FP6–004171 HEARCOM

Hearing in the Communication Society

INTEGRATED PROJECT

Information Society Technologies

D-2-8:

Evaluation of the Auditory Profile in relation to auditory rehabilitation

Contractual Date of Delivery:	M53 (+45 days)
Actual Date of Submission:	April 8, 2009
Editor:	Wouter A. Dreschler
Sub-Project/Work-Package:	SP1/WP2
Version:	2.0
Total number of pages:	41

Dissemination Level		
PO	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Deliverable D-2-8

VERSION DETAILS	
Version:	2.0
Date:	April 8th, 2009
Status:	Final, after review

CONTRIBUTOR(S) to DELIVERABLE

Partner	Name
NL-AMC	Rolph Houben
NL-AMC	Wouter A. Dreschler
DE-HZO	Matthias Vormann,
DE-HZO	Michael Schulte

DOCUMENT HISTORY

Version	Date	Responsible	Description
0.1		Rolph Houben	First Draft and Outline
0.2		Wouter Dreschler	First full version
1.0		Wouter Dreschler	Submitted for review
2.0	April 8, 2009	Wouter Dreschler	Final version based on the reviewer's suggestions

DELIVERABLE REVIEW

Version	Date	Reviewed by	Conclusion*
1.0	April 6, 2009	Prof. Tammo Houtgast	Accept
1.0	April 8, 2009	Prof. Norbert Dillier	Accept

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

Table of Contents

List of Figures

List of Tables

Acknowledgement

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

Pre-Amble

This deliverable summarizes the work from work package 2, task 4 in which the use of the HearCom Auditory Profile (developed in work package 2) for evaluating the possible benefit of newly developed hearing-aid algorithms (developed and tested in work package 5 and 7) is investigated.

In work package 2, the Auditory Profile was developed and evaluated for its applicability to audiological diagnostics. Since this Auditory Profile was shown to be powerful in describing hearing deficits, it is expected that it can be a useful tool in auditory rehabilitation as well. For instance, the profile could be of use for hearing aid evaluation, hearing aid selection, and even for hearing aid fitting.

In HearCom part SP3, work package 5, several noise reduction schemes have been developed that have been perceptually evaluated in work package 7. It is unknown whether there is a relationship between the efficacy of the WP7 noise reduction algorithms and the hearing loss of the subjects. The perceptual data presented in deliverable D-7-5 focussed on group effects for three groups that were based on the audiogram configuration only.

In the current deliverable the results are linked to a more in depth description of the subjects' hearing-loss with their results measured and described according to the Auditory Profile tests. For this, a subset of the subjects that took part in the evaluation of Work Package 7 was re-invited and tested with the test battery of the Auditory Profile. The results from the WP 7 study have been re-analyzed for the subgroup of subjects participating to these additional tests. In addition, the individual relationships between the outcome measures have been investigated.

1 Executive Summary

The Auditory Profile (AP) is a diagnostic tool for assessing the problems of auditory functioning. The AP has been developed as a "complete" and non-redundant set of diagnostic procedures for the characterization of hearing deficiencies:

The aim of the AP is that it should be used as a diagnostic tool, for use in a (specialized) hearing centre or clinic or in audiological research, for a broad population of subjects with complaints about their performance in (auditory) communication tasks. The end users of the AP are the audiology professional interested in the characteristics of the hearing of a particular client/patient. The AP characterizes the individual's auditory impairment profile in a well-standardized way across Europe, and so extends and unifies hearing care in Europe. A broad application of the AP will reveal epidemiologic data about the incidence and prevalence of complex hearing deficits. A basic set of tests in the AP is intended to be conducted for every patient, while other tests can be thought of as additional tests that can clarify issues that arise during basic testing. The AP has been validated in two extensive multi-centre studies (> 100 subjects per study).

By factor analyses on the preliminary set of data we found a clustering of test results that indicate that hearing impairment is a multi-dimensional problem. The AP is a powerful means to analyse this multi-dimensionality. The implementation of the tests on a uniform software platform will facilitate clinical application. If the AP is able to estimate the problems that individual subjects will encounter in adverse communication situations, this work may stimulate a broad clinical acceptance of such a broad innovative approach to auditory testing.

In addition, some of the AP tests can help to determine the benefits from assistive devices. The results of the multi-centre studies show that the AP allows a detailed analysis of auditory disabilities by a very broad diagnosis of auditory deficits. In many subjects problems in the auditory communication are not only caused by reduced audibility, but also by a different loudness perception, reduced supra-threshold resolution, reduced binaural cooperation, or problems in cognition. It is worthwhile to assess the strength of contributing factors in individual subjects.

This deliverable investigates whether the WP2 Auditory Profile

- provides insight into the WP7 results on speech perception in noise, listening effort and preference rating in normal hearing and hearing impaired subjects using hearing aids with the noise suppression algorithms from SP2/ work package 5;
- has predictive power for the effect of the noise reduction algorithms in the tested subjects.

The Auditory Profile comprises a unique set of (internationally) standardised and validated tests that go beyond the traditional pure-tone audiogram. The results obtained in this study (n=55) are well in agreement with the two earlier extensive HearCom validation studies. The test-centre effects disappeared after using the pre-defined language corrections for the language-dependent test items.

The individual outcome measures for the effects of different noise reduction schemes (SRTs, LES-values, and PR-scores) all reflect different aspects of change. They appear also to be related to different aspects in the Auditory Profile.

However, in the set of test algorithms and the test conditions selected, the correlation coefficients were - although significant - too small to predict the effects or make a selection on the basis of individual AP parameters in this population.

This study illustrates that the AP may be able to disclose some of the underlying causes for speech perception and/or improvements in speech perception due to signal-processing schemes, but further experimental work is necessary to prove the added value in the selection of signal-processing schemes and/or in the prediction of the perceptual benefit.

Introduction

1.1 Perceptual evaluation of HearCom noise reduction algorithms

In HearCom SP3 / work package 5 five noise reduction algorithms have been developed (see confidential deliverable D-5-2, "Effectiveness of signal enhancement techniques for various auditory profiles") that have been perceptually evaluated in SP3 / work package 7. A report on this evaluation can be found in the confidential deliverable D-7-5, "Report on the evaluation of signal enhancement techniques".

In this section, the relevant results from deliverable D-7-5 will be summarized. The deliverable addressed the following research questions:

1. Do signal processing strategies lead to similar outcomes across different test-sites, taking into account the different test environments/rooms, test materials and evaluation test platforms?
2. Have the enhancement strategies different outcomes for different auditory profiles?
3. How do the results on speech reception threshold, listening effort scaling and preference rating assessment vary across the signal processing algorithms?

The noise reduction algorithms were evaluated in a multi centre study that took place at five test sites. At each of the test sites, 10 normal hearing subjects and 20 subjects with a moderate perceptual hearing loss (with average hearing thresholds between 35 and 60 dB) were invited to participate in this study. The subjects were divided into three groups: normal hearing, sloping hearing loss, flat hearing loss.

The algorithms used various approaches to enhance the desired signal (speech) and suppress unwanted background signals (noise). The following signal enhancement algorithms were evaluated: single channel noise suppression based on perceptually optimized spectral subtraction

(SC1), Wiener-filter-based single-channel noise suppression (SC2), broadband blind source separation based on second-order statistics (BSS), spatially pre-processed speech-distortion-weighted multi-channel Wiener filtering (MWF) and binaural coherence dereverberation filter (COH). Table 1 gives an overview of the algorithms used.

Table 1: algorithms used in the perceptual evaluation

Algorithm	Description	Number of mics used
SC1	single channel spectral subtraction	2 x 1 (double monaural)
SC2	single channel Wiener-filter	2 x 1 (double monaural)
BSS	broadband blind source separation based on second-order statistics	2 (binaural)
MWF	spatially pre-processed speech-distortion-weighted multi-channel Wiener filtering	2 x 3 (double monaural)
COH	coherence dereverberation filter	2 (binaural)

The three perceptual measures that were used in the evaluation were speech reception threshold tests (SRT), listening effort scaling (LES) and preference rating (PR).

For both SRT and LES values are available for the unprocessed condition as well as for each algorithm. Since the PR measurement consisted of a comparison between a processed and an unprocessed signal, only (relative) data is available for each algorithm. This data indicates the benefit (improvement or deterioration of the algorithm relative to no processing). The absolute SRT and LES values can also be expressed as a benefit relative to the unprocessed condition. This leads to the following outcome measures for the WP7 evaluations:

- absolute SRT values for all algorithms and the unprocessed condition;
- absolute LES scores for all algorithms and the unprocessed condition for five Signal-to-noise ratios (SNRs at -10, -5, 0, +5, and +10 dB);
- $\mathrm{SRT}_{\mathrm{improvement,}}$ this is 1 value for each algorithm;

- $\text{LES}_{\text{difference}}$, available for 5 SNRs for each algorithm;
- PR scores were available for three SNRs (0, +5, and +10 dB), for each algorithm.

