FP6–004171 HEARCOM

Hearing in the Communication Society

INTEGRATED PROJECT

Information Society Technologies

D-7-2: Specifications of standard speech test / environmental conditions for Europe

Contractual Date of Delivery:	31-7-2006 + 45 days
Actual Date of Delivery:	26-9-2006
Editor:	M. Büchler, N. Dillier
Sub-Project/Work-Package:	SP3 / WP7
Version:	2.0
Total number of pages:	43

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

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

Deliverable D-7-2

VERSION DETAILS	
Version:	2.0
Date:	8.9.2006
Status:	Final

CONTRIBUTOR(S) to DELIVERABLE	
Partner	**Name**
UHZ	N. Dillier, M. Büchler
SAT	H. Puder: Technical Evaluation of Microphone Array Processing

DOCUMENT HISTORY			
Version	**Date**	**Responsible**	**Description**
0.1	4.10.2005	M. Büchler	First draft
0.2	8.2.2006	M. Büchler	Various edits
1.0	28.2.2006	N. Dillier	Various edits
1.1	23.3.2006	M. Büchler	Various edits, mainly chapter 4
1.1	27.3.2006	N. Dillier	A few more edits
1.2	13.4.2006	M. Büchler	Various edits
1.3	23.6.2006	M. Büchler	Various edits, mainly chapter 4
1.4	29.7.2006	N. Dillier	A few more edits
2.0	8.9.2006	M. Büchler	Proposed changes after review and some more updates

DELIVERABLE REVIEW			
Version	**Date**	**Reviewed by**	**Conclusion***
1.4	29.7.2006	K.Eneman, V.Hohmann	Update

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

Table of Contents

Acknowledgement

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

Pre-Amble

HearCom Workpackage (WP) 7 focuses on the evaluation of new hearing instrument algorithms with hearing impaired subjects using standardized test procedures in different European languages. The algorithms were developed and technically evaluated in WP5 (D-5-1 and D-5-2). Deliverable D-7-1 summarizes the speech materials and measurement procedures to be used for the evaluation on experimental subjects. Speech tests are also developed and presented in WP 1 and WP2, but with a different goal, which is, to assess the communication performance and the auditory impairment of a subject (D-1-2 and D-2-1). Furthermore, investigations in WP9 (D-9-1) identified the everyday situations which can be difficult and challenging for hearing impaired subjects.

This deliverable summarizes the environmental acoustical conditions in which such evaluation tests are to be carried out. Specifically, a standard setup is proposed for the evaluations of the various algorithms. The setup is motivated by the characteristics of the processing algorithms and their technical evaluations as described in D-5-1 and D-5-2, as well as by the findings of WP9. The goal of this work was to define test conditions that are well suited to demonstrate the features of the signal processing algorithms, and in addition representative for realistic everyday situations.

Investigations within WP3 (D-3-3) have already addressed the issue of real life listening conditions, such as various amounts of room reverberation. The methods and results of these investigations also served as guidelines for this deliverable.

1 Executive Summary

This report summarizes the different test setups and test environments for the evaluation of algorithms and procedures for hearing systems used by listeners in various communication situations.

Users of hearing instruments and cochlear implants often report that their hearing abilities in real life situations and everyday communication is more impaired than what would be assumed by testing them in a quiet sound treated room. The effects of room acoustics such as echoes and reverberation as well as a variety of interfering noises degrades the signal quality and reduces speech recognition considerably.

Algorithms which attempt to improve the signal quality by reducing interfering noise and reverberation therefore need to be evaluated in test environments which are representative for the difficult communication situation of the hearing impaired listeners. Presentation of target and interfering signals should be done via loudspeakers to accommodate different types of hearing systems with microphones located in the ear canal or behind the ear. The number and characteristics of the loudspeakers should allow reproducing essential characteristics of various acoustic environments in a controlled and systematic manner.

As the purpose of the evaluation tests is to assess the influence of parameter variations of the algorithms under investigation, the words and sentences should be presented such that learning and memory effects remain small and controlled. This requires either a very large number of equivalent word and sentence lists or a randomization procedure which generates unpredictable combinations of test words. The latter option has been developed and established for different European languages as described in the HearCom report D-7-1 and will be mainly used for the purpose of the algorithm evaluations.

This report is intended as a guideline to set up experimental conditions for audiological verification in realistic yet controlled test environments.

The intended audience includes professional audiologists who attempt to systematically investigate the function of specific features of hearing systems, in particular noise reduction and other complex signal processing options.

2 Introduction

2.1 Modern Hearing Systems

Digital technology in today's hearing systems (HS) has made it possible to implement highly sophisticated algorithms that aim to improve speech intelligibility and/or sound quality. The term hearing systems includes hearing instruments (HI) which provide amplification and modification of sound signals presented acoustically or through mechanical vibration of middle ear structures (middle ear implants, MEI) as well as cochlear implants (CI) which stimulate the auditory nerve electrically via electrode arrays placed into the inner ear (cochlea). Because of the particularly damaging effects of background noise on the speech intelligibility for people with hearing loss, this problem is of critical importance. Generally, the performance of such algorithms is evaluated through technical measures on the one hand and through tests with hearing impaired subjects on the other hand. For specific purposes, simulations of hearing impairments for normal hearing listeners may also be used for algorithm evaluations (Poissant et al., 2006).

In HearCom work package 5 (WP5), a number of hearing system signal enhancement approaches have been defined and implemented. Moreover, technical evaluations of these approaches have been carried out and are described in detail in D-5-1 (2005) and D-5-2 (2006). In particular, a number of test conditions and test signals have been defined for the technical evaluation of the different algorithms. The evaluation of the aided performance (i.e. on hearing impaired subjects) is a major task for WP7. A pilot study on subjets will also be carried out in WP5. The selection of test conditions will be based on the definitions described in D-5-1 and D-5-2.

Laboratory studies of advanced signal processing hearing aids typically evaluate benefit for speech recognition in noise with respect to the factors of audibility, directionality, and noise reduction. Many competing requirements must be considered when designing test materials to assess human performance with hearing aids. Among these are the general experimental design requirements of reliability, sensitivity, and validity. In addition, when quantifying speech recognition in noise, temporal characteristics (used by noise reduction algorithms) and spatial characteristics (used by directional microphone systems) must be controlled or consistently manipulated to allow for parsing of individual contributors to performance. Of special interest in this project is the ability to replicate equipment setup at different sites to produce comparable performance in a range of listening conditions. For broader use, this equipment should neither be too complex nor too expensive.

In general, such experiments can be carried out with the stimuli presented over headphones or over loudspeakers in the freefield. Furthermore, the signals can be preprocessed with the respective algorithm, or they can be processed at the time of stimulation. In the HearCom project, the algorithms to be tested will be running real-time on a digital signal processor that is connected to a hearing aid satellite. The hearing impaired test subject will wear this satellite on his head, and therefore, the experiments will be carried out in the freefield.

For the evaluation of hearing system algorithms on subjects, different state of the art methods have been used in the past. While some of the methods are comparable, differences between methods occur for example in the setup of the hardware or in the test signals. Some methods have already been defined in other HearCom work packages, but with a different focus. In WP1 and WP2, tests are defined and implemented to assess the communication performance and the auditory impairment. The performance of a subject will either be measured at the audiologist, or as self-assessment over the telephone or internet (D-1-2, 2005; D-2-1, 2005; D-2-1b, 2006). In WP7, the performance of the *algorithms* on hearing impaired subjects will be assessed, not the impairment itself. However, some of the speech reception tests used for the tests in WP1 and WP2 can also be used for WP7. They are described in deliverable D-7-1 (2005). In the present document, common procedures are defined for the performance evaluation of hearing system signal processing algorithms with experimental subjects.

2.2 Everyday Hearing Environments

In everyday life, auditory-verbal communication is very important for information and social exchange. Hearing impaired persons often have difficulties in following a conversation or understanding announcements, especially when there are competing speakers or other background noise. Not surprisingly, they report that the most relevant listening situations for them are situations in “social life” where the target speech signal is distorted by background noise, which often means by competing speakers (Fedtke et al., 1990; Gabriel, 2003). In the HearCom report D-9-1 (2005), hearing impaired subjects were asked for daily life situations where problems were encountered. The most difficult situations were public announcements and when listening to dialogues at the theatre or cinema. Background noises in these situations include speech and traffic noise, but also music. Furthermore, users of multichannel cochlear implants all experience difficulty in noisy listening situations (Fetterman and Domico, 2002).

This is the reason why many signal enhancement approaches in hearing systems aim to improve speech understanding in noise, and why speech in noise tests are commonly used for the evaluation of signal enhancement approaches, assessing the speech reception threshold SRT[1] (see D-7-1, 2005).

Other important situations for HI users named were “in the house”, “in traffic”, “in the car”, “leisure time”, “work”, “TV”, and “music”. Thus, important signal categories in these situations are again speech in noise or speech in quiet, and music. In consequence, it is also important to assess the perceived quality of the signals after they have undergone some kind of enhancement.

2.3 Hearing System Signal Enhancement Approaches

Following the reflections in section 2.2, signal enhancement in hearing systems typically consists of enhancing the target signal and attenuating any undesired background noise or jammer signal. The approaches can mainly be divided into single-channel and multi-channel techniques. Furthermore, the acoustic characteristics of the hearing system itself can lead to artefacts that may be suppressed by signal processing.

The algorithms selected for real-time implementation in WP5 cover the state of the art of such approaches:

Single-channel approaches:

- Single-channel noise reduction algorithms

Multi-channel approaches:

- Adaptive beamforming algorithms (BF)
- Blind source separation algorithm (BSS)
- Coherence based algorithms (binaural de-reverberation and noise suppression)

Artefact reduction:

- Feedback suppression algorithm

Typical tests for such evaluations are speech perception tests, localization tests, and quality judgments. The next chapter provides an overview over the most commonly used setups of such tests, with a focus on speech perception in noise.

