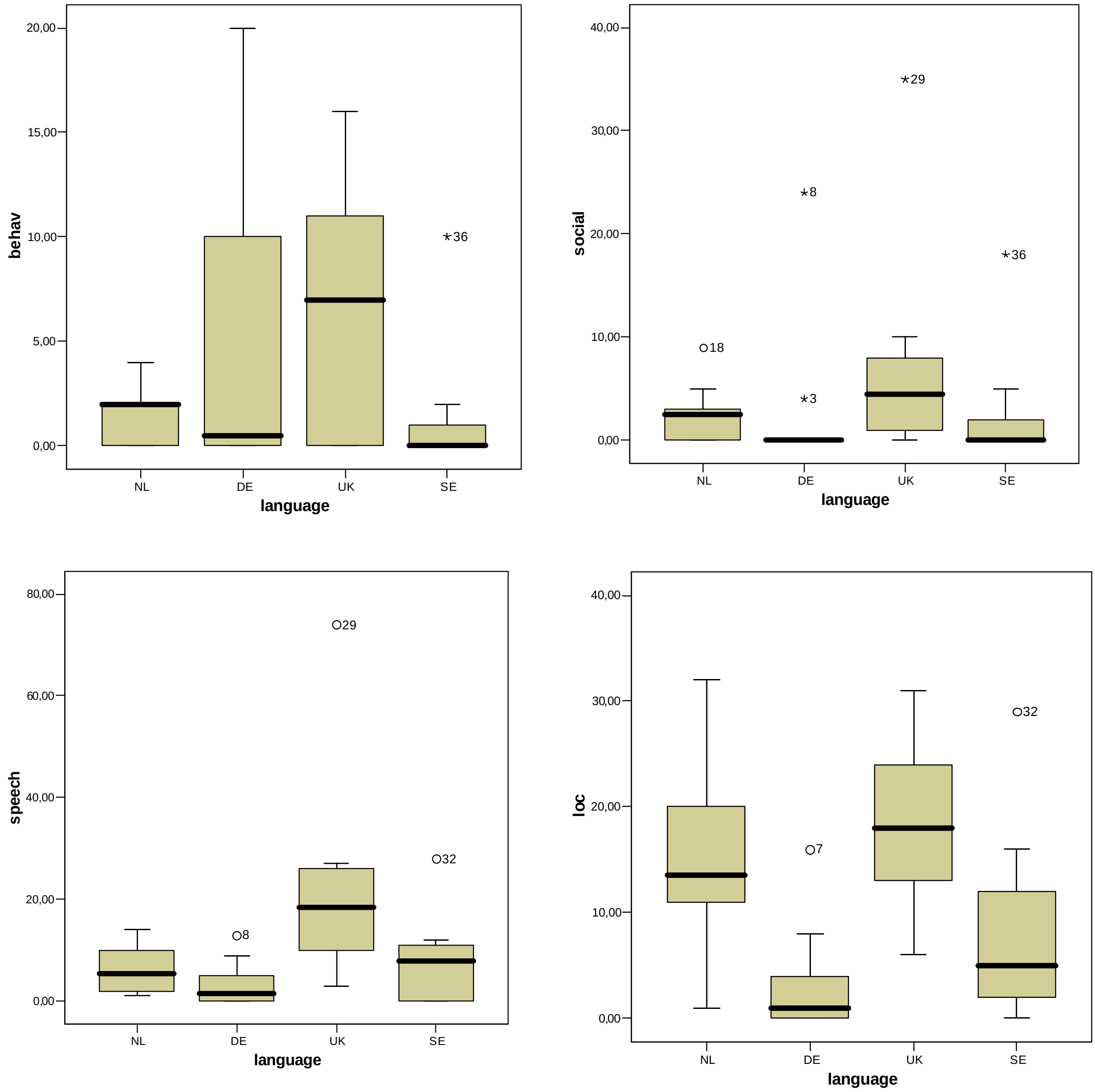

Figure 3: Gothenburg Profile results from reference study; differences between languages for the four subscales: 'behaviour and reaction' (upper left panel), 'social interaction' (upper right), 'speech perception' (lower left), and 'spatial hearing' (lower right). Please note that vertical axis scales are not identical.

2.4 Discussion, conclusions and future plans

2.4.1 ILD and BILD

Both in ILD and BILD, significant, clinically relevant differences (typically 1-2 dB) between languages are found. These differences are difficult to interpret, because the ILD and BILD are difference values (subtractions of SRTs in different conditions). The most probable explanation is related to differences in speech spectra used in the different languages that cause small, but consistent and significant differences across languages, especially in combination with possible interactions with the HRTF-filtering of speech and noise. This possibility will be investigated at ISVR by studying all filtered and unfiltered speech and noise spectra and by comparing the average BILD for a male and a female talker.

Although we do not yet completely understand the cause of the language effects, we decided at the Leuven meeting that we need to correct for these effects in the analyses of the multi-center study data. Therefore we calculated mean values (left and right sides pooled) of ILD and BILD for each language. These mean values will be subtracted from the multi-center study data to correct for the language effects.

