D-2-1: Implementation of a preliminary test set for auditory impairments

**Contractual Date of Delivery:** 31 August 2005 +45 days

**Actual Date of Delivery:** 4 October 2005

**Editor:** W. Dreschler

**Sub-Project/Work-Package:** SP1, WP2

**Version:** Final, v2.0

**Total number of pages:** 85
Deliverable D-2-1

VERSION DETAILS

Version: 2.0
Date: 4 October 2005
Status: Deliverable version 2, final version (after review)

CONTRIBUTOR(S) to DELIVERABLE

<table>
<thead>
<tr>
<th>Partner</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL-AMC</td>
<td>Wouter A. Dreschler</td>
</tr>
<tr>
<td>NL-AMC</td>
<td>Thamar van Esch</td>
</tr>
<tr>
<td>NL-VUMC</td>
<td>Johannes Lyzenga</td>
</tr>
<tr>
<td>SE-LINK</td>
<td>Birgitta Larsby</td>
</tr>
<tr>
<td>NL-AMC</td>
<td>Joyce Vliegen</td>
</tr>
<tr>
<td>NL-AMC</td>
<td>Jeroen Sol</td>
</tr>
<tr>
<td>DE-HZO</td>
<td>Kirsten Wagener</td>
</tr>
</tbody>
</table>

DOCUMENT HISTORY

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Responsible</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>01-03-2005</td>
<td>Thamar van Esch</td>
<td>Inventory of non speech tests, literature on spectro-temporal processing and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Johannes Lyzenga</td>
<td>speech in noise tests</td>
</tr>
<tr>
<td>0.2</td>
<td>18-03-2005</td>
<td>Johannes Lyzenga</td>
<td>Literature on subjective judgments and listening effort added</td>
</tr>
<tr>
<td>0.3</td>
<td>28-04-2005</td>
<td>Brigitta Larsby</td>
<td>Literature on cognitive tests added, lay-out according to template</td>
</tr>
<tr>
<td>0.4</td>
<td>15-06-2005</td>
<td>Joyce Vliegen</td>
<td>Literature on binaural hearing and phoneme perception added</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jeroen Sol</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>14-07-2005</td>
<td>Kirsten Wagener</td>
<td>Auditory profile, speech tests, implementation of tests, additions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wouter Dreschler</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>29-07-2005</td>
<td>Thamar van Esch</td>
<td>Executive summary, Introduction, Implementation, Dissemination. Overall</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wouter Dreschler</td>
<td>editorship</td>
</tr>
<tr>
<td>Version</td>
<td>Date</td>
<td>Reviewed by</td>
<td>Conclusion</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>------------------------------------</td>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td>1.1</td>
<td>August 2005</td>
<td>Jan Koopman, NL-EMC</td>
<td>Accepted with minor modifications</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Andrew Faulkner, UK-UCL</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>4 October</td>
<td></td>
<td>Final version after modifications according to the comments of the reviewers</td>
</tr>
</tbody>
</table>

* e.g. Accept, Develop, Modify, Rework, Update
# Table of Contents

1. Executive Summary ................................................................. 7

2. Introduction .................................................................................. 9
   2.1 Goals of the definition of an auditory profile .............................. 9
   2.2 Areas of interest ........................................................................ 9
   2.3 Limitations of testing in a clinical environment .......................... 10

3. Inventory of audiological tests .................................................... 11
   3.1 Inventory of clinical tests used by the partners in the consortium .... 11
       3.1.1 Inventory of non-speech tests ........................................... 11
       3.1.2 Inventory of speech tests ................................................. 12
   3.2 Review of current literature .................................................... 13
       3.2.1 Tests on audibility and loudness perception ....................... 13
       3.2.2 Tests on frequency resolution and temporal acuity ............... 15
       3.2.3 Tests on the perception of speech elements ....................... 18
       3.2.4 Tests on speech perception in noise .................................. 20
       3.2.5 Tests on spatial hearing ................................................... 25
       3.2.6 Tests on subjective judgments and communication ............. 29
       3.2.7 Tests on listening effort .................................................. 29
       3.2.8 Tests on cognitive abilities .............................................. 30

4. Preliminary selection of tests to be used in an Auditory Profile ......... 36

5. Implementation of tests ............................................................. 39
   5.1 Implementation for phase I testing ........................................... 39
       5.1.1 Spectral and temporal resolution (NL-AMC) ......................... 39
       5.1.2 Perception of speech elements (NL-AMC) ............................ 40
       5.1.3 Spatial hearing (NL-AMC) ............................................... 41
       5.1.4 Cognitive tests (SE-LINK) ............................................... 42
       5.1.5 Hardware ....................................................................... 44
   5.2 Experimental set-up for phase II testing ................................... 44

6. Dissemination and Exploitation .................................................. 45
List of Tables

<table>
<thead>
<tr>
<th>Number</th>
<th>Title</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Inventory of non-speech tests</td>
<td>3.1.1</td>
</tr>
<tr>
<td>3.2</td>
<td>Summary of the inventory of speech tests</td>
<td>3.1.2</td>
</tr>
<tr>
<td>3.3</td>
<td>Open set speech tests, validated for normal hearing listeners</td>
<td>3.2.4</td>
</tr>
<tr>
<td>3.4</td>
<td>Closed set speech tests, validated for normal hearing listeners</td>
<td>3.2.4</td>
</tr>
<tr>
<td>3.5</td>
<td>Open set speech tests, normative data for hearing impaired listeners</td>
<td>3.2.4</td>
</tr>
<tr>
<td>3.6</td>
<td>Closed set speech tests, normative data for hearing impaired listeners</td>
<td>3.2.4</td>
</tr>
<tr>
<td>4.1</td>
<td>Overview of candidate tests for the auditory profile, including goal, some details, indications, outcome, and the proposed category</td>
<td>4.0.0</td>
</tr>
<tr>
<td>4.2</td>
<td>Preliminary test set for the auditory profile (category I and II)</td>
<td>4.0.0</td>
</tr>
</tbody>
</table>

Acknowledgement

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

This report describes the composition of an auditory profile. The auditory profile presented in this deliverable is a central part of the HEARCOM-project. Characterizing the individual auditory capacities by a coherent set of data will open the following possibilities:

- It will become possible to estimate the problems that individual subjects will encounter in adverse communication situations (Subproject SP2).
- It will become possible to design signal processing strategies for compensation of auditory impairments (Subproject SP3, WP5).
- It will become possible to evaluate the benefits of signal processing in advanced listening devices (Subproject SP3, WP7).

The dissemination of the tests needed for the assessment of the auditory profile among professionals will be part of HEARCOM eServices (Subproject SP5, WP11).

The set of tests and the parameters to be derived have a preliminary character, because some knowledge is not yet available. At different sites local tests are in progress to obtain the scientific background for the final choice. However, a detailed inventory among the project partners, supplemented with a thorough literature review and a constructive consensus meeting, makes us believe that the content of chapter 4 is close to the auditory profile in its final form.

The original title was “Implementation of a preliminary test set for auditory impairments in one language”. The phrase “in one language” refers to the fact that the tests developed in this stage will be applied in local centres. Due to the fact that the psychophysical tests are conducted in the Netherlands, and the cognitive facts in Sweden, we decided to skip the second part of the original title. However, the basic idea is unchanged: separate elements from the auditory profile will be developed locally before we compose a common set of auditory tests that will be made available in more languages.

Chapter 2 defines the areas of interest: audibility and loudness perception, frequency resolution and temporal acuity, perception of speech elements, speech perception in noise, spatial hearing, subjective judgments and communication, listening effort, and cognitive abilities. Given the large scope of tests, some limitations have been used to secure that the set of tests selected will be feasible in a clinical environment.

Chapter 3 describes the results of a detailed inventory among the project partners (supplementing the information form the NATASHA project) and the targeted literature review in the areas where additional information was necessary.

Chapter 4 describes the results of the consensus meeting in Gothenburg (June 21th, 2005): a preliminary selection of tests to be used in the auditory profile. The tests cover all areas, but there are three levels of testing: standard, advanced, and specialized. The specialized tests are not considered as a part of the auditory profile.

Chapter 5 describes some details about the implementation of the tests. In phase 1, different tests will be applied locally in order to solve some of the remaining questions about the optimal applicability of some tests. After the local testing, we will conduct a large multi-center study using the final test set of the auditory profile in different languages, using the same hardware and software.
Chapter 6 describes a plan for the dissemination and exploitation of the results from this work package.

Wouter A. Dreschler, WP2-leader
2 Introduction

2.1 Goals of the definition of an auditory profile

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. The scope differs from the work conducted in WP1. Where WP1 aims at screening tests, to be used by large numbers of subjects, WP2 aims at a diagnostic profile that can be assessed in a (specialized) hearing centre or clinic. So, the end user of WP1 is each individual interested in his/her auditory performance. The end user of WP2 is the professional interested in the characteristics of the hearing of a particular client/patient.

The auditory profile should be used to characterize the individual’s auditory impairment profile in a comparable way within Europe. The auditory profile can be used to determine the individual hearing deficiency in communication and can help to determine the benefit from assistive devices. The auditory profile should include all necessary measures to describe all details and differences between different hearing impairments. On the other hand, the auditory profile should minimize redundancy between measures.

2.2 Areas of interest

The components of the auditory profile should be relevant for auditory communication performance. Usually most emphasis is given to speech perception, but the scope of the auditory profile is clearly broader: the profile should also be related to signal recognition, sound quality, spatial hearing, listening comfort, listening effort, and an adequate processing of daily sounds. A limited set of tests will never be able to cover all aspects in detail, but the aim is that the auditory profile is broad enough to cover at least the main parameters in these areas.

More in detail, the partners in WP2 selected the following eight fields for testing:

1. Audibility and loudness perception
2. Frequency resolution and temporal acuity
3. Perception of speech elements
4. Speech perception in noise
5. Spatial listening
6. Questionnaires / subjective judgments and communication
7. Listening effort
8. Cognitive abilities

In each of these fields we made an inventory of the available tests and in some specific areas an additional review of the literature was conducted (see chapter 3).
2.3 Limitations of testing in a clinical environment

As stated before, WP2 aims at a diagnostic profile that can be assessed in a (specialized) hearing centre or clinic. The scope is not primarily on medical diagnosis, but on the diagnosis of communication performance. Therefore, the test included should have a clear relation with communication issues in the broadest sense: speech perception, localisation, sound quality, listening effort, adequate handling of environmental sounds, etc. To be applicable in a clinical environment, also some extra methodological limitations have to be taken into account:

- The tests should be reliable and reproducible
- The tests should not exhibit strong learning effects
- The tests should be efficient. The time consumption of the total test battery should be within reasonable limits, e.g. less than 90 – 120 minutes of testing time.
- The test procedures should be well described
- The test result should present a clear outcome that can be interpreted by the professional end user and can be explained to the hearing impaired individual.
- The test should be applicable in a large variety of hearing impairments

One of the most problematic issues is the large number of relevant areas (see above) versus the limited testing time available. The project group likes to invest in a hierarchical structure in which phase I contains limited measurements in each of the areas of interest and phase II will be more detailed measurements in an area in which problems appear to be present or in an area that is of particular interest for an individual subject.
# 3 Inventory of audiological tests

## 3.1 Inventory of clinical tests used by the partners in the consortium

### 3.1.1 Inventory of non-speech tests

It was decided to update, actualise, and complement the inventory carried out in the EC-project NATASHA (see [http://www.phon.ucl.ac.uk/home/andyf/natasha/index.html](http://www.phon.ucl.ac.uk/home/andyf/natasha/index.html)) for the specific fields of interest (see Section 2.2). The results are presented in the table below.

<table>
<thead>
<tr>
<th>Research field</th>
<th>Test</th>
<th>Procedure</th>
<th>Stimuli</th>
<th>Method</th>
<th>Adapt. proc.</th>
<th>Equipment</th>
<th>Extra info.</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOUDNESS PERCEPTION</td>
<td>Clinical</td>
<td>UCL/MCL</td>
<td>Pure tones, 0.5 - 1 second</td>
<td>Loudness match: Fowler's test</td>
<td>Ascending from comfort level in 5 dB steps until UCL is reached</td>
<td>No</td>
<td>Audiometer (+ external CD-player)</td>
<td>Natasha pages; British Society of Audiology recommended procedures; Arlinger, 1989</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Loudness matching: Tone decay</td>
<td>Pure tones, long duration</td>
<td>Loudness matching: Acapro</td>
<td>Subject presses a button as long as the tone is audible</td>
<td>Level at the better ear is varied until equal loudness is found</td>
<td>Audiometer</td>
<td>Arlinger, 1989</td>
</tr>
<tr>
<td></td>
<td>Research</td>
<td>Loudness matching: Ten test</td>
<td>Pure tones in TEN and in speech etc.</td>
<td>Loudness matching: WHF (contour test)</td>
<td>Categorical scaling</td>
<td>Randomized order</td>
<td>PC-controlled audiometer, headphones / speakers</td>
<td>Natasha pages; Brand and Hohmann, 2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Loudness matching: Bandlimited noises</td>
<td>Pure tones in TEN and in speech etc.</td>
<td>Loudness matching: Bandlimited noises</td>
<td>Categorical scaling</td>
<td>Randomized order</td>
<td>Audiometer, insert headphones</td>
<td>Cox et al, 1997</td>
</tr>
<tr>
<td></td>
<td>Research</td>
<td>Cross-modality matching</td>
<td>All sounds</td>
<td>Cross-modality matching</td>
<td>Match sound level with line length</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## TEMPORAL RESOLUTION

<table>
<thead>
<tr>
<th>Research field</th>
<th>Test</th>
<th>Procedure</th>
<th>Stimuli</th>
<th>Method</th>
<th>Adapt. proc.</th>
<th>Equipment</th>
<th>Extra info.</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical</td>
<td>Gap detection</td>
<td>Noise bands (500 Hz bandwidth) in broadband (or highpass) noise</td>
<td>Detection of temporal gaps in noise bands</td>
<td>Two or three interval forced choice</td>
<td>PC and headphones</td>
<td>Natasha pages</td>
<td>Simple way to obtain index of temporal resolution ability</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temporal integration</td>
<td>Pure tones</td>
<td>Threshold detection of tones of several durations (10-500 ms)</td>
<td>PC and headphones</td>
<td>Natasha pages</td>
<td>Not found to be helpful for predicting speech recognition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research</td>
<td>Psychoacoustical Modulation Transfer Function</td>
<td>Pure tones in modulated noise</td>
<td>Threshold detection of brief tones in noise peaks and valleys</td>
<td>Bekesy tracking</td>
<td>PC and headphones</td>
<td>Linblad et al, 1999</td>
<td>Hearing aid fitting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Procedure</td>
<td>Tone bursts</td>
<td>Detection of the probe masked by the masking just before or just after the probe</td>
<td>ZAFC/3AFC</td>
<td>PC and headphones</td>
<td>Molin et al, 2005</td>
<td>Peripheral hearing capacity of CI users</td>
<td></td>
</tr>
</tbody>
</table>

## FREQUENCY RESOLUTION

<table>
<thead>
<tr>
<th>Research field</th>
<th>Test</th>
<th>Procedure</th>
<th>Stimuli</th>
<th>Method</th>
<th>Adapt. proc.</th>
<th>Equipment</th>
<th>Extra info.</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical</td>
<td>TEN test</td>
<td>Pure tones in Threshold Equalizing Noise</td>
<td>Threshold measurements of tones in TEN and in quiet</td>
<td>Standard Carhart-Jerger method audiometry</td>
<td>Audiometer, CD player, headphones</td>
<td>Moore et al, 2000</td>
<td>Diagnosis of dead regions</td>
<td></td>
</tr>
<tr>
<td>Research</td>
<td>Notched-noise</td>
<td>Pure tones in notch-noise</td>
<td>Threshold for tone as function of notch width (symmetric and asymmetric notches)</td>
<td>Either fixed noise variable tone or fixed tone variable noise</td>
<td>PC and headphones</td>
<td>Natasha pages</td>
<td>Preferred indicator of frequency resolution ability</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Psychoacoustical Tuning Curves</td>
<td>Pure tones and narrow-band noise maskers</td>
<td>Threshold for tone as noise level varied; measured for several masker frequencies</td>
<td>PC and headphones</td>
<td>Natasha pages</td>
<td>Diagnosis of dead regions</td>
<td>Used for measuring frequency resolution</td>
<td></td>
</tr>
</tbody>
</table>

D-2-1: Implementation of a preliminary test set for auditory impairments  
Page 11 of 85
3.1.2 Inventory of speech tests

In each country, there are numerous speech tests using different type of speech material, e.g. monosyllabic words, spondees, numbers, sentences. The validation of this material is sometimes poor. The results of an inventory are presented in Appendix A. In the table below, we selected the most common and best validated speech material in the participating countries.

<table>
<thead>
<tr>
<th>Speech material</th>
<th>Language</th>
<th>Test ID</th>
<th>SRT in noise (normal hearing)</th>
<th>Suitable for severe HI / CI</th>
<th>Available</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plomp-type sentences</strong></td>
<td>Danish</td>
<td>-</td>
<td>Not available</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dutch</td>
<td>Plomp Smoorenburg</td>
<td>-5.6 dB</td>
<td>no</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VU98 LIST</td>
<td>-5.1 dB</td>
<td>no</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-4.0 dB / -4.1 dB</td>
<td>no</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-7.8 dB</td>
<td>yes</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>ASL</td>
<td>-16.8 dB</td>
<td>To a certain degree</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>Wable</td>
<td>to do (BE-LEU)</td>
<td>?</td>
<td>End of 2005</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>Göttingen</td>
<td>-6.2 dB / -5.6 dB</td>
<td>no</td>
<td>CD, software</td>
</tr>
<tr>
<td></td>
<td>Polish</td>
<td>under construction</td>
<td>Not available</td>
<td>-</td>
<td>August 2005</td>
</tr>
<tr>
<td></td>
<td>Swedish</td>
<td>Hallgren</td>
<td>Not available</td>
<td>-</td>
<td>2005/2006</td>
</tr>
<tr>
<td><strong>OIsa-type sentences</strong></td>
<td>Danish</td>
<td>DANTALE II</td>
<td>Not available</td>
<td>yes</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>Dutch</td>
<td></td>
<td>under construction</td>
<td>Not available</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td></td>
<td>under construction</td>
<td>Not available</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td></td>
<td>under construction</td>
<td>Not available</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>Oldenburg (OIsa)</td>
<td>-7.1 dB / -6.3 dB</td>
<td>yes</td>
<td>CD, software</td>
</tr>
<tr>
<td></td>
<td>Polish</td>
<td></td>
<td>under construction</td>
<td>Not available</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Swedish</td>
<td>Hagerman</td>
<td>-7.8 dB / -8.1 dB</td>
<td>yes</td>
<td>CD</td>
</tr>
<tr>
<td><strong>CVC</strong></td>
<td>Danish</td>
<td>DANTALE</td>
<td>-8.7 dB</td>
<td>yes</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>Dutch</td>
<td>NVA</td>
<td>-10.1 dB</td>
<td>yes</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>FAAF</td>
<td>only in quiet</td>
<td>-8.6 dB</td>
<td>yes (?)</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>Fournier</td>
<td>Not available</td>
<td>-</td>
<td>CD</td>
</tr>
</tbody>
</table>

Table 3.1: Inventory of supra-threshold non-speech tests based on a questionnaire sent out to the partners of the consortium. See Appendix C for descriptions of abbreviations.