[1] The objective of this measure is to obtain the lowest level at which speech can be identified at least half the time, see section 3.5.

3 Evaluation of Algorithms on Aided Subjects: State of the Art

3.1 Single-Channel Approaches

The benefit of noise reduction algorithms is typically assessed by measuring the improvement in the SRT of a speech-in-noise test (e.g. SPIN test) in the conditions unprocessed versus processed signal.

A typical setup for the experimental evaluation of a noise canceller for hearing instruments has been used for the EU project LISCOM (Listening comfort system for hearing instruments and telephones, contract DE3005) and is described in Dahlquist et al. (2005). Signal and noise were both presented from the front (0°), and an adaptive method was used to find the SNR for 80% correct of two Plomp-type sentence tests. Verschuur et al. (submitted) evaluated the same noise canceller on CI users. For hearing aid users, this SNR is typically around 0 to 5 dB, for CI users around 10 dB. Verschuur et al. showed that a noise reduction algorithm may perform differently at different SNRs, and thus, a different benefit can be expected for CI than for HI users.

A noise reduction algorithm may need some time to reach a steady state. It is therefore recommended to either leave the noise switched on during the whole test, or for each trial to introduce a delay between the onset of the noise and the onset of the speech signal of some seconds (Nilsson et al., 2005).

Furthermore, acclimatization of subjects to the algorithm may also be an issue. Dahlquist et al. (2005) discuss this topic and suggest that valid evaluation results can be obtained even in laboratory studies that allow little time for acclimatization.

3.2 Multi-Channel Approaches

3.2.1 Configuration for Speech Perception

When talking of multi-channel approaches, it means here hearing aids with more than one microphone, allowing some kind of directional processing. The hearing aids may be unilaterally or bilaterally fitted. With the algorithms of WP5, both bilateral and binaural algorithms are evaluated. Bilateral means that processing is separate for each hearing aid, that is, no exchange of information is made between the two hearing aids. Binaural however means that there is some interaction across the two devices.

To assess the benefit of multi-microphone approaches in terms of speech perception, the intelligibility level difference (ILD) is normally measured.

Typically, a speech source is located in front, and a masker source in one condition in front, and in another condition at one side (90° or 270°). The difference in the SRT for the two conditions is the ILD (Johansson and Arlinger, 2002; D-2-1, 2005). Another possibility is to compare two hearing instrument settings under the same condition, i.e. speech from the front and noise from one side (Wouters et al., 1999). Furthermore, in some situations, the speech source may not be localized in front. For testing the situation “in a car”, the speech comes typically from 90° or 270°. A number of studies which investigated the benefit of bilateral versus unilateral hearing instruments or cochlear implants have used two loudspeakers at +/- 45° positions for speech and noise signals respectively (Laszig et al., 2004).

With the ILD, the effects of the head shadow and binaural processing are assessed together. If a separation of these effects is desired, the binaural intelligibility level difference (BILD) is measured (Johansson and Arlinger, 2002; D-2-1, 2005). Typically, the SRT with speech from 0° and a masker from the side (90° or 270°) is measured once with both ears, and once with the ear on the masker side occluded. The BILD is the difference in the SRT for the two conditions.

However, the spatial configuration of the masker can also be set up more realistically. With omnidirectional microphones, the above settings work well, but with directional microphones, they yield results that are not applicable in real-world conditions. Directional microphones in hearing systems can for example have a static cardioid character. In this case, a loudspeaker radiating noise from 90° may point directly at the rejection null of the bidirectional pattern of the microphone. In the real world, some of the competing noise would likely come from the rear, and it would not be attenuated by the microphone’s null. Because the noise level in this measurement setup would be attenuated more than in the real world, the perceived improvement in SNR would be better in the test system than in the real world, yielding misleading results. Alternately, a rear loudspeaker radiating noise may point at a lobe of the bidirectional pattern. In this case, the SNR perceived by the listener in the real-world would be better than in the test setup. In real-world environments, some of the competing noise would likely come from the sides.

Thus, a multiple-loudspeaker array is needed to reveal accurate results in testing the real-world performance of hearing systems having a directional pickup pattern. Luts et al. (2004), for example, measured the performance of an assistive microphone array for hearing instruments, and used a setup with uncorrelated noise sources at 90°, 180°, and 270°. The performance for this setup was clearly worse than for a setup with only one noise source at 90°. Ricketts (2000a) and Ricketts (2000b) showed that data collected with a single noise source can not predict real-life performance of directional microphones and used therefore a setup with five loudspeakers with uncorrelated noise. For adaptive directional microphones, one of the noise sources may even be moving during the test. Peissig and Kollmeier (1997) used an arrangement with fixed loudspeakers at 105° and 255°, and a source varying from 0° to 360°. Rickets and Henry (2002) showed that an adaptive directional microphone gives better performance compared to a non-adaptive when a competing noise was presented from the listener's sides in both a fixed noise configuration and a configuration with a single panning noise source.

Compton et al. (2004) used an array of eight loudspeakers with 45° angle (full circle) to simulate a noisy restaurant environment. They evaluated an omnidirectional versus a directional microphone on normal hearing subjects and showed that the performance of the simulated configuration was not significantly different from the performance of the corresponding real restaurant situation. The system used, called the R-SPACE, was developed for the purpose of accurately recording and reproducing/simulating real-world environments (Revit et al., 2002). Revit et al. claim that to evaluate the performance of a first-order directional microphone in the horizontal plane, an array of eight equally spaced loudspeakers does a good job under head-worn conditions that allow for minor head rotations during testing. It should be noted that the original R-SPACE system uses a radius of only 60 cm for the recording microphone array as well as the reproducing 8 loudspeakers whereas other authors recommend a radius of one to two meters for such a system.

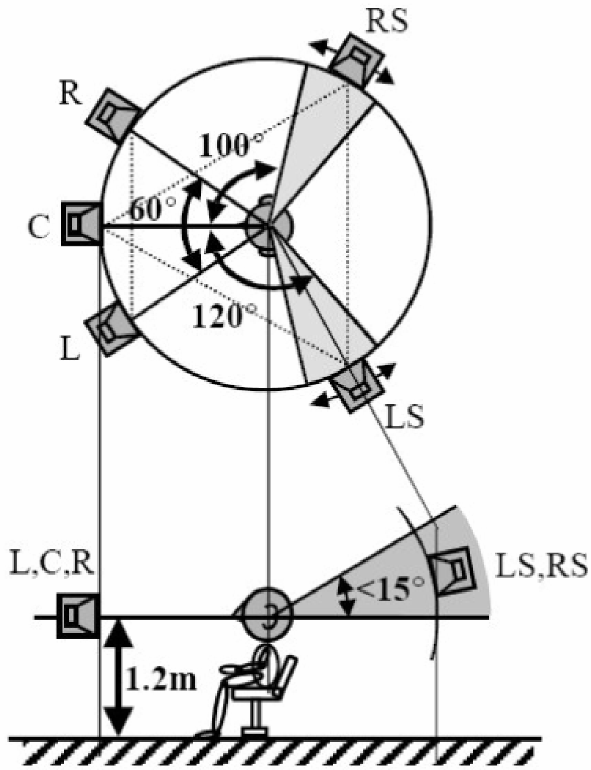

Figure 3-1: Loudspeaker setup recommended by ITU-R_BS.775-1 (1992)

On the other hand, while more complex sound fields may better represent the environments encountered in real-world situations, complex sound fields are very difficult to replicate in clinical settings. As a compromise, Nilsson et al. (2005) propose to use four loudspeakers for the masker (and one for the signal). They present a new standard test environment for the evaluation of modern hearing aid features, showing that the main factor for repeatability of the setup in different clinics is the room dimensions (see also section 3.9). They propose to standardize a test setup on a 5.1-channel surround-sound system according to Figure 3-1. Such a system could also be used to reproduce moving sound sources for the evaluation of adaptive directional microphones. Other configurations such as 6.1 and 7.1 have also been proposed (Dolby, 2006) and become more and more popular in consumer electronics and personal computers.

In a recent study, Kinkel et al. (2006) investigated the benefit of using multiple loudspeakers for hearing instrument adjustments and evaluation. They used 8-channel recordings which were obtained with a modified R-SPACE system which used 8 microphones arranged on a 2 m diameter. The 8-channel recordings were then either used directly for reproduction with an 8 loudspeaker setup or down mixed to 5.1-, 2-2- or 2-channel setups. Speech in noise tests as well as scaling tasks with questionnaires evaluating spaciousness and naturalness revealed that for practical purposes a 5.1 system may be suitable for most hearing instrument evaluation purposes. They also note that the reproducing arrangement of loudspeakers has to correspond as much as possible to the recording configuration which was used to generate the test material. For example, the 5.1 standard uses left and right loudspeakers which are placed at +/– 30° and left and right surround speakers at positions between +/– 100° to 120°. Thus, when 5 loudspeakers placed on a full circle with 45° interval (8 loudspeaker setup) are being used for playback, the perceived spaciousness is largely distorted. A 5 loudspeaker selection of a setup with 30° spacing (12 loudspeaker setup, selection for L, C, R, LS, and RS), on the other hand, would allow accurate reproduction of the standard surround signals.

Finally, in a draft guideline of the European Telecommunications Standards Institute entitled “Speech quality performance in the presence of background noise” (DEG/STQ-00038-1, 2005), it is proposed to use a setup with four loudspeakers (45°, 135°, 225°, 315°) and a subwoofer. A 5.1 surround system was also evaluated, but it was stated that due to the lack of easy to use recording techniques allowing a spatial recording of a sound field, they are not suitable for creation of a background noise database with realistic background noises and calibrated background noise simulation in a laboratory. While this statement may be true for real recordings, the widely used 5.1 and other surround systems are more and more supported by standard audio processing tools. For example, in the Adobe Audition audio processing software (formerly CoolEdit Pro), a normal stereo wave file can be converted into a 5.1 surround file. The user can choose any desired location in the sound field, and panning from one location to another over time is also possible. A similar software called vector base amplitude panning (working with an arbitrary number of loudspeakers) has been developed by Pulkki (2001). Furthermore, an extension of the popular MP3 compression format for the representation of multi-channel audio, including 5.1, has been presented recently (Herre et al., 2005).