2.4.2 Gothenburg Profile

In the Gothenburg Profile data, significant language effects were found only in the speech perception and spatial hearing subscales. These effects could be caused by translation differences, and are too large to be ignored. Although no differences were found in the other two subscales, it was decided in Leuven that it would be more consistent to correct all subscales than only two out of four. Therefore we calculated median values of each subscale (data are not distributed normally) for every language. These median values will be subtracted from the multi-centre study data to correct for the language effects.

3 Analysis of full data set

All analyses described in deliverable D-2-3 were repeated on the final data set. In this deliverable, we will only present a selection of these analyses, and leave out some analyses from which results did not change substantially by including more subjects. For a detailed description and explanation of the analyses please refer to deliverable D-2-3.

3.1 Subjects included

Data inclusion for the multi-centre study was closed by the end of March 2008. By that time the following numbers of normally-hearing and hearing-impaired subjects were included at the five centres:

#	**AMC**	**HZO**	**ISVR**	**LINK**	**VUMC**	**Total**
NH	5	5	10	5	5	30
HI	15	15	12	15	15	72
Total	20	20	22	20	20	102

Table 1: Numbers of normally-hearing and hearing-impaired listeners included in each centre.

In fact, at ISVR 20 normally-hearing subjects were measured, but data from 10 of them were used for the reference study described above and excluded from the main analyses. As can be seen, we included a few more normally-hearing (30 instead of 25) and a few less hearing-impaired listeners (72 instead of 75) than planned. Considering the four audiometric categories, we achieved the goals we set:

category	**# subjects included**	**target**
mild & flat	21 (29%)	25%
severe & flat	13 (18%)	25%
mild & sloping	21 (29%)	25%
severe & sloping	17 (23%)	25%

Table 2: Numbers of subjects (based on analysis of better ears) in the four categories. For a description of the audiometric criteria of the categories please refer to deliverable D-2-3.

It proved to be difficult to find enough subjects with asymmetrical and conductive/mixed hearing losses: we included 12 subjects (17%) with asymmetric hearing loss (target: 20%) and 60 with symmetric osses.

Finally, 9 subjects (12,5%) had a conductive or mixed hearing loss (target: 17%) and 63 subjects had a purely perceptive hearing loss. Criteria for symmetric/asymmetric and perceptive/conductive hearing losses are presented in deliverable D-2-3.

3.2 Effects of test-retest and learning

We investigated the test-retest reliability of the tests by calculating the intra-class correlation coefficient (ICC) for the total group, hearing-impaired listeners and normal-hearing listeners (see also D-2-3):

$$ICC = SD\ (between)^2 / (SD\ (between)^2 + SD\ (within)^2)$$

			ICCtotal	ICCnh	ICChi	Sdwithin
FT (dB)	spectral resolution at 500 Hz	F500	0.709	0.466	0.732	2.204
	spectral resolution at 3000 Hz	F3000	0.925	0.796	0.854	1.628
	temporal resolution at 500 Hz	T500	0.748	0.735	0.704	3.872
	temporal resolution at 3000 Hz	T3000	0.868	0.823	0.568	2.697
ListEff (%)	mean listening effort	Eff	0.897	0.771	0.863	5.107
	eff. in stationary noise, SNR+5	EffC5	0.854	0.724	0.854	8.557
	eff. in stationary noise, SNR-5	EffCmin5	0.653	0.558	0.610	6.333
	eff. in fluctuating noise, SNR+5	EffF5	0.875	0.735	0.854	7.929
	eff. In stationary noise, SNR-5	EffFmin5	0.869	0.784	0.776	7.307
Gothenburg profile (%)	speech perception	speech	0.942	0.549	0.903	6.754
	spatial hearing	loc	0.950	0.726	0.934	5.236
	social interactions	social	0.924	0.754	0.904	6.823
	behaviour and reactions	behav	0.952	0.653	0.951	5.406
Acalos (dB)	MCL at 500 Hz	MCL500	0.821	0.526	0.828	5.394
	MCL at 3000 Hz	MCL3000	0.923	0.603	0.901	3.935
	MCL for broadband noise	MCLbb	0.856	0.617	0.852	5.022
(cu/dB)	slope of curve at 500 Hz	SL500	0.826	0.288	0.828	0.067
	slope of curve at 3000 Hz	SL3000	0.854	0.246	0.817	0.090
	slope of curve for bb noise	SLbb	0.762	0.551	0.773	0.096
Plomp (dB)	SRT in quiet	SRTquiet	0.992	0.587	0.985	1.560
	SRT in stationary noise	SRTstat	0.888	0.600	0.847	1.322
	SRT in fluctuating noise	SRTfluct	0.952	0.716	0.892	1.664
binaural (dB)	ILD (release of masking)	ILD	0.849	0.579	0.795	1.052
	BILD (release of masking)	BILD	0.708	0.787	0.627	1.549
LexDec	lexical decision making test	LexDec	0.850	0.764	0.812	0.097
minimum audible angle (deg)	MAA broadband noise	MAAbb	0.863	0.662	0.873	2.969
	MAA high-pass noise	MAAhp	0.896	0.846	0.900	3.082
	MAA low-pass nosie	MAAlp	0.856	0.658	0.872	3.315

Table 3: Intraclass correlation coefficients for the total group (ICCtotal), hearing-impaired listeners (ICCHI) and normal-hearing listeners (ICCNH) and within-subject standard deviations (SDw).