<table>
<thead>
<tr>
<th>Research field</th>
<th>Test ID</th>
<th>Method</th>
<th>Adapt. proc.</th>
<th>Equipment</th>
<th>Extra info.</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPECTRO-TEMPORAL RESOLUTION</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>research</td>
<td>Stream segregation</td>
<td>Sequence of sinusoidal tone bursts (two different tones)</td>
<td>Subjects indicate when they can no longer hear two separate streams</td>
<td>Frequency difference between the two tones decreases in an exponential manner</td>
<td>PC and headphones</td>
<td>Noise and Moore, 1997</td>
</tr>
<tr>
<td></td>
<td>Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>conform Landy and Arlinger</td>
<td>Pure tones and 4 different noise masksers</td>
<td>Thresholds of tones in noises with temporal / spectral / temporal-spectral / no gaps</td>
<td>Bekesy tracking</td>
<td>PC and headphones</td>
<td>Larsby and Arlinger, 1997</td>
</tr>
<tr>
<td></td>
<td>Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>conform Hilkhuysen et al</td>
<td>Frequency sweeps in spectral or temporal noise grids</td>
<td>Detection of sweeps in grids with varying densities and SN ratios</td>
<td>Subject indicates whether he/she perceived 1,2,3 or no sweeps</td>
<td>Hilkhuysen et al, 2005</td>
<td>A fast and reliable method to measure spectral and temporal resolution</td>
</tr>
<tr>
<td><strong>BINAURAL HEARING</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CLINICAL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uittenweiler test</td>
<td>Speech</td>
<td>Different words at separate ears</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dieroff test</td>
<td>Speech</td>
<td>Freiburg word test with reverberation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Localization</td>
<td>Speech, tone and noise stimuli from various azimuth angles</td>
<td>Identity source direction</td>
<td>no</td>
<td>PC and speakers</td>
<td>Natasha pages</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RESEARCH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BMILD</td>
<td>Low frequency-tone in narrow-band noise</td>
<td>Threshold of tone in diconic and dichotic (180 degree phase difference) conditions</td>
<td>Ascending method</td>
<td>PC and headphones</td>
<td>Johansson and Arlinger, 2002</td>
</tr>
<tr>
<td></td>
<td>BILD</td>
<td>Speech (spondees)</td>
<td>Speech reception in diconic and dichotic (180 degree phase difference) conditions</td>
<td>Ascending method</td>
<td>PC and headphones</td>
<td>Peissig &amp; Kollmeier, 1997; Johansson and Arlinger, 2002</td>
</tr>
</tbody>
</table>
### Table 3.2: Summary of the inventory of speech tests (Appendix A). See Appendix C for descriptions of abbreviations

<table>
<thead>
<tr>
<th>Language</th>
<th>Test</th>
<th>Threshold (dB)</th>
<th>Available</th>
<th>Record Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>Freiburg WaKo rhyme</td>
<td>only in quiet -7.5 dB</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Polish</td>
<td>Pruszewicz</td>
<td>Not available</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Swedish</td>
<td>Magnusson</td>
<td>-2.8 dB</td>
<td>yes</td>
<td>CD</td>
</tr>
<tr>
<td>Danish</td>
<td>DANTALE</td>
<td>only in quiet</td>
<td>yes</td>
<td>CD</td>
</tr>
<tr>
<td>Dutch</td>
<td>VU LINT</td>
<td>-11.2 dB (headphone), -10.0 dB (headphone)</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>French</td>
<td>under construction</td>
<td>Not available</td>
<td>yes</td>
<td>End of 2005</td>
</tr>
<tr>
<td>German</td>
<td>Oldenburg</td>
<td>-9.3 dB (headphone)</td>
<td>yes</td>
<td>July 2005</td>
</tr>
<tr>
<td>Polish</td>
<td>-</td>
<td>Not available</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Swedish</td>
<td>under construction</td>
<td>Not available</td>
<td>yes</td>
<td>July 2005</td>
</tr>
</tbody>
</table>

#### 3.2 Review of current literature

##### 3.2.1 Tests on audibility and loudness perception

We decided not to do a further review of the current literature in the field of audibility and loudness perception, because the clinical methods used are well standardized and they can easily be implemented as soon as audibility measures or loudness perception data are needed in the auditory profile. The only exception is the occurrence of so-called “dead regions”, which may lead to an erroneous interpretation of the pure-tone audiogram. This can have a large impact on the processing power of the ear in a specific range of frequencies.

* A test for the diagnosis of dead regions in the cochlea (Moore et al., 2000)

Dead regions (a region of the basilar membrane with complete loss of inner hair cells) are not easily diagnosed from the audiogram. However, the presence or absence of dead regions may have serious implications for hearing aid fitting. Amplification over a frequency range corresponding to a dead region will not be beneficial and may even impair speech intelligibility.

In this paper, the TEN-test is introduced as a quick and simple method for the diagnosis of dead regions. The test is based upon the detection of sinusoids in the presence of a broadband noise, designed to produce almost equal masked thresholds (in dB SPL) over a wide frequency range for listeners without dead regions. An abnormally high masked threshold at a particular frequency is taken to indicate a dead region at that frequency. Results from the TEN-test are compared to Psychophysical Tuning Curves (PTCs), where a tip shifted away from the signal frequency indicates a dead region at that frequency. Generally, there was a very good correspondence between the results obtained by use of the TEN and the PTCs.

* The use of psychophysical tuning curves to explore dead regions in the cochlea (Moore and Alcantara, 2001)

This paper shows that PTCs can be used to detect and delimit dead regions, but that some caution is needed. The shape of the PTC around the tip can be affected by beat detection and signal level. Moreover, very low signal levels can lead to variable results and for signal frequencies far from the boundary of the dead region required signal levels can become too
high. Finally, PTCs are time-consuming to measure. However, using PTCs, the edge frequency of the dead region can be defined more precisely than with the TEN-test.

* Identifying dead regions in the cochlea: psychophysical tuning curves and tone detection in threshold-equalizing noise (Summers et al., 2003)

In this paper, results of PTCs and TEN-test measurements of 18 ears are compared. TEN and PTC results agreed on the presence or absence of dead regions in 10 of 18 ears (56% agreement rate). In instances where the TEN and PTC results disagreed, the TEN results suggested the presence of dead regions whereas the PTC results did not. The authors consider PTC as a gold standard and therefore conclude that the TEN test is not a reliable diagnostic tool for identification of dead regions. However, the authors did not account for the fact that PTCs can be influenced by e.g. beat detection or perception of difference tones.

* Dead regions in the cochlea: conceptual foundations, diagnosis, and clinical applications (Moore, 2004)

This paper gives a review of the principles underlying the concept of dead regions. A new definition of a dead region is proposed: a dead region is a region in the cochlea where IHCs and/or neurons are functioning so poorly that a tone producing peak basilar-membrane vibration in that region is detected by off-place listening. Both PTCs and the TEN test are described and evaluated. It is argued that the interpretation of PTCs is problematic; detection of combination tones and beats can lead to a tip at the signal frequency even when the signal frequency falls within a dead region. The TEN test also has problems, especially when applied to people with severe or profound hearing loss. However, a new version is being developed which overcomes most of these problems. Moreover, the results of the TEN test are not affected by detection of beats and combination tones.

Evidence (VCV measurements) has been reviewed indicating that people with dead regions extract little or no speech information from frequencies falling well inside a dead region. However, for people with high-frequency dead regions, there does appear to be some benefit from applying amplification up to, approximately, the edge frequency of the dead region. In general, a peak or plateau in VCV scores was found with speech with an upper cut-off frequency within one octave of the edge frequency of the dead region.

* New version of the TEN test with calibrations in dB HL (Moore et al., 2004)

A new version of the TEN test is developed which overcomes some problems that limited the clinical application of the original TEN test. All levels are now expressed in dB HL. This means that absolute thresholds only need to be measured once, as is done routinely as part of a clinical evaluation of hearing. Moreover, calibration is done in such a way that both the TEN(HL) level/ERBN and the test tone levels correspond to the values indicated on the audiometer. Finally, the TEN(HL) bandwidth is restricted, and the TEN(HL) has a very low crest factor. This allows the TEN(HL) level/ ERBN to be increased while avoiding distortion, excessive loudness, and possible further damage to hearing.

Usually, it is sufficient to conduct the TEN(HL) at only one level approximately 10 dB above the absolute threshold in the frequency region of interest. The level of the TEN(HL) can be altered for different test frequencies, it does not have to be the same for all frequencies. If the test is conducted for all 7 frequencies, using one TEN(HL) level per frequency, the testing time is approximately 5 minutes per ear.
3.2.2 Tests on frequency resolution and temporal acuity

3.2.2.1 Frequency resolution

* Frequency resolution measurements with notched noises for clinical purposes (Leeuw and Dreschler, 1994)

Development of a clinical test for measuring spectral resolution, based on the notched-noise procedure (Patterson and Nimmo-Smith, 1980). To reduce measurement time, the number of threshold measurements is minimized (5 instead of 13 notch widths) and the method of measuring thresholds is changed into a Békésy paradigm.

Recently the test is improved further (unpublished) by using the SIAM procedure (see Kaernbach, 1990 for a description of this procedure) instead of the Békésy paradigm. One filter measurement now takes approximately 10 minutes, without being less accurate than the original notched-noise procedure.

* Auditory filter nonlinearity at 2 kHz in normal hearing listeners (Rosen et al., 1998)

Fitting of filter shape measurements from 9 listeners over a wide range of levels and notch widths demonstrates that filter shape is controlled by its output rather than its input level. According to the conventional notched-noise method, thresholds are measured in fixed level noises. However, since filter shape is controlled by its output, the level of the probe should be fixed instead of masker level. Fixing the masker level leads to a derived filter shape that is some kind of average of a number of shapes, caused by the change in probe level as notches are varied.

* Relations between intelligibility of narrow-band speech and auditory functions, both in the 1 kHz frequency region (Noordhoek et al., 2001)

The relation between speech reception (bandwidth) threshold and auditory functions (including spectral resolution) is investigated. To determine spectral resolution, both upward spread of masking (USOM) and downward spread of masking (DSOM) are measured by obtaining detection thresholds of tones in pink noise and in USOM and DSOM maskers. This method has high accuracy (standard errors of 0.7% and 2.6% for USOM and DSOM respectively). Measurement time is not reported. Spectral resolution proves to be important for speech reception.

* A simplified measurement method of auditory filters for hearing-impaired listeners (Nakaichi T et al., 2003)

This paper presents and investigates a new method to measure auditory filters, based on the conventional notched-noise method. The filters are assumed to be symmetric. An ascending procedure is used instead of the original 2AFC. For each auditory filter, the dynamic range of that filter is established first. Next, only one masked threshold (a white noise with a spectral notch is used as a masker) is measured to determine the filter shape. The measurement time is approximately 3 minutes for each auditory filter. It is found that (1) there is no loss of accuracy in threshold measurements compared to the conventional method and (2) ERB data obtained using the new method correlate highly with ERB data obtained using the conventional method.
Spectro-temporal discrimination (just noticeable difference in peak to valley ratio) is measured in CI users and normal hearing listeners. The test signals consist of two parts which are spectral complements of each other (speech shaped noise) in different time orders. The CI users' results are good (not much worse than for normal hearing listeners) for 2 frequency bands, but much worse for more spectral bands. The outcomes are related to speech recognition scores of the CI users. There seems to be a trend between performance in the spectro-temporal discrimination test (with 2 frequency bands) and speech perception, but more research is needed to confirm this. Possible problems of this method are a small learning effect and difficulties to understand the task.

3.2.2.2 Temporal acuity

* The shape of the ear’s temporal window (Moore et al., 1988)

In this paper the temporal resolution of the auditory system is modelled using a temporal window (an intensity weighting function over time) analogous to the auditory filter measured in the frequency domain. To estimate the shape of this hypothetical temporal window, threshold was measured for a brief sinusoidal signal presented in a temporal gap between two bursts of noise. The duration of the gap was systematically varied and the signal was placed both symmetrically and asymmetrically within the gap. Although the temporal-window model failed to predict certain aspects of e.g. forward masking, it was shown that the model successfully accounts for the data from a variety of experiments measuring temporal resolution.

* Hearing aid fitting using psychoacoustical modulation transfer functions (Lindblad, 1993)

The Psychoacoustical Modulation Transfer Function (PMTF) is used to determine temporal resolution: thresholds of tones in 100% intensity-modulated octave-band noise bursts placed around test frequency are measured. The difference in threshold between tones in noise valleys and tones at peaks is a measure of temporal resolution. For normal hearing listeners, this threshold difference is maximal at normal speech level. This level dependence of the threshold difference results from the active nonlinearity of the inner ear (compression). Therefore, PMTF-measurements are used to choose the gain that should be applied to ensure that threshold differences are largest at normal speech level. Speech recognition improved in 7 out of 16 subjects (not significantly) using this fitting strategy.

* Amplitude-modulation detection at low- and high-audio frequencies (Eddins, 1999)

The main goal of this study is to characterize the temporal acuity of the auditory system over a broad range of audio frequencies using the amplitude-modulation detection paradigm. Moreover, the effect of stimulus bandwidth on the shape of the modulation-transfer function is investigated since earlier, some researchers found large bandwidth effect, while others did not find any dependency. Since this discrepancy was possibly due to different stimulus generation techniques, two experiments are conducted using two different stimulus generation techniques. It is found that when the stimuli were band-pass filtered after amplitude modulation (the most common technique), modulation-transfer functions varied systematically with bandwidth (especially for high-modulation frequencies and narrow-band noises), while when the standard stimulus was a quasi frequency modulated noise, the shape of the modulation-transfer-functions was independent of bandwidth. Probably, the effective modulation depth was reduced by filtering after modulation. Temporal acuity is found to be constant over a range of audio frequencies from 600 to 12800 Hz.
* Measurement of auditory temporal processing using modified masking period patterns (Eddins, 2001)

A common metric of auditory temporal processing is the difference in the threshold for a pure-tone signal masked by either unmodulated or amplitude-modulated noise (modified masking period pattern). However, threshold differences are not only results of temporal acuity, but also of frequency resolution and more complex spectro-temporal processes. In this investigation, the influences of signal frequency and masker bandwidth on modified masking period patterns are evaluated. Based on the results, one may conclude that modified masking period patterns may or may not depend upon signal frequency, depending on specific choices of masker bandwidths across frequency, masker levels and perhaps modulation frequency and type of modulation. So, modified masking period patterns depend strongly upon a number of specific stimulus parameters. The interaction among these and perhaps other stimulus variables renders this paradigm a poor choice for clinical measurements of auditory temporal resolution.

* Relations between intelligibility of narrow-band speech and auditory functions, both in the 1-kHz frequency region (Noordhoek et al., 2001)

In this study (see also section 3.2.2.1), also temporal resolution was assessed. This was done by measuring both forward and backward masking. Detection thresholds of clicks in pink noise in forward and backward maskers were obtained. This method has high accuracy (standard errors of 7.3% and 7.0% for forward and backward masking respectively) but measurement time is not reported. It was found that forward masking had a significant influence on speech perception, while backward masking did not.

3.2.2.3 Combined spectral and temporal resolution

* Frequency-temporal resolution of hearing measured by rippled noise (Supin et al., 1997)

Frequency-temporal resolution of hearing is measured in normal hearers using rippled noise stimulation in conjunction with a phase-reversal test: listeners have to detect phase reversals (interchanges of peaks and valleys in the frequency domain) for different reversal rates and ripple densities. A large learning effect is found when measurements are conducted with feedback; subjects appear to discriminate based on spectral coloration instead of reversals. It is tried to eliminate this effect by rejecting feedback. However, it cannot be shown whether this change in paradigm really eliminated this extra cue, since the long-term spectra of alternating and non-alternating stimuli are always different.

* A method for evaluating temporal, spectral and combined temporal-spectral resolution of hearing (Larsby and Arlinger, 1998)

Here, a method for evaluating temporal, spectral and combined-temporal-spectral resolution of hearing is presented. Masked thresholds of tone pulses in four different noises are measured; broadband continuous noise, noise with spectral gaps (around signal frequency), noise with temporal gaps (coinciding with the signals) and noise with both spectral and temporal gaps. A Békésy tracking procedure is used to obtain the masked thresholds. No training is required, measurement time is not reported.
**Auditory temporal and spectral resolution in normal and impaired hearing (Larsby and Arlinger, 1999)**

The method presented in their previous paper is used to measure spectral, temporal and combined resolution in two groups of listeners (normal hearing and hearing impaired). Since many hearing impaired listeners have sloping hearing losses, different masking effects at different frequencies will occur. To avoid such problems, the method is modified by replacing the wide-band pink noise with an octave band noise as masker. It is found that hearing-impaired subjects show less release of masking than normal-hearing subjects, the degree of hearing impairment is inversely related to release of masking. The test-retest reliability was reasonably good, comparable to or better than what is found for regular hearing threshold measurements using a tracking method. There is no need for training for this test.

**Fast and reliable measurements of spectral and temporal acuities in noise for naïve listeners (Hilkhuysen et al., 2004)**

This paper presents a test to measure both spectral and temporal resolution. Listeners detect sweep signals masked by temporal and spectral grid maskers. A presentation contains none to three sweeps and listeners report the number of perceived sweeps. No training is required for this task. An adaptive procedure (1-up 1-down) is used to obtain the masked thresholds. Three detection thresholds can be acquired in approximately 15 minutes and allow to estimate the spectral and temporal acuity in a particular frequency region and position in dynamic range.

**Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired and cochlear implant listeners (Henry et al., 2005)**

In this study, spectral peak resolution was investigated in normal hearing, hearing impaired and cochlear implant users. To do this, the ‘ripple phase reversal test’ (Supin et al., 1997) was used, since this test gave the possibility to directly compare spectral peak resolution among listeners with acoustic hearing (normal or impaired) and listeners with electric hearing. A significant relationship was found between spectral peak resolution and both vowel and consonant recognition scores in quiet across the three listener groups. These results suggest that efforts to improve spectral peak resolution for HI and CI users may lead to improved speech recognition.

### 3.2.3 Tests on the perception of speech elements

The perception of speech elements is usually tested with short nonsense syllables of the consonant-vowel (-consonant) type (VCV, or CVC). The recognition of consonants is of special interest because hearing-impaired persons typically have more difficulties perceiving consonants than vowels. For scoring consonant recognition percentages are used as well as patterns of consonant confusions.

Bosman (1989) tested recognition and confusion for natural CVC words. It was found that listening strategies were in agreement with audiometric configurations: hearing-impaired listeners with a high frequency hearing loss made more use of the first formant. Hearing-impaired subjects also made more intensive use of the duration of the vowels than normal-hearing listeners did. For subjects with presbyacusis and Ménière’s disease the ratio of vowel and consonant scores was much lower relative to normal-hearing subjects and subjects with noise-induced hearing loss. The perception of the initial and final consonant was cued by voicing, sonority, frication and voicing, sonority/glide, sibilance, respectively.

The perception of speech elements can also be tested using processed stimuli. Coarticulation is an important issue when processing syllables. Part of the information from consonants is
also found in the vowel of a syllable and this complicates defining the moment of transition from vowel to consonant and vice versa. In literature various techniques are used, but in general detection of the transition is supported by the use of waveform editing tools or the spectrogram.

Processed speech stimuli can be used for diagnostic reasons. In this case the signal is degraded in order to test the vulnerability of a subject’s recognition to, for example, filtering. Processing can also be applied in an attempt to improve the perception of speech elements. This processing method can be integrated in speech-algorithms of hearing aids when proved effective. An example of an attempt to improve speech perception is consonant enhancement. For consonant enhancement the C-V intensity ratio (CVR) is increased by either increasing the intensity of the consonant or decreasing the intensity of the vowel, or a combination of both. After processing stimuli are presented and scored. Processing of short utterances concentrates on three important distinctions between consonants and vowels regarding frequency, intensity and duration.

3.2.3.1 Frequency

* Stop-consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. I: The contribution of selected frequency regions (Dubno et al., 1989)

Dubno et al. (1989) investigated the contribution of the spectral shapes associated with place of articulation to consonant perception among hearing-impaired and normal-hearing listeners. They also tried to characterize the differences for both subject groups. Nonsense CV syllables were processed using five low-pass filter frequencies and six high-pass filter frequencies. Recognition and confusion were registered at various presentation levels. Differential filtering effects on stop-consonant place perception were consistent with the spectral properties of the stimuli. They found that the differences in stop-consonant place perception between normal-hearing and hearing-impaired subjects could largely be explained by the difference in audiometric configurations.

* Log-linear modeling of consonant confusion data (Bell et al., 1986)

Bell et al. (1986) applied log-linear modeling and statistical analyses on CV consonant confusion data of Miller and Nicely (1955). Low-pass filtering significantly affected error patterns when categorized by place of articulation, duration, or nasality; whereas, high-pass filtering only affected voicing and frication error patterns. Their conclusions differ from the findings of Miller and Nicely (1955) who reported the confusions for high-pass filtering to be random. The original finding that low-pass filtering and noise have similar effect on confusions was not supported by the log linear model.

3.2.3.2 Intensity

* Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing (Gordon-Salant, 1986)

Gordon-Salant (1986) found an improvement of consonant recognition for normal-hearing and hearing-impaired subjects when C-V intensity ratio (CVR) was increased with 10 dB. Similar results were obtained by Montgomery and Edge (1988) who found that consonant recognition at 65 dB SPL improved by 10\% after adjusting the consonant intensity to a CVR of 0 dB. However, at a stimulus level of 95 dB SPL the increase in recognition was relatively small.
* Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners (Kennedy et al., 1998)

Kennedy et al. (1998) studied the effect of adjusting the CVR on consonant recognition. The CVR of VC syllables was adjusted in steps of 3-6 dB by increasing the intensity of the consonant. Six types of consonant enhancement functions were obtained showing various effects on percentage consonant recognition with increasing consonant enhancement. Individualized adjustment of CVR produced improvement of consonant recognition for some subjects. For approximately half of the subjects the individualized adjustment of CVR produced little or no effect.

*The role of consonant-vowel amplitude ratio in the recognition of voiceless stop consonants by listeners with hearing impairment (Sammeth et al., 1999)

Sammeth et al. (1999) investigated the recognition of voiceless stop-consonants in isolation and in CV syllables. The CVR was increased by reducing the energy of the vowel by 6 dB and 12 dB. Presenting the consonants in isolation (infinite vowel reduction) showed a decrease in recognition score with respect to CV syllables. Results indicated that adjusting the CVR gave no significant improvement in recognition.

* Consonant perception in quiet: effect of increasing the consonant-vowel ratio with compression amplification (Hickson and Byrne, 1997)

Hickson and Byrne (1997) studied the effects of an increase in CVR on consonant perception. CVR increase was produced by syllabic compression via a hearing aid. Linear and compression amplified CV and VC nonsense syllables were presented to normal-hearing and hearing-impaired listeners by a loudspeaker. No significant difference was found in overall scores with type of amplification for both subject groups. Subjects with normal-hearing showed improved perception of voiceless fricatives with compression. For hearing-impaired listeners the perception of voiceless stop-consonants decreased with compression. They concluded that for hearing-impaired listeners the CVR is a cue to the perception of some consonants.

3.2.3.3 Duration

* Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing (Gordon-Salant, 1986)

Gordon-Salant (1986) found a small to negligible improvement of consonant recognition for normal-hearing and hearing-impaired subjects after doubling consonant duration. Montgomery and Edge (1988) found similar results at 65 dB SPL after consonant duration was doubled. The improvement in recognition was twice as large at a stimulus level of 95 dB SPL.

