Communication Acoustics in Classroom Environments
- On the Use of Assistive Listening Devices

Johan Odelius
Communication Acoustics in Classroom Environments
- *On the Use of Assistive Listening Devices*

Johan Odelius

Division of Sound and Vibration
Department of Human Work Sciences
Luleå University of Technology
SE-971 87 Luleå, Sweden
Abstract

Assistive listening devices (ALDs) are used in classrooms to assist communication for students with hearing loss. An ALD, a system of external microphones, transmits sound directly to the students’ hearing aids. The signal is coupled to the hearing aid using a radio frequency (FM) or an induction loop (IL) system. Using a switch on the hearing aid, the students can listen to the signal from the hearing aid microphone or the ALD signal received by a FM-receiver or a telecoil. An interest in the impact of ALD solutions on student communication and interaction prompted the work reported in this thesis. The thesis evaluates how the quality of classroom ALDs can be optimized in terms of the concept communication acoustics. Aspects of room acoustics, sound quality, and binaural hearing were explored.

The methodical approach was based on self-assessment using questionnaires, interviews and listening tests. The empirical data in Paper I, II, and III consisted of responses from 25 students (10-20 years old) who were attending classes for the hard-of-hearing. In Paper I, the hearing aid microphone (M) and telecoil (T) mode were assessed using a questionnaire. When the hearing aid was in T mode, audibility increased: speech intelligibility was improved and less listening effort was required. Better awareness was achieved using M mode. The students could better hear sounds in the environment around them and participate in conversations – classified as non-teaching – when the ALD was not used. An important feature of sound quality was the distinction of sounds, which is the ability to recognize additional characteristics of a speech sample, e.g., the ability to identify students by voice and judge the mood of students from their voice.

Hearing aids also offer a combined mode where the signals from the internal microphone and the telecoil/FM are mixed. In Paper II, different hearing aid mode combinations were assessed using a combined approach where the different combinations were self-rated in a questionnaire and compared in a listening test. The result supports the finding that a combination of M and T mode is a feasible compromise between audibility and awareness. The students were active in their use of different hearing aid modes and aware of the advantages and disadvantages of the alternatives. The hearing strategies varied in different classroom settings and for different degrees of hearing loss, findings that emphasize the importance of individual adjustments. In Paper III, binaural aspects of hearing and ALDs were assessed in a listening test. A binaural model was compared to an omni-directional microphone. No advantage in speech intelligibility and listening effort was found using a binaural ALD.

ALD design and characteristics can be evaluated using room acoustic modelling and auralization in different room acoustic conditions. In Paper IV and V, auralization and binaural reproduction techniques used in Paper II and III were investigated. Aspects of binaural and spatial hearing were assessed in normal-hearing subjects. Auralization is a reliable method to render a binaural listening experience in a classroom environment: the performance was equal to that of using artificial head recordings. The method used for binaural reproduction – a two-loudspeaker cross-talk cancellation system – introduces distortion in reproduced interaural differences. The binaural advantages in speech intelligibility were reduced when compared to headphone reproduction. The interaural differences were sufficiently reproduced in the frequency region of ALDs (300-4k Hz); the use of cross-talk cancellation for hearing aid and ALD evaluation is to be further studied.

High sound quality matches students’ expectations and demands. To take an active part in the communication in the classroom, students expect to hear sounds in the classroom that they perceive as adequate. However, the students with hearing loss required speech signals with significantly reduced noise and competing speech levels. Today, students have to make a compromise between audibility and awareness. Any alternative, however, could make communication in the classroom difficult. Different classroom settings and sound environments as well as individual factors of preference and degree of hearing loss affect their decision.

Keywords: communication acoustics, assistive listening devices, classes for the hard of hearing, sound quality, speech perception, spatial hearing, auralization, classroom acoustics
Preface

I want to thank all colleagues at the division of Sound of vibration for their contributions to my thesis. In particular, I would like to thank: my supervisor Örjan Johansson for your guidance and a never-ending stream of ideas and encouragement; my assistant supervisor Anders Ågren for your guidance, I am grateful that you initiated this project and gave me this opportunity; my assistant supervisor Roger Johnsson for your support, you are always there to help out; and Arne Nykänen for many interesting and helpful discussions. I also want to thank all the friends at the department of Human work sciences for their support, especially the inspirational excursions under the supervision of Anders Berghlund and Matti Rantatalo.

This work was initiated as a part of Dialogprojektet: Thank you Ann-Christine Wennergren and Erik Ivarsson. I would like to thank the Swedish Association of Hard of Hearing People (HRF), the Swedish Inheritance Fund, and Comfort Audio for their generous support. Further, the collaboration and contributions of the staff and students of Dialogprojektet, Sweden’s National Upper Secondary Schools for the Deaf and for the Hard of Hearing (RGD/RGH), and the National Agency for Special Needs Education and Schools (SPSM) are gratefully acknowledged. Håkan Bergkvist and Arne Gustafsson, thank you for your enthusiastic support and contribution to my work. I am also grateful for the support and interest from my colleagues in the Noise Network.

Finally, I would like to thank my family, Ulrika and Axel. You give me so much joy and love every day!