The WP 7 results show that SRT improvements and PR scores for the different signal processing strategies were well in agreement across the different test sites. For LES, slightly different results were obtained across the test sites. However, general trends were similar, but the size of the effects was different. For the current investigation it is important that the test sites produced similar results because here we include data from two of these sites: NL-AMC and DE-HZO, see Chapter 3.

In all measurement conditions, speech was presented from the front of the listeners (azimuth 0°). Interfering noise was added in two different configurations. However the current deliverable uses only data that was measured in the condition with noise presented at 90°, 180° and 270° (and speech at 0°). As described by Luts et al. in report D-7-5, MWF was the only algorithm that provided a significant improvement in SRT, LES and PR relative to the unprocessed condition for this pseudo-diffuse noise scenario. These effects were found across different subject groups and test sites. We realize that the experimental test scenario is not optimal for the BSS algorithm, because this implementation is only suited when the background noise comes from one point-source. Unlike the lack of improvement in SRT for both single channel noise suppression techniques SC1 and SC2, and COH, PR outcomes indicated increased preference for these algorithms compared to the unprocessed condition at all tested signal-to-noise ratios. Additionally, an improvement in listening effort was observed at 0 dB SNR.

1.2 Auditory Profile tests

The HearCom Auditory Profile (further referred to as AP) was designed to enable a uniform and well-standardized characterization of an individual’s auditory impairment across Europe. The AP can be used to determine the individual hearing deficiencies in communication and can help to indicate the deficits that need to be compensated, either by signal processing or by alternative strategies.

A multi-centre study (D-2-5) showed that the AP allows a detailed analysis of auditory deficits. In many subjects who experience problems in auditory communication, this is not only caused by reduced audibility, but also by deficits such as a changed loudness perception, reduced supra-threshold resolution, reduced binaural integration, or cognitive problems.

In a second multi-centre study (D-2-7) the tests of the AP have been validated in a second group of subjects. The results showed a clustering of test results that indicate that hearing impairment is a multi-dimensional problem and the AP is a powerful means to analyze this multi-dimensionality (see also D-2-3).

This indicates that it is worthwhile to assess the degree to which different factors contribute to problems in the auditory communication for each individual listener. As such, the AP could also yield additional insight in the individual results from the WP 7 evaluation.

The tests that comprise the AP are listed in (public) deliverable D-2-5, "Optimized, final set of impairment tests included in the AP". Table 2 summarizes the AP tests that were used in the current study. A complete description of these tests can be found in public deliverable D-2-1 ("Implementation of a preliminary test set for auditory impairments") and public deliverable D-2-1b ("Demo version of the preliminary test set for auditory impairments").

Not all AP tests were used. In line with the inclusion criteria of the WP 7 study, the pure tone audiogram consisted of air conduction thresholds only (no bone conduction). In order to reduce measurement time it was decided to omit the minimal audible angle test because the test battery already contained two other spatial hearing tests, and because the minimal audible angle was regarded as the least informative of the tests in the test battery. Measurement time was reduced further by not replicating the effort scaling test from the AP, since the WP 7 data already contained data on the subjects' listening effort (LES test), even in much more detail.

Table 2: AP tests that were used in the current study (see public deliverable D-2-5, "Optimized, final set of impairment tests included in the AP" for more detail).

Domain	**Test**
Audibility	Pure tone audiogram Air conduction Bone conduction (not available)
Loudness perception	Acalos at 0,5 and 3 kHz
Frequency-time resolution	Combined F-T test
Speech perception in noise	SRT test in quiet and noise
Spatial hearing	Intelligibility Level Difference (ILD) Binaural Intelligibility Level Difference (BILD)
Subjective judgement on communication	Gothenburg Profile
Listening effort	Effort scaling for speech in noise (only available from WP 7 data, not from new AP measurements)
Cognitive abilities	Lexical decision making

2 Data collection

The current study was done at DE-HZO and at NL-AMC. DE-HZO conducted the WP 7 evaluation tests and WP AP tests in parallel: subjects had to revisit HZO two additional times for the AP measurements during the WP 7 measurements (April-June 2008). At NL-AMC the AP tests were done after all WP 7 testing was finished. The measurements for each subject were done in a single return visit (September-December 2008).

All AP tests were done at the experimental set-up that was described in deliverable D-2-2. Deliverable D-7-5 describes the materials and methods used in the perceptual evaluation. Data was analysed using MATLAB (Release 14), MS Excel (Office 2003), and SPSS (version 16.0).

2.1 Subjects

Three groups of subjects were selected: normal hearing subjects (NH), subjects with a flat moderate (between 35 and 60 dB HL) symmetrical perceptual hearing loss (HI-F), and subjects with a sloping moderate (between 35 and 60 dB HL) symmetrical perceptual hearing loss (HI-S). Detailed subject inclusion criteria are given in deliverable D-7-5. All subjects received a financial reimbursement for their travelling expenses and a small participation fee.

Since the AP measurements at HZO were done in parallel to the WP 7, data is available for every subject that completed the WP7 experiments. For AMC, all subjects were invited for a return visit. However, not all subjects were available for this second experiment. The cited reason was a lack of time. Table 3 shows the number of subjects that completed all the AP measurements. Subsequent analyses use complete data sets only. Since not all subjects participated in the AP experiment, the data set is not balanced, both with respect to the threshold groups (NH, HI-F, HI-S) and to test site (NL-AMC, DE-HZO).

Table 3: Number of subjects per test site.

	Number of subjects	
Subject group	**NL-AMC**	**DE-HZO**
normal hearing	6	10
flat hearing loss	9	10
sloping hearing loss	10	10

2.2 Perceptual evaluation data

One of the outcomes of the multi-centre WP 7 study was that the results were largely independent of measurement site. This implies that results based on the current data set (i.e. an incomplete data set of two centres, instead of four) might perhaps be more or less representative for the whole data set. In order to check if this assumption is valid, this section compares results of the subset to that of the complete set.

2.2.1 Speech reception threshold

A repeated-measures ANOVA on SRT with test-retest and algorithm as within variables and hearing loss group and test site as between variables showed the same trends as analyses on the complete set:

- SRT was better for NH subjects than for HI-F and HI-S.
- There was a significant difference between German-speaking and Dutch-speaking test sites. This is most likely due to difference in speech materials
- Retest scores were significantly better than test scores. This indicates a learning effect.
- SRT results were significantly different for different algorithms.

Another repeated-measures ANOVA on $SRT_{improvement}$ (i.e. SRT relative to the unprocessed condition) showed the same trends as analyses on the complete set as well:

- No significant effects were found for test-retest,

- There was no significant effect for test site (DE-HZO and NL-AMC)
- No significant effect of hearing loss group was found.
- Only the factor algorithm was still significant, as expected. SC1, SC2, and COH showed no significant effect on SRT. MWF showed an significant improvement in SRT of 6.4 +/- 0.2 dB (6.6 +/- 0.2 dB in total dataset). BSS showed a significant deterioration in SRT of 1.8 +/- 0.3 dB (1.9 +/- 0.3 dB in total dataset).

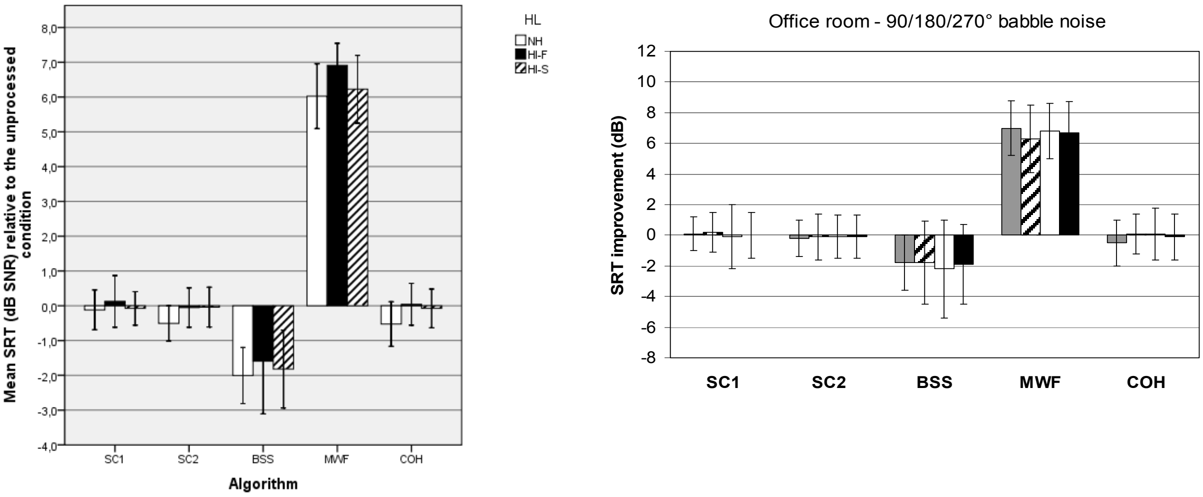

Figure 1: Mean SRT improvements relative to the unprocessed condition for the different hearing loss groups. The left panel shows averaged data of the subset of data for which the AP is measured. The right panel shows all data (Figure 5 of Deliverable D-7-5). More positive scores mean more benefit in SRT. Error bars denote standard deviation.