3.2.4 Tests on speech perception in noise

This overview of speech reception tests is intended for those who are planning to develop such a speech reception test, possibly in a new language or with different specifications than existing tests. The overview does not attempt to be complete, it only tries to summarize the similarities and dissimilarities of different speech intelligibility tests that are in much use or have been in much use. This overview concentrates on tests that use sentence material, but word tests are also much used in daily practice.
Amongst the first speech-perception tests using everyday words were the so-called rhyme tests (e.g., Fairbanks, 1958), CVC tests based on earlier CV tests (e.g., Miller and Nicely, 1955). In such tests, CVCs are presented in noise while only the first or last consonant or central vowel are varied and the other two phonemes are fixed and known to the listener. The number of alternatives for the rhyming CVCs is usually set between four and six. Lists are created by combining different, rhyming, CVC sets. Later, closed response-set versions were newly developed, or constructed by modifying open response-set rhyme tests, to assure an equal number of alternatives for each CVC and to allow more accurate phonemic balancing of the lists of CVCs used in the tests (e.g., for American English: House et al., 1965; for British English: Foster and Haggard, 1986; for German: Sotscheck, 1982; von Wallenberg and Kollmeier, 1989). Extended evaluations of such modified rhyme-tests were performed (e.g., for American English: Williams et al., 1973; for British English: Foster and Haggard, 1987; Shields and Campbell, 2001; for German: Kollmeier et al., 1992; and specifically for children: Kliem and Kollmeier, 1995; Brand et al., 1999). In addition, a two-syllable test has been developed in German (Kliem and Kollmeier, 1992). Rhyme tests are often used for speech-reception testing in children since they depend less on language development than sentence tests (e.g. for American English: Nabelek and Robinson, 1982; for German: Steffens, 2003). There are strong indications that closed response-set rhyme tests show very little learning effects (e.g. House et al., 1965; Williams et al., 1973; Lutman, 2005), but Kollmeier et al (1992) demonstrated they are not well suited for an analysis of consonant and vowel confusions using the confusion-matrix technique of Miller and Nicely (1955).

The advantage of using sentences, rather than isolated words or syllables, is that this incorporates a larger part of the language system and thus resembles everyday listening conditions more closely than using isolated utterances (e.g., Plomp and Mimpen, 1979a; Kollmeier and Wesselkamp, 1997). In Appendix A, a list is given of the most important specifications of a number of sentence tests along with a summary of the construction principles of their sentence materials. The described tests all aim to approach natural listening conditions as closely as possible.

The tests listed in Appendix A all use sentences that the listener has to reproduce (repeat) as closely as possible. But they differ in nature of scoring (keyword, keywords, all words separately, or whole sentences as a block), method (fixed-SNR or adaptive presentation methods), noise (mainly multi-talker babble or speech-shaped noises), and construction of sentence material (different optimization factors have been used in the different tests). Most of them have been developed for use with stationary noises, and have been evaluated using those types of noises. The use of fluctuating noises, in relation to stationary noises, will be discussed further below.

The majority of the listed tests use an open set of sentences, a few use a closed sentence set which, of course, leads to a closed response set. Closed response-set tests often suffer somewhat more from learning effects than open response-set tests. For example for closed response-set tests, Hagerman (1982) initially found a learning effect for normally-hearing listeners of 0.26 dB per list over four lists, whereas in a later investigation (Hagerman and Kinnefors, 1995) a lower value of 0.1 dB per list with 10 sentences each was found. However, for hearing-impaired listeners Hagerman and Kinnefors found a learning effect of 2 dB per list. For a closed response-set test, Wagener (1999a; 1999b; 1999c) found a learning effect of 0.3 dB/list over six lists with 20 sentences each (of which 1 dB occurred between the first two lists) for normally-hearing listeners. For an open response-set test, Plomp and Mimpen (1979a) found a fairly constant learning effect of about 0.1 dB per list over ten lists with 13 sentences each in normally-hearing listeners, but this value can be expected to be somewhat higher in hearing-impaired listeners (probably of the order of 0.5 dB per list). However, closed response-set tests can be used repeatedly for the same listener, as redundancy is low. This makes them suitable for lengthy testing as, for example, often occurs with cochlear-implant patients, hearing aid fitting, or in research. As the amount of speech material is relatively small compared to open response-sets, the
homogeneity of the speech material in closed response-set tests is high since all test lists consist of the same word material. Open response-set tests usually show relatively small learning effects when the test lists are used only once, and may therefore have a slightly better accuracy when only a short amount of time is available, but once a listener has heard a particular list it cannot be used again on that same listener until any memory of it has faded. In most countries the amount of speech material for open-set tests is too limited for extensive testing. In summary, open-set tests are suitable for quick and accurate assessments of speech reception, while closed-set tests may be more suited for prolonged testing on the same listener.

Compared to scoring keywords or whole sentences, scoring all the words separately requires more effort, and possibly time and expertise, of the experimenter, but this is paid back in increased measurement efficiency, often enabling quicker measurements (Brand and Kollmeier, 2002). Using all the information in the sentence, by scoring all words separately or scoring the whole sentence as a block, approximates natural listening conditions better than scoring one or a few keywords (e.g., Plomp, 1986; Gelfand et al., 1988).

Optimizing sentence lists for consistency of list-scores, by careful selection of sentences or modifying signal-to-noise ratios of individual sentences, usually leads to steeper psychometric functions. However, this is only entirely reliable for the conditions under which the optimization was performed (in all cases so far, normally-hearing listeners were used for that). For that reason, some tests evade this issue by using more arbitrary ways to equalize sentences such as equal RMS levels (e.g., Versfeld et al., 2000).

Using adaptive methods eliminates the need of a pilot to assess the correct presentation levels, and thus saves time. Modern tests produce results that are at least as accurate as those of fixed level methods (e.g., Plomp and Mimpen, 1979a; Hagerman and Kinnefors, 1995; Brand and Kollmeier, 2002).

Using a multi-talker babble approximates natural listening conditions better than speech-shaped noises, but the latter have better controlled statistics and produce accurate results in a less time (e.g., Shields and Campbell, 2001).

When the number of talkers in the background babble is low, strong fluctuations occur in the temporal envelope of the noise. This is comparable to using strongly modulated noises. Studies comparing the use of stationary and fluctuating noises usually show greater accuracy for stationary noises than for fluctuating noises, as well as larger differences between the scores of normally-hearing and hearing impaired listeners for fluctuating noises than for stationary noises (e.g., Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Takahashi and Bacon, 1992; Festen, 1993; Gustafsson and Arlinger, 1994; Eisenberg et al., 1995; Hagerman, 1997, 2002; Bacon et al., 1998; Dubno et al., 2002; Nelson et al., 2003; Summers and Molis, 2004). From these studies it appears that hearing-impaired listeners are very poor in using the speech information that is physically present in the gaps of the fluctuating noises. Other studies show that there are large interindividual differences in benefit from fluctuating noise (Wagener and Brand, 2005, Wagener et al, 2005). Part of this effect appears to be related to audibility (elevated thresholds), the remainder is often interpreted as being caused by a loss of temporal acuity and possibly by informational masking.

At present, existing sentence tests have been, or are being, translated and validated in a large number of languages, for example the HINT test (based on the English BKB test) has been translated into: Latin American Spanish (Soli et al., 2002), Canadian French (Vaillancourt et al., 2004), Japanese, Cantonese (Wong and Soli, 2005), Mandarin, Swedish (in progress), etc. A new Flemish test (van Wieringen and Wouters, 2005) is based on the Dutch Plomp-and-Mimpen test (Plomp and Mimpen, 1979a). OLSa type sentence tests are available in Swedish (Hagerman 1982), German (Wagener et al., 1999a; 1999b; 1999c),
Danish (Wagener et al., 2003). Translations to American English, British English, Dutch, and French (last three within HearCom) are in progress.

It seems superfluous to include a detailed description of all of those tests, as they are all based on tests which are described in Appendix A.

In many languages sentence lists are not yet available but word lists have been recorded and validated; i.e., Castilian Spanish (Cárdenas and Marrero, 1994), Rumanian (Dante, 1995), Hindi (De Sa, 1973), Arabic in Moroccan accent (Messouak, 1956), literary Arabic in Baghdad accent (Alusi et al., 1974), literary Arabic in Saudi Arabia accent (Ashoor and Prochazka, Jr., 1982), Arabic in Sudanese accent (Onsa, 1984), Thai, the Congolese Bantu languages Lingala and Ciluba (Muyunga, 1974), the Australian Aboriginal languages Warlpiri and Tiwi (Plant, 1990), etc.

For practical implementations, the more important specifications of a test are the availability of the test material and specification of the audiometric results for normally hearing listeners: average SRT, slope of the psychometric function, measurement accuracy or test-retest reliability. Table I gives these values for the sentence tests that have been evaluated and tested most extensively in the literature.

<table>
<thead>
<tr>
<th>Open-set tests</th>
<th>SRT</th>
<th>Slope</th>
<th>Acc.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPIN, PH sent.</td>
<td>-2.0</td>
<td>14</td>
<td>2.3</td>
<td>Gelfand et al., 1988</td>
</tr>
<tr>
<td>HINT</td>
<td>-2.9</td>
<td>-</td>
<td>0.8</td>
<td>Nilsson et al., 1994</td>
</tr>
<tr>
<td>IHR ASL</td>
<td>-8.6</td>
<td>-</td>
<td>≈ 1.0</td>
<td>Moore et al., 2001</td>
</tr>
<tr>
<td></td>
<td>-10</td>
<td>-</td>
<td>≈ 1.2</td>
<td>Alcantara et al., 2003</td>
</tr>
<tr>
<td>Plomp, female</td>
<td>-5.6</td>
<td>15-20</td>
<td>0.9</td>
<td>Plomp and Mimpen, 1979a</td>
</tr>
<tr>
<td></td>
<td>-4.5</td>
<td>15.9</td>
<td>1.1</td>
<td>Versfeld et al., 2000</td>
</tr>
<tr>
<td>Plomp, male</td>
<td>-5.1</td>
<td>18</td>
<td>1.8</td>
<td>Smoorenburg, 1992</td>
</tr>
<tr>
<td></td>
<td>-3.7</td>
<td>17.7</td>
<td>1.0</td>
<td>Versfeld et al., 2000</td>
</tr>
<tr>
<td>VU98, female</td>
<td>-4.1</td>
<td>16.6</td>
<td>1.1</td>
<td>Versfeld et al., 2000</td>
</tr>
<tr>
<td>VU98, male</td>
<td>-4.0</td>
<td>15.2</td>
<td>1.1</td>
<td>Versfeld et al., 2000</td>
</tr>
<tr>
<td>Göttingen sent.</td>
<td>-6.2</td>
<td>19</td>
<td>0.3</td>
<td>Kollmeier and Wesselkamp, 1997</td>
</tr>
</tbody>
</table>

Table 3.3: Open set speech tests, validated for normal hearing listeners. See Appendix C for descriptions of abbreviations and Appendix A for information about the tests.
<table>
<thead>
<tr>
<th><strong>Closed-set tests</strong></th>
<th>SRT</th>
<th>Slope</th>
<th>Acc.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hagerman sent.</td>
<td>-8.1</td>
<td>25</td>
<td>0.44</td>
<td>Fixed SN, Hagerman, 1982; Hagerman, 1984</td>
</tr>
<tr>
<td></td>
<td>-7.8</td>
<td>-</td>
<td>0.6</td>
<td>Adaptive, Hagerman and Kinnefors, 1995</td>
</tr>
<tr>
<td>Oldenburg sent.</td>
<td>-7.1</td>
<td>17</td>
<td>0.2</td>
<td>Wagener et al., 1999a; Wagener et al., 1999b; Wagener et al., 1999c</td>
</tr>
<tr>
<td>Dantale II</td>
<td>-8.43</td>
<td>13.2</td>
<td>0.95</td>
<td>Wagener et al., 2003</td>
</tr>
</tbody>
</table>

*Table 3.4: Closed set speech tests, validated for normal hearing listeners. The accuracies need to be multiplied by √2 before comparing them to the values in table 3.3. See Appendix C for descriptions of abbreviations and Appendix A for more information about the tests.*

<table>
<thead>
<tr>
<th><strong>Open-set tests</strong></th>
<th>SRT</th>
<th>Slope</th>
<th>Acc.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPIN, PH sent.</td>
<td>3.0</td>
<td>-</td>
<td>-</td>
<td>Gelfand et al., 1988</td>
</tr>
<tr>
<td></td>
<td>-1.0</td>
<td>-</td>
<td>≈3.0*</td>
<td>72 dB SPL, Dubno et al., 1984</td>
</tr>
<tr>
<td>SPIN, PL sent.</td>
<td>1.0</td>
<td>-</td>
<td>≈5.0*</td>
<td>72 dB SPL, Dubno et al., 1984</td>
</tr>
<tr>
<td>IHR ASL</td>
<td>-5.6</td>
<td>-</td>
<td>≈2.0*</td>
<td>Moore et al., 2001</td>
</tr>
<tr>
<td></td>
<td>-5.8</td>
<td>-</td>
<td>≈3.0*</td>
<td>Alcantara et al., 2003</td>
</tr>
<tr>
<td>Plomp, female</td>
<td>-3.5</td>
<td>-</td>
<td>≈1.5*</td>
<td>Nn, Bronkhorst and Plomp, 1990</td>
</tr>
<tr>
<td>Plomp, male</td>
<td>-3.9</td>
<td>-</td>
<td>0.9</td>
<td>Mixed grp., Smoorenburg, 1992</td>
</tr>
<tr>
<td>Göttlingen sent.</td>
<td>-0.5</td>
<td>13</td>
<td>0.9</td>
<td>Brand and Kollmeier, 2002</td>
</tr>
<tr>
<td>Audivox sent.</td>
<td>-0.7</td>
<td>-</td>
<td>1.1</td>
<td>Wable, 2005</td>
</tr>
</tbody>
</table>

*Table 3.5: Open set speech tests, normative data for hearing impaired listeners. For descriptions of abbreviations, see Appendix C, and for more information about the tests, see Appendix A.*
### Closed-set tests

<table>
<thead>
<tr>
<th></th>
<th>SRT</th>
<th>Slope</th>
<th>Acc.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hagerman sent.</td>
<td>-0.1</td>
<td>20</td>
<td>0.9</td>
<td>Hagerman, 1984</td>
</tr>
<tr>
<td>Hagerman sent.</td>
<td>-2.2</td>
<td>-</td>
<td>0.7</td>
<td>Hagerman and Kinnefors, 1995</td>
</tr>
<tr>
<td>Oldenburg sent.</td>
<td>-2.9</td>
<td>14</td>
<td>0.7</td>
<td>Brand and Kollmeier, 2002</td>
</tr>
</tbody>
</table>

*Table 3.6: Closed set speech tests, normative data for hearing impaired listeners. The accuracies need to be multiplied by \( \sqrt{2} \) before comparing them to the values in table 3.5. SRT and accuracy in dB, slope in %/dB. For descriptions of abbreviations, see Appendix C, and for more information about the tests see Appendix A.*

* means that these accuracies have been estimated rather roughly.

The tests in the tables show a considerable spread in standard SRT values and slopes of the psychometric functions. The differences in SRT values can partly be explained by the applied scoring methods. The SRTs of sentence tests with sentence scoring are higher than the SRTs of sentence tests with word scoring. For example, the standard SRTs for the SPIN (which uses twelve-talker babble rather than speech-shapes noise) and HINT tests are relatively high. But even though Nilsson et al. (1994) compared their stimulus material extensively to that of Plomp and Mimpfen (1979a), they failed to find any clear indications for their deviant SRT values. A similar unexplained difference is seen between the slope values of the Oldenburg and the Dantale II sentences, which were both constructed in the same careful way. One possible explanation can be seen from comparing the results for the Dutch sentence materials. The SRTs for the Plomp female sentences are lower than those for the other lists and speakers. This could be an effect of speaker quality, as an exceptionally good speaker was used for the Plomp female sentences. This option would also explain the very low SRTs for the IHR ALS sentences, as an exceptionally good speaker was used to record those sentences as well. Though these speculations are in good agreement with the experiences with daily use of the tests, at present they are not yet backed up by any hard evidence.

Appendix A contains lists of the most important specifications of a number of important speech tests, along with a summary of the construction principles of their sentence materials. This inventory does not claim to be complete, word-list and sentence-list tests are continuously being developed in new languages.

### 3.2.5 Tests on spatial hearing

#### 3.2.5.1 Free-field localisation

For free-field sound localisation, most existing set-ups consist of a sphere or a hemisphere of about 5 – 11 equally spaced speakers, mostly only in azimuth (at 0° elevation, defined as the level of the subject's ears). The subject is seated in the centre of the (hemi)sphere and has to indicate (verbally or on a response box) from which speaker the sound is emanating. Sounds vary from white or pink noise bursts to spoken sentences, with or without background noise. In most tests the subjects are instructed to keep their head still in the straight-ahead direction. This is important, as there is some controversy whether head movements might facilitate sound localisation. Most tests also apply small variations in stimulus level (roving), to reduce level cues.
* Kobler and Rosenhall, 2002, “Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification”

<table>
<thead>
<tr>
<th>Speaker setup</th>
<th>8 speakers: 360 deg in 45 deg steps in the horizontal plane (1.7 m radius)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimuli</td>
<td>Swedish sentences consisting of five words</td>
</tr>
<tr>
<td>Sound level</td>
<td>75 dB SPL</td>
</tr>
<tr>
<td>Background noise</td>
<td>speech-shaped noise, signal-to-noise ratio: +4dB presented from all 7 other speakers</td>
</tr>
<tr>
<td>Response method</td>
<td>repeat speech signal and point out target loudspeaker</td>
</tr>
<tr>
<td>Scoring</td>
<td>percentage correctly identified loudspeakers</td>
</tr>
<tr>
<td>Room</td>
<td>sound-insulated room</td>
</tr>
<tr>
<td>Head movement</td>
<td>head still at straight ahead position</td>
</tr>
</tbody>
</table>

* Lorenzi et al., 1999, “Sound localization in noise in hearing impaired listeners”

<table>
<thead>
<tr>
<th>Speaker setup</th>
<th>11 speakers: -90 to +90 deg in 18 deg steps in the horizontal plane (1.25 m radius)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimuli</td>
<td>train of 23-µs pulses with a repetition rate of 100 Hz and duration of 300 ms broadband (low-pass filtered at 11 kHz) low pass (low-pass filtered at 1.6 kHz)</td>
</tr>
<tr>
<td>Sound level</td>
<td>70 dB SPL</td>
</tr>
<tr>
<td>Background noise</td>
<td>900-ms white noise, 52, 58, 64, 70, 76, 79 dB SPL, signal-to-noise ratio: -9 - +18 dB presented at -90, 0, or +90 deg</td>
</tr>
<tr>
<td>Response method</td>
<td>response box (with a matrix of buttons configured like the speaker matrix)</td>
</tr>
<tr>
<td>Scoring</td>
<td>stimulus/response plot, rms localization error, consistency (r2)</td>
</tr>
<tr>
<td>Room</td>
<td>sound-treated room</td>
</tr>
<tr>
<td>Head movement</td>
<td>head still at straight ahead position (head movement allowed between trials)</td>
</tr>
</tbody>
</table>

* Noble et al., 1994, “Effects on sound localization of configuration and type of hearing impairment”

* Noble et al., 1997 “Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners”

<table>
<thead>
<tr>
<th>Speaker setup</th>
<th>20 speakers: -90 to +90 deg in 18 deg steps in the horizontal plane and -72 to +90 deg in 18 deg steps in the vertical plane (1.22 m radius)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimuli</td>
<td>pink noise, pulses of 150 ms, 50 ms inter-pulse interval, total duration of 0.9 s</td>
</tr>
<tr>
<td>Sound level</td>
<td>MCL and ½ MCL (± 3 dB level roving)</td>
</tr>
<tr>
<td>Background noise</td>
<td>no background noise</td>
</tr>
<tr>
<td>Conditions</td>
<td>frontal horizontal plane and median vertical plane: subject is sitting facing the loudspeaker array</td>
</tr>
<tr>
<td></td>
<td>lateral horizontal and vertical plane: subject has turned 90 deg to the left, so that the loudspeaker array is lateral</td>
</tr>
<tr>
<td>Response method</td>
<td>verbally identify target loudspeaker</td>
</tr>
<tr>
<td>Scoring</td>
<td>percentage correct, localization error</td>
</tr>
<tr>
<td>Room</td>
<td>anechoic room</td>
</tr>
</tbody>
</table>
Head movement subjects were free to move their head/torso while the stimulus was active.