Luleå, February 2010

Johan Odelius
Thesis

The thesis is based on the work reported in the following papers:

**Paper I** Odelius, J., and Johansson, Ö.: Self-assessment of classroom assistive listening devices. *Accepted for publication in International Journal of Audiology.*


# Table of contents

1. **Introduction** ................................................................................................................. ............. 1  
   1.1. Hearing technology ........................................................................................................ ....... 1  
   1.2. Classes for the hard-of-hearing ........................................................................................... .. 3  
   1.3. Objectives ................................................................................................................ .............. 4  

2. **Theoretical background** ....................................................................................................... ...5  
   2.1. Room acoustics ............................................................................................................ ......... 5  
   2.2. Sound quality ............................................................................................................. ........... 8  
   2.3. Binaural and spatial hearing.............................................................................................. .. 11  

3. **Summary of papers** ............................................................................................................ ....17  
   3.1. Methodical considerations ................................................................................................. . 17  
   3.2. Paper I: Self-assessment of assistive listening devices....................................................... 18  
   3.3. Paper II: Assessments of hearing aid microphone and FM/telecoil mode 
       combinations ....................................................................................................................... 19  
   3.4. Paper III: Effects on speech intelligibility using a binaural assistive listening device ...... 21  
   3.5. Paper IV: Self-assessment of speech intelligibility listening to binaural recordings........... 21  
   3.6. Paper V: Binaural speech intelligibility in free-field and reverberation: A 
       comparison of headphone and cross-talk cancellation reproduction .................................. 23  

4. **Discussion** ................................................................................................................... .............25  
   4.1. Speech quality features of ALDs ........................................................................................ 25  
   4.2. Hearing strategies – The combined ALD and hearing aid microphone mode ............... 26  
   4.3. Binaural advantages ............................................................................................................ 27  
   4.4. Validation of the auralization process ................................................................................. 27  
   4.5. Generalization of the results ............................................................................................... 28  

5. **Conclusion** ................................................................................................................... ...........31  

6. **Further Research** ............................................................................................................. ......33  

7. **References** ................................................................................................................... ............35
1. Introduction

In classrooms, reverberation and noise can make speech difficult to hear. Early room reflections add to the useful speech level, whereas late reflections add to the detrimental noise level. The ratio between the level of the speech signal and the noise signal (the signal-to-noise ratio, SNR) must be sufficient for the intelligibility of speech. A minimum SNR of 15 dB is often recommended, but rarely observed in classrooms (Houtgast, 1981; Ross, 1992; Bistafa & Bradley, 2000; Seep et al., 2000; Hodgson, 2002; Lundquist, 2003; Crandell & Smaldino, 2004). Students with hearing loss – even though they wear hearing aids – need even better listening conditions (Nabelek & Pickett, 1974a; Finizio-Hieber & Tillman, 1978). The extra SNR required is in the range of 10-20 dB and depends on the degree of hearing loss and the situation; the largest difference is obtained for a spatially separated interfering speech source (Nabelek & Robinson, 1982; Moore, 2003).

To obtain sufficient SNRs, external microphone systems can help students with hearing loss. The sound received by the external microphone is transmitted directly to the students’ hearing aids. Placing a microphone close to the mouth significantly reduces the amount of noise and reverberation, a strategy that improves speech recognition (e.g., Hawkins, 1984; Nabelek et al., 1986; Boothroyd & Iglehart, 1998; Boothroyd, 2004). These microphone and sound transmission systems are denoted assistive listening devices (ALD). In the literature, ALDs are also referred to as auditory training systems and classroom amplification systems.

Speech intelligibility is one key aspect when discussing the benefits of ALDs. Because schools rely on group work and dialogue-oriented learning, communication means more than merely listening: communication in a school setting requires interacting and participating with many people in many listening situations. To cover a broad range of qualities, this thesis analyzes ALDs with respect to the concept communication acoustics (Blauert, 2005). The concept covers areas of acoustics that relate to modern communication technologies, such as hearing aid technology.

Areas of communication acoustics – room acoustics, sound quality, and binaural hearing and technology – will be described in Chapter 2. Before the objectives of this thesis are outlined, the sections below will give an introduction to hearing technology and classes for the hard-of-hearing.

1.1. Hearing technology

An ALD acts as a complement to a hearing aid. A hearing aid compensates for the reduced sensitivity, dynamic range, and frequency and temporal resolution associated with hearing loss (Moore, 1996; Sandlin, 2000; Dillon, 2001). Additional functionalities, such as directional microphones and noise reduction algorithms, are also used to increase speech intelligibility in difficult listening situations. Figure 1 illustrates the basic components of a hearing aid. The microphone(s) receives the acoustic input that is processed and amplified and then presented to the user as one of several stimuli:

a) acoustic stimuli by a loudspeaker (receiver) coupled to the ear canal;

b) vibration stimuli, such as into the mastoid bone behind the ear, a bone anchored hearing aid (BAHA); and

c) electric stimuli directly in the cochlea, a cochlea implant (CI).

There are two principal designs of hearing aids: behind-the-ear (BTE) and in-the-ear (ITE). For a BTE hearing aid, the components are mounted in a case behind the ear and the sound is conveyed acoustically via a tube to a custom earmould. An ITE hearing aid completely or partly fills the concha. Smaller variants of ITE are referred to as in-the-canal and completely-in-the-canal hearing aids. If not specified, hearing aids will hereafter denote a type (a) BTE hearing aid, which is also the most commonly used.
Figure 1. A basic scheme of a hearing aid.

The ALD can be seen as extra hearing aid microphones that can be placed close to the sound source. Two dominant technologies – radio frequency (FM) and electromagnetic induction (induction loop, IL) – are then used to couple the signal from the external microphones to the hearing aid. A variant of ALDs, sound-field amplification systems, present the sound to the students using loudspeakers; these systems are not addressed in this thesis. There are many ALD designs, technical specifications, and performance characteristics (Ross, 1992; Lewis, 1994a; Lewis 1994b; Crandell & Smaldino, 2000; Dillon, 2001; Flexer, 2004). Three designs of classroom ALDs are shown in Figure 2. An external wireless microphone (e.g., handheld, clip-on, and headset) transmits using FM radio in the following ways:

a) to a stationary FM receiver where the signal then is transmitted using a room IL and picked up by a telecoil in the hearing aid;

b) to a personal-worn FM receiver that is coupled to the hearing aid using a mini IL (neck loop or silhouette inductor) or by wire connected to the electrical/audio input of the hearing aid; and

c) to the hearing aid directly using an appended (or built-in) mini-FM receiver.