2.2.2 Listening Effort Scaling

For the complete data set, the effect of test site was significant at all signal-to-noise ratios except at -10 dB SNR (p <0.005). The Dutch-speaking subjects (measured at BE-LEU and NL-AMC) reported more listening effort than the German-speaking subjects (measured at DE-HZO and CH-UHZ), see Figure 2, left panel. The LES scores are in parallel with the SRT scores. Indicating that the effect of site is (at least partly) due to the difference in speech intelligibility of the subject groups at those sites. No difference was observed between both hearing-impaired subject groups for SRT as well as LES scores. Subjects from DE-HZO reported the lowest effort, those from NL-AMC the highest. This is fortunate, because these two centers thus provide the complete range of LES values for incorporation in our correlation analyses.

For the improvement in LES relative to unprocessed the sizes of the improvement was slightly different across test sites, with similar overall trends. The subset showed the same trends, and can thus be interpreted as being roughly indicative of the overall results, see Figure 2, right panel.

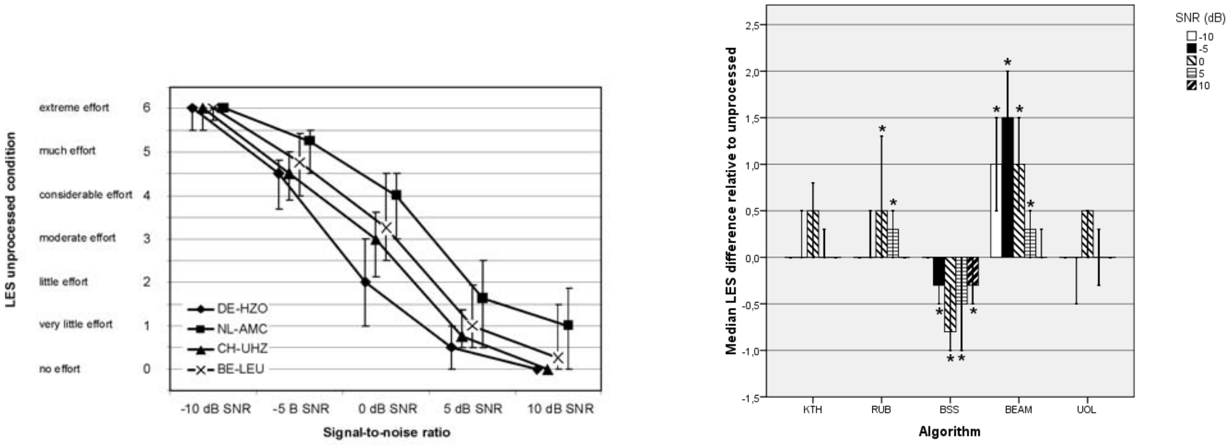

Figure 2: Left panel: Median LES values for the unprocessed condition at different signal-to-noise ratios: effect of test site. Error bars represent first and third quartile. From D-7-5 (Figure 10).
Right panel: LES scores presented relative to the unprocessed condition. Data is averaged over both centers and all three hearing loss groups. Error bars represent 95% confidence interval.
More positive scores mean more improvement in LES score.

MWF required less effort than the unprocessed condition at the lower signal-to-noise ratio's (-10 to 0 dB SNR). At the higher signal-to-noise ratio's (the easier conditions) MWF did not decrease the listening effort. BSS showed a negative effect on the listening effort, mainly at the higher SNR, although variability was large. No general trend was observed for SC1, SC2 and COH to provide benefit in listening effort.

2.2.3 Preference Rating

In deliverable D-7-5, the preference rating results were first analyzed in terms of percentage of wins (that is, the percentage of times a presented sound sample was chosen over other sound samples). This analysis was insightful and intuitive. However, more experimental information is available since the subjects were not only asked to compare a processed sound sample to an unprocessed sample; they were also asked to rate the degree of preference for the sample that they had chosen over the other sample. The procedure is explained in detail in D-7-5. This preference rating consisted of 5 levels (from 'very much better' (5) to 'very slightly better’ (1). Interpretation of this raw score is as follows: a positive score (values 1 to 5) indicate that the sample was more preferred than the unprocessed condition. A negative score (-1 to -5) means that the unprocessed condition was preferred. Since the subjects had to chose between the two samples (forced choice paradigm), the answer ‘no difference’ could not occur. However, the arithmetic average of the raw scores can be zero, indicating that on average the subjects had no preference with respect to the unprocessed condition.

In the present correlation analysis we used these raw preference scores. More sophisticated measures and models are available (for instance see deliverable D-7-5 for a recalculation of the data onto a Linear Gaussion Scale), but since we are primarily interested in finding measures that best relate the individual preferences to the AP measures, the raw PR scores are best suited. Figure 3 shows the results from the preference rating for each algorithm and for each signal to noise ratio. Note that the results from the entire data set (left panel) are given in win counts, and that the results from the subset are presented as the raw preference values. The figure shows that for both sets, SC1, SC2, MWF and COH were more preferred than the unprocessed condition, at all three tested SNR levels. For BSS the subjects preferred the unprocessed condition more often. This is indicated by the low number of average win counts (left panel, all data) and the negative average PR score (right panel, data from NL-AMC and DE-HZO only). The general trends were the same for the three SNRs, and for each subject group.

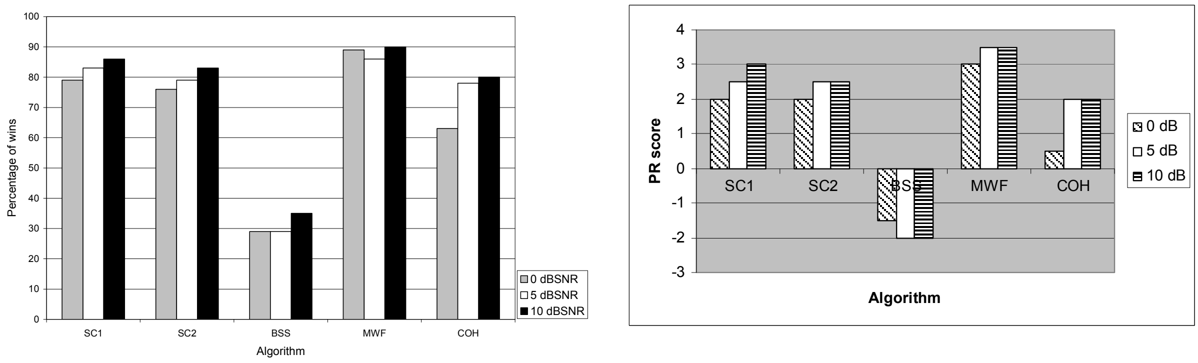

Figure 3: Results from the preference rating test at three different signal to noise ratios. The left panel shows data from the complete set (in win counts) , the right panel of the subset (in raw scores). Higher win counts (left panel) or more positive scores (right panel) indicate a higher average subjective preference.

2.2.4 LES^{SRT} and PR^{SRT} as composite variables

Both LES and PR were measured at several signal to noise ratios (SNRs). In contrast, SRT was measured with an adaptive procedure and has a single SNR as outcome measure. Whereas the adaptive SRT gives a direct indication of the performance of the subject, the LES and PR scores suffer from floor and ceiling effects because they were obtained at fixed levels with respect to the available speech information (fixed SNRs), irrespective of the performance of the subjects at those levels. Such data are well suited for a cross-sectional investigation into the best algorithm, but are less applicable for an analysis of individual results. It would be useful to obtain LES and PR values that can be interpreted directly as a measure that reflects the subjects. Therefore, we calculated composite values that reflect the overall performance for each subject and for each algorithm. For this purpose, we interpolated and extrapolated the LES and PR values at an SNR that corresponds to that of the measured SRT for the individual subject in the unprocessed condition (i.e. the aided SRT).