🟩	> 0.9	excellent
🟩	0.8 - 0.9	good
🟨	0.7 - 0.8	
🟧	0.6 - 0.7	moderate
🟧	0.5 - 0.6	
🟥	< 0.5	poor

ICCs of all outcome measures (except audiogram) for the total group (ICCtotal), for hearing-impaired listeners (ICCHI) and for normal-hearing listeners (ICCNH) are shown in table 4. ICCs are calculated of left-ear variables (if applicable), and corrected values (concerning language-dependent tests). In the calculation of between-subject standard deviations only test values (no retest) were used.

Overall, results are similar to those presented in deliverable D-2-3. Again, no clinically relevant learning effects were found: for each test, the difference between the averaged (over subjects) test and retest outcomes was much smaller than the within-subject standard deviation. One of the major differences compared to the preliminary analysis is that the test-retest variability of the temporal resolution measures (T500 and T3000) proves to be a bit better now.

Contrary to what we did in the preliminary analyses we now decided to use test values rather than averages of test and retest in further analyses, because this is more clinically relevant. Of course we used the information about effects of test-retest and learning when deciding about including or excluding parameters in the final auditory profile.

3.3 Analyses of per-ear variables

Before we analysed the relations between per-ear variables, we investigated their distributions to see if they deviated from normal distributions. We did this by performing Kolmogorov-Smirnov tests and by visual inspection. We found that all variables except air-bone gap (ABG) were distributed approximately normally. In the present analyses, we transformed the ABGs using a Blom transformation (Blom, G. 1958. Statistical estimates and transformed beta variables. New York: John Wiley and Sons). In future work we will analyse possible differences between subjects with sensorineural and conductive hearing losses. Table 4 shows relations between per-ear measurements (of the better ear of each subject) in the group of hearing-impaired listeners by means of Pearson’s correlation coefficients. Each cell in the table shows Pearson’s correlation value (r, first line) and its significance (p, second line). For an explanation of variable abbreviations, see table 3. Moreover, also audiogram variables are added, i.e. air-condition thresholds (AC) at 500 and 3000 Hz, pure-tone averages of low (500,1000,2000 Hz, ‘pTAl’) and high (1000,2000,4000 Hz, ‘PTAh’), slope of the audiogram (threshold difference between 500 and 4000 Hz) and air-bone gap (ABG, average of ABG at 500 and 1000 Hz).

		Audiogram						Acalos						FT			
		AC500	AC3000	PTAl	PTAh	slope	nABG	MCL500	MCL3000	MCLbb	SL500	SL3000	SLbb	F500	T500	F3000	T3000
Audiogram	AC500	1.000															
	AC3000	0.436 0.000	1.000														
	PTAl	0.904 0.000	0.607 0.000	1.000													
	PTAh	0.704 0.000	0.821 0.000	0.903 0.000	1.000												
	slope	-0.599 0.000	0.342 0.003	-0.406 0.000	0.025 0.832	1.000											
	nABG	0.225 0.057	-0.078 0.514	0.096 0.421	-0.017 0.887	-0.263 0.026	1.000										
Acalos	MCL500	0.608 0.000	0.184 0.122	0.553 0.000	0.419 0.000	-0.418 0.000	0.245 0.038	1.000									
	MCL3000	0.478 0.000	0.628 0.000	0.555 0.000	0.638 0.000	0.057 0.636	0.105 0.382	0.592 0.000	1.000								
	MCLbb	0.608 0.000	0.382 0.001	0.663 0.000	0.614 0.000	-0.249 0.036	0.098 0.415	0.649 0.000	0.561 0.000	1.000							
	SL500	0.501 0.000	0.156 0.190	0.523 0.000	0.401 0.000	-0.354 0.002	-0.148 0.216	-0.032 0.790	-0.034 0.779	0.294 0.013	1.000						
	SL3000	0.009 0.942	0.421 0.000	0.194 0.103	0.299 0.011	0.198 0.096	-0.144 0.229	-0.244 0.039	0.009 0.938	-0.116 0.333	0.225 0.058	1.000					
	SLbb	0.395 0.001	0.147 0.218	0.353 0.002	0.266 0.024	-0.237 0.045	-0.064 0.592	0.101 0.401	0.020 0.865	-0.074 0.537	0.510 0.000	0.206 0.082	1.000				
FT	F500	0.301 0.012	0.209 0.084	0.370 0.002	0.347 0.003	-0.107 0.382	0.041 0.740	0.119 0.330	0.260 0.031	0.365 0.002	0.338 0.004	0.056 0.645	0.069 0.573	1.000			
	T500	0.331 0.006	0.148 0.224	0.236 0.051	0.149 0.221	-0.216 0.075	0.087 0.477	0.133 0.277	0.195 0.108	0.256 0.035	0.228 0.059	-0.103 0.401	0.172 0.157	0.371 0.002	1.000		
	F3000	-0.188 0.121	0.426 0.000	-0.053 0.665	0.164 0.178	0.477 0.000	-0.345 0.004	-0.127 0.297	0.313 0.009	0.041 0.742	-0.117 0.338	0.230 0.058	-0.124 0.311	0.104 0.395	-0.014 0.909	1.000	
	T3000	-0.122 0.319	0.251 0.038	-0.021 0.863	0.109 0.372	0.290 0.016	-0.242 0.045	-0.292 0.015	-0.078 0.526	-0.099 0.421	-0.016 0.894	0.210 0.083	-0.065 0.598	0.070 0.568	-0.046 0.708	0.432 0.000	1.000
SRT	SRTstat	0.231 0.051	0.588 0.000	0.324 0.006	0.515 0.000	0.349 0.003	0.070 0.558	0.020 0.870	0.372 0.001	0.257 0.030	0.034 0.775	0.224 0.058	0.148 0.214	0.373 0.002	0.250 0.038	0.262 0.030	0.155 0.205
	SRTfluct	0.372 0.001	0.608 0.000	0.490 0.000	0.623 0.000	0.194 0.102	-0.037 0.757	0.010 0.936	0.332 0.004	0.285 0.016	0.239 0.043	0.297 0.011	0.229 0.053	0.503 0.000	0.238 0.049	0.262 0.030	0.332 0.005