Figure 2. Different designs of classroom ALDs.

An alternative to FM is infrared technology (cf. ALD design b): the user wears an IR receiver that is coupled to the hearing aid using a mini IL. An alternative is also to connect the external microphones by wire to a room IL amplifier, e.g., linked table microphones. Recently, the radio transmission systems were exclusively analogue, but now digital radio systems have been introduced. As with hearing aids, ALDs can have additional functionality such as directional microphones and noise reduction algorithms.

The dominant design of ALDs in Swedish classes for the hard-of-hearing is a stationary system coupled to the hearing aid using room IL. Teachers can use wireless microphones and students can use either wireless or linked table microphones. An alternative ALD design is a centrally placed microphone used on a table similar to a teleconference system. For a more detailed account on ALDs in Swedish schools see Gustafsson (2009). In Europe, hearing aids most often have a built-in telecoil (some 85-90%). In the USA, only about 30-40% of the hearing aids include a telecoil, and FM coupling systems are used to a greater extent (Ross, 2002). Comparisons between FM and IL
systems have shown equal benefit regardless of the technique (Nabelek et al., 1986; Noe et al., 1997).

Using a switch on the hearing aid, the students can either listen to the signal from the hearing aid microphone(s) or the signal from the ALD received by the FM-receiver or telecoil. The alternatives are denoted the hearing aid microphone (M) mode, the telecoil (T) mode, and the FM mode. Many hearing aids also offer additional modes where the signals from M and T/FM are mixed: the M+T or M+FM mode. Because the hearing aids (most often) differ in software design rather than in hardware design, several different programmes including M+T setups can be used. It is rather a question of how the audiologist programmes the aid and the effectiveness of the user interface.

Listening with two ears (binaural hearing) increases speech intelligibility and the ability to localize the person talking. Wearing two hearing aids increase the possibilities to gain from these advantages. Because ALDs provide monophonic signals, the opportunity to gain advantages of binaural hearing are reduced listening with the hearing aid in T/FM mode (Dillon, 2001).

1.2. Classes for the hard-of-hearing

Classroom environments vary. In addition, there are several types of schools for students with hearing loss. In Sweden, 84% of the deaf and hard of hearing students attend regular classes and 16% attend special classes (HRF, 2007). The teaching in these special classes can be in sign language or spoken Swedish; the latter is referred to as classes for the hard-of-hearing. In general, these classes have fewer students – on average 7-8 students (Wennergren, 2008) – compared to regular classes and are acoustically treated and have ALDs that are regularly controlled for. In these classes, students’ degree of hearing loss can vary from mild to profound.

In order for students who use ALDs to take an active part in learning dialogues, the ALDs’ possibility to handle peer-to-peer communication has been emphasized. This is done, for example, by increasing number of microphones in the classroom. Tvingstedt (1993) notes that the hearing-impaired students in regular classes interact to a greater extent with the teacher and to a lesser extent with peers than normal-hearing students. Wennergren (2004, 2008) investigated how the design of ALDs affects peer interaction. An ALD design with a centrally placed microphone on a table was compared to wireless or linked table microphones. One project was designed to change the learning environment in classes for the hard-of-hearing where a more dialogue-oriented and participatory learning environment was promoted (Wennergren, 2007). Questions were raised with respect to how students use ALDs and how ALDs affect student interaction. These questions made it apparent that there was a need to evaluate and develop these communication systems.

The different auditory scenarios in a dialogue-oriented learning environment when compared to a lecture setting are also important to consider with regard to which features of sound quality students assign ALDs and hearing aids. Dillon (2001) discusses such a dilemma: the clearest signal from the teacher is received using FM mode, a combined mode is of intermediate clarity, and the worst signal is received using M mode. However, children seem to prefer the reverse order (Cotton, 1988). The children feel detached from the environment around them when they use FM. Lewis and Eiten (2003) showed that students prefer the FM signal relative to the microphone signal when listening to the teacher. The hearing aid microphone signal was preferred when listening to others (not the teacher) in the classroom.
1.3. **Objectives**

Educators and researchers have become interested in how ALDs influence student interactions. This concern prompts the research in this thesis that aims to evaluate how the quality of classroom ALDs can be optimized in terms of communication acoustics. The following research questions were raised:

1. What features of sound quality do students assign classroom ALDs with regard to the hearing aid microphone mode?
2. Which hearing strategies do students employ, how do students use ALDs and which strategies are important for their choice?
3. Does a combination of the hearing aid microphone mode and the T/FM mode use the best or worse parts of each mode?
4. Is the loss of bilateral advantages using ALDs a key factor for students use and preference of operating hearing aid mode?

The studies reported were limited to stationary room IL systems in Swedish classes for the hard-of-hearing.
2. Theoretical background

Communication acoustics is a concept that is composed of several areas that relate aspects of human-to-human and machine-to-human communication. The interest in these areas has increased since the introduction of digital signal processing and recording. As a result, psychological and physiological models are increasingly exploited in technical applications (Blauert, 2005). With regard to the evaluation and development of assistive listening devices (ALDs), three aspects of communication quality will be addressed: (2.1) room acoustics, described with an emphasis on acoustic quality related to the intelligibility of speech; (2.2) sound quality, where a special focus is on speech intelligibility; and (2.3) binaural hearing and techniques for recording and reproducing binaural signals.