LES^{SRT} was calculated as follows. If the aided SRT-value was within the range of SNRs used for obtaining the scores (-10 to +10 dB), LES$_{SRT}$ values were calculated by interpolation of the values at neighbouring SNRs. For SRT-values under -10 dB and above 10 dB, LES$_{SRT}$ was extrapolated using the trend lines derived from the nearest SNR-intervals [-10 to -5 dB] or [+5 to +10 dB], respectively. For the calculation of PR$_{SRT}$ a similar procedure was used, with the exception that PR data was available for 0, +5, and + 10 dB only.

The added advantage of these composite variables are a reduction in the large number of variables for LES (5 SNRs) and PR (3 SNRs). We expect that both LES and PR will be associated also to the overall intelligibility level. Therefore, we have to compensate for the overall intelligibility level in order to separate the extra information provided by LES and PR from the SNR-dependent components. This way we expect to get parameters that are primarily related to listening effort and preference, respectively. However, intelligibility matching can also reduce the range of LES and SRT values since LES^{SRT} and PR^{SRT} reflect the effort and the preference at the point of 50% speech intelligibility only. Thus, very easy (high SNR) and very difficult (low SNR) conditions are not represented in LES^{SRT} and PR^{SRT}. Moreover, deliverable D-5-2 showed that the processing by several of the HearCom noise reduction algorithms depends on SNR. This means that LES^{SRT} and PR^{SRT}, and SRT, reflect the individual's speech intelligibility, and not equal processing of the algorithms. Still it is expected that LES^{SRT} and PR^{SRT} correlate with the LES and PR values obtained at fixed SNRs. Joint significant correlations for all algorithms were:

- LES^{SRT} correlated significantly with LES at -5 dB SNR (except for the COH-algorithm) (correlations for SC1, SC2 and BSS around 0.4, for MWF 0.78);
- The benefits in LES^{SRT} correlated significantly with the benefits at SNRs of -5 and 0 dB (correlations between 0.42 and 0.68).
- PR^{SRT} correlated significantly with PR at SNR=0 dB (correlations between 0.77 and 0.88).

In D-5-2 it was shown that the largest calculated improvement in the speech intelligibility index (SII) was independent of the algorithm used and was found at a signal to noise ratio of +5 or +10 dB.

Table 4: Outcome parameters that are selected for comparison to the AP parameters

Parameter	**Description**
SRT$_{unprocessed}$	Speech Reception Threshold for the unprocessed condition (aided SRT), raw scores
ΔSRT	Speech Reception Threshold relative to that for the unprocessed condition (SRT$_{unprocessed}$)
LES$^{SRT}_{unprocessed}$	Listening Effort Scaling calculated at SNR=SRT$_{unprocessed}$
ΔLES^{SRT}	Listening Effort Scaling calculated at SNR=SRT for an algorithm relative to LES$^{SRT}_{unprocessed}$
PR$^{SRT}_{unprocessed}$	Preference Rating calculated at SNR=SRT$_{unprocessed}$
ΔPR^{SRT}	Preference Rating calculated at SNR=SRT$_{unprocessed}$

Therefore, in addition to LES^{SRT}/PR^{SRT}, the analyses were also done with the original data at fixed SNR. Special attention was given to SNR=+5 and +10 dB. For clarity, only results obtained for LES^{SRT}/PR^{SRT} are presented, and where needed we also analyzed the original variables.

2.3 Auditory Profile

The AP-results of our subjects were compared to the AP-results in the HearCom validation study (deliverable D-2-3 and D-2-5).

2.3.1 Inclusion criteria

First of all, in the AP validation study, inclusion criteria for the hearing impaired subjects were broader than for our WP7-AP subjects. The AP validation study included subjects with more different hearing loss configurations, such as hearing losses with a PTA above 60 dB HL, conductive and combined hearing losses, and subjects with asymmetric hearing losses. Thus, our WP7-AP subjects meet the AP inclusion criteria, but the hearing losses are more homogeneous.

2.3.2 Language correction

In the AP validation study, the test results were corrected for language differences between test sites. The same corrections were applied for test results that could have been affected by language or translation differences, such as SRT, (B)ILD, Gothenburg Profile and Lexical Decision test. With paired t-tests (Bonferroni corrected) we verified that the language correction also applies on the new dataset. Mean differences in test results were no longer significant or at least less significant when the language correction was applied. We thus decided to use language corrected AP results in all further analyses.

2.3.3 Selection AP outcome parameters

We selected the most relevant parameters from the AP test results, related to the pure-tone audiogram, loudness perception, speech-in-noise, binaural functioning, subjective speech perception, frequency- and time resolution, cognition, and age. For tests using narrow-band signals, we selected parameters for the low frequencies (500 Hz) and for the high frequencies (3000 Hz). The selection of parameters is presented in Table 5.

Table 5: Overview of outcome parameters for the Auditory Profile.

AP-parameter	**Derived from**	**Description parameter**
AC500	Audiogram	Air conduction at 500 Hz
AC3000	Audiogram	Air conduction at 3000 Hz
PTA	Audiogram	Pure tone average (from frequencies 500 to 4000 Hz)
SlopeAudio	Audiogram	Difference between minimum and maximum threshold (from frequencies 500 to 4000 Hz)
Slope500	Acalos	Slope at 500 Hz (loudness curve)
Slope3000	Acalos	Slope at 3000 Hz (loudness curve)
Slope-bb	Acalos	Slope for broadband noise (loudness curve)
SRTcontinue	SRT	Speech Reception Threshold in stationary noise
SRTfluct	SRT	Speech Reception Threshold in fluctuating noise
ILD	ILD	Intelligibility Level Difference
BILD	BILD	Binaural Intelligibility Level Difference
GPspeech	Gotheburg Profile	Subjective judgement of intelligibility
F500	FT-test	Frequency Resolution at 500 Hz
F3000	FT-test	Frequency Resolution at 3000 Hz
T500	FT-test	Temporal Resolution at 500 Hz
T3000	FT-test	Temporal Resolution at 3000 Hz
LD	Lexical Decision Test	Lexical Decision Score
Age	Subject Information	Age in years

3 The Auditory Profile in relation to noise reduction

This section investigates the results from the perceptual evaluation in the light of the newly measured subject data according to the AP. The previous sections described the variables that were selected for the investigation. These variables were used in the correlation analyses that are described below.

Reported correlations with SRT represent the results from parametric statistics (Pearson's correlation coefficient) and values obtained with LES and PR were obtained with non parametric statistics (Spearman's correlation coefficient).

SRT is measured both in the WP7 experiment and in the additional AP measurements. Differences between the measurements are for instance the fact that they are measured on different time periods, the use of a different background babble, a dummy hearing aid (WP7) or a headphone (AP), and frequency dependent amplification (WP7) or broadband amplification (AP).

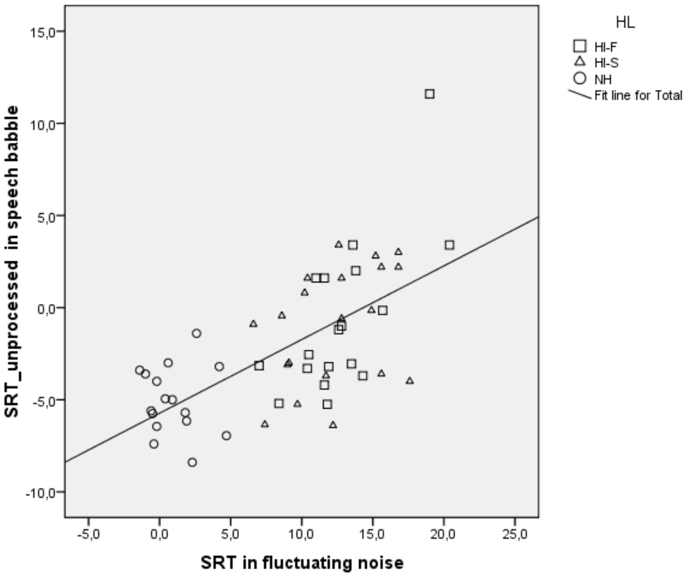

Figure 4: SRT data from WP7 plotted against SRT data obtained for AP for the same subjects. Each dot represents a single subject.

Although there are differences in the circumstances between the measurements, it is expected that the results will correlate well. Figure 4 shows the aided SRT as measured in WP 7 as a function of the SRT from the AP. The correlation between these measurements was highly significant ($r=0.65$, $p<6*10^{-8}$) and explained about 42% of the variance. Nevertheless, the correlation is considerably lower than in a real test-retest situation (where we found coefficients of about 0.9), probably due to the difference in sound presentation between the two experiments (dummy hearing aids versus headphones).