3.2.2 Localization

It has been shown that localization performance of hearing aid users can be affected when using directional as opposed to omnidirectional microphones (Van den Bogaert et al., 2006). It may therefore be valuable to assess the influence on localization of some of the WP5 algorithms.

The localization performance is most important in the horizontal plane, and thus, localization tests normally are restricted to this plane, see (D-2-1, 2005). However, there are also instances where localization is evaluated in the third dimension.

The score assessed is either the mean or root-mean-square error in degrees, or the minimum audible angle (just noticeable difference in azimuth). The minimal achievable resolution differs a lot for different subjects groups: Normal hearing subjects can resolve 1-3°, hearing impaired subjects (bilaterally fitted) 4-7°, and monaural CI users only around 70°.

The achievable resolution depends also on the frequency of the signal. Thus, the sound material must be chosen based on this fact.

Localization tests will be further discussed and elaborated in deliverable D-7-3 (2006).

3.3 Speech Material

The HearCom report D-7-1 (2005) described the three types of speech tests to be considered for WP7: Sentence tests, CVC (consonant-vowel-consonant) tests, and number triplet tests.

There are good reasons why different test types are employed. For example, sentence tests are well suited to represent everyday communication, but training effects may occur that make it difficult to use the same test several times on the same subject. Some sentence tests may also be too difficult for CI users. CVC tests may not correspond to everyday communication, but they are quite common in clinical practice for diagnostic purposes.

The advantages and disadvantages of the different tests are further discussed in D-7-1.

3.4 Masker Types

There are three types of noise that are particularly interfering with speech intelligibility: (1) random noise with an intensity-frequency spectrum similar to that of speech, (2) a second interfering voice or multiple voices (speech babble), and (3) substantial room reverberation. Note that for the evaluation of the listening comfort, other noise types may be important (see section 3.7). The first two noise types are discussed together, followed by a section about reverberation.

3.4.1 Spectrotemporal Masker Properties

Drullman in (D-7-1, 2005) suggests to use speech-shaped noise for speech-in-noise tests rather than fluctuating noise or multitalker babble. This is supported by a publication of Drullman & Bronkhorst (2004), where it was shown that speech-shaped noise leads to a smaller variance in the SRT results than with for example one interfering talker. With the masker being only one or a few talkers, it is likely that the subject listens to the signal in the dips of the temporal envelope of the interference.

Furthermore, Festen & Plomp (1990) and Stickney et al. (2004) and various other authors showed that for hearing impaired and CI subjects, masking release is not improved when the interfering noise is modulated, compared to unmodulated noise. The low sensation level at which the impaired listeners receive the masker seems a major determinant. Reduced temporal and spectral resolutions in the impaired ear are other factors for the absence of unmasking.

In addition, a study by Wouters (1999) suggested that one speech-in-noise condition may yield enough information in the evaluation of directional microphones. Speech-weighted noise, restaurant and traffic noise were used as background noise, all leading to the same SRT improvement of a directional relative to an omnidirectional configuration.

Finally, Brungart & Simpson (2002) showed that distance-dependent changes (near the head) in the interaural difference cues are much more important for a speech masker than for a speech-shaped noise masker.

From this point of view, unmodulated speech-shaped noise seems best suited as masker in speech tests. Apart from that, Byrne et al. (1994) showed that the long-term average speech spectrums of many languages are similar. Thus, unmodulated speech-shaped noise can be language independent.

On the other hand, the studies cited above also showed that some of the hearing impaired subjects took advantage of masker modulation, while others with a similar hearing loss did not. Thus, the SRT in modulated noise could be a better indicator for the expected deterioration of individual hearing than the SRT in unmodulated noise. Furthermore, the type of noise can have significant influence on the performance of single-channel noise reduction algorithms or adaptive directional microphones. In single-channel approaches, the algorithm acts on the more-or-less steady-state elements of a signal. In contrast, for blind source separation (BSS) algorithms, usually the nonstationarity of the signals is exploited to improve separation performance. Therefore, BSS algorithms will yield better results for fluctuating noise. For the evaluation of a single-channel noise canceller, Dahlquist et al. (2005) used stationary random noise and speech babble, and the noise type had a significant influence on the performance. However, although the SRT differed between the noise types for both the unprocessed and the processed condition, the *relative* differences between the unprocessed and the processed conditions did not differ consistently for the two noises.

Another point of view is that in many real-world listening conditions, the level and the spectrum of interfering noise fluctuate from moment to moment. Therefore, tests of hearing in noise that use stationary noises lack a potentially important degree of face validity relative to real-world, fluctuating noise conditions. From this point of view, it seems reasonable to use the masker that occurs most in everyday life and that is most challenging for the listener (see section 2.2), that is, a masker that has similar temporal and spectral properties as the target signal, which is speech babble. A second characteristic of real-world environments that can affect perception involves the correlation of sources. For any symmetrical signal configuration surrounding a listener (i.e., no delay between signals reaching the listener’s position), if the signals from all sources are correlated (i.e., they have the same waveform), then the sound will be perceived as coming from the centre of the head. In contrast, when noise sources are uncorrelated, they appear distinct and separated.

Nilsson et al. (2005) try to account for both of these points (see also section 3.2.1). They provide two different types of uncorrelated noise: Four separate tracks of speech-shaped noise, delivered through four loudspeakers, and combinations of recordings of uncorrelated multiple-talker speech, with four, eight, twelve and sixteen individual talkers. The latter masking condition adds the variable of temporal modulation across different numbers of talkers. During evaluation of the material, it was however demonstrated that the influence of the room size is much bigger with multitalker babble than with uncorrelated random noise (see section 3.9), and it was claimed that the cost of using speech as an experimental masker is the loss of control associated with room dimensions. This would again suggest that unmodulated noise shall be used.

On the ICRA noise CD, unmodulated as well as modulated speech-shaped noises can be found (Dreschler et al., 2001). ICRA 1 is the unmodulated standard, ICRA 5 the modulated standard for a male talker.

Certain tests come with their own noise, such as the Oldenburg sentence test, whose background noise is a speech-simulating continuous noise that exactly matches the long-term spectrum of the sentence material, generated from 30 statistically controlled overlaps of the test words used (Wagener et al., 1999). This noise has also been used to create a diffuse noise field for the evaluation of noise cancellers in hearing aids (Stephan et al., 2006). Four uncorrelated versions of the noise were presented through four loudspeakers (45°, 135°, 225°, and 315°).

Finally, for the above mentioned ITU draft guideline (DEG/STQ-00038-1, 2005), a background noise database has been put together. Binaural recordings of 30 s length include the following situations: Inside a car, a bus, a train and an airplane; traffic noise; public places, such as pub, shopping centre, sports, station, cafeteria; workplace, such as office, jackhammer, machine shop; living room; children in a kindergarten and a schoolyard.

3.4.2 Reverberation

It has been pointed out above that reverberation can have a strong effect on speech intelligibility. This is especially the case in the presence of background noise. In experiments with CI simulations, for example, speech perception in noise deteriorated rapidly with increasing reverberation. In quiet, only the simulation below 12 vocoder channels was affected (Poissant et al., 2006).

It seems therefore important to apply different amounts of reverberation to the signal or both signal and masker when evaluating the hearing aid algorithms. Reverberation may be introduced into the signals by convolving them with room impulse functions or using multimicrophone recordings in a real reverberant room, or in a controlled simulation setup such as the Communications Acoustic Simulator (Kommunikations-Akustik-Simulator KAS, HoerTech Oldenburg).

If a room simulation software is used, it is important that the relative delays between the microphones are correct. Otherwise, it may lead to falsified results when multichannel hearing aid algorithms are used.

A system for computing simulations for multi-channel loudspeaker setups from measured room impulse responses has been presented by Merimaa & Pulkki (2005) and Pulkki & Merimaa (2006). The Spatial Impulse

Response Rendering (SIRR) system works exclusively with Waves IR-360 Discrete Surround plug-in and has been successfully tested for 5.0 and 7.0 loudspeaker setups.

Another commercially available tool is the FIReverb Suite from CATT-Acoustic, see also Dalenbäck (2002). With the PureVerb tool, natural-sounding finite impulse responses can be created based on parameters from many real halls. Then, reverberation can be added to a signal with the 8x8 channel MultiVolver convolver. The layout of the up to 8 speakers can be chosen arbitrarily.

A high-level software called ODEON (Rindel, 2000) can simulate the interior acoustics of buildings where, from the geometry and properties of surfaces, acoustics can be calculated, illustrated and listened to. It includes auralisation, binaural as well as surround.

3.5 Threshold Estimation Procedures

In most of the state of the art evaluations discussed above, the outcome was the SRT for specific test conditions (e.g., speech from the front, and noise from the side). One of the standard procedures for assessing the SRT is an adaptive test, where typically the noise level is kept constant, and the signal level is varied stepwise up and down, until a point is reached where 50 % of the answers are correct (Levitt, 1971; Müller, 1992; Wagener et al., 1999).

Such an adaptive up-down procedure is simple, robust, and efficient. The point of interest (the SNR where 50 % of the responses are correct) is directly measured. Another approach is to assess the scores at different fixed SNRs, and to fit a psychometric curve to this points. This has however the disadvantage that a large proportion of the observations are placed at some distance from the region of interest. Also, a preliminary experiment is often necessary in order to get some idea of the required range. With an *adaptive* procedure, the relevant range will automatically be found, and possible gradual drifts during the test can be compensated.