2.1. Room acoustics

Since the work of W.C. Sabine (1922), reverberation time has been the most important factor for room acoustic quality. However, the experience of sound in a room is due to several factors; that is, the experience of sounds varies between rooms with the same reverberation time. To better describe the sound experience and the intelligibility of speech in rooms, different parts of the room impulse response (rather than only a linear regression of its slope, the reverberation time) needs to be analyzed.

2.1.1. Listening in rooms and the importance of early reflections

The sound energy in a room can be divided into direct and reverberant sounds. The direct sound energy decreases with the square of the distance, whereas the sound energy of the reverberant sound field depends on the room’s dimensions and amount of absorption. The distance where the energy of the direct sound and the reverberant field are equal defines the critical distance: 
\[ r_c = 0.1\sqrt{V/\pi T} \]
(Cremer & Müller, 1978), where \( T \) is the reverberation time, \( V \) the volume, and \( I \) the source directivity ratio. When considering ALDs, it is important to consider where the microphone is relative to the speaker; the signal-to-noise ratio (SNR) is increased 6 dB for every halving of distance when inside the critical distance.

The direct sound has an important role in sound perception and speech intelligibility in rooms. The precedence effect refers to a group of phenomena concerning the perception and localization between a direct sound and its reflections (Wallach et al., 1949; Haas, 1951; Freymam et al., 1991; Hartman, 1997; Litovsky et al., 1999; Blauert, 2001). The precedence effect contributes to localization (the localization dominance of the first wave-front) and to speech intelligibility (discrimination suppression of reflections), e.g., the Haas effect and de-reverberation.

The Haas effect refers to the advantage rather than the interference of early reflections to speech perception (Haas, 1951). Speech intelligibility is fairly constant for a single reflection within 20 ms, both for normal-hearing and hearing-impaired individuals (Nabelek & Robinette, 1978). According to the authors, the principles of room acoustics apply both hearing-impaired and normal-hearing individuals. The same conclusion was drawn by Bradley et al. (2003): “increased early reflection energy has the same effect on speech intelligibility scores as an equal increase in the direct sound energy. This was true for both nonimpaired listeners and for listeners with mild to moderate hearing threshold shifts.”

The speech is sensitive to reverberation characteristics. The effect of de-reverberation (de-colouration) is clearly heard when comparing a recording of sound to listening in the actual sound field; the effect depends on receiving the sound without any distortions (Hartman, 1997). Although there is monaural de-reverberation, the binaural contribution to de-reverberation is emphasized;
e.g., binaural recording improves clarity in reverberant environments (Koenig, 1950). The de-reverberation process is reduced for individuals who wear hearing aids (Hartman, 1997).

2.1.2. Acoustic predictors of speech intelligibility

After Haas’ studies on the importance of early reflections, several measures to quantify speech intelligibility based on the ratio between early and late reflections were developed. Thiele (1953) defines definition (deutlichkeit) as the relationship between the energy of the first 50 ms and the total energy. Instead of using 50 ms as a distinction between early and late reflections, 80 ms was suggested for music, which led to the definition of clarity ($C_{80}$) (Reichardt et al., 1975): the energy of the early reflections relative to the late reflections. Clarity is frequently used with regard to speech intelligibility, but with the integration time of 50 ms ($C_{50}$). Similar measures have also been suggested by Lochner and Burger (1958, 1964); a weighting function was used instead of a sharp limit between early and late reflections.

Lochner and Burger (1964) also suggested a signal to total noise ratio where level of the ambient noise was added to the late reflections. A similar measure denoted as the useful-to-detrimental index ($U$) was later defined by Bradley (1998). Useful is the direct signal energy and the early reflections (typical 50 ms) and detrimental is both the late reflections and the energy ratio between the noise and the speech. There are several other available predictors of speech intelligibility defined as a function of both reverberation and noise level. Knudsen (1929) described various acoustic factors affecting speech perception. Peutz (1971) correlated acoustic quantities with percent correct repeated phonemes and words and defined the articulation loss of consonants ($A_{\text{cons}}$). The loss of consonants increases with the square of the distance and the reverberation time and decreases with the room volume. However, only up to a certain distance – defined as $r_{\text{cl}} = 0.2\sqrt{\frac{V}{T}}$, about 2.5 times the critical distance for a half spherical source ($\Gamma = 2$) – beyond this distance is the $A_{\text{cons}}$ constant. The effect of the noise level on $A_{\text{cons}}$ was presented in graphs (Peutz, 1971).