* Seeber et al., 2004, "Localization ability with bimodal hearing aids and bilateral cochlear implants”

Speaker setup 11 speakers: -50 to +50 deg in 10 deg steps in the horizontal plane (1.95 m radius)
Stimuli Gaussian white noise, five pulses of 30 ms separated by 70 ms pauses
Sound level 64-76 dB SPL (randomly varied in 3-dB steps to prevent level cues)
Background noise no background noise
Conditions bimodal subjects: CI only, HA only, CI + HA
bilateral subjects: left CI, right CI, or binaural CI
Response method adjust laser spot to match the position of the perceived sound
Scoring localization error (as a function of stimulus location), gain, bias linear fit
Room darkened anechoic room
Head movement head still at straight ahead position

* Tyler et al., 2002 “Patients utilizing a hearing aid and a cochlear implant: speech perception and localization”

Speaker setup 3 speakers: -45, 0, 45 deg in the horizontal plane
Stimuli four 200-ms bursts of speech noise, 200 ms inter-burst interval
Sound level 73, 77, 83 dB SPL
Background noise no background noise
Response method indicate target loudspeaker (left or right)
Scoring percentage correct
Head movement head still at straight ahead position

* van Esch et al., 2005, "Horizontal directional hearing in subjects with unilateral conductive hearing loss”

Speaker setup 9 speakers: -120 to +120 deg in 30 deg steps in the horizontal plane (1.3 m radius) (the extreme positions were not used to avoid edge effects)
Stimuli third-octave band at 500 or 3000 Hz of about 1 sec
Sound level 40-70 dB SPL
Background noise no background noise
Response method indicate target loudspeaker
Scoring mean absolute localization error
Room standard sound treated room
high reverberance room
low reverberance room
Head movement head still at straight ahead position

* van Hoesel and Tyler, 2003, “Speech perception, localization, and lateralization with bilateral cochlear implants”

Speaker setup 8 speakers: -54 to +54 deg in 15.5 deg steps in the horizontal plane (1.4 m radius)
Stimuli: four 170-ms pink-noise bursts, 50 ms inter-burst interval, total duration of 830 ms
Sound level: 65 dB SPL (± 4 dB level roving)
Background noise: no background noise
Conditions: left CI, right CI, or binaural CI
Response method: indicate target loudspeaker
Scoring: stimulus/response plot, rms error
Room: anechoic room
Head movement: head still at straight ahead position

* Wilmington et al., 1994, “Binaural processing after corrected congenital unilateral conductive hearing loss”

Speaker setup: 7 speakers: -26.5 to +26.5 deg (0, ±9.8, ±19.1, ±26.5 deg, horizontally aligned) the extreme positions were not used to avoid edge effects
Stimuli: trains of 10 clicks
Sound level: 30 dB HL
Background noise: no background noise
Response method: indicate target loudspeaker
Scoring: mean absolute localization error
Room: double-walled, double-floored, sound-attenuating room
Head movement: head still at straight ahead position

3.2.5.2 Virtual localisation

For virtual localisation with headphones, head-related transfer functions (HRTFs) are used to filter the sounds. HRTFs are usually measured with a miniature microphone placed in the ear canal near the eardrum while sounds that vary in azimuth and elevation are being presented. But ‘generic’ HRTFs can also be measured with a dummy head, so that it is not necessary to measure HRTFs for each subject. Sounds filtered with the appropriate HRTF for a certain direction and presented through headphones will be perceived as originating from that direction outside the head.

The big advantage of virtual localisation over free-field localisation is that it is not necessary to have a separate room with a set-up of several speakers. Virtual localisation behaviour can be tested in a standard audiometry room. Moreover, confounding factors like sound-field calibration and head movements are eliminated. The disadvantage is first that, as this test is done over headphones, it cannot be done with hearing aids. Second, dummy head HRTFs are ‘average’ HRTFs and they are not optimal for all subjects. Therefore the sound image resulting from filtering with these HRTFs might vary slightly for different subjects. Middlebrooks (1999b; 1999a) has done extensive research on virtual localisation with own-ear HRTFs and other-ear HRTFs. He found that localisation was less accurate in the other-ear condition and listeners showed systematic undershoots or overshoots in the horizontal direction when listening through HRTFs from smaller or larger listeners, respectively. However, these errors were small, and were mostly caused by the smallest listeners localising with HRTFs from the largest listeners and vice versa, so for localisation with a limited number of speakers, the dummy-head HRTFs are probably entirely adequate.

* A test of virtual auditory localization (Besing and Koehnke, 1995)

Besing and Koehnke (1995) developed a virtual localisation test for clinical use with 9 virtual speakers at -90 to +90 deg in 22.5 deg steps. They used three-word phrases as stimuli, convolved with KEMAR dummy-head HRTFs, measured both in an anechoic and a reverberant room. They found that the test was easy to administer to both children and
adults and was reliable and sensitive to the effects of auditory pathology. They used the same procedure to test speech intelligibility in noise (Koehnke and Besing, 1996), with three speakers, at -90, 0, and +90 deg.

<table>
<thead>
<tr>
<th>Speaker setup</th>
<th>9 speakers: -90 to +90 deg in 22.5 deg steps (simulated locations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimuli</td>
<td>three-word phrases ('mark the spot', 'fly by night', 'crack the walnut', 'come here later') convolved with KEMAR dummy-head HRTFs</td>
</tr>
<tr>
<td>Sound level</td>
<td>70 dB SPL (at straight ahead)</td>
</tr>
<tr>
<td>Background noise</td>
<td>no background noise</td>
</tr>
<tr>
<td>Response method</td>
<td>press 1-9 on computer keyboard to identify the target speaker</td>
</tr>
<tr>
<td>Scoring</td>
<td>percentage correct, mean error, rms error</td>
</tr>
</tbody>
</table>

### 3.2.6 Tests on subjective judgments and communication

Appendix B gives an overview of self-report questionnaires concerning hearing disability and handicap is based on the comprehensive overview given by Bentler and Kramer (2000), extended with the later SSQ and IOI questionnaires. The overview mainly focuses on questionnaires of which publications are available in English and that have been validated. Some questionnaires specifically target the elderly, e.g. the Communication Scale for Older Adults (CSOA), Nursing Home Hearing Handicap Index (NHHI), or the prelingually deafened, e.g., the Communication Self Assessment Scale Inventory for Deaf Adults (CSDA). Some concentrate on staff or significant others, e.g., Nursing Home Hearing Handicap Index (NHHI), and the International Outcome Inventory for Hearing Aids for significant Others (IOI-AH-SO). A large number of questionnaires, intended to assess hearing disability, mainly concentrate on speech-intelligibility problems. A number of inventories include other aspects of disability, such as localization problems and the perception of non-speech sounds: the Hearing Measure Scale (HMS), the Gothenburg Profile (GP), the Amsterdam Inventory, and the Speech, Spatial, and Qualities of Hearing Scale (SSQ). Some lists also target stress or listening effort: the Communication Profile for the Hearing Impaired (CPHI), the Hearing Disabilities and Handicap Scale (HDHS), and the Speech, Spatial, and Qualities of Hearing Scale (SSQ), or tinnitus: the Hearing Measure Scale (HMS), the Gothenburg Profile (GP), and the Oldenburg Profile (OP). Many, but not all, questionnaires provide a measure of the experienced handicap next to a measure of disability. Further information to facilitate making the correct choice of questionnaire for a particular purpose may be found in Noble (1998) and Hyde (2000).

### 3.2.7 Tests on listening effort

Under listening conditions that allow hearing-impaired individuals to understand speech and successfully participate in a conversation, this may still require more effort than it does for normally-hearing individuals. Increased listening effort, and the subsequent stress, create an increased working load for hearing-impaired individuals in relation to normally-hearing ones, so the former are likely to tire more quickly. Self-reported measures of the increased effort for hearing-impaired listeners can be obtained using a questionnaire that contains an item or items to investigate this aspect of hearing, e.g. the Speech, Spatial, and Qualities of Hearing Scale (SSQ), introduced by Gatehouse and Noble (2004). This overview concentrates on more objective measures of listening effort. Such measures have been obtained using dual-tasking techniques and recordings of pupil dilatations.
3.2.7.1 Dual tasking

A technique that is often used to measure effort is the application of dual tasking, which involves a primary and a secondary task. The mental load of the primary, in our case, speech-listening task is thought to hinder the execution of the secondary task, thus reducing scores on the secondary task when the primary one becomes more and more difficult. This reduction in scores on the secondary task is taken as a measure of the effort that was used in the primary task. The secondary task will be chosen so that it involves mental activity and an active subject response (e.g., reaction time test, Downs, 1982; memory task, Rakerd et al., 1996; response to random presentations, Hick and Tharpe, 2002). For the assessment of listening effort this technique has been used for normally-hearing adults by e.g., Broadbent (1958), Rabbitt (1966), Downs and Crum (1978), Feuerstein (1992), and Rakerd et al (1996), and for hearing-impaired adults by Downs (1982), and Rakerd et al (1996). More recently, this technique has been applied by Hick and Tharpe (2002) for both normally-hearing and hearing-impaired children with ages between 5 and 11 years. The studies with normally-hearing adults showed that the scores on the secondary task can be affected by the difficulty of the primary task even when the scores on the primary task are unchanged. For hearing-impaired listeners, Rakerd et al (1996) showed that such individuals use more effort when listening to speech in noise than normally hearing listeners. Hick and Tharpe (2002) found corresponding results in children. Downs (1982) found that the effort exerted by hearing-impaired adults can be decreased by the application of hearing aids. In summary, the dual-tasking technique has been applied successfully to assess listening effort in normally-hearing and hearing-impaired children and adults.

3.2.7.2 Pupil dilatation

To obtain an objective measure of listening effort, pupil dilatation has been used successfully by Kramer et al (1997). Pupil dilatation has been reported to be a sensitive measure of the mental effort required by tasks (Hess and Polt, 1964; Janisse, 1977; Heemstra, 1988). In addition, pupil diameters change significantly in memory tasks using words or sentences (Wright and Kahneman, 1971; Kahneman and Beatty, 1966), and during speech production (Hoeks, 1995). Kramer et al (1997) found that recorded pupil dilatation increased more strongly with reductions in speech-to-noise ratios in hearing-impaired listeners than in normally-hearing ones. In addition, they found that pupil dilatation in hearing-impaired listeners correlated strongly with self-reported handicap measures (Kramer et al., 1998). So, though pupil dilatation does not solely depend on effort (Hess, 1975; Janisse, 1977), and measuring them requires a rather extensive setup, it does appear to correlate strongly with listening effort.

3.2.8 Tests on cognitive abilities

This overview on cognitive tests and speech comprehension does not attempt to be complete. First a summery of the cognitive mechanisms involved in speech processing will be given followed by assumptions and general reasoning underlying available cognitive tests. Finally descriptions of some tests and results from those tests will be given.
3.2.8.1 The speech process and cognitive functions

The speech process yields not only the sensation of an incoming stimulus, but also its processing and interpretation in the context of previous experiences. Information about how things are related and categorized, for example contextual, lexical, syntactic and semantic information, is stored in a long-term memory. Controlling top-down processes work in parallel with stimulus-driven bottom-up processes in every information-processing stage. Thus, hearing includes both audition and cognition.

The information available from perception and cognitive processing has to be available for conscious manipulation. In theoretical models (e.g. Baddeley, 2003) the working memory is responsible for the active part in language comprehension, as well as for the transfer of information into long-term memory. The central unit in the working memory (the central executive) is in control of directing attention and mental resources and uses slave systems for limited short-term storage of information (the phonological loop and the visuospatial sketchpad). The dynamic situations in many speech comprehension tasks, where earlier words in a sentence or phrase must be held in memory while subsequent ones are studied for syntactic and semantic relations, put high demands on good working memory abilities especially if the speech signal is limited or distorted. The working memory thus involves both storage and processing.

Any kind of distortion or limitation of an incoming stimulus, e.g. in difficult listening situations, such as noise and reverberation, or because of a hearing impairment, makes the process more dependent on top-down processing. The situation becomes more cognitively demanding than normally. Functions like rapid access to semantic and lexical knowledge, working memory capacity, selective attention, and high speed of information processing become thus more critical for the understanding of spoken language.

It has been verified that some cognitive functions deteriorate as a consequence of hearing impairment in severely hearing-impaired and deaf people. More precisely, it has been proven that the phonological ability declines when auditory stimulation is reduced during a longer period (Andersson and Lyxell, 1998; Andersson, 2001).

It has also been shown that cognitive functions decline with increasing age (Baltes and Lindenberger, 1997; Park, 1999). Poorer results on highly demanding working memory tasks as well as decreased speed of performance have been noticed in the elderly (Birren et al., 1980; Birren and Fisher, 1995; Rönnberg, 1990; Salthouse, 1996; Hallgren et al., 2001b; Li et al., 2001; Wingfield and Tun, 2001). Both sensory motor and mental processes slow down at increasing age affecting all memory systems. The slowing tempo might play a major role in speech understanding difficulties that occur with aging; it could influence the speed with which both top-down and bottom up processes can operate on fluent speech.

The extra problems with speech understanding in the elderly population, especially in noisy situations, are often discussed in connection with a poorer peripheral hearing. Previous studies have reported that peripheral auditory impairments account for most of the variance in speech recognition scores among older listeners (e.g. van Rooij and Plomp, 1992; Schneider et al., 2000). But a decline in peripheral hearing is not the only explanation. Decline in higher order abilities, cognitive and/or central auditory functions has also been discussed as an explanation to the problems in the elderly (e.g. Jerger et al., 1989; Humes, 1996; Hallgren et al., 2001a; Pichora-Fuller, 2003). For example, age-related changes in working memory and processing speed have been suggested as the mechanism responsible for diminished speech understanding ability in older adults (Pichora-Fuller, 1997; Sommers, 1997; Wingfield and Tun, 2001).
The importance of context in speech comprehension for the elderly population has been studied by Pichora-Fuller et al., 1995 and Wingfield and Tun, 2001. By using recall tasks with low- and high-probability SPIN (Speech In Noise) sentences they showed that slowed processing, diminished memory capacity as well as impaired hearing affect speech comprehension. They reported that older adults performed more poorly than younger adults when little contextual information was available in the speech signal. However, recall improved when contextual content increased, suggesting that the availability of top-down information in the message aids speech comprehension in older adults.

In dichotic speech testing, where the two ears are stimulated simultaneously with different speech sounds, the role of cognition has been verified. Hallgren et al., 2001a for example, showed that in elderly subjects there is an overall decrement with age in both dichotic and cognitive test results. Furthermore, when using directed report of stimuli, there were significant correlations between the performance in the left focusing condition and the cognitive parameters speed of performance and working memory capacity. These correlations were not seen in the right focusing condition. Also Jerger et al., 1994 and Alden et al., 1997 demonstrated that older subjects have reduced ability to perceive stimuli presented to the left ear.

It should be noted that cognitive abilities are never the entire explanation to the problems with speech comprehension in the elderly but always a part of the explanation. However, the relative effect on speech understanding of the different ageing processes (peripheral/central auditory/cognitive) has been difficult to determine, because of the problems in isolating the different components.

3.2.8.2 Cognitive tests – general considerations

Traditionally hearing evaluation and rehabilitation has been based on the pure tone audiogram which primarily represents the peripheral hearing in stimulus driven bottom-up processing. This assumes that two persons with the same audiogram have the same communicative ability, which is not the case. It is also important to consider the ability to make use of cognitive processing in order to compensate for a limited and distorted sensory signal.

It is a well-known fact that many cognitive functions decline with age in the later part of life.

The first studies on age-related cognitive decline and speech recognition focused on general, global cognitive functions and idealized listening conditions (review see Working Group on Speech Understanding and Aging., 1988; Sommers, 1997). Studies have been published about for example the ability to use semantic context (Wingfield et al., 1994) and the role of lexical discrimination in speech recognition (Sommers, 1996). It is only during the last decade that focus has been changed to speech-specific cognitive capacities in natural listening environments such as noise.

Different noise sources put different demands on cognitive skills in the individual. In the complex process of speech understanding the listener depends on peripheral hearing as well as central auditory and cognitive functions. Several studies have shown the importance of cognitive skills in speech processing tasks (Gatehouse et al., 2003; Lunner, 2003; Lyxell et al., 2003; Pichora-Fuller, 2003). Working memory capacity, speed of verbal information processing, and phonological skills are cognitive components that are critical. For speech processing in noise these cognitive functions are likely to be especially important since the noise partly masks the speech signal.
How working memory relates to speech understanding in individuals with hearing loss has been studied (Hallgren et al., 2001a; Lunner, 2003; Larsby et al., 2005), but has received surprisingly little attention in the research literature (Lyxell et al., 2003).

Cognitive functions are also important in order to make use of amplification in modern hearing aids with advanced signal processing. In recent studies it has been convincingly argued that individual cognitive prerequisites interact with different signal processing algorithms in determining the benefit obtained from hearing aids (Gatehouse et al., 2003; Lunner, 2003).

3.2.8.3 Cognitive tests

Van Rooij et al., 1989 and van Rooij and Plomp, 1992 studied auditive and cognitive factors in speech perception. The cognitive test battery included tests of simple and choice reaction time, a memory scanning test, a lexical-access test, a test of verbal comprehension, memory tests, a test of divided attention ability, and an IQ test. The tests where chosen because they are involved in the online processing of speech. In order to check on the effect of modality, auditive and visual versions of some of the cognitive tests were included. The authors concluded that the most important cognitive correlates of speech perception performance appeared to be processing speed and sensorimotor speed.

TIPS/SVIPS

To assess different aspects of cognition in speech processing, a cognitive test-battery called TIPS (Text Information Processing System, Ausmeel, 1988), was developed at the department of psychology, Linköping University. The TIPS tests capture functions of working memory capacity, selective attention, phonological ability and speed of information processing. Tests included can be divided into three main categories of tests, namely:

1. Verbal information processing speed
   - Semantic decision test, where the task is to decide whether a word belongs to a certain pre-defined semantic category or not
   - Lexical decision test, where the task is to judge whether a combination of three letters is a real word or a non-word
   - Name matching test, where the task is to judge whether two presented letters are the same (e.g., A - A) or not (A - B)
   - Physical matching test, where the task is to judge whether two presented letters has the same physical shape (e.g., A - A) or not (A - a)

2. Phonological processing
   - Rhyme-judgment 1, where the task is to decide whether two presented words rhyme or not
   - Rhyme-judgment 2, where the task is to decide whether two presented bisyllabic non-words rhyme or not
   - Rhyme-judgment 3, where the task is to decide whether two presented monosyllabic non-words rhyme or not
• Rhyme-judgment 4, where the task is to decide whether two presented bisyllabic words - one “real” word and one a non-word - rhyme or not

In the TIPS tests above both accuracy and speed of performance are assessed.

3. Working memory capacity

• The reading span test. The subject’s task is to comprehend sentences and to recall either the first or the final words of a presented sequence of sentences in correct serial order (Baddeley et al., 1985). The words are presented in a word-by-word fashion. Half of the sentences are absurd (e.g., “The train sang a song”), and half are normal sentences (e.g., “the girl brushed her teeth”). The subjects’ task is to respond “yes” (for a normal sentence) or “no” (for an absurd sentence) after the presentation of each sentence. After a sequence of sentences (three, four, five or six sentences) the test leader indicates that the subject should start to recall either the first or the final words for each presented sentence in the sequence.

Results from the TIPS tests have been shown to be related to various kinds of communicative performance for populations of hearing-impaired subjects (e.g. Lyxell and Andersson, 1998; Lyxell and Holmberg, 2000; Andersson, 2001; Hallgren et al., 2001a; Lunner, 2003; Lyxell et al., 1998). Significant correlations between results in the reading span test and speech recognition in noise with the Hagerman sentences (Hagerman, 1982; Hagerman and Kinnefors, 1995) have been shown by Lunner, 2003. These studies have used text stimuli for measurement of cognitive functions. Thus, the modality specific parts in speech comprehension are not present. In order to obtain a more complete picture of a person’s cognitive abilities a system for presentation of some of the tests in TIPS in auditory and audiovisual modalities was developed (SVIPS, Speech and Visual Information Processing System, Hallgren et al., 2001b).

The introduction of SVIPS tests as a complement to TIPS opened for studies of natural communication in normal hearing and hearing-impaired subjects both in easy and in adverse listening situations as was suggested by Ronnberg et al., 2000. The visual contribution in speech processing can be quantified and modality specific functions can be identified and quantified. The SVIPS test battery has been described and evaluated in young and elderly, normal-hearing and hearing-impaired subjects both in quiet and in different noise backgrounds (Hallgren et al., 2001b; Larsby et al., 2005). Furthermore, the SVIPS battery has been used to study speech understanding in quiet and noise with and without hearing aids (Hallgren et al., 2005).

Visual-digit monitoring task and visual-letter monitoring task - “Vigilance tests”

The visual-digit monitoring task was developed to predict performance in cochlear implant users (Knutson et al., 1991). The visual letter monitoring task was developed at the MRC Institute of Hearing for use in their evaluation of cochlear implantation in the UK. The MRC material has been translated into Danish and Swedish versions.

In the visual-digit monitoring task the subject has to watch the computer monitor as single-digit numbers are presented one at a time. When the displayed number reflects a pattern of even-odd-even numbers, the subject is to respond on the keyboard. The subject’s task is to maintain the last two digits in working memory and respond when those digits and the currently displayed digit produce the specified pattern. The numbers are presented at a rate of either one per second or one every 2 seconds, so subjects must respond rapidly. The visual-digit monitoring task correlated significantly with audiological measures of sentences, consonants, vowels and phonemes in words (Knutson et al., 1991).

In the visual-letter monitoring task (Gatehouse et al., 2003) the subject observes a stream of single letters presented sequentially on a computer screen, and while observing the items, the subject shall perform a semantic and a memory task. The letters are arranged
alternately vowels (V) and consonants (C), and the subject’s task is to identify CVC sequences that form a real word. The subject is instructed to press a button whenever he identifies a CVC sequence as being a real word. The semantic task is to find the word and the memory task is to keep track of previous letters, to make it possible to decide whether a CVC sequence forms a real word. The use of the vigilance testing in relation to hearing aids is described by Behrens et al., 2004; Sundewall and Behrens, 2004; Vestergaard, 2004. Gatehouse et al. (2003), showed the vigilance test to have predictive leverage on hearing aid outcome.
4 Preliminary selection of tests to be used in an Auditory Profile

Based on the information collected in chapter 3, a consensus meeting was organized in Gothenburg (June 21st, 2005) to discuss the selection of tests that should be included in the auditory profile. First of all, we selected the most relevant tests in each domain, specifying the goal, details, indication, and outcomes (see Table below).