3.1.1 AP parameters and SRT results

Table 6 summarizes the relevant results for the SRT data. In the second column significant correlations with the overall performance $\mathrm{SRT}_{\mathrm{unprocessed}}$ are shown. In column 3 until column 8 significant correlations with the SRT improvement due to the different noise reduction algorithms are indicated.

Table 6: Summary and description of relevant results of correlation analysis on SRT. AP measures are only shown if any significant correlation explained 10% of the variance or more. No significant correlation (at 5% level, no correction for multiple comparisons) is indicated with “ns”. Correlations that are type set in italic indicate an explained variance <10% for, and type set in bold indicates a correlation that explained ≥ 10%. See Appendix A1 for all values and test fields.

	SRT	**ΔSRT**				
	unprocessed	**SC1**	**SC2**	**BSS**	**MWF**	**COH**
Audiogram	low loss @ .5k, 3k, and PTA	ns	ns	ns	ns	ns
Loudness slopes	shallow slopes @ .5k (and bb)	*(shallow slopes @ 3k)*	ns	steep slopes @ .5k and bb	ns	ns
F-resolution	good F-res @ 3k	ns	good F-res @.5k	ns	ns	ns
T-resolution	good T-res @ 3k	ns	ns	ns	ns	ns
SRT in noise	cont / fluct noise	ns	ns	ns	ns	ns
ILD / BILD	ns	*(good BILD)*	ns	*(poor ILD)*	ns	*(good BILD)*
GP - speech	subj. good SRT	ns	ns	ns	ns	ns
Cognition	good cognition	ns	poor cognition	ns	ns	ns
Age	younger subjects	ns	ns	ns	ns	*(older subjects)*

Variables that explained at least 10% of the variance (corresponding to a Pearson's correlation coefficient of 0.31 or higher) are shown, the other variables were omitted. In Appendix A1 the details of the results are given.

Unsurprisingly, the results indicate that the SRT in the unprocessed condition (this is the overall aided speech performance) was significantly better for young subjects. It was also better for subjects with a smaller audiometric loss (at .5 and 3 kHz). However, these audiometric parameters were not significantly related to the SRT-improvements due to the use of the noise-reduction algorithms.

Looking at the algorithms, table 6 shows that the improvements in SRT due to the use of the algorithms resulted in only a few significant correlations:

- SC2 showed a significantly higher SRT-improvement in subjects with good frequency resolution at 0.5 kHz and/or poor cognition (as measured by the lexical decision test). The correlation coefficients were 0.32 and 0.42 respectively.
- BSS showed a significantly higher SRT-improvement in subjects with recruitment at 0.5 kHz (Acalos at 0.5 kHz) and broadband noise (Acalos wide band). The correlation coefficients were 0.37 and 0.34 respectively.

To give an indication of the relationships described above, Figure 5 and 6 shows scatter plots of the data for SC2 and BSS respectively. The symbols that are used in the figures correspond to the audiogram categories that were used in WP 7 (normal hearing, a flat hearing loss, and a sloping hearing loss).

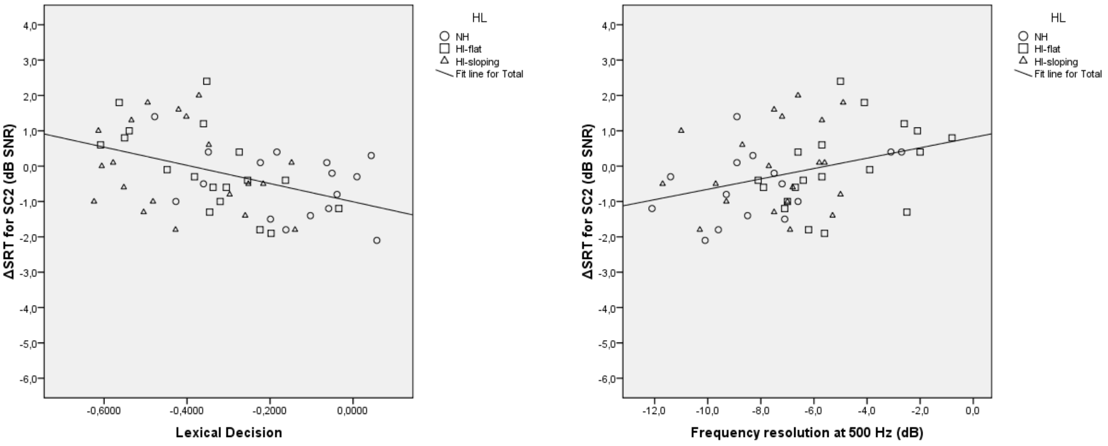

Figure 5: Correlation between the measured improvement in SRT for SC2 and the outcome of the lexical decision test (left panel) and frequency resolution at 500 Hz (right panel).

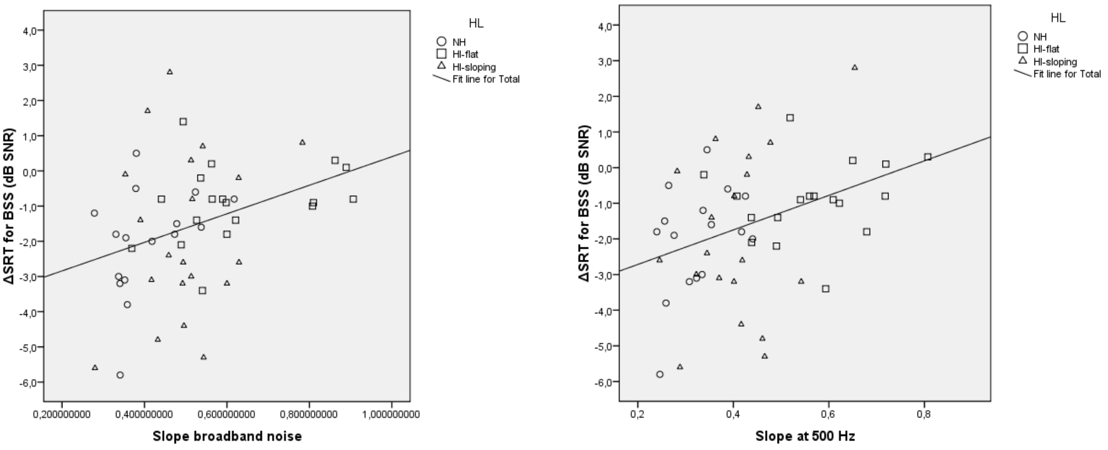

Figure 6: Correlation between the measured improvement in SRT for algorithm BSS and the slope of the loudness scaling test (Acalos) at 500 Hz (left panel) and broadband (right panel).

The average results (section 2.2) showed that SC2 caused no significant difference in SRT relative to the unprocessed condition and that BSS showed a degradation in SRT (-1,8 +/- 0.3 dB). From figure 5 it can be seen that the difference in benefit between the subjects is fairly large (from -2 to +2 dB SNR for SC2 and even larger for BSS). However the correlation is most likely not strong enough to yield a good predictor for the benefit of SC2 or for the degradation of BSS in this noise condition.

The algorithm that yielded the largest average effect on SRT was MWF. We found no strong correlations between the large benefit of MWF and AP variables. Figure 7 gives the strongest (but still weak) correlation: frequency resolution at 500 Hz (r=0.25, explained variance is 6%).

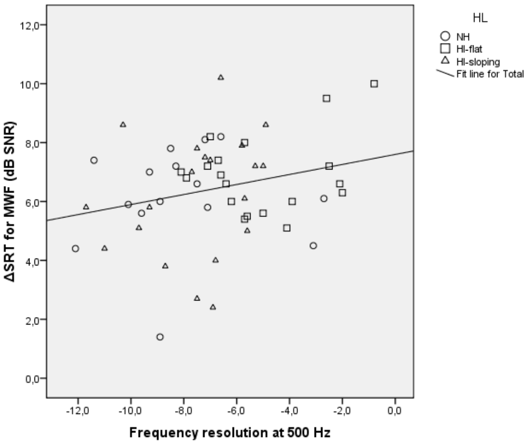

Figure 7: ΔSRT for MWF as a function of frequency resolution at 500 Hz.

3.1.2 AP parameters and Effort Scaling (LES$^{\mathrm{SRT}}$)

Table 7 shows the relevant results for the LES$^{\mathrm{SRT}}$ data. Aided LES$^{\mathrm{SRT}}$ results appeared significantly better (less effort) for subjects with a poor SRT in stationary and fluctuating noise. This applies to a lesser extent to the pure-tone threshold at 0.5 kHz, subjective speech perception (GP), and cognition.