French and Steinberg (1947) introduced a method for determining the speech perception based on physical parameters. The method was the basis for the Articulation Index (AI). The method has been improved, primarily through the work of Kryter (1962), and the standard for AI has been revised (ANSI, 1997). The name was also revised to the Speech Intelligibility Index (SII). The calculation of SII can be divided into three stages: (1) estimation of effective signal-to-noise ratio for different frequency bands, including effects such as spread of masking; (2) a linear transformation of the effective signal-to-noise ratio (within ±15 dB) per frequency band to an audibility index with a range between 0 and 1; and (3) calculating a frequency-weighted average of the audibility index. Steeneken and Houtgast (1980) developed a different method for measuring speech intelligibility, the Speech Transmission Index (STI; IEC, 2003). The AI was, according to the authors, a method well suited for communication channels with distortion in the frequency domain (e.g., disruptive noise), but it was not well suited for harmonic distortion or distortion in the time domain (e.g., reverberation). STI is based on the Modulation Transfer Function (MTF) that describes the frequency-specific transfer function of the envelope of a test signal (designed to resemble spectral and temporal characteristics of speech) through a room or other types of communication channels. Hence, the method takes into account both reverberation and background noise. Based on the MTF, a signal-to-noise ratio for each frequency band is calculated. The method is analogous to the computation of SII. Several simplification of STI exist (IEC, 2003), such as the Rapid Speech Transmission Index (RASTI), that are adequate for room acoustic measurements (Houtgast & Steeneken, 1984). Relevant to the application of STI is Schroeder’s (1981) definition of the complex MTF, which can be calculated from the system’s impulse response for a linear passive system. This approach uses the noise level separately.
Although different approaches, the different predictors for speech intelligibility in rooms are strongly correlated (Bradley, 1998). Bradley (1986) also investigated how appropriate some of these acoustic measures were as a predictor of speech intelligibility in classrooms. Both the useful-to-detrimental index (U50, 1 kHz) and STI were correlated to percentage of correctly repeated words. However, Bradley (1986) also concludes that in small rooms, such as classrooms, early-to-late reverberant ratios can be predicted by the room reverberation time. The combination of SNRs and reverberation times can predict speech intelligibility with only a little less accuracy than U50 and STI.

Modifications of SII and STI that account for effects due to hearing loss, binaural hearing, articulation, and audio-visual effects (e.g., lip reading) have been addressed in different studies. An approximate audio-visual correction for SII is included in the standard (ANSI, 1997). Performance and modifications of AI (early version of SII) and STI for hearing-impaired listeners was discussed by Humes et al. (1986). According to Holube and Kollmeier (1996), AI and STI (including different modifications) are good predictors of speech intelligibility on a group average level, but may fail to predict the result of a single subject with a certain hearing loss. A new predictor for binaural speech intelligibility based on the SII for both normal-hearing and hearing-impaired was presented by Beutelmann and Brand (2006). Wijngaarden and Drullman (2008) described a binaural version of STI. The authors also concluded that a better-ear STI was useful to the standardized STI.

2.1.3. Room acoustic modelling and auralization

Room acoustic modelling is based on geometric acoustics and is used to predict energy-based measures such as reverberation time and clarity. Using these predictions, different acoustic treatments can be evaluated to determine what classroom acoustic design will best benefit the students. Another application is auralization – an analogue to visualization – and is defined by Kleiner et al. (1993) as “the process of rendering audible, by physical or mathematical modelling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modelled space”. In addition to the room acoustic measures, the acousticians and students can listen to and compare the different room acoustic designs.

In geometric acoustics, the wave propagation is approximated by rays under the condition that the wavelength of the sound is small compared to the dimensions of boundary areas and obstacles (Cremer & Müller, 1978; Blauert & Xiang, 2008). Room acoustic modelling is widely applied in room acoustic design and several software programs have been developed (see review in Savioja et al., 2002). In addition to ray-based methods, computational heavy way-based methods are suitable for simulation of low frequencies, e.g., finite element methods and boundary elements methods (Kleiner et al., 1993; Savioja, 1999). In addition, statistical modelling, such as statistical energy analysis, can be used to predict noise level in coupled systems (Savioja, 1999).

Two commonly used ray-based methods are the image source model (ISM) and ray-tracing (Figure 3). When using ISM, an image source is created as a geometric image in the reflecting surface. Because the number of sources increases exponentially, the method is used for early reflections. A short-coming is also the limited ability to handle diffuse reflections (Dalenbäck, 1995). Ray-tracing uses a large number of rays originating from the source in predefined or randomized directions. Absorption and scattering coefficients for each surface characterize each ray’s way to a receiver volume. Fewer computations are needed when compared to ISM, but the method creates an unnatural reflection growth (Dalenbäck, 1999). A variant to ray-tracing is cone-tracing, where a ray trajectory represents a potential reflection only if the receiver point falls within the reflected cone.
Figure 3. Principles of (a) the image source model; and (b) ray-tracing (after Kleiner et al., 1993).

The software CATT-Acoustics¹ have been used in the reported studies in this thesis. CATT features a hybrid model, referred to as randomized tail-corrected cone-tracing (RTC) (Dalenbäck, 1999). The method is based on cone-tracing where diffuse reflection is managed in a similar way as ray-tracing and ISM is used for the direct sound and the early reflections. The ISM (early part) is handled deterministically and the ray/cone-tracing (late part) is computed in octave bands 125-4k Hz (extrapolation up to 16 kHz). The auralization chain consists of three parts (Dalenbäck, 1995):

Prediction. Octave band echograms are predicted using RTC given the room geometry, surface properties (absorptions and scattering coefficients), source directivity, and head direction.

Post-processing. Each reflection-path is post-processed where different receiver models can be applied, e.g., microphone polar patterns, binaural model, and ambisonic models. The binaural models are defined by a matrix of measured head-related impulse responses from different directions. The result is the predicted room impulse responses (RIR).

Convolution. The RIRs are used for convolution with anechoic audio material. There is also the possibility to add filters, such as head-phone equalization.

In an early verification of auralization, Kleiner (1981) found lower speech intelligibility scores listening to an auralization based on calculated echograms when compared to listening in the actual sound filed (on-site), whereas no differences were found between an auralization based on measured echograms and on-site listening. A key part in the verification of the auralization process is that the conditions of geometrical acoustics are met, because the prediction of the room impulse responses is based on those assumptions. Yang and Hodgson (2007) conclude that if the room is not too absorptive or noisy, speech intelligibility tests using room acoustic modelling and auralization (CATT-Acoustics) are reliable.