The adaptive procedure can however be quite time consuming. The time required for a complete speech reception threshold measurement by, for example, Müller (1992), is 20 minutes. A different procedure has been presented for example by Peissig and Kollmeier (1997). The subject adjusted the level of a test sentence himself, to a value which subjectively corresponded to 50 % intelligibility of the sentence. The advantage of this subjective assessment was that the subject required on the average about 7 s to adjust the signal-to-noise ratio in a given situation. Even if this measurement was repeated four to eight times, the reduction in measurement time was considerable. Another advantage was that the subject responded directly to the computer via keyboard (which may these days be a touchscreen). It should be noted, however, that the subjectively assessed thresholds were highly dependent on the subject's criterion. Thus, comparatively high interindividual standard deviations were observed. However, the relative differences between two assessed thresholds for different test conditions showed interindividual differences of below 1 dB. These deviations were in the same order of magnitude as the variability in the SRT data reported by for example Plomp and Mimpen (1981). Furthermore, the subjective intelligibility assessment methods appear to "compress" the maximum observable effect of varying an independent variable (e.g., azimuth of interfering sound source) on the observed speech intelligibility.

Hence, if a quick procedure is desired and one is willing to accept the described drawbacks, subjective assessment could be a valuable alternative to automated adaptive tests.

3.6 Testing De-reverberation

A straightforward approach to evaluate a de-reverberation algorithm is to test speech intelligibility in rooms with different characteristics, as has been discussed in the previous section.

In a recent study (Gabriel, 2005), two rooms were simulated in a room using virtual acoustics: a non-reverberant living room and a large reverberant room. The speech material (CVC) was presented in quiet at 55 dB SPL. In both rooms, the speech test was conducted in two hearing programs, normal and EchoBlock.

If a more demanding situation is desired, some background noise could be introduced that is reverberated in the same way as the signal.

Note however that the algorithms of the category "coherence based" can filter out any incoherent signal components and thus work as a binaural noise filter in general, as for example in diffuse noise conditions. This category shall therefore not only be tested as proposed by Gabriel, but also in diffuse noise (see section 4.4).

3.7 Testing Feedback Suppression

At present, no method is known to assess the benefit of feedback suppression in terms of speech perception through tests with experimental subjects.

In modern hearing instruments, the feedback threshold is measured in situ during the fitting, and the gain is adapted based on this measurement to prevent feedback in everyday life. A feedback canceller is an additional tool that comes only dynamically into play in special situations, for example when feedback is provoked through sound reflections from an object that is near the ear (hand, wall, telephone handset).

One possibility to measure the impact of a feedback canceller could be to assess speech intelligibility in a situation where feedback occurs, with and without feedback-reduction. In addition, the subjective comfort of the algorithm for hearing aid users could be assessed.

3.8 Subjective Listening Comfort Tests

An increased SNR and improved speech intelligibility are the main goal of signal enhancement algorithms. However, it is also very important that such algorithms preserve or even improve the sound quality. Consequently, the subjective listening comfort has to be evaluated in addition to objective speech perception, and background noises other than speech-like noises may be important.

A method to assess the perceived sound quality for a noise canceller compared to the unprocessed signal was for example developed in the EU project LISCOM (Dahlquist et al., 2005). Four types of noises were tested: Stationary random noise, speech babble, street noise, and cafeteria noise. In another study (Ricketts and Hornsby, 2005), the sound quality of processed versus unprocessed was assessed through paired comparison tests.

Listening comfort tests are not a major concern in this report. The HearCom report D-7-4 (2007) will provide an overview of this topic and suitable tests will be developed and presented.

3.9 Test Room Requirements and Layout

The physical dimensions of the test room, the acoustical characteristics of the walls, floor and ceiling as well as the layout of the sound equipment may all have an influence on reproduced audio and inevitably on the test results.

The wall to wall (and floor to ceiling) distances should differ, and simple ratios of these distances should be avoided, to distribute any room resonances across frequency.

To minimize reverberation, the room should be equipped with sound attenuating material on the floor, walls and ceiling. Ideally, absorption coefficients should be in the same range as for audiometric test booths. In order to reduce the influence of external noise, the noise floor in the room should be less than 30 dB SPL (A).

Regarding the loudspeaker setup, a number of recommendations can for example be found in (Genelec, 2003). The loudspeaker setup should be symmetrical, so that reflections in the time domain are identical from both left and right half of the room. The loudspeakers should be placed at least 1.1 m away from the wall behind, so that the cancellation frequency goes down below the usual 85 Hz cut-off of each main loudspeaker. It is important to place the loudspeakers precisely, as deviations of only few cm forwards or backwards may already give noticeable differences in arrival time at lower frequencies.

If a computer screen is placed in front of the subject, it should be either acoustically transparent or placed low enough so that the front loudspeaker is above the screen (not below). Generally, any reflecting surfaces between the loudspeakers and the listening position should be removed or minimized, to prevent early reflections that can smear the coherence of spatial information.

As mentioned above, Nilsson et al. (2005) propose a new standardized setup. An initial setup was evaluated at four different sites, each equipped with an audiometric test room large enough to accommodate the five-loudspeaker setup (a minimum inside dimension of 2 m by 2 m). It was shown that the use of multitalker-babble as masker yields scores that are related to the length and width of the room. When using a diffuse uncorrelated speech-shaped noise field, the horizontal dimensions could be well controlled. However, the results were then affected by the height of the room (1.98 m at three sites, and 2.44 m at one site). Nilsson et al. suggest to introduce an additional uncorrelated noise source in the vertical dimension in order to eliminate any interaction between SRT and room size.

3.10 First Conclusions

Based on the considerations in this chapter, some first conclusions can be drawn:

1) A single-channel noise reduction algorithm can be tested following the procedure described by Dahlquist (2005). Unmodulated as well as modulated noises should be used whenever possible.

2) For the evaluation of multi-channel algorithms, a configuration with more than two loudspeakers seems mandatory. A 5.1 surround-system could be sufficient for this task, unless additional speakers would be needed for the production of a diffuse sound field in the third dimension.

3) It seems important to test the algorithms under realistic conditions. Furthermore, the multi-channel algorithms require modulated maskers to work correctly.

4) It is important to evaluate the algorithms when the signal (and optionally a masker) is reverberated. De-reverberation algorithms can be tested following Gabriel (2005).

5) In an optimal configuration, all test rooms would have the same dimensions. As this is hardly feasible, similar reverberation times and room heights are desirable. If the latter can not be achieved, either additional noise sources in the third dimension will be necessary, or the absolute test results will not be directly comparable from site to site (especially when using modulated maskers as proposed under point 3).

6) Further evaluations will be needed to verify test-retest reproducibility, intra-site as well as inter-site.

In the next section, a standardized setup is proposed that shall further be evaluated.

4 Proposed Setup

4.1 Introduction

For the purpose of evaluating the signal processing algorithms developed in WP5 with experimental subjects, three sites (University Leuven, Hörzentrum Oldenburg, University Hospital Zurich) have been selected to perform controlled free field test trials in flexible and multiple configurations ranging from one to twelve loudspeaker setups with simultaneous playback of a speech stimuli and a wide variety of interfering signals. The setup proposed in this chapter will further be evaluated at these three sites.

Moreover, the proposed setup may be used in the future for evaluations at other sites. There are two reasons why direct absolute comparison of the results from site to site will be rather difficult: Different room characteristics (see section 4.5), and different languages for the tests (see D-7-1, 2005). Evaluations made in WP3 (D-3-3, 2006) have also shown that direct absolute comparisons cannot be made. It should however be ensured that *relative* comparisons remain possible.

Comparisons of the results should also be possible to a certain extent between the different algorithms. Thus, the test conditions shall be the same for all algorithms where this makes sense.

4.2 Sound Sources

Existing standards shall be used whenever possible for the sound system and configuration of the loudspeakers. It is proposed to further evaluate the use of a standard 5.1 or 7.1 surround system (Nilsson et al., 2005; Dolby, 2006). Especially for the widely used 5.1 system, a variety of signals and signal processing software is available. Furthermore, the costs of hard- and software are moderate.

Another advantage of the 5.1 setup is that a subset of the speakers needed for a localization test may be used, for example of a full circle with 12 speakers that are 30° apart.

4.3 Preparation of Sound Signals

4.3.1 Signal Processing Software

With the widely used Adobe Audition software (former CoolEdit Pro), a stereo wave file can be converted to a 5.1 surround-sound field, by choosing any desired location in the field, and panning from one location to the other over time is also possible.

If a different loudspeaker setup than 5.1 should be required, the Vector Base Amplitude Panning software from Pulkki (2001) can produce virtual sound sources with arbitrary loudspeaker setups.

Hamasaki et al. (2005) state that with a 5.1 setup, simple amplitude panning over five channels can sometimes not provide smooth rotation of a sound image around the listener. They propose a (prototype of a) system called Integrated Surround Sound Panning System (ISSPS), where additional three virtual loudspeakers are introduced before panning, two at the side and one at the back. It is claimed that with ISSPS, “lateralization” is much better feasible than with the conventional 5.1 setup.

Finally, sound data stored in a format equal or higher to CD quality may need considerable amount of disc space. If a distribution over the internet is desired, a different format may be required. Herre et al. (2005) have introduced an MP3 standard for 5.1 surround sound, which reduces the size of sound files considerably and makes efficient distribution and reproduction possible. However, as the MP3 format tries to compress based on perceptual criteria, this may pose a problem to multichannel signal processing algorithms, and an uncompressed format may be preferred.

4.3.2 Speech Signals

Speech material as described in D-7-1 (2005) will be used, presented in quiet and in noise.

4.3.3 Maskers

Unmodulated as well as modulated speech-shaped maskers (ICRA noises or test-specific noises such as the OLSA-noise) are available, coming either from one source, or uncorrelated from several loudspeakers, producing a diffuse sound field. For the OLSA-noise, Stephan et al. (2006) have recently gained experience with producing a diffuse field.

In addition, special real-life conditions may be considered and are available, as for example conversation in restaurant, background music from several directions, street noise coming from another direction, interfering voices and kitchen noises from other directions.

The exact specifications of the required masking noises depend also on the algorithms under test and will be discussed in section 4.4.