Table 4: Correlations between per-ear measurements for HI. Each cell displays Pearson's correlation coefficient with its significance (p). Significant correlations are marked green ($p<0.01$) and yellow ($p<0.05$).

		Acalos						FT				SRT	
		MCL500	MCL3000	MCLbb	SL500	SL3000	SLbb	F500	T500	F3000	T3000	SRTstat	SRTfluct
Acalos	MCL500	1.000 .											
	MCL3000	0.464 0.000	1.000 .										
	MCLbb	0.546 0.000	0.278 0.020	1.000 .									
	SL500	-0.240 0.044	-0.411 0.000	0.066 0.588	1.000 .								
	SL3000	-0.426 0.000	-0.247 0.038	-0.399 0.001	0.120 0.320	1.000 .							
	SLbb	-0.012 0.918	-0.201 0.093	-0.313 0.008	0.457 0.000	0.138 0.252	1.000 .						
FT	F500	-0.031 0.802	0.053 0.670	0.205 0.097	0.231 0.058	-0.053 0.667	-0.026 0.834	1.000 .					
	T500	0.078 0.525	0.131 0.285	0.211 0.087	0.186 0.129	-0.156 0.203	0.139 0.258	0.344 0.004	1.000 .				
	F3000	-0.219 0.073	0.274 0.024	-0.077 0.534	-0.202 0.098	0.192 0.117	-0.176 0.151	0.051 0.681	-0.039 0.749	1.000 .			
	T3000	-0.374 0.002	-0.192 0.116	-0.212 0.085	-0.066 0.593	0.187 0.127	-0.098 0.428	0.034 0.781	-0.063 0.609	0.422 0.000	1.000 .		
SRT	SRTstat	-0.251 0.034	0.066 0.583	-0.087 0.473	-0.219 0.066	0.086 0.476	0.014 0.908	0.241 0.048	0.205 0.094	0.210 0.086	0.116 0.348	1.000 .	
	SRTfluct	-0.354 0.002	-0.110 0.362	-0.159 0.188	-0.015 0.902	0.148 0.218	0.084 0.486	0.391 0.001	0.188 0.125	0.207 0.090	0.340 0.005	0.751 0.000	1.000 .

Table 5:Partial correlations between per-ear measurements for HI (control variable: PTA(1,2,4)). Same layout as table 4.

Clearly, there is a lot of correspondence between the audiogram and the other tests (left columns). The bottom rows show that there are also several significant correlations between the SRTs (in stationary and especially in fluctuating noise) and the other per-ear tests (Acalos, FT test), besides the dependence on the audiogram. However, if we calculate partial correlations between per-ear variables, controlled for audibility (PTA(1,2,4)) not all correlations remain significant, as can be seen in table 5. However, even after cancelling out audibility, SRTs (both in stationary and fluctuating noise) correlate significantly with MCL and spectral resolution at 500 Hz. As expected, temporal resolution at 3000 Hz proves to be more important for SRT in fluctuating noise than for SRT in stationary noise: only the correlation with SRT in fluctuating noise is significant.