2.2. Sound quality

Quality of transmission technology can be described as sound-transmission quality where maximal output quality means no distortion from input (Blauert & Jekosch, 2003). This implies that the ALD has a maximum quality if the speech signal is received and reproduced by the hearing aid without

¹ http://www.catt.se
any other sounds and reflections coming from the room. The ALD quality can be described by quantifying the distortion in the transmission from the speech source to the hearing aid. Different models exist to investigate the quality of speech transmission in telecommunications (for an overview see Möller, 2005). A method to rate the transmission quality with respect to speech intelligibility is the Speech Transmission Index (STI) (cf. Section 2.1.2).

Communication quality is sometimes used to mean transmission quality. That is, communication is seen as transmission of messages where it is important to analyze the obstacles between the sender and receiver. An alternative semiotic perspective is that communication produces and exchanges meanings (Fiske, 1990). Including a semiotic perspective, speech quality is defined as follows:

“The result of assessing all the recognized and nameable features and feature values of a speech sample under examination, in term of its suitability to fulfil the expectations of all the recognized and nameable features and feature values of individual expectations and/or social demands and/or demands.” (Jekosch, 2005. p. 6).

On the basis of speech quality, the adequacy of the output is to be assessed where individual and social factors are important. Accordingly speech quality varies between context, between individuals, and over time. Self-assessment is hence an important tool for the subjective judgements of adequacy. Several features of hearing are to be considered in order to evaluate the qualities of ALDs. Sound attributes of sound-reproducing systems, e.g., loudspeakers, headphones, and hearing aids have been investigated by Gabrielsson and Sjögren (1979) and Gabrielsson et al. (1988). Together with an overall impression of quality seven perceptual dimensions were revealed from adjective ratings, these were: fullness, loudness, brightness, softness, nearness, spaciousness, and clarity.

2.2.1. Self-assessment of hearing

The use self-assessment has also been argued when assessing auditory disability (High et al., 1964; Schow & Gatehouse, 1990). Several questionnaires have been developed to assess disability and handicap (see review by Schow & Gatehouse, 1990). Using the questionnaire Amsterdam Inventory for Auditory Disability and Handicap, Kramer et al. (1995) identified five basic auditory disability factors: 1) distinction of sounds, 2) intelligibility in noise, 3) auditory localization, 4) intelligibility in quiet, and 5) detection of sounds. Detection of sounds reflects situations where the information about the acoustical environment is important, whereas distinction of sounds refers to the timbre of the perceived sound, e.g., distinguishing between male and female voices. The importance of perceived acoustical environments relates to the previous defined perceptual dimensions of nearness and spaciousness. In addition, Blauert and Jekosch (2003) discuss the plausibility of the reproduced sound where the feeling of presences is one component. A speech signal transmitted and reproduced by a hearing aid without any information about the room (reverberation and other sound sources) may not be perceived as adequate. Dillon (2001) discusses the negative feeling of detachment from one’s environment. The inability to detect off-axis sounds has also been discussed with regard to directional microphones in hearing aids (e.g., Stadler & Rabinowitz, 1993). The discussion also relates to the definition of noise as an unwanted sound (Kryter, 1994).

Another questionnaire developed to assess a broad range of hearing functions was the Speech, Spatial, and Qualities of Hearing scale (SSQ) (Gatehouse & Noble 2004). The range of hearing is divided into three sections: Speech hearing, Spatial hearing, and Other Qualities. Speech hearing covers various listening comprehension situations of differing difficulty. The Spatial hearing section concerns the ability to locate sound and the externality of sounds. The section Other Qualities is a grouping of several issues: segregation of sounds, recognition of voices, identification of other people’s moods, listening effort, and judgments of clarity and naturalness. SSQ shares items with other questionnaires such as the Amsterdam Inventory for Auditory Disability and Handicap (Kramer et al., 1995), although it places more emphasis on the issue of binaural hearing (Gatehouse
& Noble, 2004). The SSQ has also been used to assess the advantages of wearing hearing aids in both ears instead of only in one ear (bilateral advantage) (Noble & Gatehouse, 2006).

2.2.2. Speech intelligibility

A key part in the described questionnaires was speech intelligibility. Through a semiotic approach, speech intelligibility is the dominant dimension of speech quality (Jekosch, 2005). However, preference and high speech intelligibility are sometimes contradictory factors. This can be seen in the low-frequency region that has been shown to be important for hearing aid sound quality; at the same time, it resulted in a reduction in intelligibility (see review by Gabrielson et al., 1988). Sound quality judgments of the seven perceptual dimensions (fullness, loudness, brightness, softness, nearness, spaciousness, and clarity) are better at distinguishing between sound-reproducing systems when compared to speech intelligibility (recognition of PB words and sentences in noise) (Gabrielsson et al. 1988). The systems were characterized by five frequency responses and both speech and music stimuli were used. The result was consistent for normal-hearing and hearing-impaired listeners. Self-assessment of clarity of speech was able to distinguish between the different frequency responses where speech intelligibility was not.

Speech intelligibility is often quantified in recognition tests, where the number of correctly repeated phonemes, words, or sentences is measured in the presence of masking sounds. Several tests have been developed (see review in Jekosch, 2005). The tests are also referred to as speech intelligibility test; however, only segmental intelligibility is measured (Jekosch, 2005). The notation recognition test is used here. Instead of measuring the percentage of correctly repeated words, the SNR between the target and the masker can be measured at a certain detection threshold, referred to as the speech-reception threshold (SRT) in noise. A commonly used detection threshold is 50%; this is due to better convergence assessing recognition at the steepest part of the intelligibility curve in noise. An example of the relation between percentage recognition/intelligibility and the SNR is shown in Figure 4. A Swedish sentence recognition test was developed by Hagerman (1982) and an adaptive procedure was described by Hagerman and Kinnefors (1994). For an international cross-validation of sentence recognition test see Wagener et al. (2007).