4.3.4 Reverberation

Reverberation may be introduced into the signals by convolving them with room impulse functions using the SIRR system (Merimaa and Pulkki, 2005; Pulkki and Merimaa, 2006), or a different tool, such as the FIReverb Suite or ODEON. Another approach is to use multimicrophone recordings in a simulation setup, such as the Communications Acoustic Simulator (KAS, HoerTech Oldenburg): Dry recordings of speech and other sounds can be used to generate sound environments with different reverberation and room characteristics. The sound signals are then recorded using a modified R-Space recording system with for example 8 or 12 microphones arranged on a circle of 2 to 3 m diameter. These recordings are then played using 8 or 12 loudspeaker setups or mixed-down versions for a 5.1 setup. Reverberation times ranging from less than half a second (sound treated room) to more than 4 seconds (church, train station) can be achieved in real time in the KAS.

In any case, it seems mandatory to further verify that a simulated reverberant room using a 5.1 setup yields similar results than a real reverberant room.

4.4 Proposed Test Conditions

4.4.1 Overview

In WP5 (D-5-1, 2005; D-5-2, 2006), the algorithms have been technically evaluated under a variety of conditions, such as different azimuth of signal and masker, different maskers, different room characteristics, etc. For the evaluation with hearing impaired subjects, one has to stick to the most important and most realistic situations on the one hand, and on the other hand to those situations for which a benefit can be expected at all with a specific algorithm. Otherwise, the number of required test sessions would very quickly become enormously high with about nine algorithms under test. At the time of writing, the evaluations in WP5 were still progressing, including pilot tests with hearing impaired subjects. Thus, the final test conditions may differ from the ones proposed here.

In Table 1, an overview is given over the proposed conditions for the different algorithms, except for the feedback suppression algorithm, which will require a different approach. Conditions for a fixed directional microphone are shown to complete the list, although such a setting is not part of the WP 5 algorithms.

The desired signal is always speech coming from the front, except in the car speech coming from the side. Where it makes sense, the same conditions shall be used for different algorithms. Because reverberation is one of the most important factors that influence the performance of multichannel algorithms, some reverberation shall be added to the point-sources, to simulate an everyday living-room. The specifications of an IEC Listening Room (IEC-60268-13, 1998) may serve as a basis for this simulation. In the diffuse conditions, reverberation according to the room is also added to the signal (not to the masker), i.e. cafeteria and car (traffic is in the free field). The evaluations in WP5 have shown so far that for the BSS and adaptive beamformer, most benefit can be expected in the point-source condition. Single-channel and coherence-based algorithms on the other hand be more effective in diffuse noise fields.

If a 5.1 setup is used, the point-source masker signal could also come from 120°, or 120° and 240° respectively. This could make the soundfield more realistic, since azimuths of 90° or 180° will only be virtually represented (see also pilot test in section 4.6). From a point of view of creating a soundfield as realistic as possible, there does not seem to be much difference between the two possibilities.

Test Condition				**Algorithm Type**				
Class	Signal Azim.	Masker Type	Masker Azim.	Single Ch.	BSS	Fixed BF	Adapt. BF	Coher. based
Point-source (living-room)	0°	Speech-shaped	90°	o[1]	o	o	o	o[1]
	0°	Modul. speech-shaped	90°	–	+	+	+	o[1]
	0°	Music	90°	o[1]	+	+	+	–
	0°	Modul. speech-shaped + Modul. speech-shaped	90° 180°	–	–	o	o	–
	0°	Modul. speech-shaped + Music	90° 180°	–	–	o	o	–
Diffuse	0°	Cafeteria	all	+	+[1]	+	+[1]	+
	0°	Traffic	all	+	o[1]	o	o[1]	+
	90°	Car	all	+	o[1]	o	o[1]	+
Combined	0°	Cafeteria + Modu. Speech-shaped	all 90°	–	+	+	+	–
	0°	Cafeteria + Music	all 90°	–	o	o	o	–
	0°	Cafeteria + Modu. Speech-shaped + Music	all 90° 180°	–	–	o	o	–
Strong Reverberation	0°	Low reverberation (as in sound booth)	–	o[2]	o[2]	o[2]	o[2]	o[2]
	0°	Medium rev. room	all	–	–	–	–	o
	0°	Big rev. hall	all	–	–	–	–	+

[1]Only minor improvement expected at time of writing
[2]May reveal processing artefacts that reduce intelligibility compared to unprocessed
+ Must
o Optional
– Not tested

Table 1: Proposed test conditions for the different algorithms. In the point-source condition, reverberation of a living-room is simulated. In the other conditions, the signals are reverberated as indicated in the masker type field (Cafeteria, Car, Hall).

4.4.2 Threshold Estimation Procedure

It has been pointed out in section 3.5 that adaptive procedures after Levitt (1971) are better suited than measurements with fixed stimuli. However, with several algorithms to be tested under various conditions, the time needed may be extraordinary. The approach of Peissig and Kollmeier (1997) may be chosen instead if a faster procedure is desired, and if the drawbacks are acceptable.

4.4.3 Single-Channel Approaches

For the evaluation of single-channel noise reduction systems, a method following Dahlquist et al. (2005) is suggested. The most relevant maskers are cafeteria noise with dish-rattling and footstep sounds, traffic noise, or in-car noise.

The same loudspeaker and room setup can be used as for the multi-channel approaches. In theory, only one of the loudspeakers would be needed for testing, but if the results shall be directly comparable to the multimicrophone approaches, it may make sense to use exactly the same setup.

The recommendations of Nilsson et al. (2005) regarding the adaptation time of the algorithm are also important, and thus, the masking noise shall either be switched on during the whole test, or at least five seconds before each trial.

4.4.4 Blind Source Separation and Beamformer

The evaluation of these multi-channel approaches requires a setup with several sound sources and diffuse sound fields, as described above and summarized in Table 1. One of the major upcoming tasks in WP7 is to elaborate these setups and procedures further. This needs to be done in close collaboration with the WP5 partners, since the expected performance of the algorithms may also have some influence on the setup, and since pilot tests are also planned in WP5.

4.4.5 De-Reverberation

It is proposed to follow the procedure proposed by Gabriel (2005) for a first setup. CVC tests shall be carried out in quiet, in two different virtual rooms. That is, the speech signal is convolved with different room impulse responses, such as medium reverberated room, and church-like hall. The reference may be a test with no reverberation at all, or the results with the de-reverberation algorithm switched off. The virtually reverberated signals can be reproduced through the same loudspeaker setup as is used for the other algorithms (e.g. 5.1 configuration).

In addition, a point-source with modulated speech-shaped noise and a diffuse cafeteria noise may be tested, if a benefit is expected in these situations.

4.4.6 Feedback Suppression

The howling or whistling of feedback will probably have an effect on speech intelligibility and speech quality. This could be measured by adapting the gain in the feedback algorithm, to measure the maximal level or gain at which 50 % of the speech is understood or oscillations are observed.

It has to be studied further how the gain can be adapted in the real-time implementation of WP5. In any case, experiments on an artificial head are preferred over direct tests with subjects because of safety issues. Measurements will then be carried out via recordings and headphones.

4.4.7 User Response

Whenever possible, the user response should be made directly into the PC via mouse or touchscreen. This is feasible for CVC and number triplets. For Plomp and OLSA type of testing, the University of Leuven is developing an interface that allows unsupervised testing. It incorporates simple self-correction features and automatic scoring.

4.5 Test Room and Loudspeaker Layout

It will hardly be feasible to reproduce the same room characteristics at different test sites. Of course, a number of basic considerations as described in section 3.9 must be followed. However, the test results will probably not be reproducible from site to site, and thus not be directly *absolutely* comparable, but only *relatively*.

Acoustic considerations for positioning the loudspeakers in real rooms with limited dimensions have been described in section 3.9 (from Genelec, 2003) and are not repeated here.

For the speaker layout of a 5.1 setup after ITU-R_BS.775-1 (1992, Figure 3-1), a minimum radius of 2 m is proposed in ITU-R_BS.1116-1 (1994). This may be difficult to achieve. However, this standard was defined based on the assumption that a larger area around the centre of about 2x1.4 m needs to receive a similar sound field, as shown in Figure 4-1. With only one subject sitting exactly in the centre, the radius may be reduced considerably.

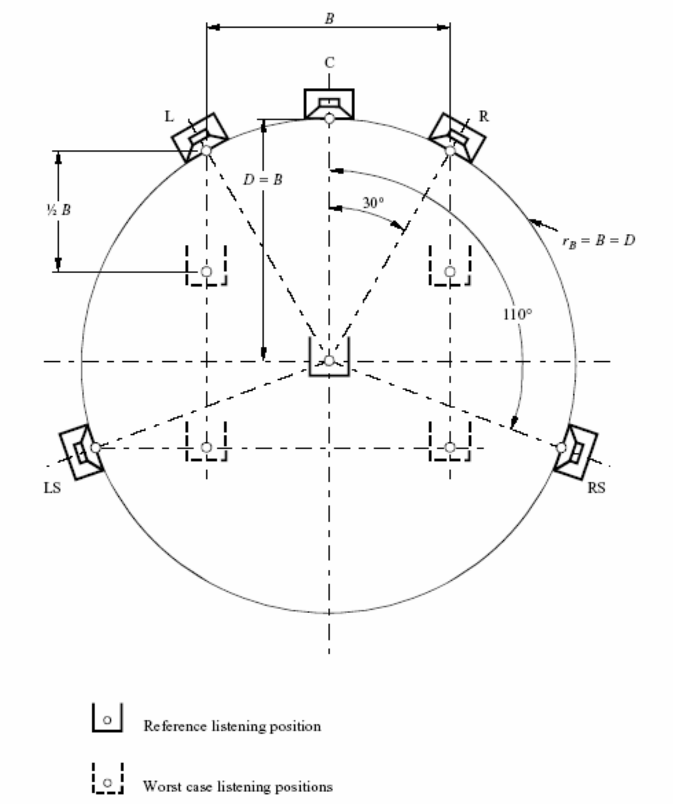

Figure 4-1: Listening arrangement as recommended by ITU-R_BS.1116-1 (1994)

On the other hand, it is proposed by Zwicker and Zollner (1987) to stay below the critical distance (the distance from the sound source at which the intensities of the direct sound field and the indirect field are equal). The critical distance depends on the directivity factor of the sound source, the absorption coefficients of the floor, walls, ceiling, and objects in the room (or the reverberation time), and the room dimensions. A radius of 1 to 1.6 m is typically well suited to stay below the critical distance.