We also repeated the regression analyses like in deliverable D-2-3 to predict SRTs in noise from other per-ear variables (refer to D-2-3 for details about this analysis). Here we only present results for the group of hearing-impaired listeners, see table 6. These results reflect the same effects as the correlations: besides audibility there are other factors that predict SRT, especially in fluctuating noise. In general, SRT in fluctuating noise is better predictable than SRT in stationary noise (also for subgroups of listeners, data not shown). Besides the per-ear variables, we also included ‘Lexical Decision’ (lexdec) as an independent variable, since it might be expected to play a role in speech perception in noise.. However, it was found that lexdec was not a predictor for SRTs in stationary and fluctuating noise.

Dependent	Predictors	Beta	R^2 (model)
SRTstat	AC3000	0.534	0.411
	F500	0.261	
SRTfluct	AC3000	0.376	0.618
	F500	0.403	
	age	0.233	
	T3000	0.191	
	SLbb	0.177	

Table 6: Prediction of the SRT in noise by regression analysis. See text for details and table 3 for explanations of abbreviations.

Results of the rotated factor-analyses for the full group of hearing-impaired listeners are presented in table 7. See deliverable D-2-3 for details about the analysis. As can be seen, results are very similar to those presented in D-2-3.

Based on all per-ear analyses we decided to include the following variables in further analyses:

- Audiogram: threshold at 3 kHz and slope
- FT test: F500 and T3000
- Acalos: MCL500 and SL500
- Speech: SRTstat and SRTfluct

Compared to the selected parameters in deliverable D-2-3 F3000 has been replaced by T3000. This is done because T3000 proved to be more important for speech perception, despite its poorer test-retest reliability.

	Component			
	1	2	3	4
slopeb	**0.779**			
F3000b	**0.746**			
T3000b	0.612			
MCL3000b		**0.884**		
MCL500b	-0.437	**0.789**		
MCLbbb		**0.729**		0.414
SLbbb			**0.829**	
SL500b			**0.774**	
SL3000b	0.542		0.492	
F500b				**0.811**
T500b				**0.701**
AC3000b	0.609	0.624		
nABG	-0.460			
SRTfluctb	0.525			0.502
AC500b		0.636	0.535	
Explained variance:	20.369	19.822	14.099	12.810
Interpretation:	high-freq processing	audibility	recruitment	low-freq processing

Table 7: Results of factor analysis of per-ear variables: rotated component matrix. Factor loadings below 0.4 are suppressed; values above 0.7 are printed bold. Same abbreviations as in table 3.

3.4 Analyses of per-subject parameters

From here on, we will call variables that are not measured per ear ‘per subject’ instead of ‘binaural’, because age and lexical decision are clearly no ‘binaural’ parameters. We started this analysis with an exploration of the distributions. We tested normality both by Kolmogorov-Smirnov tests and by visual inspection. We found that all variables except the MAA variables were distributed approximately normally. Consequently, we transformed the MAA variables using Blom (adding ‘n’ to the variable names).

We recalculated the correlations between per-subject parameters for the hearing-impaired listeners (data not shown). Correlations are a little different than in deliverable D-2-3, due to e.g. use of test (vs. mean of test and retest) values, including more subjects, and using corrected (B)ILD values.

	Component		
	1	2	3
age			**0.811**
LexDec			**-0.805**
SRTq		**0.715**	
ILDb		**0.725**	
BILDb		0.691	0.442
nMAAbb	**0.856**		
nMAAlp	**0.862**		
nMAAhp	**0.846**		
Explained			
variance:	29.790	20.584	19.631
interpretation:	MAA	speech	age/cogn.

Table 8: Rotated component matrix of the factor analysis of 'per-subject' parameters. Same layout as table 7. The determinant of the correlation is 0.071, KMO (sampling adequacy) = 0.627 and Bartlett's test significance = 0.000. The total variance explained by these 3 factors is 70 %.

A major difference compared to the results shown in D-2-3 is a significant correlation between age and lexical decision (r=-0.432, p<0.001). However, the overall results are very similar and are also reflected in the rotated component matrix of the factor analysis (see table 8). The factor analysis was performed in the same way as in deliverable D-2-3 and shows again three groups of variables clustering MAA, speech tests and age. New compared to the results in D-2-3, is that the third factor now also incorporates lexical decision, like we expected from the correlation between age and lexical decision.

Based on the analyses of the per-subject parameters we decided to include age, lexical decision, ILD and MAAbb in further analyses (no changes compared to D-2-3)

3.5 Analyses of communication-performance parameters

First we investigated the normality of the distributions of the communication-performance parameters by Kolmogorov-Smirnov tests and visual inspection. It was decided to transform the Gothenburg-Profile variables because of their skewed distributions. Distributions of the listening-effort results were approximately normal.

Next, Pearson's correlations were calculated between the communication-performance parameters and the selected 'per-ear' and 'per-subject' variables. These correlations are shown in table 9. As can be seen, results are similar to those presented in deliverable D-2-5 and yield the same conclusions: there is little correspondence between the communication-performance parameters and the per-ear and 'binaural' tests, except for the speech-perception tests. Listening effort in fluctuating noise shows more correspondence with other tests than listening effort in continuous noise, and the speech perception and spatial hearing subscales of the Gothenburg Profile are better predictable than the other two subscales.