The low detection thresholds (40-50%) used is not a reasonable listening situation To capture a more natural situation, such as a listener in a noisy environment trying to understand a spoken message, the just-follow-conversation (JFC) method has been suggested (Hygge et al., 1992, Larsby & Arlinger, 1994). The JFC test is performed by letting a listener adjust the level of speech or masker until they are just able to follow what is being said. On average, the difference is approximately 10 dB (Larsby & Arlinger, 1994; Magnusson et al., 2001), i.e., significantly higher word recognition (cf. Figure 4). Speech intelligibility in scenarios where the sound sources are separated in space is described in Section 2.3.1, where the effect of hearing loss to aspects of speech intelligibility also will be addressed.

In the context of this thesis, it is important to consider that hearing includes both audition and cognition (Hällgren, 2005). Speech recognition is not necessarily reduced in difficult listening situations; i.e., the audition performance is not decreased. However, speech understanding will require more cognitive processing (top-down), e.g., working memory capacity and semantic and lexical knowledge. Measuring listening effort is thus important when attempting to capture the cognitive load (Hällgren et al., 2005). The flat intelligibility curve at high recognition (Figure 4) can be assessed by measuring cognitive load. There is also a difference between to hear and to remember what was said. The effect of noise on memory has been shown even though the words were 100% recognized (Kjellberg et al., 2008).
2.3. Binaural and spatial hearing

Listening with two ears instead of one ear is central for sound perception. There are both localization and speech intelligibility advantages gained through binaural listening. Binaural advantages are made possible wearing two hearing aids instead of one. This is referred to as bilateral advantages. Evaluating the benefits using ALDs’ binaural advantages is a factor to consider. Because ALDs today provide monophonic signals, advantages from dichotic listening (a different sound is presented to each ear) cannot be expected (Dillon, 2001). Localization and speech intelligibility aspects of spatial hearing and how these aspects are affected by a hearing loss are addressed below. In Section 2.3.2, binaural technology, such as the synthesising and reproduction of binaural signals, is described.

2.3.1. Localization and binaural speech intelligibility

The sound emitted from a source in space will arrive to the ears at different levels (interaural level difference, ILD) and times (interaural time difference, ITD). ITD and ILD for different horizontal angles are shown in Figure 5. The ITD is higher at low frequencies (below 1.5 kHz) because the phase cue becomes ambiguous when frequency increases (Wightman & Kistler, 1997). For high frequencies, the sensitivity of ongoing ITDs is carried by the envelope (Bernstein & Trabjotis, 2002). The ILD is prominent for high frequencies; there is no effect of head shadow when the dimensions are small in comparison to the wavelengths. Judgement of angles in the vertical plane is made possible by spectral cues (above 5 kHz) due to the outer ear (pinna) (Shaw, 1997). These cues add to the binaural spectral differences as well as to monaural cues. To externalize sounds, i.e., to perceive the sound as to be outside the head, both interaural differences and pinna resonances are important. Room reflections also add to the ability to externalize sounds as well as to avoid front-back reversals (Blauert, 2001; Dillon, 2001). Externalization and front-back reversal are important aspects of binaural technology due to the sensitivity of high-frequency errors. Blauert (2001) also addresses additional factors important for the perception of sound in space, such as head movements and visual inputs.
Figure 5. Interaural difference measured for an artificial head. (a) The interaural time difference (ITD) for different azimuths: (dotted) measured (line) polynomial fit. (b) The interaural level difference (ILD) plotted against azimuth and frequency.

In addition to the ability to localize a person talking, our ability to listen to one person when others are speaking at the same time is very high, a phenomenon known as the cocktail party effect (Cherry, 1953). Cherry presented different speech signals to each ear, and the subjects could easily focus on one ear. Studies on speech intelligibility in multi-talkers environments are reviewed by Bronkhorst (2000). It is also important to distinguish between energetic and informational masking. Informational masking can be seen as a central or higher-level interference, whereas energetic masking is due to a distortion of the signal’s peripheral representation (Brungart, 2001; Brungart et al, 2001). It has been shown that the advantage due to spatial separation of sources is greater for
informational masking than for energetic masking (Arbogast et al., 2002). The ability to separate sounds and perceptual grouping (an auditory stream) are further explored by Bregman (1990).

The cocktail party effect is often presented in scenarios where the speech and the masker are spatially separated. The improvement gained in speech intelligibility is attributed to two factors (Zurek, 1990): head shadow and binaural processing (also known as binaural integration and binaural squelch). The head shadow effect is due to the possibility to attend to one ear with the better SNR. The effect is a result of the ILDs that are prominent for high frequencies explained by the wavelength and size of the head. Binaural processing relies on the brain taking advantage of the interaural differences in both time and level.

A common descriptor of the advantage in speech intelligibility for spatially separated sources is the intelligibility level difference, defined as the difference in SRT when speech and noise are presented from the front (S0N0) and when speech is presented from the front and the noise is presented for the side (S0N90). In free-field, this difference is about 8-10 dB (Bronkhorst, 2000; Wagener et al., 2007). To separate the contribution of head shadow and binaural processing, SRT can be measured in the setup S0N90 by blocking the ear directed towards the noise source; i.e., there is no extra information to gain from the noisy ear. The difference between the blocked monaural case and its binaural equivalence quantifies the binaural processing and is denoted the binaural intelligibility level difference. Typically 3-4 dB is gained due to binaural processing (Wagener et al., 2007). The effects of head shadow and binaural processing are not necessarily additive effects. Bronkhorst and Plomp (1988) studied the individual contribution of ITDs and ILDs. They found intelligibility level differences of 5.0 dB for ITD-only, 7.8 dB for ILD-only, and 10.1 dB when stimulus contained both ITD and ILD. There is also an advantage when the same signal is presented to both ears when compared to monaural listening. The effect is named diotic summation or binaural redundancy and the advantage is about 1-2 dB (Bronkhorst & Plomp, 1989; Dillon, 2001).