4.6 Pilot Experiment: Reverberation & Surround

4.6.1 Introduction

In section 4.4.1, it was proposed to simulate a slightly reverberant room for the point-source condition. It was however not clear to what extent speech perception is affected when such a room is simulated, compared to a quasi anechoic condition. Beutelmann & Brand (2006) have investigated this aspect already for an anechoic room, an office room, and a cafeteria, Results have been described in the HearCom deliverable D-3-3 (2006). Speech was presented from the front, and noise from various angles. It was shown that the SRT is significantly worse in reverberant conditions, when the noise was presented from the side. Tests were carried out with headphones.

For the WP7 evaluations of hearing instrument algorithms, free field tests are required and sound will have to be presented via loudspeakers in a test room with specific acoustic characteristics. It can be expected that the influence of the test room characteristics may lead to different results for free-field tests in comparison with headphone tests. Therefore, a pilot experiment with normal hearing subjects in normal and reverberant conditions was performed. The experiment was also done in order to explore the technical details involved when simulating reverberation and playing the signal over a surround setup and to demonstrate that this kind of test setup was feasible and useful for the purpose of WP7.

4.6.2 Method

The adaptive Oldenburg sentence test in noise was carried out to assess the 50 % SRT in two conditions:

1. Quasi anechoic condition, Signal from the front (0°), Noise from the side (90°)

2. Slightly reverberant condition, Signal from the front (0°), Noise from the side (90°)

In the quasi anechoic condition, the signal was presented from the front speaker only, and the noise from a speaker at 90°. The distance to the listener was 1.60 m. The test room was not perfectly anechoic but had a rather low reverberation time constant T_{60} of less than 0.2 seconds (ambient noise level less than 25 dB SPL A).

In the reverberant condition, both the signal and the noise were preprocessed with the FIReverb Suite, i.e. convolved with a room response and mapped to a 5.0 surround setup. The simulated room characteristics chosen had specifications similar to an IEC Listening Room (IEC-60268-13, 1998):

- Length x Width x Height = 6.7 x 4.2 x 2.8 m
- T_{60}: 500 ms (125 Hz), 400 ms (≥ 250 Hz)
- Listener position: Centre of room
- Signal source: 0°, 1 m distance to listener
- Noise source: 90°, 1 m distance to listener
- Surround setup: 5.0 as shown in Figure 3-1, but with a radius of 1.6 m (rear speakers in the horizontal plane at 120°/240°).

The masking noise was the speech-shaped Oldenburg noise, presented at 65 dB SPL A (measured at the listener's position). The level of the speech signal was adjusted adaptively in 1 dB steps according to the subjects verbal responses which were scored by the experimenter as number of words correctly recognized. The control software for this experiment (Lai and Dillier, 2002) was not able to vary the sound levels for all five loudspeakers simultaneously. Therefore, manual adaptive level control had to be performed for the reverberant condition.

Seven normal hearing subjects from 29 to 59 years were tested. Conditions 1 and 2 were used in randomized order in the experiment. A training list of 10 sentences was presented before testing each condition. The test list contained 30 sentences for each condition. A one down one up procedure was used (Levitt, 1971) that converged well before the end of the list.

4.6.3 Results & Discussion

The results are summarized in Figure 4-2. For most subjects, there was a significant difference between the SRTs measured in the two conditions. The average difference in SNR was about 2 dB.

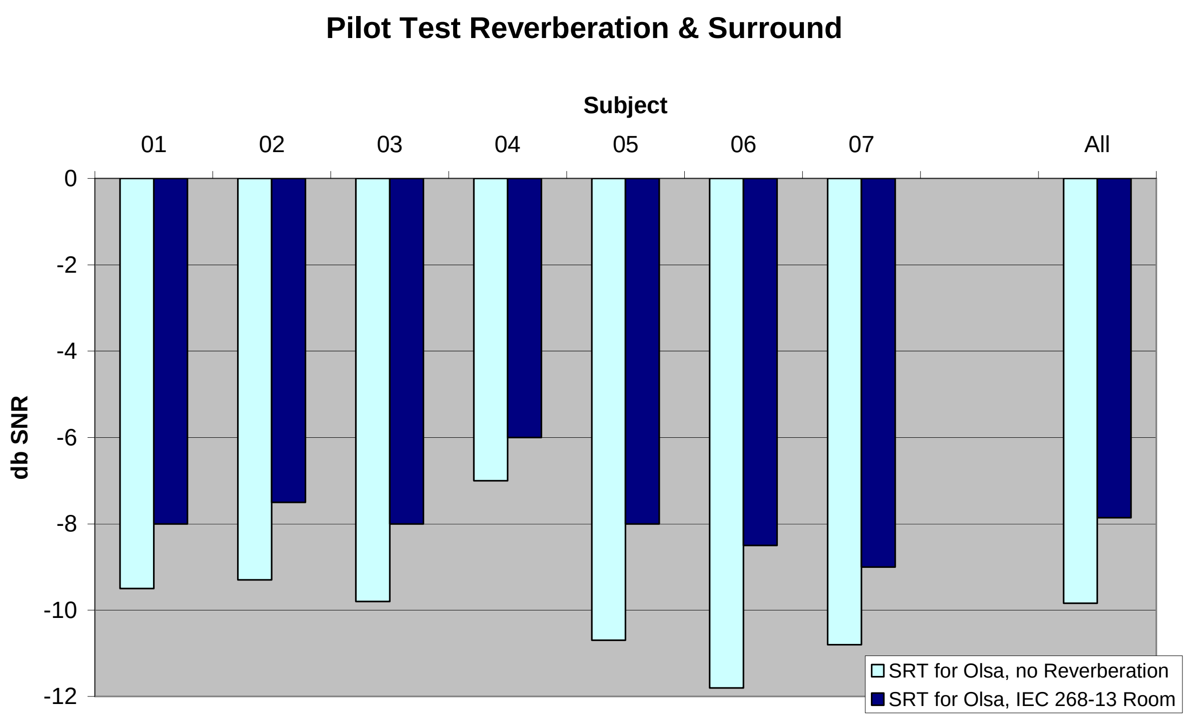

Figure 4-2: Pilot test results for seven normal hearing subjects. The average difference for the two conditions of the SRT was about 2 dB.

The results are consistent with what was expected. Even a small amount of room reverberation can have a significant impact on speech intelligibility in difficult situations (i.e., speech in noise). It is assumed that for hearing impaired subjects, the impact can be even more pronounced. As reverberation times of 0.5 seconds (and more) are quite common in real life environments it seems obvious that the effect of reverberation has to be considered for the evaluation of hearing instrument algorithms in a realistic environment.

The values measured in the quasi anechoic condition differ from what was measured by Beutelmann & Brand (2006). For a true anechoic condition (listening with headphones), they measured an SRT of -17 to -19 dB at a noise angle of 90° (see also Figure 3-2 in D-3-3). This is better than the -10 dB measured in this pilot experiment. Whether the differences between the Beutelmann & Brand and our results can be explained by the differences in presentation mode (headphone versus loudspeaker setup) or by differences in the selection of experimental subjects or other experimental conditions remains to be investigated. Possible variables which may influence the results are the age distribution of the subjects and their familiarity with the test language. In another experiment using cochlear implant subjects with and without FM systems with the same loudspeaker test setup, it was found that 50 % SRT values of -18 dB were obtained for the FM condition in some subjects. In those experiments, the loudspeaker at 90° was placed at only 1 m distance from the ear of the listener and the front loudspeaker at 2 m. Thus, the exact locations and distances of loudspeakers are important factors and should always be specified in the description of an experimental setup. To investigate differences between sound reproduction setups a control experiment could be carried out using the same group of subjects being exposed to both headphone and loudspeaker presentation modes.

Regarding the virtual reverberation and the surround setup, subjects found the simulation quite natural. Only the directivity of the noise source was reported not to be perfect. Some subjects found that the position of the noise source was a bit more in the direction of the rear right speaker of the 5.0 setup (120°), rather than at 90°. This agrees somehow with what Hamasaki et al. (2005) have already claimed (see section 4.3.1). A setup with more speakers (7.0 or 7.1) might be advantageous. Currently, most sound processing software available at present is restricted to a 5.1 setup only. Thus, a 5 speaker setup is a straightforward solution for today to keep things simple and make the equipment readily available for a broad range of potential users. For future evaluations, however, a more sophisticated setup might be preferred.

4.6.4 Conclusion

The results from the pilot experiment indicate that it is very desirable to add some reverberation also to the test conditions with pure point-sources, as proposed in section 4.4.

The virtual room reverberation software (FIReverb) has proven to be an effective tool for this task.

Free field tests results will always be influenced by the characteristics of the test room and the relative position of the loudspeakers, even without virtual reverberation. Thus, variations in test results between different test laboratories may be due to variations in acoustic room characteristics.

5 Dissemination and Exploitation

The proposed test procedures and setup will be used in the evaluation phase for the algorithms developed in WP5 with normal hearing and hearing impaired subjects.

Results will be presented at conferences and in peer reviewed journal publications.

A sample sound database will be made available on the HearCom portal with test material processed to simulate different background conditions or room characteristics with varying amounts of reverberation.

6 Conclusions

The test conditions and setup options which were described in this report cover a whole range of possibilities for the evaluation of signal processing and signal enhancement algorithms in existing and future hearing system algorithms.