Similar conclusions can be drawn from the regression analyses (data not shown) in which we tried to predict communication performance from 'per-ear' and 'per-subjects´ parameters: SRTs are the most important predictors; effort in fluctuating noise is better predictable than speech in stationary noise and from the Gothenburg profile the speech perception and spatial hearing subscales are best predictable.

		Gothenburg Profile				Listening Effort			
		nspeech	nloc	nbehav	nsocial	EffC5	EffCmin5	EffF5	EffFmin5
Audiogram	AC3000b	0.429	0.307	0.165	0.221	0.346	0.126	0.411	0.467
		0.000	0.009	0.166	0.062	0.003	0.290	0.000	0.000
	AC3000dif	-0.006	0.176	0.016	0.004	-0.047	-0.054	-0.139	-0.197
		0.959	0.139	0.896	0.973	0.693	0.652	0.244	0.097
	slopeb	-0.112	-0.135	-0.080	-0.014	0.071	0.053	-0.021	-0.032
		0.350	0.259	0.506	0.905	0.551	0.656	0.864	0.787
	slopedif	0.004	0.054	-0.136	-0.130	0.030	0.011	-0.084	-0.118
		0.976	0.653	0.254	0.276	0.800	0.925	0.484	0.322
Acalos	MCL500b	0.376	0.269	0.144	0.223	0.073	0.071	0.239	0.248
		0.001	0.023	0.228	0.060	0.541	0.555	0.043	0.036
	MCL500dif	0.018	0.054	0.153	0.016	0.227	-0.078	0.261	0.045
		0.883	0.655	0.201	0.897	0.055	0.515	0.027	0.708
	SL500b	0.330	0.353	0.167	0.039	0.013	-0.090	0.156	0.196
		0.005	0.002	0.161	0.748	0.913	0.454	0.191	0.099
	SL500dif	0.018	0.072	0.209	0.241	0.059	0.079	0.114	0.141
		0.884	0.547	0.078	0.041	0.625	0.512	0.339	0.236
Fttest	F500b	0.379	0.293	0.216	0.259	0.078	0.126	0.095	0.167
		0.001	0.014	0.075	0.032	0.522	0.304	0.439	0.169
	F500dif	-0.051	0.021	0.075	-0.030	0.119	0.049	0.137	0.127
		0.682	0.862	0.543	0.805	0.335	0.691	0.265	0.301
	T3000b	0.161	0.044	-0.050	0.055	-0.134	0.113	-0.159	0.101
		0.186	0.722	0.685	0.652	0.273	0.355	0.192	0.411
	T3000dif	-0.011	0.117	0.050	-0.012	0.000	0.017	0.051	-0.106
		0.930	0.346	0.689	0.921	0.999	0.889	0.682	0.394
Speech	SRTstatb	0.361	0.247	0.377	0.347	0.395	0.304	0.386	0.354
		0.002	0.036	0.001	0.003	0.001	0.009	0.001	0.002
	SRTstatdif	0.058	0.137	-0.027	0.033	0.125	0.025	0.149	0.064
		0.629	0.252	0.820	0.782	0.297	0.836	0.210	0.592
	SRTfluctb	0.439	0.358	0.347	0.384	0.424	0.252	0.411	0.442
		0.000	0.002	0.003	0.001	0.000	0.033	0.000	0.000
	SRTfluctdif	0.084	0.214	0.057	0.042	-0.115	0.035	0.020	0.094
		0.484	0.071	0.636	0.725	0.334	0.773	0.868	0.434
age	age	-0.054	-0.213	0.062	0.173	0.239	-0.118	0.128	-0.058
		0.650	0.072	0.606	0.146	0.043	0.325	0.284	0.626
Speech binaural	SRTq	0.626	0.581	0.379	0.339	0.192	0.142	0.457	0.503
		0.000	0.000	0.001	0.004	0.109	0.239	0.000	0.000
	ILDb	0.266	0.435	0.195	0.192	0.219	0.048	0.272	0.337
		0.024	0.000	0.101	0.107	0.065	0.690	0.021	0.004
	ILDdif	0.099	-0.177	0.014	0.084	0.124	0.025	0.073	0.028
		0.407	0.137	0.906	0.483	0.301	0.836	0.542	0.816
MAA	nMAAbb	0.152	0.136	0.182	0.204	0.035	0.081	0.051	-0.007
		0.204	0.259	0.130	0.087	0.771	0.502	0.673	0.957
LexDec	LexDec	-0.046	0.159	0.007	0.041	0.177	0.192	0.029	0.121
		0.704	0.181	0.953	0.730	0.136	0.107	0.811	0.311

Table 9: Correlations between selected per-ear and selected 'binaural' measures, and communication-performance measures in hearing-impaired listeners. Same layout as table 4.