Only the improvements in LES$^{\mathrm{SRT}}$ due to the use of MWF noise reduction showed significant correlations. However, the majority of correlations were quite weak. The strongest correlation indicated that more benefit was found for subjects with a small audiometric loss at 3 kHz ($r=0.32$). This is caused by the fact that normal hearing subjects had a slightly larger benefit than hearing impaired subjects, see figure 8.

Note however that the separation between normal hearing and hearing impaired subjects (visible in Figure 8) is due to an effect in audiometric threshold and not due to an effect in ΔLES$^{\mathrm{SRT}}$. Moreover, both the effect of subject group and the interaction effect of subject group and algorithm were not significant in a repeated measures analysis.

Table 7: Summary and description of relevant results of correlation analysis on LES$^{\mathrm{SRT}}$. AP measures are only shown if any significant correlation explained 10% of the variance or more. No significant correlation (at 5% level, no correction for multiple comparisons) is indicated with “ns”. Correlations that are type set in italic indicate an explained variance <10% for, and type set in bold indicates a correlation that explained $\geq$ 10%. See Appendix A2 for all values and test fields

Effort Scaling: LES$^{\mathrm{SRT}}$	**Overall performance LES$^{\mathrm{SRT}}$ unprocessed**	**ΔLES$^{\mathrm{SRT}}$ SC1**	**ΔLES$^{\mathrm{SRT}}$ SC2**	**ΔLES$^{\mathrm{SRT}}$ BSS**	**ΔLES$^{\mathrm{SRT}}$ MWF**	**ΔLES$^{\mathrm{SRT}}$ COH**
Audiogram	.5k	ns	ns	ns	**mild loss @3k**	ns
Loudness slopes	@ .5k and bb	ns	ns	ns	*(shallow slope @ 3k)*	ns
F-resolution	-	ns	ns	ns	ns	ns
T-resolution	-	ns	ns	ns	ns	ns
SRT in noise	cont / fluct	ns	ns	ns	ns	ns
ILD / BILD	-	ns	ns	ns	ns	ns
GP - speech	subj. poor SRT	ns	ns	ns	*(subj. good SRT)*	ns
Cognition	X	ns	ns	ns	ns	ns
Age	-	ns	ns	ns	*(younger subjects)*	ns

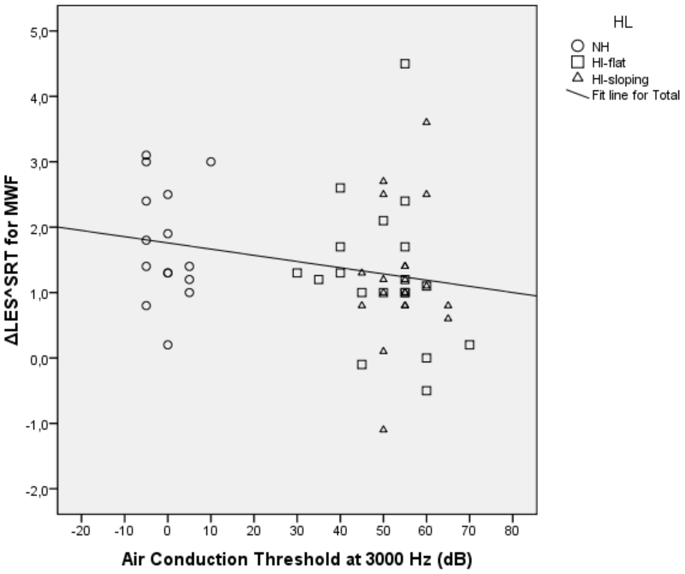

Figure 8: ΔLES^{SRT} for MWF as a function of audiometric loss at 3 kHz.

3.1.3 AP parameters and Preference Rating (PR^{SRT})

Table 8 shows results for PR. Since all data was obtained relative to the unprocessed condition only PRimprovement is shown in the table. No significant correlations were found for MWF. MWF is the algorithm with the highest improvement on PR, but this improvement did not significantly correlate with the AP measures.

Table 8: Summary and description of relevant results of correlation analysis on PR^{SRT}. AP measures are only shown if any significant correlation explained 10% of the variance or more. No significant correlation (at 5% level, no correction for multiple comparisons) is indicated with "ns". Correlations that are type set in italic indicate an explained variance <10% for, and type set in bold indicates a correlation that explained ≥ 10%. See Appendix A3 for all values and test fields

	ΔPR^{SRT}				
	SC1	**SC2**	**BSS**	**MWF**	**COH**
Audiogram	ns	high loss @ .5k	high loss @.5k	ns	high losses
Loudness slopes	ns	ns	ns	ns	higher slope @ 3k
F-resolution	poor F-res @.5k	ns	*(poor F-res @.5k)*	ns	ns
T-resolution	ns	ns	ns	ns	ns
SRT in noise	ns	ns	poor SRT cont&fluct	ns	*(poor SRT in fluct)*
ILD / BILD	ns	ns	ns	ns	poor ILD/BILD
GP - speech	ns	ns	ns	ns	subj. poor SRT
Cognition	ns	ns	ns	ns	ns
Age	ns	ns	older subjects	ns	older subjects

All other algorithms showed significant correlations with one or more AP-test results and improvements in PR^{SRT} outcomes due to noise reduction. Significant correlations between higher preferences for an algorithm over the unprocessed condition were found:

- Subjects with higher audiometric losses (at 0.5 kHz) had a significantly higher preference for SC2, BSS and COH (r=0.33, 0.38 and 0.45, respectively).

- Subjects with more recruitment at 3 kHz (steeper slopes in Acalos at 3kHz) had a significantly higher preference for COH (r=0.38).
- Subjects with a poor frequency resolution at 0.5 kHz had a significantly higher preference for SC1 (r=0.35).
- As indicated earlier, there were hardly any subjects who preferred BSS for our experimental conditions. The negative effects were relatively mild in subjects with a poor SRT in stationary and fluctuating noise (r=0.33 and r=0.44).
- Subjects with a poor SRT in stationary and fluctuating noise tended to have a significantly higher preference for and for COH (r=0.31).
- Subjects with a poor ILD/BILD had a significantly higher preference for COH (r=0,41 and r=0,45, respectively)
- Subjects with a poor subjective (GP-speech) had a significantly higher preference for COH (r=0,33).
- Older subjects had a higher preference for COH (r=0.35) and less problems with BSS (r=0,32).

For the algorithm with the largest average improvement (MWF), no significant correlations were found. Although some correlations with the other algorithms were statistically significant, they might be of limited practical use. For instance, Figure 9 shows the significant correlation between PR^{SRT} for COH and BILD. A careless look at the figure would suggest that COH is more preferred by subjects with a higher BILD. A more careful investigation shows that it is simply less preferred by normal hearing subjects.

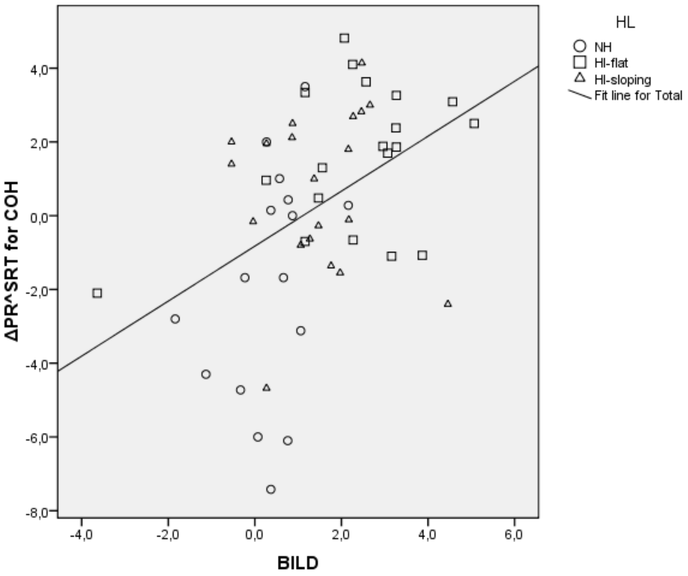

Figure 9: ΔPR^{SRT} for COH as a function of BILD. .

4 Discussion

This report describes a first attempt to assess the added value of the AP for auditory rehabilitation. The choice of the signal processing schemes was more or less dictated by the signal processing schemes and evaluation procedures chosen in Work Package 7, in order to build a bridge between the work conducted in the Sub Projects SP1 and Sp3.

In general terms the results of the new tests were in good agreement with the results obtained in the two multi-centre studies that were conducted to develop and validate the AP. Also, the results from the subgroup of subjects that were willing to participate in this follow-up study were in close agreement with the results of the total group, as described in report D-2-5. After a correction factor for language differences, we found a good correspondence between the two participating labs. All these findings indicate that this study can be regarded as representative, despite the relatively small number of participants.