<table>
<thead>
<tr>
<th>Test</th>
<th>Goal</th>
<th>Details</th>
<th>Indication</th>
<th>Outcome</th>
<th>CAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Audibility</td>
<td>Pure-tone audiogram AC</td>
<td>Hearing threshold AC</td>
<td>25/5/1/2/3/4/6/8 kHz</td>
<td>Basic</td>
<td>Frequency-dependent hearing loss</td>
</tr>
<tr>
<td></td>
<td>Pure-tone audiogram BC</td>
<td>Hearing threshold BC</td>
<td>25/5/1/2/3 kHz</td>
<td>Basic</td>
<td>Separation conductive/sensorineural</td>
</tr>
<tr>
<td></td>
<td>TEN-HL test</td>
<td>Determine dead regions</td>
<td>25/5/1/2/3/4 kHz</td>
<td>High losses or steep audiometric slopes</td>
<td>Indication for the presence of dead regions</td>
</tr>
<tr>
<td>2. Loudness</td>
<td>ACALOS test</td>
<td>Loudness scaling</td>
<td>see Brandi/Hohmann, using 1/3 oct. LNN at .5 and 3 kHz</td>
<td>Basic</td>
<td>Dynamic range (THR - UCL) and slope of the loudness</td>
</tr>
<tr>
<td></td>
<td>ACALOS test</td>
<td>Loudness scaling</td>
<td>more frequencies</td>
<td>in case on recruitment or for a higher resolution</td>
<td>II</td>
</tr>
<tr>
<td>3. F-T resolution</td>
<td>F-T test</td>
<td>Degree of suprathreshold resolution</td>
<td>see Hitchhines and Larsby &amp; Arlinger; test at .5 and 3 kHz</td>
<td>Suprathreshold problems</td>
<td>Separation of problems in F- or T-domain</td>
</tr>
<tr>
<td></td>
<td>Notched-noise test</td>
<td>Determine auditory filter</td>
<td>depending on F-T test and audiogram</td>
<td>Problems in the F-domain</td>
<td>Estimates of the frequency resolution of the ear</td>
</tr>
<tr>
<td></td>
<td>Temporal-window test</td>
<td>Determine temporal window of the ear</td>
<td>depending on F-T test and audiogram</td>
<td>Problems in the T-domain</td>
<td>Estimates of the temporal resolution of the ear</td>
</tr>
<tr>
<td>4. Speech perception</td>
<td>SRT-test with natural sentences</td>
<td>Determine SRTs in quiet and in noise</td>
<td>in quiet and in noise (ICRA-1 and ICRA-5)</td>
<td>Basic in case of single determination of speech</td>
<td>SRT in quiet and in noise and unmasking due to noise</td>
</tr>
<tr>
<td></td>
<td>SRT-test with MATRIX-sentences</td>
<td>Determine SRTs in quiet and in noise</td>
<td>in quiet and in noise (ICRA-1 and ICRA-5)</td>
<td>Basic in case of repeated determination of speech</td>
<td>SRT in quiet and in noise and unmasking due to noise</td>
</tr>
<tr>
<td></td>
<td>VCV-test</td>
<td>Analyse consonant confusions</td>
<td>processed speech elements in quiet and in noise (ICRA-1)</td>
<td>in case of poor speech perception</td>
<td>Confusion patterns</td>
</tr>
<tr>
<td>5. Spatial hearing</td>
<td>ILD-test (headphones)</td>
<td>Determine head shadow effect</td>
<td>for speech at MCL</td>
<td>Basic</td>
<td>Estimated gain of head shadow for speech</td>
</tr>
<tr>
<td></td>
<td>BILD-test (headphones)</td>
<td>Determine binaural advantage</td>
<td>for speech at MCL</td>
<td>Problems in binaural processing</td>
<td>Estimated binaural advantage for speech</td>
</tr>
<tr>
<td></td>
<td>JND-azimuth test (headphones)</td>
<td>Use of localization cues</td>
<td>NB-noises at .5 and 3 kHz</td>
<td>Problems in localization cues</td>
<td>Ability to use of localization cues</td>
</tr>
<tr>
<td></td>
<td>ILD-test (free field)</td>
<td>Determine head shadow effect</td>
<td>free-field speech at MCL</td>
<td>Deviant findings in headphones tests on ICRA-5</td>
<td>Real gain of head shadow</td>
</tr>
<tr>
<td></td>
<td>BILD-test (free field)</td>
<td>Determine binaural advantage</td>
<td>free-field speech at MCL</td>
<td>Deviant findings in headphones tests on ICRA-5</td>
<td>Real binaural advantage</td>
</tr>
<tr>
<td></td>
<td>Localization test (free field)</td>
<td>Horizontal localization</td>
<td>ICRA-1 noise in a 8 - 16 speaker set-up (see WP-6)</td>
<td>Deviant findings in headphones tests</td>
<td>Real horizontal localization ability</td>
</tr>
<tr>
<td>6. Subjective judgements</td>
<td>OTGP-questionnaire</td>
<td>Measure handicap/disability/benefit</td>
<td>Basic inventory</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GMABP</td>
<td>Measure improvements from hearing aids</td>
<td>In case of hearing aid fitting</td>
<td>User-specific improvements and residual problems</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td>SSQ</td>
<td>Extensive review of disabilities</td>
<td>if detailed inventory is wanted</td>
<td>Parameters of problems in dynamic daily-life situations</td>
<td>III</td>
</tr>
<tr>
<td>7. Listening effort</td>
<td>Effort scaling</td>
<td>Scaling of subjective effort</td>
<td>For speech in ICRA-1 and ICRA-5 at S/N 0 and S/N +10</td>
<td>Additional to speech testing</td>
<td>I</td>
</tr>
<tr>
<td>8. Cognitive function</td>
<td>Lexical decision making</td>
<td></td>
<td>Basic</td>
<td>word recognition</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td>TRT-test</td>
<td>Determine use of redundancy</td>
<td>vis. recognition of sent. in fluctuating background patterns</td>
<td>Visual analogue of SRT-test</td>
<td>Use of redundancy</td>
</tr>
<tr>
<td></td>
<td>Vigilance (CVC) test</td>
<td>slow/fast version</td>
<td>in case of impaired cognitive abilities</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reading Span test</td>
<td></td>
<td>in case of impaired cognitive abilities</td>
<td>III</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Overview of candidate tests for the auditory profile, including goal, some details, indications, outcome, and the proposed category. For descriptions of abbreviations, see Appendix C.
The partners within WP2 agreed on determining the auditory profile in three different categories: ‘Standard’ (I: used for each patient), ‘Advanced’ (II: HearCom inventory), and ‘Specialized’ (III: II plus additional measurements). The category belonging to each test is indicated in the last column. It was decided that the auditory profile consists of the tests in categories I and II. Test in category III are useful additional measurements, but they will not be implemented in the context of HEARCOM. Therefore, they have been printed in italic.

The details of the tests listed in the table are given in the review part of this deliverable (section 3).

As can be seen in the table 4.1, there are some uncertainties. Choices regarding these questions will be made based on phase-I testing in the different labs (see chapter 5.1). Subsequently, the test set for phase II testing will be defined (see chapter 5.2). These tests will be used in a multi-centre study within HearCom to determine the final auditory profile.

So, the preliminary auditory profile contains the following tests:

<table>
<thead>
<tr>
<th>I: Standard</th>
<th>II: Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Audibility</td>
<td></td>
</tr>
<tr>
<td>Audiogram (AC+BC)</td>
<td></td>
</tr>
<tr>
<td>2. Loudness</td>
<td></td>
</tr>
<tr>
<td>Acalos at 0.5 and 3 kHz</td>
<td>Acalos at other frequencies</td>
</tr>
<tr>
<td>3. Frequency-time resolution</td>
<td>Combined F-T test (to be selected)</td>
</tr>
<tr>
<td>4. Speech perception</td>
<td></td>
</tr>
<tr>
<td>Speech in quiet and noise (SRT or OLSA)</td>
<td></td>
</tr>
<tr>
<td>5. Spatial hearing</td>
<td></td>
</tr>
<tr>
<td>ILD-test</td>
<td>BILD-test</td>
</tr>
<tr>
<td>JND-azimuth test</td>
<td></td>
</tr>
<tr>
<td>6. Subjective judgement, communication and listening effort</td>
<td>Oldenburg Inventory/Gothenburg profile</td>
</tr>
<tr>
<td>7. Listening effort</td>
<td></td>
</tr>
<tr>
<td>Effort scaling for speech in noise</td>
<td></td>
</tr>
<tr>
<td>Cognitive abilities</td>
<td></td>
</tr>
<tr>
<td>Lexical decision making or TRT</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Preliminary test set for the auditory profile (category I and II). For descriptions of abbreviations, see Appendix C.

Regarding the speech perception tests, a choice between an SRT-test with natural sentences (a Plomp-type test) or an OLSa-type test will be made based on availability in the concerned language. The OLSa-type tests have the advantage of being more similar across languages (less variety in norm values), while Plomp-type tests are more suitable for quick measurements due to smaller learning effects. Therefore, in case both tests are available, it is suggested to use a Plomp-type test when only a short measurement time is available, and an OLSa-type test for prolonged testing on the same listener.

It should be stressed that the auditory profile described above is primarily focussed on the diagnosis of auditory functioning. For the purpose of auditory rehabilitation, some extra tests may be needed in order to select, fit, and evaluate hearing aids. For the subjective judgments, specific questionnaires have been developed to evaluate the benefit of hearing aids, e.g. the Glasgow Hearing Aid Benefit Profile (GHABP). These aspects will be considered in the work of WP6.
A second difference between tests for diagnostic purposes and tests for the purpose of hearing aid evaluation is that the latter tests are usually performed in free-field conditions. This requires special facilities, to be described in WP6.
5 Implementation of tests

Before we define the final test set, different tests will be applied locally in order to solve some of the remaining questions about the optimal applicability of some tests. The implementations of these tests and the protocols to be used have been described in Section 5.1. After the local testing (for the psychoacoustical tests in the Netherlands and for the cognitive tests in Sweden), we will conduct a large multi-center study using the final test set of the auditory profile in different languages, using the same hardware and software (see Section 5.2).

5.1 Implementation for phase I testing

5.1.1 Spectral and temporal resolution (NL-AMC)

5.1.1.1 Introduction

At NL-AMC, an extensive experiment will be performed to investigate which test(s) of spectral and temporal resolution should become part of the auditory profile.

Recently, two tests have been developed which measure spectral and temporal resolution simultaneously (Hilkhuyzen et al., 2004, Larsby and Arlinger, 1998; 1999). These tests give separate measures for spectral and temporal resolution, contrary to e.g. the rippled noise test (Supin et al., 1997). Possibly, such a combination test is an efficient way to screen for spectral and temporal deficits. The measurement time (an important criterion in clinical settings) of a combination test is expected to be shorter than the measurement time of two separate tests. This makes these tests interesting candidates for the auditory profile. However, it is not yet known to which extend these tests are actually feasible and relevant for clinical use. That will be the main focus of our experiment. The tests will be conducted in a group of 6 normal-hearing and at least 12 hearing-impaired listeners. Results of the two ‘combination tests’ will be compared to results of more conventional spectral and temporal resolution tests, to check for reliability. A notched noise method (Leeuw and Dreschler, 1994) will be used to measure spectral resolution. For temporal resolution two tests will be used: the conventional gap detection measurement (see below) and a temporal window measurement (Moore et al., 1988), which measures at the same time scale as the two combination tests do. In addition, results will be compared to speech perception measurements (SRT-test, see below), to examine the relevance of the tests in relation to speech perception. Finally, measurement times, accuracy’s and test-retest reliabilities of both tests will be determined.

5.1.1.2 Protocol

All measurements will be performed monaurally at the better ear at 0.5 and 3 kHz. To exclude loudness effects as much as possible, all measurements will be conducted at equal loudness, MCL. To retrieve individual MCLs, a loudness scaling measurement is added. Results of this measurement will also be used to find the relative contributions of distorted loudness perception and spectral and temporal resolution abilities to reduced speech perception.
1) **Loudness scaling: Acalos**
Categorical loudness scaling (Brand and Hohmann, 2002) with 1/3 octave narrowband noises (centre frequencies 0.5 and 3 kHz) and broadband noise (all low-noise noises) of 1 second duration.

2) **Speech Reception Threshold test (SRT)**
Open set sentences in fluctuating (completely modulated) and continuous noise. Because of the restricted dynamic range of hearing impaired listeners, the combined level of the speech and noise signals is fixed (instead of the noise level), and the signal to noise ratio will be adapted to estimate the SRT.

3) **Spectral resolution: auditory filter measurements**
A clinical method, developed by Leeuw and Dreschler (1994) and based on the classical notched noise procedure (Patterson and Nimmo-Smith, 1980) will be used as a measure of spectral resolution. The adaptive procedure has been changed from Békésy tracking to a SIAM procedure.

4) **Temporal resolution: temporal window measurements**
Temporal windows are estimated using the method described by Moore et al (1988). Subjects have to detect a tone pulse in a temporal gap between two noise bursts. Both symmetrical and asymmetrical conditions will be measured.

5) **Temporal resolution: gap detection**
Detection of a temporal gap in 1/3 octave narrowband noise. Splatter will be masked by a broadband noise mask. To minimize loudness fluctuations within the noise bursts, low-noise noises will be used. Like in the temporal-window test, a 2AFC 2-up 1-down adaptive procedure is used.

6) **Combination test 1: Hilkhuyzen et al. 2005**
Detection of spectral sweeps in masking noise grids. Sweeps are presented in three different noise masks; continuous noise, spectral grids and temporal grids. None to three sweeps are presented in each trial and the subject indicates the perceived number of sweeps.

7) **Combination test 2: Larsby and Arlinger 1998, 1999**
Detection of a pulsed tone signal in different noise masks: noise with no gaps, spectral gaps, temporal gaps or spectro-temporal gaps. Detection thresholds are measured using Békésy tracking.

5.1.2 Perception of speech elements (NL-AMC)

5.1.2.1 Introduction
At NL-AMC, an experiment will be performed to investigate consonant recognition. Several researchers investigated the possibilities of increasing the consonant-vowel ratio (CVR) to improve the perception of consonants by hearing-impaired listeners. Some found possible improvement after individual CVR adjustment (Kennedy et al., 1998) whereas others (Sammeth et al., 1999) found no significant improvement in voiceless stop-consonant recognition. Instead of increasing the CVR to improve perception we will investigate the effects of reducing the intensity of the consonant on confusion patterns. Possible differences between normal-hearing and hearing-impaired listeners concerning listening strategy will be evaluated using individual consonant confusion matrices. Besides reduction of the consonant
energy the separate effect of filtering will be studied. The effect of filtering on stop-
consonant recognition was studied before by Dubno et al. (1989) who found that differential
filtering effects were consistent with the spectral properties of the stimuli. The current study
investigates if this consistency applies to all consonants.

5.1.2.2 Protocol

Nonsense VCV syllables are presented through headphones in stationary noise (which is
added after processing). Conditions listed below are tested for a signal-to-noise ratio of 0 dB
and +6 dB. To exclude loudness effects as much as possible, all measurements will be
conducted at equal loudness (see protocol 5.1.1.2). Listeners take part in a test and re-test
session each lasting approximately 80 minutes.

1) Natural stimuli
Nonsense VCV syllables (spoken by a woman) will be used, the total number of
unprocessed stimuli is 51: 17 consonants and the vowels /a/, /o/, /i/.

2) Consonant reduction
Stimuli are processed to create a level reduction of -6 dB or -12 dB for the
consonants only.

3) Filtering
For low-pass filtering cut-off frequencies of 1,0 kHz and 1,6 kHz are applied. For
high-pass filtering cut-off frequencies of 1,6 kHz and 2,0 kHz are chosen.

5.1.3 Spatial hearing (NL-AMC)

5.1.3.1 Introduction

At NL-AMC, an experiment will be performed to measure binaural listening performance in
hearing-impaired listeners. First, a clinical test will be designed to investigate to which
extent localisation cues are available for hearing-impaired listeners. As this test should be
made available to all partners, it should be easy to implement in different labs and clinics.
Therefore, the possibility of a virtual localisation test is investigated. In the virtual
localisation test the stimuli are filtered by HRTFs (Head Related Transfer Functions – the
direction dependent transfer functions from loudspeaker to eardrum) of different azimuth
directions. The stimuli are presented via headphones, but their apparent source location will
be somewhere outside the subject, with a direction that depends on the HRTF with which
the stimulus is filtered. Both a free-field and a virtual localisation test will be developed. The
virtual test will be based on HRTF measurements in the free-field set-up. By comparing the
results of both tests within subjects, we can check whether the virtual test gives consistent
and reliable results. If the results of the virtual localisation tests are in agreement with the
results of the free-field tests, we can proceed to develop a headphone test to measure the
just noticeable difference (JND) in azimuth source angle (minimal audible angle, MAA).

Second, the advantage of separating speech and noise is measured in the Intelligibility Level
Difference (ILD) test and the binaural advantage of listening with two ears instead of with
one, is measured in the Binaural Intelligibility Level (BILD) test.
5.1.3.2 Protocol

Stimuli (pink noise or speech-shaped noise) are presented both in a free-field set-up and via headphones. Stimulus level is 70 dB SPL (normal conversational level). Level roving (small variations in sound level from trial to trial) is applied to make sure sound level is not used as a localisation cue. This cue consists of loudness variations between the different loudspeakers and the head shadow effect. A range of 10 dB (± 5 dB) is probably enough to eliminate loudness variation between speakers and to reduce the effectiveness of loudness as a cue.

1) Free-field localisation
   Localisation of noise stimuli in a free-field set-up. Listeners are instructed to look straight ahead and keep their head still, to exclude possible effects of head movement on localisation.

2) Free-field (binaural) intelligibility level difference: (B)ILD
   In order to estimate the effects of a spatial separation of speech and noise on speech perception, we measure Intelligibility Level Difference (ILD) test and the binaural advantage of listening with two ears instead of with one (Binaural Intelligibility Level (BILD)).

3) Virtual localisation
   Localisation of noise stimuli, filtered with HRTFs measured in the free-field set-up. The stimuli are presented via headphones.

4) Virtual (binaural) intelligibility level difference: (B)ILD
   A virtual realisation of test 2) via headphones.

5) Minimal audible angle (MAA)
   Virtual localisation stimuli are presented over headphones and an adaptive, two-alternative forced-choice procedure is used to measure the JND of sound-source azimuth.

5.1.4 Cognitive tests (SE-LINK)

5.1.4.1 Introduction

The purpose of the experiments, conducted at SE-LINK is:

a) to study which cognitive skills are important for speech recognition/comprehension as measured in the Hagerman speech test and in the Swedish Hearing In Noise Test (HINT).

b) to study the cognitive factors affecting the outcome of the result in the Vigilance test (CVC).
5.1.4.2 First results

Based on the results of a first experiment in 24 subjects, the following conclusions were arrived at:

- Among the three cognitive tests (reading span, CVC and verbal ability), only the reading span test was found to correlate significantly with the S/N in the Hagerman test. The correlations were significant with Hagerman noise and reversed speech as background but not with speech as background.
- The cognitive measures in the reading span, CVC and verbal ability tests were not found to correlate with one another.
- Effort scores in the CVC and verbal ability tests were found to correlate with cognitive measures in the same test but not with other cognitive measures.
- The highest score of perceived effort was seen in the reading span test.

5.1.4.3 Protocol for a follow-up experiment

In a follow-up experiment the following tests were used:

1) Speech recognition was measured with the **Hagerman speech test**. The speech signal was fixed at 70 dB SPL (C-weighted). The noise level was adjusted adaptively in an interleaved method to reach 50 % and 80 % correct responses. The S/N values for 50% and 80% correct responses, respectively, were calculated and used as outcome measures.

2) **HINT** (Hearing In Noise Test): Speech recognition was measured with the Swedish HINT sentences (Hällgren et al, manuscript). The speech level was fixed at 70 dB SPL (C-weighted). The noise signal was adjusted adaptively in two different procedures, to reach 50% correct sentences and to reach 60% correctly recognized keywords. The S/N values for the two procedures were calculated and used as outcome measures.

3) **Vigilance test (CVC):** Letters were presented on a computer screen in front of the test subjects one at a time. Every second letter is a consonant and every other a vowel. The letters were presented with an inter-stimulus interval of one (fast) or two (slow) seconds. The subject’s task was to press the space-bar on the computer every time three letters in a sequence of consonant-vowel-consonant (C-V-C) constituted a real word. The test items were twenty real C-V-C words for each inter-stimulus interval, i.e. a total of forty. Outcome measures were the numbers of existing words recognized in the conditions CVC d’slow and CVC d’fast, to be combined to CVC d’average.

4) **Reading span test:** The subject’s task was to comprehend sentences and to recall either the initial or the final words of a presented sequence of sentences in correct serial order. The words in each sentence are presented in a word-by-word fashion. Half of the sentences are absurd and half are normal sentences. The subjects’ task was to respond “yes” (for a normal sentence) or “no” (for an absurd sentence) after the presentation of each sentence. After a sequence of sentences the test leader indicated that the subject should recall either the initial or the final word of each presented sentence in the sequence. Outcome measure was the total number of correctly identified initial and final words.

5) **Lexical decision making:** The subject’s task was to judge whether a string of three letters constituted a real Swedish word or not. Half of the non-words sounded like a
real word when pronounced, whereas the remaining half did not. Stimuli were presented on the screen and the subjects pressed predefined keys, one for "no" and one for "yes". Both number of correct answers and reaction times were measured.