4 The clinical relevance

4.1 For auditory diagnosis

In discussions about the clinical relevance for auditory diagnosis, we met at least some doubt about the relevance of all these tests. The reason is that we do not know:

- How critical the problems not directly related to audibility are
- How problems other than audibility can be identified
- How often these complex problems occur

This is not commonly known, because thus far a systematic approach such as will be facilitated by the Auditory Profile has not been available in the clinic. In a way, it is a “chicken and egg” problem.

The Auditory Profile will produce large data sets, even if the AP will only be used by a small percentage of the clinics in the countries of the language areas covered by the Auditory Profile (thus far). One of the big advantages of an AP is that such data will become available in the near future.

4.2 For hearing aid selection and hearing aid fitting

For application of the AP in rehabilitative Audiology, an important question is whether the AP can be used for a classification of hearing impairment in a way that there are consequences for the choices to be made during auditory rehabilitation. Although this type of application needs to be investigated in a new validation study, it may be worthwhile to start with some “well-educated” guesses, party based on clinical experience and partly on the outcomes with the preliminary version of the AP.

The AP may be expected to be able to discriminate “mainstream” subjects that have effects of loss of audibility, loss of discrimination etc. that can be predicted from the audiogram and subjects that are outside the mainstream, e.g. abnormally great loss of discrimination, abnormally reduced dynamic range, abnormally reduced frequency resolution, and/or temporal resolution. In other subjects, problems in binaural integration or problems with cognition may appear. In principle, the AP may be expected to support rehabilitative audiology with information about supra-threshold deficits, binaural processing, and cognition. However, more research in this field is needed before strong claims can be made.

5 Classification of subjects in the test population

If classification of hearing problems is important for the application of the AP in auditory rehabilitation, it may be considered to change the structure of the AP into a hierarchical structure of tests with a decision-tree structure.

In the consensus meeting in Leuven, we hypothesized that there are well-described categories of hearing-impaired listeners. In future work we will conduct a cluster analysis on the data of the first multi-centre study to find evidence that these categories exist. In a certain percentage of the subjects with a hearing impairment, audibility, associated with mild supra-threshold processing deficits, is the main problem. In others, excessive problems in supra-threshold processing exist. A third group of subjects are subjects with reduced binaural integration. Finally, a fourth group of subjects are subjects with reduced cognitive functions. The auditory profile will allow us to distinguish between the above-mentioned groups.

6 A hierarchy of test items in the AP

Based on the outcomes of the first multi-center study, a first draft for a hierarchical approach has been designed. In this paragraph, we indicate how such a hierarchical approach could work.

6.1 Phase 1: the complexity of the auditory disabilities

The goal of phase 1 is to determine if the hearing-impaired client belongs to the category “mainstream”, indicating that their supra-threshold deficits can be predicted from the pure-tone audiogram. Or, alternatively, that the hearing loss is more complex and should be tested in more detail. Phase 1 testing comprises:

- Pure-tone audiogram with adequate masking
- SRT in quiet
- SRT in fluctuating noise
- GP speech and localization

From the research conducted in multi-center study 1 we will derive a set of criteria that discriminate between “mainstream” and “complex”. Such criteria can be derived from model predictions (e.g. the SII-model of Rhebergen, Versfeld, and Dreschler[1]), given the pure-tone audiogram.

The most important criteria are:

- SRTfluct is poorer than expected based on SII modeling
- GP speech poorer than expected from the pure-tone hearing loss and from the SRT in quiet
- GP localization poorer than ..., etcetera

---

[1] Rhebergen KS, Versfeld NJ and Dreschler WA (2006).
“Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise,” J. Acoust. Soc. Am. 106 (6), 3988-3997

One of the goals of a renewed analysis of the results of the multi-centre study is to investigate whether such discrimination in classes is feasible and what criteria are most discriminative.

6.2 Phase 2: the origin of the complexity

The tests to be conducted in phase 2 have to depend on the outcome in phase 1. We have to design different scenarios of testing and for illustration some examples will be given here:

In case that the SRT in noise is poorer than expected, it may be worthwhile to start with the combined spectral and temporal resolution test (Larsby and Arlinger, 1998) in order to investigate the possible role of (reduced) frequency resolution or temporal acuity as an explanation of the poor speech results. A next step can be the analysis of the loudness coding by means of the Acalos test. If no explanation can be found from these psychophysical tests, the cognitive function may be investigated by means of the lexical decision test or psychological testing/therapy.

In case that the speech perception is subjectively poor relative to the SRT outcomes, a likely approach is to start with testing the binaural co-operation by means of ILD/BILD tests. If a reduced binaural integration is not the likely cause, again the cognitive function may be investigated by means of the lexical decision test.

In case of subjective complaints with localization, without an adequate explanation from the asymmetry in the pure-tone audiogram, we propose performing further analyses by means of MAA-tests and ILD/BILD testing.