For the first time and for this study, two area's in the HearCom project, Diagnostics and Rehabilitation, were brought together to obtain these correlations. The experimental question was: *What, if any, is the relationship between the efficacy of the WP7T5 noise suppression algorithms and the results of the AP in normal hearing and hearing impaired adults? Can the results from the AP be used in rehabilitation of hearing impaired with hearing aids.*

The patterns of significant correlation coefficients indicate that a number of AP parameters are associated with different aspects of speech perception in noise and listening effort and can influence preference data. In these data sets, most correlation coefficients are relatively low, but this can be induced by the test precision in relation to the inter-individual variability. Figure 4 illustrates that even for similar tests like the SRT test in fluctuating noise the correlation coefficients are far below 1.

Nevertheless, a interesting pattern of significant correlations was found and this hints to the fact that AP parameters may be important for specific noise reduction strategies.

For the SRT outcome values (see Table 6), it is striking that the absolute results, obtained for unprocessed signals, show many more significant correlations than the relative results, i.e. the ΔSRT values for the different algorithms relative to "unprocessed". An obvious reason is that the relative effects are small for all noise reduction schemes, except MWF (see Figure 1). However, it is especially for MWF where none of the parameters of the AP shows a significant correlation. We expect that this is caused by the fact that the inter-individual variability in the ΔSRT values due to MWF is very small. It is conceivable that MWF can be regarded as a front-end signal processing technique that enhances SNR and improves SRTs for all subjects, irrespective their hearing loss, other audiological parameters, or age.

For the LES^{SRT} outcome values (see Table 7) the absolute results, obtained for unprocessed signals, show a clearly interpretable pattern of correlations with the AP parameters derived from the pure-tone audiogram, loudness perception, and objective and subjective speech perception. For the relative outcome measures, ΔLES^{SRT}, it is striking that we found only significant correlations with the LES-improvements due to MWF. This suggests that the inter-individual changes in LES scores due to different signal processing techniques are related more strongly to auditory perception than the 50% SRT-thresholds are. But given the relatively low correlation coefficients there is no strong evidence.

For the PR^{SRT} outcome values (see Table 8) we have only relative values. Again the effects of AP parameters on preferences due to MWF is absent, probably because the MWF signals are preferred by almost every subject, irrespective their hearing loss, other audiological parameters, and age. For the other noise-reduction schemes correlation patterns were found that indicate that either the pure-tone audiogram, or the low-frequency frequency resolution play a role in the single-channel algorithms and in BSS. The binaural parameters derived from ILD and BILD are relevant for the preference of COH, and age is important for the preference scores of BSS and CH.

In retrospect, it is clear that the selection of test algorithms and test conditions cannot be considered as an optimal choice to test the relevance of the AP for auditory rehabilitation. This is partly caused by the fact that the average effects are very small (if present) in four of the noise-reduction schemes relative to unprocessed. In the remaining algorithm (MWF) the superiority was that large that it was found in all subjects, almost to the same degree.

Given the test precision in relation to the small inter-individual variation, high correlations could not be found and thus the pattern of correlation coefficients is too weak to be applied for prediction of the effects. For the same reason, the AP will not be very robust in the selection of one of these noise-reduction algorithms for an individual hearing-aid user. In this study, the most successful strategy for this specific test condition seems to be to choose the MWF algorithm or use a combination of algorithms, including MWF.

However, despite the relatively low test precision, we found interesting patterns of correlations that suggest that the AP is able to disclose some of the underlying causes for speech perception and/or improvements in speech perception due to signal-processing schemes. Although too weak to be used for predictions in this study, the AP parameters may do a better job for other types of signal processing and for schemes and test conditions that yield stronger effects and higher inter-individual variability.

In retrospect to the results of D-7-5, it is also important to note that the AP, if it had been used for all subjects, would not have suggested completely different subject groups as have been used now (derived from the pure-tone audiogram).

5 Dissemination and Exploitation

As discussed earlier (see D-13-4) the HearCom diagnosis and rehabilitation tests in the different languages will be made available to the professional community: Researchers/developers working in the areas audiology, hearing aids, audio equipment, and telecommunications; Audiometer suppliers for integration within equipment; Clinics and Hearing aid acousticians (directly or through professional organizations such as EFAS, EUHA).

The dissemination will be supported by providing stand-alone versions of the test materials and procedures on CD/CD-ROM/DVD, including background information. Also, the tests can be distributed via Internet (Portal, WP 11). Here the materials can be ordered and free demo versions are available as downloads.

The AP is part of the OMA workstation (Oldenburg Measurement Applications) consisting of mainstream hardware with specific software. The OMA workstation (HörTech) consists of software that is marketed to professionals:

- Software including tests distributed via dongle,
- Hardware: Mainstream computer; high-quality mainstream sound-card, mainstream response device,
- Support for different audiometers
- HearCom-developed plug-ins for OMA + basic OMA software

Marketing will be done in different EU countries + (outside) by HörTech (DE-HTCH) and country-specific partners (following the existent experiences in Germany). The market demand for this advanced audiological testing software will be increased by the wide dissemination of the screening tests: Test “failures” will (besides being provided with general information on how to seek specific help) preferentially be referred to audiological centres or individual experts that use any of the professional tests for speech in noise developed in HearCom.

5.1 Dissemination

This deliverable is a confidential report. The authors aim however, at submitting the work presented in this deliverable to a scientific journal.

5.2 Ethical issues

All participants that participated to the tests were volunteers and were allowed to step down at any time without disprofit. The calibration and hard limiting in the measurement systems guaranteed that the tests could be performed without risk of hearing damage.

6 Conclusions

The Auditory Profile comprises a unique set of (internationally) standardised and validated tests that go beyond the traditional pure-tone audiogram. The results obtained in this study (n=55) are well in agreement with the two earlier extensive HearCom validation studies. The test-centre effects disappeared after using the pre-defined language corrections for the language-dependent test items.

The individual outcome measures for the effects of different noise reduction schemes (SRTs, LES-values, and PR-scores) all reflect different aspects of change. They appear also to be related to different aspects in the Auditory Profile.

However, in the set of test algorithms and the test conditions selected, the correlation coefficients were - although significant - too small to predict the effects or make a selection on the basis of individual AP parameters in this population.

This study illustrates that the AP may be able to disclose some of the underlying causes for speech perception and/or improvements in speech perception due to signal-processing schemes, but further experimental work is necessary to prove the added value in the selection of signal-processing schemes and/or in the prediction of the perceptual benefit.

7 Literature

HearCom D-2-1: Implementation of a preliminary test set for auditory impairments, Final version, October 2005.

HearCom D-2-1b: Demo version of the preliminary test set for auditory impairments. Version 2.0, May 2006.

HearCom D-2-2: Procedures for the tests included in the AP in four languages, Final version, August 2006.

HearCom D-2-3: Report about the results of the multi-centre evaluation of the AP. Final version, October 2007.

HearCom D-2-5: Optimized, final set of impairment tests included in the AP. Final version, July 2008.

HearCom D-5-2: Effectiveness of signal enhancement techniques for various APs, Final version, September 2006

HearCom D-7-5: Report on the evaluation of signal enhancement techniques, Final version, December 2008.