6) **Physical matching:** The subject’s task was to judge whether two simultaneously presented letters had the same physical shape (e.g., A - A) or not (A - a). Stimuli were presented on the screen and the subjects pressed predefined keys, one for "no" and one for "yes". Both number of correct answers and reaction times were measured.

7) **Wordspan:** Test of serial record of monosyllabic words (3 letters). The subject’s task was to repeat a series of words presented one by one on the computer screen. After a sequence of words (3-8) the test leader indicated that the subject should start to recall all words presented since last recall. A sequence of n words was repeated three times. Outcome measure was the total number of correctly recalled words.

8) **Test of verbal ability:** The verbal ability was assessed by giving the subjects a paper with twenty-seven groups of words, each group including 5 words. The subject’s task was to choose and underline the two words among the five which were mutually opposite. The subjects were allowed a maximum of four and a half minutes to solve the task. Outcome measure was the number of correctly identified opposite pairs.

5.1.4.4 Preliminary results:

Among the cognitive tests especially the reading span test and the test of lexical access were found to correlate significantly with the S/N in the Hagerman test and in the HINT test.

Good working memory capacity facilitates the outcome of the CVC test whereas lexical knowledge seems to affect the outcome of the test negatively.

5.1.5 Hardware

All tests will be performed using a PC with soundcard (and, depending on the soundcard, an external amplifier) and headphones, in a sound attenuating room. All experiments are programmed in Delphi.

5.2 Experimental set-up for phase II testing

All tests of the preliminary auditory profile should be available to all partners within WP2 in order to perform a multi-centre clinical study with these tests. The results of the multi-centre study will be analysed in terms of redundant measures. Therefore, the outcome of these measurements is a proposal for the final auditory profile to be used as a characterization of hearing deficiencies.

For the multi-centre study all different tests may be stand-alone programs, MatLab based programs, or programs based on the ‘Oldenburg Measurement Applications’. All tests should use standard hardware.

After the decision about the final auditory profile there should be some effort to integrate the different test elements in a common measurement program base.
6 Dissemination and Exploitation

- The contents of this report will not be disseminated outside the Project.

- The QA plan is for use by the members of the consortium.

- After completion of the tests, the programs can be downloaded from the HEARCOM website by all project partners.

- The preliminary protocol will be discussed in the plenary session of the HEARCOM conference on September 21st, 2005.

- Part of the tests can be translated to internet-applications by the co-workers in SP5.

- Exploitation of the tests to the end users (professionals in audiology) will be considered after completion of the project.
Appendix A, speech in noise tests

* Speech in noise sentence test (SPIN test), Kalikow et al., 1977; Gelfand et al., 1988

The test has the following specifications:
- Language is American English
- Sentence lengths are between 5 and 8 words and 6 and 8 syllables
- Single word responses are score, the last word is the keyword, always a monosyllabic noun
- A male speaker is used
- Noise is twelve-talker babble
- Sentences and noise are stored on analogue tape
- Speech material: 8 lists of 25 sentences, with high (PH) or low (PL) predictability
- Each list contains 12 PH and 13 PL sentences, or the reverse
- PH and PL sentences occur in a pseudo-randomized order
- Scores are calculated in percent correct
- Scores are calculated by averaging the results of the PH and PL sentences of one list separately
- At 0dB SN average scores for are 88% and 38% for PH and PL sentences, respectively
- The slopes (at the steepest points) were 14 and 8%/dB for PH and PL sentences, respectively
- This was measured for NH listeners at an 80 dB SPL presentation level
- Gelfand et al., (1988) used only the PH sentences and scored whole sentences to measure SRTs
- For NH using adaptive procedure they found: SRT is -2.0 (SD of 2.3) dB

Stimulus construction:
- A male speaker was used
- Sentences were phonetically balanced
- Sentence lengths were between 5 and 8 words and 6 and 8 syllables
- Last sentence word was a monosyllabic noun which was the response target (key word)
- Initial sentence groups: low, medium, and high predictability: PL, PM, and PH
- All keywords had medium frequency (very low and very high frequencies were avoided)
- Key word predictability was verified in a “pencil and paper” test
- Verified groups contained different numbers of sentences; PL: 479; PM: 95; PH: 574
- Supplemented with 574 unpredictable “artificial” sentences: P0
- The same male speaker was used to record the remaining sentences
- Remaining after speech reception validation: PH: 382; PL: 327; P0: 285; PM: 0
- PH sentences subjected to test of phonemic representability and balance
- Phonemic content was used to select 10 balanced 25-item lists for PH and PL
- 80 NH listeners participated in test, 16 for each one of 5 conditions
- Speech and noise were presented at 80dB SPL (0 dB SNR)
- Used in hearing test: 10 lists of 50 sentences with 25 from PH and 25 from PL, balanced for score
- For each list, 2 25-item lists are composed with sentences with pseudo-randomized predictability
- These two lists contain the same 25 keywords, 13 for PH and 12 for PL or the reverse
- Speech scores are calculated by averaging the results of one 25-item list
- After testing of the speech scores 8 of the 10 sentences lists met the demand of homogeneity
Further standardization of the SPIN test, Bilger et al., 1984

The test has the following specifications:
- Language is American English
- Sentence lengths are between 5 and 8 words and 6 and 8 syllables
- Single word responses are scored
- The last word is the keyword, always a monosyllabic noun
- A male speaker is used
- Noise is twelve-talker babble
- Sentences and noise are stored on analogue tape
- Speech material: the original 10 lists of 25 sentences, with high (PL) or low (PL) predictability
- Each list contains 12 PH and 13 PL sentences, or the reverse
- PH and PL sentences have a pseudo-randomized order (never more than two PH or PL in a row)
- The lists are presented at 8 dB SNR and 50 dB above individual threshold
- Scores are calculated in percent correct
- Scores are calculated by averaging the results of the PH and PL sentences of one list separately
- Average scores are 43.6% (list SD = 1.0) and 22.9% (list SD = 2.8) for PH and PL, respectively
- Overall (PH and PL), the average score was 33.3% correct (list SD = 1.6)
- This was measured for HI listeners
- The lists failed to meet targeted reliability for HI listeners

Test construction:
- The speech material of the SPIN test was used (Kalikow et al., 1977)
- All of the 10 original 25-item lists were used
- A male speaker is used and the noise is a twelve-talker babble
- 128 HI listeners were used to evaluate the lists
- The skill using contextual information was assessed in pen-and-paper test
- Half of the listeners were tested using head phones, the other half listened in a sound field
- Half were tested in one session, the other half in two separate session (2 to 4 weeks apart)
- The first block contained one list of each pair, the second block the other
- List orders were reversed for different group, leading (with the previous splits) to 8 groups overall
- Lists were presented at 8 dB SNR and 50 dB above individual threshold
- Stimuli were presented monaurally, sometimes using the better and sometimes using the poorer ear
- Each target word was scored twice, in a strict manner, by two independent markers
- The average score was over both PH and PL sentences was 33.3% correct (SD over lists of 1.6)
- Split up, scores were 43.6% (SD = 1.0) and 22.9% (SD = 2.8) for PH and PL, respectively
- The results were independent of presentation manner (transducer), number of visits, and list order
- The effects of marker and the subject-marker interaction were significant but very small
- Further analysis was performed with the data collapsed for transducer, visit, and order
- Four groups of lists were found, from easy to hard: (6), (1, 2, 4, 5, 10), (3, 7, 9), and (8)
- These variations were mainly caused by variations in the PL sentences (list SD = 2.8)
- The scores for PH sentences were much more consistent (list SD = 1.0)
- Overall, the lists did not meet the targeted reliability for HI listeners
Dutch speech reception test (Plomp sentences), Plomp and Mimpen, 1979a

The test has the following specifications:
- Language is standard Dutch
- 10 phonemically matched lists of 13 everyday sentences
- Sentences contain 8 or 9 syllables, no words with four or more syllables
- A female speaker is used
- Later a set of 10 lists was added for a male speaker (Smoorenburg, 1992)
- Material is available at a 15256-Hz sample rate and 16-bit resolution
- Material is available at a 44100-Hz sample rate and 16-bit resolution
- Speech-shaped noise is used
- Whole sentences are scored, all words must be correct
- An adaptive procedure with +/-2-dB steps is used, depending on the previous response
- The first sentence is started well below threshold and its SNR is increased in 2-dB step until it is understood correctly
- SRT scores are averaged over the last 10 responses
- They are calculated in dB(A) or SPL for in quiet and in dB SN in noise
- Mean monaural SRT in noise is -5.1 dB, with an SD of 0.9 dB (Smoorenburg, 1992)
- Mean monaural SRT in quiet is 19 dB(A) (Plomp and Mimpen, 1979b)

Stimulus construction:
- Starting point: 220 everyday sentences of 8 or 9 syllables
- Sentences had no words with more than three syllables
- After judging by audiologists and speech therapists 170 remained
- Speech-shaped noise was used (from a third-octave analysis of the long-term speech spectrum)
- A female speaker without regional dialect was used
- In addition, the same sentences were later recorded using a male speaker (Smoorenburg, 1992)
- Sentences were recorded on analogue tape
- Sentence scores were checked using 20 NH listeners
- Sentences were presented at an average level of 50 dB SPL
- And at a fixed SNR corresponding to an average score of 50% correct
- The levels of 23 sentences, with above average level and very high scores, were reduced by 2dB
- The levels of 23 sentences, with below average level and very low scores, were raised by 2dB
- The corrected sentences were presented to 20 new listeners
- 40 sentences with more than 16, or less than 4, correct responses were eliminated
- About half of the remaining sentences received a new +/-1 dB correction according to the scores
- The remaining 130 sentences were roughly equally intelligible and had a level SD of 1.3 dB
- These were grouped in 10 phonetically balanced lists of 13 sentences
- The speech-shaped noise was regenerated using the long-term spectrum of all 130 sentences
- The lists were evaluated using 10 NH listeners
- Sentences were presented in quiet or in a 50dB-SPL speech-shaped noise
- For headphones, the mean monaural SRT in noise was -5.9 (SD = 0.9) dB
- The slope of the psychometric function was 20% per dB at its steepest
- For free-field presentation, the mean SRT in quiet was -19 dB (Plomp and Mimpen, 1979b)
- For the male speaker SRTs were evaluated by Smoorenburg (1992)
- For headphones, the mean monaural SRT in noise was -5.1 (SD = 1.8) dB
- This STR was measured at fixed, speech-shaped, noise level of 65 dB (A)
- For free-field presentation, mean SRT in quiet was -15.8 (SD = 2.3) dB

*Alternative Dutch speech reception test (VU98 sentences), Versfeld et al., 2000*

The test has the following specifications:
- Language is standard Dutch
- 39 phonemically balanced lists of 13 everyday sentences
- Sentences contain 8 or 9 syllables, no words with four or more syllables
- A male and a female speaker are used
- A 44100-Hz sample rate and 16-bit resolution are used
- Speech-shaped noise is used
- Whole sentences are scored, all words must be correct
- An adaptive procedure with +/-2-dB steps is used, depending on the previous response
- The first sentence started well below threshold and its SNR is increased in 2-dB step until understood correctly
- A maximum likelihood method is used to estimate the SRT values
- They are calculated in dB(A) or SPL in quiet and in dB SN in noise
- Mean monaural SRT in noise is -4.1 dB, with an SD of 1.1 dB

Stimulus construction:
- Sentences were selected from large databases (newspapers)
- Sentences with 8 or 9 syllables without punctuation marks were selected
- They did not contain words with more than three syllables
- The resulting 35000 sentences were checked on grammar and content
- The remaining 1500 sentences were judged by audiologists and speech therapists
- This left 1311 sentences that were recorded
- Test recordings were made using 25 speakers, male and female
- Recordings were judged by speech therapist on voice and reading quality
- For the four remaining speakers (two male and two female) the sentences were recorded
- In addition video recordings were made for e.g. lip reading courses
- Recorded sentences were checked and equated in RMS value
- For each speaker a speech-shaped noise was constructed (2048-point FIR)
- Four listeners checked clarity and unambiguousness when played in quiet
- This left 1272 sentences, their RMS values were set to one fixed value
- 48 NH listeners were used to evaluate intelligibility of the 1272 sentences
- All sentences were played for male and female speaker at -1 and -4 dB SN
- A balanced block design combined the resulting four conditions with four groups of subjects
- Speech noise was presented at 70 dB SPL from 500 ms before to 500 ms after the sentence
- Subjects entered their responses on a computer, they were checked by hand afterwards
- Overall, SRTs were found to be -3.4 dB for the female and -2.9 dB for the male speaker
- For both speakers, 509 different sentences with steep psychometric functions were selected
- Both sentence groups were distributed into 39 13-item lists with balanced phonemic content
- The resulting average SRTs were -3.9 dB with a slope of 15% per dB for both speakers
- Twelve NH listeners were used to evaluate 8 lists for each speaker
- The noise was presented at a fixed level of 70 dB SPL
- A maximum likelihood method was used to estimate the SRT values
- The average SRT was -4.1 dB with a test-retest reliability (SD) of 1.1 dB
* Speech in noise sentence test (BKB sentences), Bench et al., 1979

The test has the following specifications:
- Language is British English
- 21 lists of 16 sentences with up to seven syllables
- Sentences are balanced for grammar and first word
- They use grammatical constructions and vocabulary as used by children
- Each list has 50 keywords, approximately 3 per sentence
- The sentences were spoken by a female speaker
- They are available on audio cassette
- The speech reception score is calculated from the number of correctly reproduced keywords
- Speech reception scores are expressed in percent-correct

Stimulus construction:
- Utterances of 240 children between 8 and 15 years old were transcribed
- Sentences were checked for commonly used of vocabulary and grammar
- Sentence length was limited to 7 syllables
- The sentences were grouped into 21 16-item lists
- The list were balanced for grammatical constructions and first words
- In each list 50 keywords were assigned to score speech reception (approximately 3 per sentence)
- These list have been recorded on tape (using a female speaker)
- 11 NH and 16 partially-hearing children were used to evaluate the lists
- Each child heard 5 lists and 5 randomized lists of keywords (from 5 other lists) in quiet
- Each list or group of 50 keywords had a different presentation level
- Correctly responded keywords were scored
- Average scores were higher for sentences than for randomized keywords
- Scores of the BKB sentences did not correlate with the Illinois test of psycholinguistic abilities

* Test for auditory/visual SRTs (IHR ASL sentences), MacLeod and Summerfield, 1990

The adaptive sentence list (ASL) test has the following specifications:
- Language is British English, the test was developed for auditory and audio-visual presentation
- It contains 10 lists of 15 sentences of homogeneous audibility and lip-reading difficulty
- Sentences contain on average 5 words, 3 keywords are scored (loosely)
- A sentence is scored as correct when all 3 keywords are repeated correctly
- A male speaker reading the sentences in front of a camera was stored on video tape
- White noise, low-pass filtered at 10 kHz, is used, presented at a fixed level of 60 dB (A)
- Both auditory (A) and audio-visual (AV) thresholds are measured
- For the AV condition subjects see and hear the video recording
- For the A condition only the soundtrack is played (also available on CD)
- An adaptive procedure with +/-2-dB steps is used, depending on the previous response
- The first sentence is started well below threshold and its SNR is increased in 2-dB step until it is understood correctly
- Scores are averaged over the last 10 responses
- They are calculated in dB SN for the SRT in noise at the 50%-correct point for all three keywords
- Mean monaural auditory (A) SRT in white noise is -16.8 (SD = 1.2) dB
- Mean audio-visual (AV) SRT in white noise is -23.2 (SD = 2.3) dB
- The visual benefit is 6.4 dB, with an SD of 2.4 dB
- SRTs in speech-shaped noise were measured by Moore et al. (2001) and Alcántara et al. (2003)
- They used groups consisting of NH listeners with ages from 41 to 77 years
- The first gave SRTs of -11 (SD = 0.3) and -10 (SD = 1.2) dB at noise levels of 60 and 75dB SPL
- The second gave SRT of -8.6 (SD = 1.0) dB at a noise level of 65 dB SPL

Stimulus construction:
- The sentences were based on the BKB sentences (Bench et al., 1979)
- Sentences contained, on average, five words with 3 or 4 keywords
- The intention was to measure both auditory (A) and audio-visual (AV) thresholds
- Video recordings of the BKB sentences had been made by Rosen and Corcoran (1982)
- From corresponding lip-reading scores visual difficulty indexes were calculated for each sentence
- The 3 or 4 keywords per sentence were scored in a loose manner
- Sentences with three keywords with not too high visual scores (< 50%) were selected
- Sentences should not contain more than two keywords beginning with visually distinct consonants
- Keywords without homophonous alternatives should be avoided
- Sentence predictability should be neutral
- 330 new sentences were constructed, grouped in 22 lists of 15 items
- 24 NH and normally-seeing listeners were used to evaluate the sentences
- White noise, low-pass filtered at 10 kHz, is used
- 12 subjects received SNRs of -19 and -26 dB for A and AV conditions, respectively (group 1)
- 12 subjects received SNRs of -20 and -28 dB for A and AV conditions, respectively (group 2)
- The 3 keywords per sentence were scored in a loose manner, scores are given in %-word correct
- The noise was presented at a fixed level of 60 dB (A)
- The stimulus presentation was monaural
- An SNR of 0 dB implied a 60 dB (A) average peak level for the speech (fast-settling setting)
- For group 1, scores were 64 and 54% correct for A and AV conditions (SDs of 23 and 23%), resp.
- For group 2, scores were 58 and 44% correct for A and AV conditions (SDs of 22 and 21%), resp.
- 126 sentences were eliminated on the bases of more than 1.5 SD deviation of A or AV thresholds
- 39 sentences with largely different A and AV difficulty were eliminated
- 10 new 15-item lists were constructed balancing for syntactic structures and average scores
- 20 NH and normally-seeing listeners were used to evaluate the sentences
- Lip-reading abilities were assessed using short sentence lists (MacLeod and Summerfield, 1987)
- A and AV SRTs were measured with the adaptive procedure of Plomp and Mimpen, 1979a)
- Mean A SRT was -16.8 dB (SD = 1.2 dB), with a reliability SD of 1.2 dB
- Mean AV SRT was -23.2 dB (SD = 2.3 dB), with a reliability SD of 2.0 dB
- Visual benefit was 6.4 dB (SD = 2.4 dB), with a reliability SD of 2.3 dB
- List 1 gave a lower AV SRT than the other lists, its A SRT was normal
- List level corrections to equalize scores are given but not implemented
- No practice effects were found
- Slopes of the psychometric functions were 9.9 and 7.4% for A and AV condition, respectively
* Closed set Swedish sentences for speech intelligibility test, Hagerman, 1982; Hagerman, 1984

The test has the following specifications:
- Language is Swedish
- 11 lists of 10 sentences with a fixed structure and consisting of 5 words
- Each list holds 10 permutations with the structure: name, verb, number, adjective, object
- The set of speech material is “closed,” so learning effects may play a role, but redundancy is low
- A female speaker is used
- Speech-shaped noise is used (the long-term spectrum of the sentences)
- All five words are scored
- A list is presented at a fixed SNR
- 50% correct occurs for SNR of -8 dB and for a level in quiet of 22 dB (NH)

Stimulus construction:
- A set of 10 sentences consisting of 5 words (interchangeable between sentences) was constructed
- The sentences had the structure: name, verb, number, adjective, object, each with 10 alternatives
- A female speaker was used in the recordings
- The words were pronounced with silent gaps so the permutations could easily be cut and pasted
- Speech-shaped noise was constructed (long-term spectrum all sentences)
- Slightly amplitude modulated with a low-pass filtered noise (2.1 Hz cut-off)
- 13 lists of permutations of the 10 sentences and 13 noise samples were made and stored on tape
- NH listeners were used to evaluate the lists
- For steeper psychometric functions, the levels of some words were adapted with up to 1.3 dB
- 2 groups of 10 NH listeners were used to evaluate new lists
- The lists were tested for homogeneity, reliability, psychometric functions, and level effects
- Because it is a closed set test the effects of learning were investigated too
- The threshold in noise was found to be -8.1 (test-retest SD = 0.44) dB SN
- The threshold in quiet was found to be 21.6 dB SPL, with an SD of 2.5 dB
- Speech reception in noise scores were maximal at speech level: 55 dB SPL
- Extra testing with 9 NH listeners led to rejection of two lists
- These nine listeners showed an average increase in scores from 50 to 60% over 9 lists
- Further testing was performed with 80 HI listeners (Hagerman, 1984)
- After a single list, scores for sentences in quiet were higher than for isolated words
- Speech reception scores in noise were up to 15 dB worse than for NH listeners
- For HI listeners the repeatability SD was about 1 dB
- Psychometric functions were flatter for HI than for NH listeners
- These slopes were roughly 20 and 25% per sentence, respectively
- Averaged over all responses, the slopes were 9 and 14% per dB, respectively
- SRT scores in noise were correlated to subjective scores
- (At 82%, the number of independent units per sentence was $j = 4.8$, at 28% correct: $j = 2.9$)

* Adaptive Swedish test for speech intelligibility in noise test, Hagerman and Kinnefors, 1995