The choice of the appropriate test setup for a given evaluation task depends somewhat on the details of the algorithm under test and its intended application. Simulated virtual acoustical environments, moving and spatially distributed sound sources can be generated and used for testing purposes in controlled situations approaching real life conditions.

Standardization of these test setups with consumer electronics components and readily available software seems to be a viable option for small and medium size audiological centres. However, the level of accuracy of simulated sound fields and the desired test-retest reproducibility with such systems will have to be chosen and judged depending on the processing options and specific tasks which are to be evaluated.

7 References

Beutelmann, R., and Brand, T. (**2006**). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J Acoust Soc Am **in press**.

Brungart, D. S., and Simpson, B. D. (**2002**). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," J. Acoust. Soc. Am. **112**(2), 664-676.

Burkhard, M. D., and Sachs, R. M. (**1972**). "KEMAR the Knowles Electronics Manikin for Acoustic Research, Report No. 20032-1," (Industrial Research Products, Inc., Elk Village, Illinois).

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Heto, R., Kei, J., Lui, C., Kiessling, J., Kotby, M. N., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westermann, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**(4), 2108-2120.

Compton-Conley, C. L., Neuman, A. C., Killion, M. C., and Levitt, H. (**2004**). "Performance of directional microphones for hearing aids: real-world versus simulation," Journal of the American Academy of Audiology **15**(6), 440-455.

D-1-2 (**2005**). "Report on the proposed set of communication performance tests," (HearCom Deliverable).

D-2-1 (**2005**). "Implementation of a preliminary test set for auditory impairments in one language," (HearCom Deliverable).

D-2-1b (**2006**). "Demo version of the preliminary test set for auditory impairments," (HearCom Deliverable).

D-3-3 (**2006**). "Evaluation and comparison of speech intelligibility models for normal hearing listeners," (HearCom Deliverable).

D-5-1 (**2005**). "Sub-set of signal enhancement techniques operational on PC system," (HearCom Deliverable).

D-5-2 (**2006**). "Report on the effectiveness of signal-enhancement algorithms for various auditory-profile groups," (HearCom Deliverable).

D-7-1 (**2005**). "Speech reception tests for different languages," (HearCom Deliverable).

D-7-3 (**2006**). "Specification for spatial hearing," (HearCom Deliverable).

D-7-4 (**2007**). "Test for listening effort," (HearCom Deliverable).

D-9-1 (**2005**). "User requirements specification for assistive applications on a common platform," (HearCom Deliverable).

Dahlquist, M., Lutman, M. E., Wood, S., and Leijon, A. (**2005**). "Methodology for quantifying perceptual effects from noise suppression systems," Int. J. Audiol. **44**(12), 721-732.

Dalenbäck, B.-I. (**2002**). "The Rediscovery of Diffuse Reflection in Room Acoustics Prediction," in *ASA* (Cancoun).

DEG/STQ-00038-1 (**2005**). "Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database," in *Draft ETSI Guide* (European Telecommunications Standards Institute, Technical Committee STQ), pp. 1-46.

Dolby (**2006**). "Home Entertainment: Room Layout and Speaker Setup," in *Dolby laboratories website, www.dolby.com* (Dolby Laboratories, Inc.).

Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). "ICRA noises: Artificial noise signals with speech-like spectrum and temporal properties for hearing instrument assessment," Audiology **40**, 148-157.

Drullman, R., and Bronkhorst, A. W. (**2004**). "Speech perception and talker segregation: effects of level, pitch, and tactile support with multiple simultaneous talkers," J. Acoust. Soc. Am. **116**(5), 3090-3098.

Fedtke, T., Fuder, G., Hamann, D., and Haubold, J. (**1990**). "Natürliche Klangbilder," in *Neue Technologien in der Hörgeräte-Akustik - Herausforderung an die Audiologie* (Plath, P., Materialsammlung vom 5. Multidisziplinären Kolloquium der GEERS-Stiftung am 12. und 13. März 1990 in Bonn), pp. 116-136.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**(4), 1725-1736.

Fetterman, B. L., and Domico, E. H. (**2002**). "Speech recognition in background noise of cochlear implant patients," Otolaryngol. Head. Neck. Surg. **126**, 257-263.

Gabriel, B. (**2003**). "Research Report 20030029," (Hörzentrum Oldenburg, Germany).

Gabriel, B. (**2005**). "EchoBlock: Proven to be highly effective in reverberant situations," (Phonak Field Study News).

Genelec (**2003**). "Placement and Room Acoustics Interaction," Resolution Magazine, Sweet Spot **2**(8), 54-55.

Hamasaki, K., Komiyama, S., Hiyama, K., and Okubo, H. (**2005**). "5.1 and 22.2 Multichannel Sound Productions Using an Integrated Surround Sound Panning System," in *2005 NAB BEC*, pp. 382-387.

Herre, J., Hilpert, J., Ertel, C., Hoelzer, A., Spenger, C., Disch, S., and Linzmeier, K. (**2005**). "An Introduction To MP3 Surround," (Fraunhofer Institute for Integrated Circuits IIS, Erlangen), pp. 1-9.

IEC-60268-13 (**1998**). "Sound system equipment - Part 13: Listening tests on loudspeakers (2nd edition)," edited by I. E. Commission (Geneva, Switzerland).

ITU-R_BS.775-1 (**1992**). "Multichannel stereophonic sound system with and without accompanying picture," (International Telecommunications Union, Geneva).

ITU-R_BS.1116-1 (**1994**). "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," (International Telecommunications Union, Geneva).

Johansson, M. S., and Arlinger, S. D. (**2002**). "Binaural masking level difference for speech signals in noise," Int J Audiol **41**(5), 279-284.

Kinkel, M., Rose, S., and Mahn, T. (**2006**). "Wie viele Wiedergabekanäle braucht die „interaktive Hörgeräteanpassung?" DGA2006 **9**.

Lai, W., and Dillier, N. (**2002**). "MACarena: a flexible computer-based speech testing environment," 7th International Cochlear Implant Conference, Manchester, www.unizh.ch/orl/MACarena_Poster.pdf.

Laszig, R., Aschendorff, A., Stecker, M., Müller-Deile, J., Maune, S., Dillier, N., Weber, B., Hey, M., Begall, K., Lenarz, T., Battmer, R. D., Böhm, M., Steffens, T., Strutz, J., Linder, T., Probst, R., Allum, J., Westhofen, M., and Doering, W. (**2004**). "Benefits of Bilateral Electrical Stimulation with the Nucleus Cochlear Implant in Adults: Six Months Post-operative Results," Otology and Neurotology **25**, 958-968.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," Journal of the Acoustical Society of America **49**(2, Suppl 2), 467-477.

Luts, H., Maj, J. B., Soede, W., and Wouters, J. (**2004**). "Better Speech Perception in Noise With an Assistive Multimicrophone Array for Hearing Aids," Ear Hear **25**(5), 411-420.

Merimaa, J., and Pulkki, V. (**2005**). "Spatial Impulse Response Rendering I: Analysis and Synthesis," Journal of the Audio Engineering Society **53**(12).

Müller, C. (**1992**). "Perzeptive Analyse und Weiterentwicklung eines Reimtestverfahrens für die Sprachaudiometrie," (Dissertation, Universität Göttingen, Germany).

Nilsson, M., Ghent, R. M., Bray, V., and Harris, R. (**2005**). "Development of a test environment to evaluate performance of modern hearing aid features.[see comment][erratum appears in J Am Acad Audiol. 2005 Feb;16(2):123]," Journal of the American Academy of Audiology **16**(1), 27-41.

Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J Acoust Soc Am **101**(3), 1660-1670.

Plomp, R., and Mimpen, A. M. (**1981**). "Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences," Acoustica **48**, 325-328.

Poissant, S. F., Whitmal III, N. A., and Freyman, R. L. (**2006**). "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations," J Acoust Soc Am **119**(3), 1606-1615.

Pulkki, V. (**2001**). "Spatial sound generation and perception by amplitude panning techniques," in *Laboratory of Acoustics and Audio Signal Processing* (Helsinki University of Technology, Helsinki).

Pulkki, V., and Merimaa, J. (**2006**). "Spatial Impulse Response Rendering II: Reproduction of Diffuse Sound and Listening Tests," Journal of the Audio Engineering Society **54**(1).

Revit, L. J., Schulein, R. B., and Julstrom, S. (**2002**). "Toward accurate assessment of real-world hearing aid benefit," Hear Rev **9**, 34-38, 51.

Ricketts, T. (**2000a**). "The impact of head angle on monaural and binaural performance with directional and omnidirectional hearing aids," Ear & Hearing **21**(4), 318-328.

Ricketts, T. (**2000b**). "Impact of noise source configuration on directional hearing aid benefit and performance," Ear & Hearing **21**(3), 194-205.

Ricketts, T., and Henry, P. (**2002**). "Evaluation of an adaptive, directional-microphone hearing aid," International Journal of Audiology **41**(2), 100-112.

Ricketts, T. A., and Hornsby, B. W. (**2005**). "Sound quality measures for speech in noise through a commercial hearing aid implementing digital noise reduction," Journal of the American Academy of Audiology **16**(5), 270-277.

Rindel, J. H. (**2000**). "The use of computer modeling in room acoustics," Journal of Vibroengineering **3**(4), 219-224.

Stephan, K., Weichbold, V., and Welzl-Müller, K. (**2006**). "Oldenburger Satztest: Anwendungen zur Erfolgskontrolle bei Hörgeräteversorgung mit Störschallunterdrückung," in *Jahrestagung der Deutschen Gesellschaft für Audiologie* (Köln).

Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**(2), 1081-1091.

Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., and Wouters, J. (**2006**). "Horizontal localization with bilateral hearing aids: Without is better than with," J Acoust Soc Am **119**(1), 515-526.

Verschuur, C. A., Lutman, M. E., and Wahat, H. (**submitted**). "Evaluation of a nonlinear spectral subtraction noise suppression scheme in cochlear implant users," J. Acoust. Soc. Am.