Appendix A: Correlation Coefficients

A.1 SRT

		identity raw scores	**KTH spectral subtraction**	**RUB wiener filter**	**BSS**	**BEAM**	**UOL**
AP-parameter		SRT-av	SRT-benefit	SRT-benefit	SRT-benefit	SRT-benefit	SRT-benefit
AC500	Correlation	,416**	0,182	0,204	0,169	0,180	0,213
	Sig. (2-tailed)	,002	0,188	0,138	0,221	0,192	0,122
AC3000	Correlation	,446**	0,002	0,197	0,097	0,169	0,136
	Sig. (2-tailed)	,001	0,990	0,153	0,485	0,221	0,329
PTA	Correlation	,448**	0,026	0,261	0,091	0,152	0,141
	Sig. (2-tailed)	,001	0,853	0,057	0,511	0,271	0,309
SlopeAudi	Correlation	,250	-0,211	0,011	-0,069	-0,020	-0,031
	Sig. (2-tailed)	,068	0,125	0,934	0,619	0,888	0,822
Slope500	Correlation	,336*	0,150	0,049	0,366**	0,113	0,146
	Sig. (2-tailed)	,013	0,278	0,726	0,007	0,415	0,291
Slope3000	Correlation	,150	-0,274*	0,040	-0,092	-0,072	-0,021
	Sig. (2-tailed)	,278	0,045	0,772	0,508	0,606	0,882
Slope-bb	Correlation	,278*	0,047	0,130	0,338*	-0,077	0,148
	Sig. (2-tailed)	,042	0,737	0,349	0,012	0,579	0,287
SRTcontin	Correlation	,655**	0,084	0,135	0,026	0,064	0,097
	Sig. (2-tailed)	,000	0,544	0,327	0,850	0,643	0,483
SRTfluct	Correlation	,649**	0,086	0,222	0,036	0,112	0,161
	Sig. (2-tailed)	,000	0,537	0,107	0,794	0,422	0,245
ILD	Correlation	,196	-0,003	0,060	0,269*	-0,055	0,076
	Sig. (2-tailed)	,151	0,981	0,664	0,047	0,688	0,580
BILD	Correlation	-,086	-0,274*	0,014	0,201	-0,204	-0,305*
	Sig. (2-tailed)	,533	0,037	0,921	0,141	0,136	0,024
Gpspeech	Correlation	,349**	0,054	0,237	0,130	0,066	0,186
	Sig. (2-tailed)	,009	0,695	0,082	0,346	0,630	0,174
F500	Correlation	,224	0,221	0,324*	0,036	0,246	0,048
	Sig. (2-tailed)	,103	0,108	0,017	0,795	0,072	0,730
F3000	Correlation	,470**	0,076	0,016	-0,020	0,112	0,163
	Sig. (2-tailed)	,000	0,585	0,906	0,883	0,419	0,238
T500	Correlation	,298*	0,102	0,164	0,113	0,140	0,037
	Sig. (2-tailed)	,028	0,465	0,235	0,415	0,313	0,792
T3000	Correlation	,440**	0,063	0,207	-0,009	0,039	0,219
	Sig. (2-tailed)	,001	0,652	0,133	0,948	0,781	0,111
LD	Correlation	-,440**	-0,107	-0,424**	-0,122	0,097	-0,169
	Sig. (2-tailed)	,001	0,441	0,001	0,379	0,485	0,222
Age	Correlation	,567**	0,112	0,103	0,096	0,216	0,279*
	Sig. (2-tailed)	,000	0,416	0,452	0,485	0,114	0,039

A.2 LES^{SRT}

		identity raw scores	**KTH spectral subtraction**	**RUB wiener filter**	**BSS**	**BEAM**	**UOL**
AP-parameter		LES-SRT	LES-srt benefit	LES-srt benefit	LES-srt benefit	LES-srt benefit	LES-srt benefit
AC500	Correlation	-,278*	0,031	-0,168	-0,100	-0,224	-0,034
	Sig. (2-tailed)	,042	0,823	0,225	0,472	0,104	0,806
AC3000	Correlation	-,231	0,022	-0,097	-0,122	-0,323*	-0,082
	Sig. (2-tailed)	,093	0,875	0,485	0,380	0,017	0,553
PTA	Correlation	-,225	-0,030	-0,100	-0,040	-0,236	-0,040
	Sig. (2-tailed)	,102	0,832	0,473	0,776	0,086	0,775
SlopeAudi	Correlation	-,148	0,220	0,222	0,062	-0,232	0,022
	Sig. (2-tailed)	,284	0,110	0,107	0,657	0,091	0,877
Slope500	Correlation	-,065	0,118	-0,036	-0,066	-0,180	-0,016
	Sig. (2-tailed)	,640	0,395	0,798	0,637	0,192	0,909
Slope3000	Correlation	-,034	0,226	0,132	-0,242	-0,288*	-0,038
	Sig. (2-tailed)	,805	0,101	0,343	0,079	0,034	0,785
Slope-bb	Correlation	,049	0,135	0,098	0,043	-0,173	0,052
	Sig. (2-tailed)	,723	0,331	0,480	0,756	0,211	0,707
SRTcontin	Correlation	-,468**	0,086	0,122	-0,138	-0,215	0,084
	Sig. (2-tailed)	,000	0,533	0,373	0,314	0,116	0,543
SRTfluct	Correlation	-,353**	0,146	0,057	-0,062	-0,154	0,143
	Sig. (2-tailed)	,009	0,291	0,682	0,654	0,267	0,304
ILD	Correlation	-,001	0,186	-0,011	-0,008	-0,177	-0,095
	Sig. (2-tailed)	,992	0,174	0,934	0,953	0,196	0,489
BILD	Correlation	-,058	0,094	0,024	-0,037	-0,078	-0,018
	Sig. (2-tailed)	,672	0,496	0,862	0,786	0,574	0,897
Gpspeech	Correlation	-,305*	0,047	-0,194	-0,227	-0,311*	-0,121
	Sig. (2-tailed)	,024	0,732	0,157	0,096	0,021	0,378
F500	Correlation	-,248	-0,003	-0,126	-0,140	0,020	0,010
	Sig. (2-tailed)	,070	0,984	0,366	0,313	0,884	0,943
F3000	Correlation	-,090	0,259	0,229	0,055	-0,207	0,116
	Sig. (2-tailed)	,516	0,058	0,095	0,694	0,134	0,404
T500	Correlation	-,066	0,067	-0,194	-0,035	-0,183	0,066
	Sig. (2-tailed)	,638	0,631	0,161	0,801	0,186	0,636
T3000	Correlation	-,178	0,174	0,071	-0,142	-0,099	0,071
	Sig. (2-tailed)	,198	0,207	0,610	0,306	0,477	0,611
LD	Correlation	,288*	-0,155	-0,153	0,056	0,190	-0,126
	Sig. (2-tailed)	,034	0,262	0,270	0,685	0,169	0,366
Age	Correlation	-,230	0,178	0,128	0,045	-0,311*	0,116
	Sig. (2-tailed)	,091	0,194	0,353	0,746	0,025	0,398

A.3 PR^{SRT}

		KTH spectral subtraction	**RUB wiener filter**	**BSS**	**BEAM**	**UOL**
AP-parameter		PR-srt benefit	PR-srt benefit	PR-srt benefit	PR-srt benefit	PR-srt ben
AC500	Correlation	0,238	0,331*	0,380**	0,097	0,453**
	Sig. (2-tailed)	0,083	0,014	0,005	0,485	0,001
AC3000	Correlation	0,180	0,037	0,213	0,036	0,497**
	Sig. (2-tailed)	0,194	0,789	0,122	0,794	0,000
PTA	Correlation	0,221	0,132	0,287*	0,130	0,458**
	Sig. (2-tailed)	0,109	0,341	0,035	0,348	0,000
SlopeAudi	Correlation	0,064	-0,249	0,093	-0,076	0,244
	Sig. (2-tailed)	0,646	0,069	0,503	0,587	0,075
Slope500	Correlation	0,034	0,227	0,236	-0,046	0,252
	Sig. (2-tailed)	0,806	0,099	0,085	0,739	0,066
Slope3000	Correlation	0,213	0,119	0,084	0,101	0,380**
	Sig. (2-tailed)	0,123	0,391	0,544	0,467	0,005
Slope-bb	Correlation	-0,113	0,253	0,093	-0,031	0,266
	Sig. (2-tailed)	0,414	0,065	0,502	0,825	0,052
SRTcontin	Correlation	0,070	0,136	0,329*	-0,080	0,259
	Sig. (2-tailed)	0,613	0,323	0,014	0,562	0,057
SRTfluct	Correlation	0,120	0,223	0,441**	0,013	0,311*
	Sig. (2-tailed)	0,387	0,105	0,001	0,928	0,022
ILD	Correlation	0,192	0,195	0,144	0,113	0,414**
	Sig. (2-tailed)	0,161	0,153	0,295	0,409	0,002
BILD	Correlation	0,155	0,146	0,154	0,108	0,451**
	Sig. (2-tailed)	0,259	0,289	0,262	0,433	0,001
Gpspeech	Correlation	0,210	0,199	0,259	0,050	0,333*
	Sig. (2-tailed)	0,124	0,144	0,056	0,719	0,013
F500	Correlation	0,349**	0,231	0,268*	0,007	0,087
	Sig. (2-tailed)	0,010	0,093	0,050	0,962	0,534
F3000	Correlation	0,157	0,150	0,022	0,031	0,237
	Sig. (2-tailed)	0,257	0,279	0,875	0,826	0,084
T500	Correlation	0,131	0,232	0,116	0,127	0,056
	Sig. (2-tailed)	0,346	0,091	0,404	0,360	0,690
T3000	Correlation	0,149	0,261	0,037	0,089	0,239
	Sig. (2-tailed)	0,284	0,057	0,791	0,523	0,082
LD	Correlation	0,009	-0,036	-0,237	-0,015	0,022
	Sig. (2-tailed)	0,947	0,798	0,085	0,915	0,873
Age	Correlation	0,000	0,063	0,319*	-0,069	0,349**
	Sig. (2-tailed)	0,998	0,650	0,017	0,615	0,009