6.3 Phase 3: detailed diagnosis

Phase 3 is devoted to detailed testing in a specific area. These tests can be applied if the analysis in phase 2 points in a certain direction, but if additional information is required. Also, phase 3 can be applied for specific groups, e.g.

- subjects with second language problems
- subjects with asymmetric hearing losses

7 Final set to be validated in MCS-2

At the last HearCom meeting in Leuven, it was decided to shorten the Auditory Profile (AP) based on the analysis of the first study (AP1) and our strategic approach to developing the AP for future dissemination and exploitation. Essentially, AP1 takes too long to perform (90 minutes plus rest breaks according to the initial timings, but probably longer in practice) and some of the measures are redundant because they correlate highly with one another. Furthermore, there were some unsatisfactory features of the AP1 tests (e.g. the temporal resolution measure at 3 kHz appears to demonstrate a “floor” effect such that hearing-impaired participants score similarly; the listening effort measures of normal hearing and hearing-impaired participants overlap more than expected – maybe due to adjusted expectations; unexplained language differences in the Gothenburg Profile possibly due to semantic, cultural and attitude differences between countries).

The intention is to decide on AP2 and then run a further multicentre study across the five test centres. The further study would be necessary even if the AP were unaltered, because the analysis of AP1 included a multitude of planned comparisons and hence has very limited statistical power. The further study will enable us to test hypotheses that have been developed from AP1.

Table 1 compares the components of AP1 and proposed AP2, with brief reasons for the changes.

Test	**AP1**	**AP2**	**Comments**
Audibility (pure-tone thresholds: a-c 0.25, 0.5, 1, 2, 3,4, 6, 8 kHz; b-c 0.25, 0.5, 1, 2, 3kHz)	Yes	Yes	Assumed to be measured clinically in all cases before AP
Loudness perception (ACALOS 0.5, 3 kHz, speech-shaped noise)	Yes	Yes	Unchanged
Frequency and temporal resolution (F-T test at 0.5 and 3 kHz)	Yes	Modified	The modified version should avoid floor effect.
Speech perception (everyday sentences diotically in quiet, stationary and fluctuating noise)	Yes	Modified	A fixed noise level is considered rather than a level based on MCL
Binaural processing (combination of ILD and BILD tests using Matrix sentence materials)	Yes	Modified	A fixed noise level is considered rather than a level based on MCL
Localisation (Minimum Audible Angle test: low-pass, high-pass and broadband)	Yes	No	Removed due to long test times. The test is available for detailed diagnosis

Self-reported hearing ability (Gothenburg profile)	Yes	Still in discussion	Possibly semantic or cultural difficulties, but the subjective information is important
Listening effort (HearCom listening effort test)	Yes	No	Removed due possible effects of expectations and subjectivity of response scale
Cognitive ability (Lexical decision test)	Yes	Yes	Although not correlated with other variables in AP1, retained in principle (unchanged). This test can be used for detailed diagnosis

Table 1 Comparison of AP1 and proposed AP2

Test duration and replications

The durations of the proposed AP2 tests are shown in Table 2. These times are the nominal durations from the AP1 protocol, which had a total duration of 90 minutes. For AP1, all tests were repeated to assess the test-retest reliability of the tests. We need to consider carefully whether it is necessary to include repeats for AP2, bearing in mind the resource implications. The strategy proposed here is to replicate in the second multi-center study only the tests that have changed.

Test (min)	**Measurement time (minutes)**
ACALOS	10
Speech perception	15
ILD/BILD	15
F-T test	15
Lexical decision making test	5
Total	**60**

Table 2 Duration of testing for AP2

For some tests we need more information about the level dependency. There is some evidence to suggest that speech-in-noise tests are not affected greatly by presentation level. We are not aware of evidence for ILD/BILD being affected by presentation level, but we would suspect not. However, there is a very strong possibility that the F-T test will be affected by presentation level. Therefore, it is proposed that the F-T test is repeated at different presentation levels (two test runs in total). This will increase the nominal test time by 15 minutes for the AP2 study (but not for clinical use of the profile).

8 Conclusions

The work described in this deliverable comprises detailed analyses of the results of the first multi-centre study. Clear pattern can be identified, which indicate that in a number of subjects auditory disabilities are not only determined by reduced audibility.

Systematic differences in the results for different languages have been analysed and are corrected for by aid of the results of extra reference testing. Nevertheless, not all questions can be resolved on the basis of this large dataset. This applies in particular for the applicability of the auditory profile for hearing aid selection and hearing aid fitting. In parallel to the next phase of an experimental evaluation of the auditory profile in a second multi-centre study, some aspects of this issue will be investigated for noise reduction, to be evaluated in WP7.

A reduction of testing time, necessary for clinical applicability, has been realized. Also, we formulated some examples of a hierarchical approach that will allow an even more drastic reduction of testing time for the subjects in which the disabilities are mainly due to reduced audibility.

The possible improvements may be expected to be favourable for the clinical applicability of the Auditory Profile and thus for its acceptance by clinicians in different countries.