The test has the following specifications:
- Language is Swedish
- 11 lists of 10 sentences with a fixed structure and consisting of 5 words
- Each list holds 10 permutations with the structure: name, verb, number, adjective, object
- The set of speech material is “closed,” so learning effects may play a role, but redundancy is low
- A practice list is given first to suppress learning effects
- A female speaker is used
- All five words are scored
- An adaptive procedure with up/down steps is used, depending on the previous response
- Test in noise:
  - Speech-shaped noise is used (long-term spectrum of the sentences)
  - Sentences are presented at a fixed comfortable level, noise level is varied
  - The step size for the test in noise is +/-1, 2, or 3 dB, depending on the number of correct words
  - The first sentence starts at an SN of -8 dB
  - Scores are averaged over the last 10 responses of the list
  - Scores are calculated in dB SN for 50% correct
  - Mean SRT in noise is -7.8 (test-retest SD = 0.6) dB (NH)
- Test in quiet:
  - The step size for the test in quiet is twice that for the test in noise
  - The first sentence starts 20 dB above the average absolute threshold at 0.5, 1, and 2 kHz
  - Scores are averaged over the last 6 responses of the list
  - Scores are calculated in dB SPL for 40% correct
  - Mean SRT in quiet is about 21 (SD of 1.6) dB SPL (NH)

Test construction:
- For the stimuli, see construction of closed-set Swedish sentences (above)
- An adaptive procedure with up/down steps is used, depending on the previous word score
- Test in noise:
  - Speech-shaped noise was used (long-term spectrum of the sentences)
  - Sentences were presented at fixed comfortable level, noise level varied
  - The step size for the test in noise was +/-1, 2, or 3 dB, depending on the number of correct words
  - The first sentence started at an SN of -8 dB
  - Five lists were measured
  - Scores were averaged over the last 10 responses of the list
  - Scores were calculated in dB SN for the 50% correct point
  - 10 NH and 39 HI listeners were used
  - Mean SRT in noise is -7.8 (test-retest SD = 0.6) dB (NH)
  - HI listeners had SDs of 0.8 and 1.2 dB for a better and a poorer group
  - Learning effects were small
- Test in quiet:
  - The step size for the test in quiet is twice that for the test in noise
  - The first sentence starts 20 dB above the average absolute threshold at 0.5, 1, and 2 kHz
  - Three lists were measured
  - Scores were averaged over the last 6 responses of the list
  - Scores were calculated in dB SPL for 40% correct, and extrapolated to the 50% correct point
  - 10 NH and 39 HI listeners used in evaluation
  - Mean SRT in quiet is about 21 (SD of 1.6) dB SPL (NH)
  - HI listeners had SDs of about 4 dB
  - Learning effects were very small for NH, and about 2 dB per list for HI
**Hearing in noise sentence test (HINT sentences), Nilsson et al., 1994**

The test has the following specifications:
- Language is American English
- 25 phonemically matched lists with 10 sentences, plus 3 practice lists of 12 sentences
- A male speaker is used
- A 20161-Hz sample rate and 16-bit resolution are used
- Speech-shaped noise is used
- Whole sentences are scored, all words must be correct (relaxed criteria)
- An adaptive procedure with +/-2-dB steps is used, depending on the previous response
- The first sentence started below threshold and its SNR is increased in 2-dB step until understood correctly
- SRT scores are averaged over the last 10 responses
- Scores are calculated in dB(A) in quiet and in dB SN in noise
- Sentences were presented in quiet or in a 72dB(A) speech-shaped noise
- Mean SRT in noise was -2.92 dB, with an SD of 0.78 dB
- Mean SRT in quiet was 23.9 dB(A) with an SD of 3.5 dB

**Stimulus construction:**
- The sentences were based on the BKB sentences (Bench et al., 1979)
- These were revised to Americanize them and to equate sentence lengths
- Sentences were edited for high on naturalness and equal numbers of past and present tenses
- Only sentences with high naturalness were recorded
- A 20161-Hz sample rate and 16-bit resolution were used
- Sentences were recorded roughly at a speaking level of 70dB SPL
- A male speaker was used
- RMS levels ranged from 65 to 74 dB and were equated to 67 dB SPL
- Speech-shaped noise was generated from the long-term spectrum of the sentences (78-point FIR)
- Fixed SNR scores were equated in 7 iterations of level adjustments
- Six to eight listeners were used in each iteration, for -3 or -4 dB SN
- Scoring was on word-by-word basis, very strict at first, later more relaxed
- RMS adjustments in the last iteration were smaller or equal to 0.5 dB
- Of original 336 sentences, 252 were scaled in level
- After scaling the average sentence level was 66.7dB SPL
- Sentences were grouped to 21 lists of 12 on the basis of matched list phonemic content
- List reliability was tested with adaptive method using seventeen subjects
- The first sentence started below threshold and its SNR was increased in 2-dB step until correct
- Subsequent sentences were presented at +/-2 dB SNR depending on correctness of previous one
- Scoring was on whole sentences in a somewhat relaxed way (i.e. pas/present tense “is” “was,” etc)
- Sentences were presented in quiet or in a 72dB(A) speech-shaped noise
- Mean SRT in noise was -2.92 dB, with an SD of 0.78 dB
- Mean SRT in quiet was 23.9 dB(A) with an SD of 3.5 dB
- Sentence material rearranged into 25 list of 10 sentences with practically identical reliability

**German sentence in noise test (Göttingen sentences), Kollmeier and Wesselkamp, 1997**

The test has the following specifications:
Language is German
- 20 phonemically balanced lists of 10 everyday sentences
- A male non-professional speaker is used
- A 44100-Hz sample rate and 16-bit resolution are used
- Speech-shaped noise is presented at a fixed level of 65 dB SPL
- Sentences are presented at either -4 or -8 dB SNR
- Each word is scored and weighted individually to derive a sentence score in percent correct
- The average SRT is -6.2 (SD = 0.3) dB, with a slope of 19% per dB
- Alternatively, whole sentences can be scored at some loss of list consistency (giving an SD of 1 dB)

Stimulus construction:
- Starting point were 400 sentences, including 100 “Marburg” sentences (Niemeyer, 1967)
- After testing for grammatical correctness and contemporary contents 324 sentences remained
- A male speaker was used
- The recorded sentences were digitally stored with a 25kHz sample rate and a 16 bit resolution
- Sentences were evaluated using NH listeners
- A speech-shaped noise at a fixed level of 65 dB SPL was used
- In a pilot the SNR producing approximately 50% correct was determined using four listeners
- For 12 listeners all sentences were scored (including scores for each word) at that particular SNR
- Based on those results the sentences were grouped into 6 sets of approximately equal intelligibility
- A second and third group of listeners provided two more scores, all between 20 and 80%
- The three scores were used for a 3-point estimation of the psychometric function of each sentence
- This gave separate SNR values for 50% correct for all sentences (SRTs)
- Weighting factors were determined for each word in sentences at the SRT
- Weighting factors can be used to derive sentence score from word scores
- They were used to check the original SNRs of the measurements
- For sentences with outlying SRTs, equalizing level correction was applied
- Subsequently 200 sentences with scores between 30% and 70% at SRT were selected
- These were grouped into 20 10-item lists optimizing for phonemes occurrence and number, and for word number, word chance distribution and sentence chance distribution
- From the word scores, average SRT of the lists was -6.1 (SD = 0.2) dB
- When scoring whole sentences this SD was approximately 1 dB
- 20 NH listeners were used to evaluate the lists
- Sentences were presented at -4 or -8 dB SNR, all words were scored
- The average SRT of the lists was found to be -6.2 dB with an SD of 0.3 dB
- The average slope of the psychometric function, near SRT, was 19%/dB
- At -4 dB SN, the number of independent units per sentence was j = 2.38, at -8 dB SN: j = 1.95
- Subjective intelligibility scores with this material were to a degree less selective and consistent

* Closed set German sentence in noise test (Oldenburg sentences), Wagener et al., 1999a; 1999b; 1999c

The test has the following specifications:
- Language is German
- 10 lists of 10 sentences with a fixed structure and consisting of 5 words
- From these lists, 120 60-item lists were created (in addition 2 practice lists are available)
- Each list holds 10 permutations with the structure: name, verb, number, adjective, object
- All permutations are constructed with the correct coarticulation
- The set of speech material is "closed," so learning effects may play a role, but redundancy is low
- A non-professional male speaker is used, the speech rate is average
- A 25000-Hz sample rate and a 16-bit resolution are used
- Speech-shaped noise is used at a fixed level of 65 dB SPL
- All five words are scored
- With fixed SNRs of -5 and -9 dB:
  - The average SRT of the lists was found to be -7.1 (SD = 0.2) dB
  - The average slope of the psychometric function, near SRT, is 17%/dB

**Stimulus construction:**

- A set of 10 sentences consisting of 5 words (interchangeable between sentences) was constructed
- The sentences had the structure: name, verb, number, adjective, object, each with 10 alternatives
- They constituted the average phonemic content of the German language
- This set of speech material is "closed," so learning effects may play a role, but redundancy is low
- Permutations were constructed using multiple (10 per item) recordings that include coarticulation by using forced prosody using a PSOLA method (pitch synchronous overlap-and-add)
- 17 listeners were used to judge the naturalness of lists of permutations
- These test showed that the coarticulation method was closest to the natural pronunciation
- A non-professionals male speaker (Göttingen list) was, speaking at a normal, average, speech rate
- A 44100-Hz sample rate and a 16-bit resolution were used
- 10 permutations of the 10 sentences were recorded, that include all possible coarticulation cases
- The level was equated over the length of the sentences
- 18 lists of 10 sentences which included coarticulation were constructed
- The resolution was reduced to 25-kHz sample rate with 16-bit resolution
- Speech-shaped noise was constructed
- The evaluation used 12 highly experienced NH listeners
- To suppress learning effects, they received training before data collection
- The noise was presented at a fixed level of 60 dB SPL
- Speech and noise were presented diotically
- SN ratios between -4 and -12 dB (later -14 dB) were used in a blocked manner
- All five words were scored
- Words with scores deviating more than 0.5 dB from average are corrected in level (correction less or equal to +/- 2dB)
- 6 lists were excluded on the bases of low threshold estimation precision
- 2 lists were excluded on the bases of large SD for word scores
- The threshold for words was found to be -8.4 dB SN with an SD of 1.1 dB (after the corrections)
- The slope of the psychometric function for words was found to be 17%/dB (after the corrections)
- 20 NH listeners were used to evaluate the 10 lists (plus 2 practice lists)
- The noise was presented at a fixed level of 65 dB SPL
- Speech and noise were presented diotically
In the first part an adaptive method with variable step size was used (Brand and Kollmeier, 2002).

After the second list, a learning effect of about 0.3 dB per list was found.

Over 6 lists, the averaged SRT dropped from -5.2 to -7.2 dB.

Overall, the SRT was -6.5 dB.

In the second part fixed SNR values of -5 and -9 dB were used.

Word scores were 80.7 and 21.7% at SNRs of -5 and -9 dB, respectively.

At -5 dB SN, the number of independent units per sentence was $j = 4.3$, at -9 dB SN: $j = 3.2$.

The average SRT of the lists was found to be -7.1 (SD = 0.2) dB.

The average slope of the psychometric function, near the SRT, is 17%/dB.

*Adaptive versions of the Göttingen and Oldenburg speech tests, Brand and Kollmeier, 2002*

The test has the following specifications:

- Language is German.
- The sentence material is taken from the open-set Göttingen, or closed-set Oldenburg, speech tests.
- A non-professional male speaker is used.
- A 25000-Hz sample rate and 16-bit resolution are used.
- Speech shaped noise is used.
- Word scores are used.
- A maximum likelihood method is used to estimate SRT and slope.
- Step size and direction depend on the word score of the previous sentence.
- The rule can be chosen to estimate the sweet point (50% correct) or optimal pair (20 and 80%).
- Adaptive method converging at 50% correct (slope estimates inaccurate):
  - Göttingen sentences:
    - The SRTs are -5.6 (SD = 0.5) and -0.5 (SD = 0.9) dB for NH and HI, resp.
  - Oldenburg sentences:
    - The SRTs are -6.3 (SD = 0.7) and -2.9 (SD = 0.7) dB for NH and HI, resp.
- Adaptive method converging at 20 and 80% correct
  - Göttingen sentences:
    - The SRTs are -5.5 (SD = 0.8) and -0.8 (SD = 1.1) dB for NH and HI, resp.
    - Slopes are 16 (SD = 2.6) and 13 (SD = 3.1) %/dB for NH and HI, resp.
  - Oldenburg sentences:
    - The SRTs are -6.1 (SD = 0.6) and -2.4 (SD = 0.8) dB for NH and HI, resp.
    - Slopes are 16 (SD = 3.4) and 14 (SD = 4.1) %/dB for NH and HI, resp.

Test construction:

- For the stimuli, see construction of Göttingen and Oldenburg tests (above).
- The aim was a concurrent estimation of SRT and slope of the psychometric function (in %/dB).
- Optimal pair of response probabilities for this was determined theoretically:
  - They were 0.19 and 0.81 for the logistic ("limited" dependent variable) psychometric function.
- The adaptive procedure was designed to converge at target probabilities.
  - This procedure was based on Hagerman and Kinnefors, 1995.
  - This procedure was tested using Monte Carlo simulations.
  - A maximum likelihood method was used to estimate SRT and slope (in %/dB) values concurrently.
  - Step size and direction depended on word score of the previous sentence.
  - The step size could take any value (within reason), and decreased with increasing sentence number.
  - The rule could be adapted to estimate the sweet point (50% correct) or optimal pair (20 and 80%).
- The method was optimized using Monte Carlo simulations
- 10 NH and 11 HI listeners were used to evaluate the tests
- The noise level was fixed at 65 dB SPL for the NH listeners
- The noise level was set to medium loudness for the HI listeners
- Göttingen sentences converging at 50% correct:
  - SRTs were -5.6 (SD = 0.5) and -0.5 (SD = 0.9) dB for NH and HI, resp.
- Oldenburg sentences converging at 50% correct:
  - SRTs were -6.3 (SD = 0.7) and -2.9 (SD = 0.7) dB for NH and HI, resp.
- For convergence at 20 and 80%, the SRTs were very similar and the SDs were slightly larger
- Göttingen sentences converging at 20 and 80% correct:
  - Slopes were 16 (SD = 2.6) and 13 (SD = 3.1) %/dB for NH and HI, resp.
- Oldenburg sentences converging at 20 and 80% correct:
  - Slopes were 16 (SD = 3.4) and 14 (SD = 4.1) %/dB for NH and HI, resp.
- For the slope estimate the recorded bias was about half the size of the SD

* Dantale II, a Danish closed-set sentence test, Wagener et al., 2003

The test has the following specifications:
- Language is Danish
- 16 lists of 10 sentences with a fixed structure and consisting of 5 words are used
- Each list holds 10 permutations with the structure: name, verb, number, adjective, object
- The set of speech material is “closed,” so learning effects may play a role, but redundancy is low
- A female speaker is used
- Speech-shaped noise is used (long-term spectrum of the sentences)
- All five words are scored
- A maximum likelihood method is used to estimate SRT and slope
- Step size and direction depend on the word score of the previous sentence
- Adaptive method converging at 50% correct, Brand and Kollmeier (2002)
- Validation was performed for 60 NH listeners.
- The SRT is -8.43 (SD = 0.95) (training effect of 2.2) dB with a slope of 13.2% (SD = 1.9%)

Stimulus construction:
- A set of 10 sentences consisting of 5 words (interchangeable between sentences) was constructed
- The sentences had the structure: name, verb, numeral, adjective, object, each with 10 alternatives
- The words were chosen in a phonetically balanced way from a list of 5000 high frequency words
- Five sentences had the present and five had the past tense
- A female speaker was used in the recordings
- A 44100-Hz sample rate and a 16-bit resolution were used
- All words were pronounced several times to account for coarticulation
- 10 randomly chosen lists of 10 permutations of the 10 sentences were made and stored digitally
- 10 NH listeners were used for a first evaluation of those lists
- The words were adapted according to the outcome of that evaluation
- 25 new lists of 10 permutations of the 10 sentences were made and stored digitally
- Speech-shaped noise was constructed (long-term spectrum of the sentences in those lists)
- The noise was switch on 500 ms before the sentence started and continued 500 ms after it ended
- 16 NH listeners were used to evaluate the 25 lists
- All words were scored
Speech and noise were presented monaurally
- The noise was presented at a fixed level of 65 dB SPL
- All lists were presented a 10 different SN ratios (-18 to 0 dB, 2-dB step), chosen at random
- For steeper psychometric functions, the levels of some words were adapted with up to 4 dB
- Two groups of 30 NH listeners were used to evaluate the adapted lists
- In the evaluation, 10 lists of 20 sentences were used
- Sentences were presented monaurally with a 65-dB SPL noise level
- All words were scored
- Two SN ratios were used (-10 and -6 dB) for two groups, in a counter balanced way
- Before the evaluation 8 lists were presented in an adaptive procedure (Brand and Kollmeier, 2002)
- The resultant SRTs gave a learning effect of 2.2 dB over 8 lists
- The evaluation gave SRTs in noise of -8.38 (SD = 0.16 across lists) dB
- The corresponding slopes at the 50% point were 12.6 (SD = 0.8) %/dB
- SRTs in noise were -8.43 (SD = 0.95 across subjects) dB
- The corresponding slopes at the 50% point were 13.2 (SD = 1.9) %/dB

* Combescure sentences, a set of French sentences, Combescure, 1981; Apoux et al., 2004

The test has the following specifications:
- The language is French
- 20 phonetically balanced lists of 10 sentences (Combescure, 1981)
- 10 lists with high familiarity scores were selected (Apoux et al., 2004) (to prevent that the lists contained old-fashioned sounding language)
- A 44100-Hz sample rate, a 16-bit resolution, a 5-kHz bandwidth are used
- White noise is used, both stationary and fluctuating according to the envelope of low-pass filtered white noise (16-Hz cut off)
- SN ratios are calculated from the overall RMS values of the sentences and noise bursts
- Fixed SN ratios were used, with a total stimulus level of 70 dBA
- In total 41 measurements with NH listeners were recorded
- Stationary noise:
  - On average (weighted), they scored 37% correct at -18 dB SN
- Fluctuating noise:
  - On average (weighted), they scored 76% correct at -30 dB SN

Stimulus construction:
- Starting point: 200 everyday sentences from Combescure (1981)
- Starting point were 20 phonetically balanced lists
- 10 lists were selected for high familiarity scores in 15 NH listeners (to prevent that the lists contained old-fashioned sounding language)
- These 10 lists were recorded by a male native speaker
- 44100-Hz sample rate, a 16-bit resolution, a 5-kHz bandwidth were used
- Stationary and fluctuating white noises were constructed
- The fluctuating noise had the low-pass (16-Hz cut off) filtered envelope of a white noise burst
- SN ratios were calculated from the overall RMS values of the sentences and noise bursts
- Fixed SN ratios were used: -18 and -30 dB for stationary and fluctuating noise, respectively
- These SN ratios had been determined in a preliminary experiment
- The mixed stimuli were presented at a fixed level of 70 dBA
- All words were scored
- Speech and noise were presented binaurally
- In the first experiment, 8 NH listeners were used, aged from 21 to 24
- They scored 45% of words correct for stationary noise at -18 dB SN
- They scored 80% of words correct for fluctuating noise at -30 dB SN
- In the second experiment, 18 NH listeners were used, aged from 20 to 46
  - They scored 40% of words correct for stationary noise at -18 dB SN
  - They scored 57% of words correct for fluctuating noise at -30 dB SN
- In the third experiment, 15 NH listeners were used, aged from 20 to 29
  - They scored 28% of words correct for stationary noise at -18 dB SN
  - They scored 97% of words correct for fluctuating noise at -30 dB SN

* Audivox sentences, a set of French sentences, Wable, 2005

The test has the following specifications:
- Language is French
- The test contains 19 lists of 8 sentences
- A male speaker was used to record the sentences
- A 44100-Hz sample rate and a 16-bit resolution are used
- Speech-shaped noise is used
- An adaptive procedure is used with 1-dB steps (2-dB at first two turns)
- The SRT is averaged over the last 9 presentation levels (including the planned level at the end)
- The starting SN ratio is 0 dB SN for HI listeners
- Sentences are presented at a comfortable overall level that gives optimal word scores
- Validation was performed with HI listeners
- The average SRT was -0.7 dB, with a test-retest reliability of 1.1 dB
- The inter-list variability was 0.7 dB, the slope near threshold was approximately 20%/dB