Wagener, K., Kühnel, V., and Kollmeier, B. (**1999**). "Development and evaluation of a German sentence test: I. Design of the Oldenburger sentence test," Z Audiol **38**, 4-15.

Wouters, J., Litiere, L., and van Wieringen, A. (**1999**). "Speech intelligibility in noisy environments with one- and two-microphone hearing aids," Audiology **38**(2), 91-98.

Zwicker, E., and Zollner, M. (**1987**). *Elektroakustik* (Springer, Berlin).

8 Appendix A: Technical Evaluation of Microphone Array Processing

This document is about evaluations on aided subjects, but we add some information here about technical evaluations of directional microphone systems, because this approach could also be modified for the use with subjects.

One common approach to evaluate the performance of a directional microphone system used in a hearing aid is to measure the directivity pattern with one sound source rotating around the device. These directivity patterns allow a meaningful and realistic evaluation of the performance of static directional microphone systems since the SNR is well-defined. Measuring the directivity pattern of an adaptive directional microphone with one rotating sound source yields a falsified result since the notch adaptation follows the direction of the sound source. Thereby always the maximum possible attenuation is obtained.

The challenge is to develop an analysis tool that can overcome the above mentioned limitation of the evaluation to static microphone arrays. The evaluation of adaptive microphone arrays requires a measurement technique that determines the power of a target and one or more jammer signals separately during the simultaneous presentation of these sources since the adaptation and hence the performance of the microphone array depends on all sources.

In the following, one setup for the simultaneous SNR measurement of two coinstantaneous sources is introduced.

In Figure 8-1 the setup is depicted for the microphone array evaluation with a hearing aid mounted on a Kemar manikin (Burkhard and Sachs, 1972). The positioning of the sources is arbitrary.

The two source signals $x_1[k]$ and $x_2[k]$ are generated by filtering a white noise with specially designed comb filters in order to have a discriminable spectrum. Figure 8-2 shows the spectra of the comb filter frequency responses $H_1(e^{j\Omega})$ and $H_2(e^{j\Omega})$.

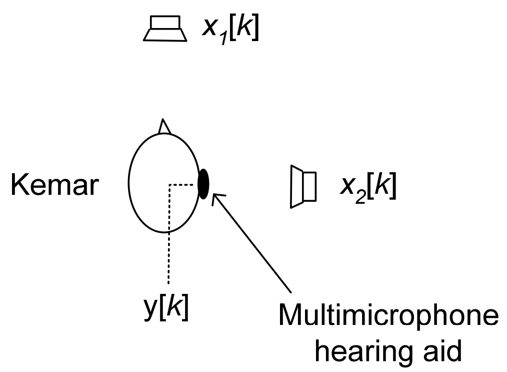

Figure 8-1: Hearing aid microphone array evaluation setup

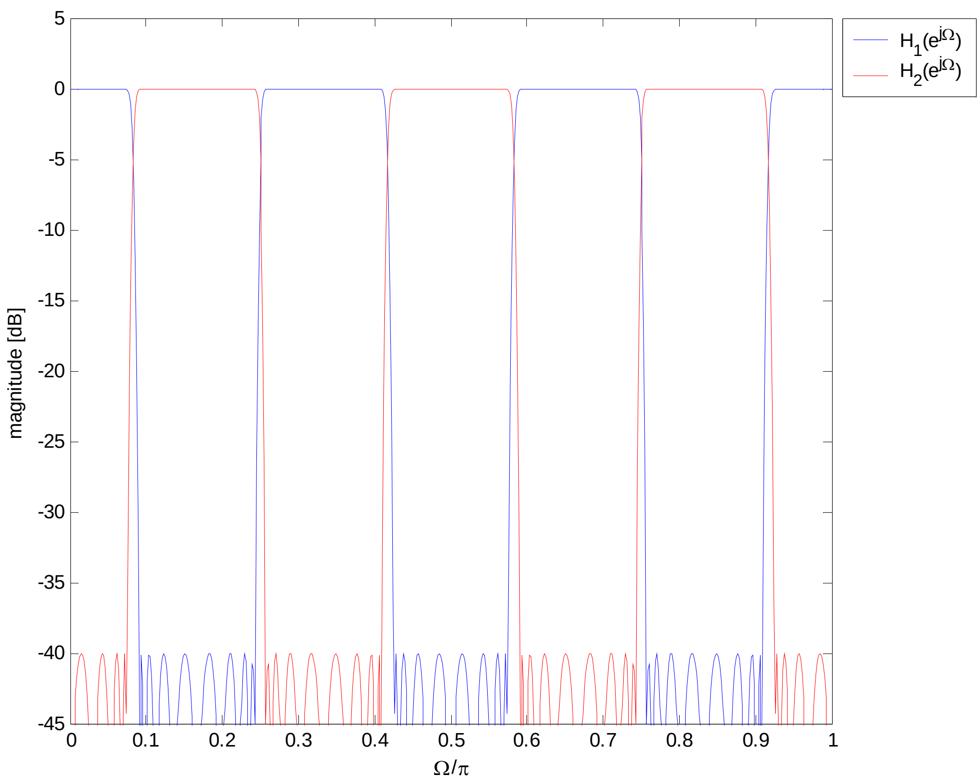

Figure 8-2: Comb filter frequency responses for source signal generation

The hearing aid output $y[k]$ contains a mixture of the source signals that are altered by the hearing aid processing. Filtering the hearing aid output $y[k]$ with the above described filters separates the two hearing aid processed signals $y_1[k]$ and $y_2[k]$ from the mixture. Hence one is able to determine the power of each source separately which first allows the identification of the magnitude of the transfer functions between each source and the output signal and second the SNR calculation.

Please note that due to the interleaved comb filter design the power calculation is carried out at interleaved frequency regions for the two sources respectively. The power density spectrum for each processed source is obtained by interpolation between these frequency regions. As modern hearing aids often use multi-channel processing one should select the frequency resolution of the comb filters such that the width of one comb is less than the half channel bandwidth. Thereby one ensures that both source signals contain energy in each hearing aid channel.

The SNR calculation of the hearing aid output is then carried out with:

$$SNR\left(e^{j\Omega}\right) = 10\log\left\{\frac{\left|Y_1\left(e^{j\Omega}\right)\right|^2}{\left|Y_2\left(e^{j\Omega}\right)\right|^2}\right\}.$$

In Figure 8-3 the above described method is depicted.

In order to quantify the performance of microphone array processing the SNR calculation should be carried out once for the hearing aid in omni-directional processing mode and once for the hearing aid in directional processing mode. The benefit of the directional processing is then obtained by the difference of these signal to noise ratios:

$$SNR_{Benefit}\left(e^{j\Omega}\right) = 10\log\left\{\frac{\left|Y_{1,dir}\left(e^{j\Omega}\right)\right|^2}{\left|Y_{2,dir}\left(e^{j\Omega}\right)\right|^2}\right\} - 10\log\left\{\frac{\left|Y_{1,omni}\left(e^{j\Omega}\right)\right|^2}{\left|Y_{2,omni}\left(e^{j\Omega}\right)\right|^2}\right\}.$$

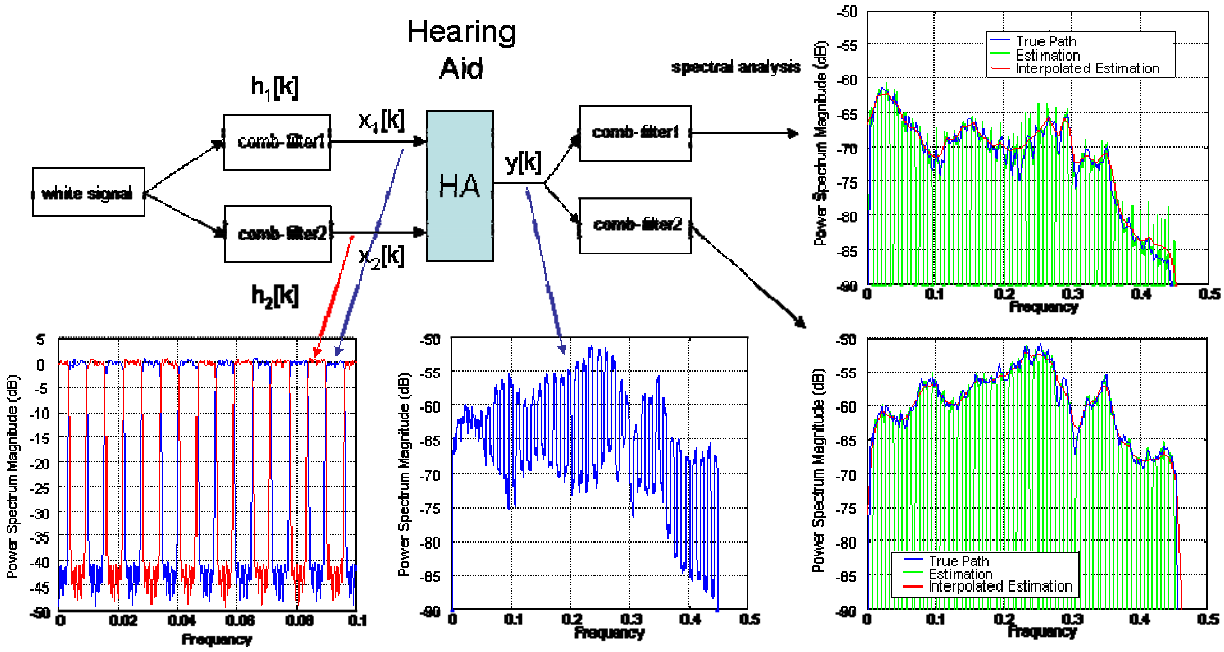

Figure 8-3: Microphone array evaluation

The proposed method for the evaluation of microphone array processing is not limited to two sources. When extending the number of sources one has to take care that the combs for each source signal are narrow-band enough so that all source signals exhibit signal components in each channel of the hearing aid.