Stimulus construction:
- 582 sentences were constructed using high-frequency words
  - These sentences were recorded using a male speaker
  - A 44100-Hz sample rate and a 16-bit resolution were used
  - After checking the recordings for pronunciation, noise, and clipping, 500 sentences remained
  - These sentences were equated in level to the same RMS value
  - 150 concatenated sentences used to generate 60 s speech-shaped noise
  - In the presentations, the noise started 300 ms before, and ended 300 ms after, the sentence
  - SN ratios were calculated across the regions where sentences and noise were superimposed
  - 193 HI listeners, average age of 73 (22 to 94), listened to 500 sentences
  - Their average loss was: PTA (0.5, 1, 2 kHz) = 42 dB (20 to 83 dB)
  - Sentences were presented at 0 dB SN, at a comfortable overall level that gave optimal word scores
  - Total: 15 listeners per sentence and 25 to 30 sentences per listener
  - Whole sentences were scores as correct or incorrect (maximum score of 15 per sentence)
  - 200 sentences with scores between 10 and 13 out of 15 were selected (on average 76% correct)
  - 240 HI listeners, average age of 74 (30 to 94), listened to 200 sentences
  - Their average loss was: PTA (0.5, 1, 2 kHz) = 42 dB (20 to 82 dB)
  - In two separate conditions, they did or did not use their hearing aids
  - Sentences were presented at 0 or -2 dB SN, at a comfortable level that gave optimal word scores
  - With 12 listeners per sentence for aided and unaided conditions, and 40 sentences per listener
Whole sentences were scored as correct or incorrect (maximum score of 24 per sentence).
- Aided, scores were 70 and 26% correct for 0 and -2 dB SN, respectively.
- Unaided, scores were 64 and 26% correct for 0 and -2 dB SN, respectively.
- So, the slope near -1 dB SN was approximately 20%/dB.
- 152 sentences were selected for maximum homogeneity.
- They had scores more than 2 apart for 0 or -2 dB SN, and less than 4 for aided/unaided.
- 19 lists of 8 sentences were chosen to minimize score differences for the 4 measured conditions.
- 33 HI listeners were used for the evaluation of the lists.
- They had an average age of 77 (46 to 90 years).
- Their average loss was: PTA (0.5, 1, 2 kHz) = 38 dB (27 to 70 dB).
- 10 subjects listened to all the sentences.
- 23 subjects listened to a subset of the sentences, their scores were used to evaluate test reliability.
- They had an average age of 78 (52 to 90 years).
- Their average loss was: PTA (0.5, 1, 2 kHz) = 40 dB (27 to 70 dB).
- No significant effect of list was found, the inter-list variability was 0.7 dB.
- The average SRT was -0.7 (SD = 1.0 dB, test-retest reliability is 1.1 dB.)
Appendix B, properties of questionnaires

* APHAB (American English, Dutch, French, German, Italian, Magyar, Norwegian, Polish, Portuguese, Slovenian, Spanish, Finnish, Turkish)

Abbreviated Profile of Hearing Aid Benefit
- To quantify disability from hearing loss and its reduction with hearing aids
  - 24 items of 4 subscales
    o ease of communication
    o reverberation
    o background noise
    o aversive ness of sound
  - Validation: 128 elderly adults (90m, 38f) with mild-to-moderate flat loss
  - Translations downloadable at: http://www.ausp.memphis.edu/harl/aphab.html
  - Cox and Alexander, 1995; Paul and Cox, 1995

* Amsterdam Inventory (Dutch, English, Spanish, Swedish)

Amsterdam Inventory for auditory Disability and Handicap
- To identify factors in hearing disability in daily life and assess its associated handicap
  - 30 items, 6 factors
    o detection of sounds
    o distinction of sounds
    o auditory localization
    o speech intelligibility in quiet
    o speech intelligibility in noise
    o intolerance of noise
  - Validation (Dutch version): 274 adults (16-66 years)
  - Kramer et al., 1995; Kramer et al., 1996; Kramer et al., 1998

* COSI (Australian English)

Client Oriented Scale of Improvement
- To identify up to five areas of listening difficulty and degree of hearing-aid benefit, compared to standard population
  - 16 standardized listening situation
  - Validation: 1770 adults with hearing loss (56% first time users)
  - Dillon et al., 1997; Dillon et al., 1999

* CPHI (American English)

Communication Profile for the Hearing Impaired
- To provide a systematic and comprehensive assessment of communication problems
  - 145 items dealing with four areas
    o communication performance
    o communication environment
    o communication strategies
    o personal adjustment
  - Validation: 827 patients at Walter Reed Army Medical Center
  - Demorest and Erdman, 1987
* **CSOA (American English)**

Communication Scale for Older Adults
- To investigate the effects of aural rehabilitation on daily life, evaluates communication strategies and feelings about hearing loss
- 145 items dealing with four areas
  - communication strategies (41 items)
  - communication attitudes (31 items)
- Validation: 135 adults (60-88 years)
- Kaplan et al., 1997

* **CSDA (American English)**

(Revised) Communication Self Assessment Scale Inventory for Deaf Adults
- To assess communication ability of prelingually deafened adults
- 125 items with four scales
  - difficult communication situation
  - importance of each situation
  - communication strategies
  - communication attitudes
- Validation: 242 prelingually deafened young adults (college and prep school age)
- Kaplan et al., 1995

* **DSCF-M (American English)**

Denver Scale of Communication Function – Modified
- To assess communication skills of senior citizens
- 34 items dealing with four areas
  - attitudes toward peers
  - socialization
  - communication
  - difficult listening situations
- Validation: 12 adults (70-92 years)
- Kaplan et al., 1978

* **GHABP (British English)**

Glasgow Hearing Aid Benefit Profile
- To assess the aspects of disability, handicap, and hearing-aid benefit
- 4 prespecified, 4 subject-specifed, items across six dimensions
  - disability
  - handicap
  - hearing-aid use
  - benefit
  - satisfaction
  - residual disability
- Validation: 293 adults (median age: 69 years)
- Gatehouse, 1999

* **GP (Swedish)**

Gothenburg Profile
- To measure the experienced hearing disability and handicap using a self-report inventory
- 20 items with 2 scales
  - Experienced disability
    - being able to hear speech (5 items)
- being able to localize sounds (5 items)
  - Experienced handicap
    - impact of hearing impairment (5 items)
    - how you perform and react (5 items)
- Validation: 924 persons (14-91 years)
- Arlinger et al., 1998; Ringdahl et al., 1998

* HANA (American English)

Hearing Aid Needs Assessment
- To examine the relation between perceived communication needs and actual benefit from hearing aids
- 11 items with 3 ratings
  - how often (hardly ever, occasionally, frequently)
  - how much trouble (very little, some, very much)
  - how much help expected (very little, some, very much)
- Validation: 82 adults (54m, 28f; 25 previous users, 57 new users)
- Schum, 1999

* HAPI (American English)

Hearing Aid Performance Inventory
- To assess effectiveness of amplification in a variety of listening situations
- 64 items with 12 bipolar features (e.g., with/without visual cues)
- Validation: 128 hearing-aid users (119m, 9f)
- Walden et al., 1984

* HAUQ (Australian English)

Hearing Aid Users Questionnaire
- To detect problems that may affect ability to use, and benefit from, hearing aids
- 11 items to assess use, benefit, problems, and satisfaction
- Validation: 4421 adults with hearing loss
- Dillon et al., 1999

* HARQ (British English)

Hearing Attitudes in Rehabilitation Questionnaire
- To measure attitude toward hearing impairment and provision of a hearing aid, in older people
- 40 items with 7 scales measuring attitude toward
  - hearing impairment
    - personal distress (9 items)
    - hearing-loss stigma (5 items)
    - minimization of hearing loss (6 items)
  - hearing aid
    - hearing-aid stigma (9 items)
    - aid not wanted (5 items)
    - positive expectation of aid (3 items)
- Validation: 140 adults (36-96 years)
- Hallam and Brooks, 1996
* **HCA (Swedish)**

**Hearing Coping Assessment**
- To assess patients’ own view of his/her ability to cope with hearing impairment
  - 21 multiple-choice items concerning
    - problem-focused coping (disability) (12 items)
    - emotion-focused coping (handicap) (9 items)
- Validation: 114 hearing-aid users (44-88 years)
  - Andersson et al., 1995

* **HDHS (French, Swedish)**

**Hearing Disabilities and Handicap Scale**
- To measure the severity of the most common hearing disabilities and handicaps from hearing loss
  - 20 items with 3 factors concerning
    - speech perception
    - non-speech sounds
    - handicap
- Validation: 242 adults (20-75 years)
  - Hétu et al., 1994; Hallberg, 1998

* **HHDI (Dutch)**

**Hearing Handicap and Disability Inventory**
- To measure the consequences of hearing loss in elderly indicating rehabilitative needs and effects of aural rehabilitation
  - 40 items with 4 scales concerning
    - performance (10 items)
    - emotional response (10 items)
    - social withdrawal (10 items)
    - perceived reactions of others (10 items)
- Validation: 262 elderly (60-95 years)
  - van den Brink, 1995

* **HHIA (American English)**

**Hearing Handicap Inventory for Adults**
- To quantify perceived handicap. Can also be used to assess benefit of hearing aids by measuring pre and post fitting of the aid
  - 25 items with 2 subscales concerning
    - emotional consequences
    - social and situational effects
- Validation: 28 adults (29-59 years)
  - Newman et al., 1990; Newman et al., 1991

* **HHIE (American English)**

**Hearing Handicap Inventory for the Elderly**
- To measure the perceived effects of hearing loss on non-institutionalized elderly
  - 25 items with 2 subscales concerning
    - emotional consequences
    - social and situational effects
- Validation: 47 non-institutionalized elderly adults, of which 20 measured face-to-face, and 27 using pencil and paper
  - Ventry and Weinstein, 1982; Weinstein et al., 1986
* **HHS (American English)**

Hearing Handicap Scale
- To quantify disadvantage caused by hearing loss in everyday listening. Can also be used to measure disability
  - 20 items
  - Validation: 24 new hearing-aid users pre and post fitting (8m, 16f; 56-91 years)
  - High et al., 1964; Tannahill, 1979

* **HMS (British/Australian English)**

Hearing Measure Scale
- To assess auditory disability
  - 42 items covering 7 areas
    - speech hearing
    - acuity for non-speech sounds
    - localization
    - emotional response
    - speech distortion
    - tinnitus
    - personal opinion
  - Validation: 27 adult males (chippers in a foundry)
  - Noble and Atherley, 1970

* **HPI (American English)**

Hearing Performance Inventory
- To assess hearing-related performance in problem areas experienced in everyday conditions
  - 158 items covering
    - understanding speech
    - intensity
    - response to auditory failure
    - social
    - personal
    - occupational
  - Validation: 190 adults
  - Giolas et al., 1979

* **HPI-R (American English)**

Hearing Performance Inventory - Revisited
- To assess communication difficulties of individuals with hearing loss
  - 90 items covering
    - understanding speech
    - intensity
    - response to auditory failure
    - social
    - personal
    - occupational
  - Validation: 354 adults (typical cross section of community)
  - Lamb et al., 1983
* HPISPL (American English)

Hearing Performance Inventory for Severe to Profound Loss
- To assess communication difficulties of individuals with severe to profound hearing loss
- 58 items with 6 subscales covering
  - understanding speech with visual cues
  - intensity
  - response to auditory failure
  - environmental sounds
  - understanding speech with no visual cues
  - personal
  - 16-item rest category
- Validation: 50 adults with severe to profound hearing loss
  - Owens and Raggio, 1988

* HPI (American English)

Hearing Problem Inventory
- To assess benefit of hearing aid and counselling. To identify influences on patients’ perception, their problems, and their hearing aid use
- 50 items
  - emotional reaction to hearing loss
  - effect of hearing loss on everyday activities
  - signal and environmental influences
  - use of visual cues
  - use, fit, and care of hearing aid
- Validation: 329 adults pre and post fitting
  - 5 hearing-loss groups
  - 4 age groups
  - 2 experience groups
  - Hutton, 1980

* IOI-HA (British/Australian English, Chinese, Danish, Dutch, Finnish, French, German, Greek, Hebrew, Italian, Japanese, Norwegian, Polish, Portuguese [Brazilian], Russian, Serbian, Sinhalese, Slovian, Spanish, Swedish, Welch)

International Outcome Inventory for Hearing Aids
- To evaluate the effectiveness of hearing-aid treatments. Designed to be applied in different languages and settings
- 7 items with a 5-point scale covering:
  - daily use
  - benefit
  - residual activity limitations
  - satisfaction
  - residual participation restrictions
  - impact on others
  - quality of life
- Validation:
  - British English: 172 adult hearing-aid users (100m, 72f; 26-98 years)
  - Wales (English): 161 adult hearing-aid users (76m, 85f; 40-92 years; 34 experienced users, 117 new users)
  - Dutch: 505 hearing aid users (280m, 225f; 15-97 years; 210 uni-, 295 bilateral)
- Translations downloadable at: http://www.ausp.memphis.edu/harl/ioiha.html
  - Cox and Alexander, 2002; Kramer et al., 2002; Stephens, 2002; Cox et al., 2003; Cox et al., 2002
Noble (Noble, 2002) proposed adaptations of the IOI are available for significant others (IOI-HA-SO) and alternative intervention, such as TVs, telephones, tinnitus maskers, training, counselling, advice on tactics and strategies, or surgery (IOI-AI). These questionnaires have not yet been validated.

*Magnitude Estimation (Egyptian Arab and British English)*

Magnitude Estimation Technique
- To obtain patients’ estimates of their overall disability and handicap
- 1 item concerning this on a scale of 0 to 100
- Validation:
  - London: 69 individuals (16-81 years)
  - Cairo: 39 individuals (19-60 years)
- Habib and Hinchcliffe, 1978

*McCarthy-Alpiner lists (American English)*

McCarthy-Alpiner Scale of Hearing Handicap
- To assess handicapping effects of hearing loss by individual and by family members (results to be used in family counselling)
- 100 items, reduced to 2 forms of 34 items over 3 areas:
  - social effects of adult hearing loss
  - psychological effects of adult hearing loss
  - vocational effects of adult hearing loss
- Validation: 100 adults (21-50 years)
- McCarthy and Alpiner, 1983

*NHII (American English)*

Nursing Home Hearing Handicap Index
- To assess hearing handicap
- 2 versions of 10 items, one for residents, one for staff
- Validation: 105 residents and staff of four nursing homes
- Schow and Nerbonne, 1977; Schow and Nerbonne, 1976

*OI (German)*

Oldenburg Inventory
- To assess individuals subjective hearing handicap and to measure hearing performance in everyday situations
- 21 items, 5 factors:
  - intelligibility in quiet (5 items)
  - intelligibility in noise (5 items)
  - directional hearing (3 items)
  - tinnitus (3 items)
  - psycho-social handicap (5 items)
- Validation: 83 adults (21-89 years)

*OPQ (British English)*

Open-ended Problems Questionnaire (for clinical use)
- To assess what hearing-impaired individuals consider to be their main problems arising from their hearing impairment
- 1 open item has been used with different purposes in different groups
  - a group of cochlear implantees (N=19)
  - a group of significant others
to investigate positive aspects of hearing (N=50)
o as an outcome measure of hearing-aid fitting
o to assess disability and handicap (N=500)
- Barcham and Stephens, 1980; Stephens et al., 1998

* PAL (American English)

Profile of Aided Loudness
- To determine if loudness restoration has been accomplished by hearing rehabilitation
- 12 items across categories: soft, average, and loud
- Validation: normalized using 41 adults with normal hearing (26-65 years)
- Mueller and Palmer, 1998; Palmer et al., 1999

* PHAB (American English)

Profile of Hearing Aid Benefit
- To generate a measure of hearing aid benefit from aided-unaided difference score
  (developed as research tool)
- 66 items in 7 subscales
  o familiar talkers
  o ease of communication
  o reverberation
  o reduced cues
  o background noise
  o aversiveness of sounds
  o distortion of sounds
- Validation: 42 hearing-aid users
- Cox et al., 1991; Cox and Rivera, 1992

* PHAP (American English)

Profile of Hearing Aid Performance
- To generate a measure of hearing aid performance rather than benefit (developed as research tool)
- 66 items to measure two aspects of performance with hearing aids
  o speech communication in typical workday situations
  o reactions to loudness or quality of environmental sounds
- Validation:
  o 225 subjects in selection of items
  o 76 subjects in internal consistency
  o 30 subjects in test-retest reliability
- Cox and Gilmore, 1990

* SADL (American English)

Satisfaction with Amplification in Daily Life
- To quantify satisfaction with hearing aid or aids using a self-report inventory
- 15 items in 4 subscales
  o positive effects (6 items)
  o service and costs (3 items)
  o negative features (3 items)
  o personal image (2 items)
- Validation: 104 adults
- Cox and Alexander, 1999

* SSQ (British/Australian English)
Speech, Spatial, and Qualities of Hearing Scale
- To measure a range of hearing disabilities across several domains, using a self-report inventory
  - 50 items in 9 subscales
    o hearing speech in a variety of competing contexts
    o the directional, distance and movement components of spatial hearing
    o the abilities to segregate sounds
    o the abilities to attend to simultaneous speech streams
    o qualities of hearing experience regarding ease of listening
    o qualities of hearing experience regarding naturalness
    o clarity and identifiability of different speakers
    o clarity and identifiability of different musical pieces and instruments
    o clarity and identifiability of different everyday sounds
- Validation: 153 new clinic clients
- Noble and Gatehouse, 2004; Gatehouse and Noble, 2004
# Appendix C, abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2AFC</td>
<td>two alternative forced choice</td>
</tr>
<tr>
<td>3AFC</td>
<td>three alternative forced choice</td>
</tr>
<tr>
<td>3I3AFC</td>
<td>three interval three alternative forced choice</td>
</tr>
<tr>
<td>3rd - oct band</td>
<td>third octave band</td>
</tr>
<tr>
<td>AC</td>
<td>air conduction</td>
</tr>
<tr>
<td>Acc.</td>
<td>accuracy</td>
</tr>
<tr>
<td>Adapt. proc.</td>
<td>adaptive procedure</td>
</tr>
<tr>
<td>ASL</td>
<td>adaptive sentence list</td>
</tr>
<tr>
<td>BC</td>
<td>bone conduction</td>
</tr>
<tr>
<td>BILD</td>
<td>binaural intelligibility level difference</td>
</tr>
<tr>
<td>BMLD</td>
<td>binaural masking level difference</td>
</tr>
<tr>
<td>CI</td>
<td>cochlear implant</td>
</tr>
<tr>
<td>CVC</td>
<td>consonant-vowel-consonant</td>
</tr>
<tr>
<td>CVR</td>
<td>consonant vowel intensity ratio</td>
</tr>
<tr>
<td>DSOM</td>
<td>downward spread of masking</td>
</tr>
<tr>
<td>ERB</td>
<td>equivalent rectangular bandwidth</td>
</tr>
<tr>
<td>ERBN</td>
<td>equivalent rectangular bandwidth of the auditory filter at 1 kHz</td>
</tr>
<tr>
<td>Extra info.</td>
<td>extra information</td>
</tr>
<tr>
<td>f</td>
<td>female</td>
</tr>
<tr>
<td>F-T</td>
<td>frequency-temporal</td>
</tr>
<tr>
<td>GHABP</td>
<td>Glasgow hearing aid benefit profile</td>
</tr>
<tr>
<td>GP</td>
<td>Gothenburg profile</td>
</tr>
<tr>
<td>HA</td>
<td>hearing aid</td>
</tr>
<tr>
<td>HI</td>
<td>hearing impaired</td>
</tr>
<tr>
<td>HL</td>
<td>hearing level</td>
</tr>
<tr>
<td>HRTF</td>
<td>head related transfer function</td>
</tr>
<tr>
<td>ICRA</td>
<td>international collegium of rehabilitative audiology</td>
</tr>
<tr>
<td>IHAFF</td>
<td>independent hearing aid fitting forum</td>
</tr>
<tr>
<td>IHC</td>
<td>inner hair cell</td>
</tr>
<tr>
<td>ILD</td>
<td>intelligibility level difference</td>
</tr>
<tr>
<td>jnd</td>
<td>just noticeable difference</td>
</tr>
<tr>
<td>LNN</td>
<td>low-noise noise</td>
</tr>
<tr>
<td>m</td>
<td>male</td>
</tr>
<tr>
<td>MAA</td>
<td>minimal audible angle</td>
</tr>
<tr>
<td>MCL</td>
<td>most comfortable loudness level</td>
</tr>
<tr>
<td>OHC</td>
<td>outer hair cell</td>
</tr>
<tr>
<td>OI</td>
<td>Oldenburg inventory</td>
</tr>
<tr>
<td>PDR</td>
<td>position dynamic range</td>
</tr>
<tr>
<td>PH</td>
<td>high predictability</td>
</tr>
<tr>
<td>PL</td>
<td>low predictability</td>
</tr>
<tr>
<td>PM</td>
<td>medium predictability</td>
</tr>
<tr>
<td>PMTF</td>
<td>psychophysical modulation transfer function</td>
</tr>
<tr>
<td>PTC</td>
<td>psychophysical tuning curve</td>
</tr>
<tr>
<td>RMS</td>
<td>root mean square</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>sent.</td>
<td>sentences</td>
</tr>
<tr>
<td>sev.</td>
<td>severe</td>
</tr>
<tr>
<td>SIAM</td>
<td>single interval adjustment matrix</td>
</tr>
<tr>
<td>SN or SNR</td>
<td>signal to noise ratio</td>
</tr>
<tr>
<td>SP</td>
<td>sub project</td>
</tr>
<tr>
<td>SPL</td>
<td>sound pressure level</td>
</tr>
<tr>
<td>SRT</td>
<td>speech reception threshold</td>
</tr>
<tr>
<td>SSQ</td>
<td>Speech Spatial Qualities</td>
</tr>
<tr>
<td>TEN test</td>
<td>threshold equalizing noise test</td>
</tr>
<tr>
<td>THR</td>
<td>threshold</td>
</tr>
<tr>
<td>TMTF</td>
<td>temporal modulation transfer function</td>
</tr>
<tr>
<td>TRT</td>
<td>textual SRT</td>
</tr>
<tr>
<td>UCL</td>
<td>uncomfortable loudness level</td>
</tr>
<tr>
<td>USOM</td>
<td>upward spread of masking</td>
</tr>
<tr>
<td>VCV</td>
<td>vowel-consonant-vowel</td>
</tr>
<tr>
<td>vis.</td>
<td>visual</td>
</tr>
<tr>
<td>WHF</td>
<td>Wurzburg hearing field</td>
</tr>
<tr>
<td>WP</td>
<td>work package</td>
</tr>
</tbody>
</table>
Literature


Lutman ME (2005) Personal communication.


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