In more realistic reverberant environments, listener performance will deteriorate and the benefit in speech intelligibility from spatial separation will decrease (Bronkhorst, 2000). The placements of sources are important to consider (Shinn-Cunningham et al., 2005). The effect of the head shadow relies on the listener to be in the direct field of the source, i.e., within the critical distance. The effect of reverberation to the binaural speech intelligibility has been investigated in several studies (see review in Bronkhorst, 2000). The studies are difficult to compare due to the variations in acoustical parameters as well as source positions. The intelligibility level difference has been found to decrease in room conditions in the range of 5-8 dB (Bronkhorst & Plomp, 1990; Koehnke & Besing, 1996). The binaural intelligibility level difference (source placed ±30 degrees) also decreases with longer reverberation times: from 5 to 3 dB when the reverberation time was increased from 0.3 s to 0.6 s (Nabelek & Pickett, 1974b). Different approaches that incorporate binaural processing in measures that predict speech intelligibility have been undertaken (Zurek, 1990; Blauert, 2001). See also Section 2.1.

The effects of hearing loss

The head shadow effect relies on ILD for high frequencies and binaural processing relies on both ILD and ITD, where ITD is largest for low frequencies. Because a sensorineural hearing loss is most pronounced in high frequencies, the advantage due to the head shadow effect is reduced. Bronkhorst and Plomp (1989) found that the binaural gain when the masker was changed from 0 deg to 90 degrees was between 2.6-5.1 dB less for the hearing-impaired subjects. When the results for ILD-only and ITD-only stimuli are considered separately, it was shown that the difference between the two groups was for the ILD-only case. There was also a great variability in the gain due to ILD among the hearing-impaired listeners (between 0 dB to normal values of 7-8 dB or even more). Festen and Plomp (1986) also showed that placing the microphone behind the ear (BTE hearing aid) decreases head shadow by 2-3 dB for the S0N90 setup. Table 1 shows the differences in
SRT between normal-hearing and hearing-impaired listeners (Moore, 2003). The small advantages of binaural redundancy (hearing the same signal in both ears) have been argued to be more pronounced for individuals with hearing loss (Cox & Bisset, 1984; Day et al., 1988).

Concerning localization, a high-frequency loss decreases the ability to localize sound sources in the vertical plane. Horizontal localization, however, is not affected. Even for low-frequency hearing loss (below 50 dB HL) the ability is hardly affected because of the high energy levels for low-frequency vowels as compared to high-frequency consonants (Byrne et al., 1992 as cited by Dillon, 2001).

Table 1. Typical amounts by which the speech-reception thresholds (SRT) in the presence of background sounds is greater for hearing-impaired than for normal hearing listeners for various types of background sounds and listening situations (Moore, 2003).

<table>
<thead>
<tr>
<th>Type of background</th>
<th>Listening situation</th>
<th>Deficit in SRT (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech-shaped noise</td>
<td>Speech + background in front, unaided</td>
<td>2.5–7</td>
</tr>
<tr>
<td>Speech-shaped noise</td>
<td>Speech + background in front, aided</td>
<td>2-6</td>
</tr>
<tr>
<td>Single talker</td>
<td>Speech + background in front, unaided</td>
<td>6–12</td>
</tr>
<tr>
<td>Single talker</td>
<td>Speech + background in front, aided</td>
<td>4-10</td>
</tr>
<tr>
<td>Single talker</td>
<td>Speech + background spatially separated</td>
<td>12–19</td>
</tr>
</tbody>
</table>

Bilateral advantage

Binaural advantages have been an issue regarding bilateral versus unilateral hearing aid fitting – the bilateral advantage. Speech intelligibility advantages have been shown in several studies, however, some studies block the unaided ear and thus measuring binaural advantage rather than bilateral advantage (Dillon, 2001). Because the unaided ear cannot contribute, bilateral advantage is greatest for people with severe hearing loss (Festen & Plomp, 1986). In an experiment where a speech target was arbitrarily presented in the horizontal plane in the presence of a diffuse noise masker, Köbler and Rosenhall (2002) showed bilateral advantage of both localization and speech recognition for listeners with mild-to-moderate hearing loss.

Using the Speech, Spatial, and Qualities of Hearing Scale (SSQ), Noble and Gatehouse (2006) showed that bilateral advantage was considerable in difficult multi speaker situations and of less importance in more simple situations (where a unilateral advantage was found). There are also bilateral advantages in sound quality. Balfour and Hawkins (1992) showed bilateral preference for overall impression and the seven previously described perceptual dimensions (fullness, loudness, brightness, softness, nearness, spaciousness and clarity; Gabrielsson & Sjögren, 1979). The strongest preference for a bilateral capacity was found for overall impression, fullness, and spaciousness. Bilateral advantage with regard to sound quality was not as evident in the SSQ study (Noble & Gatehouse, 2006).

2.3.2. Binaural synthesis, recording and reproduction

In binaural synthesis, measured head-related transfer functions (HRTFs) are used to add binaural cues as well as monaural cues to the sound (Blauert, 1997; Hammershøi & Møller, 2005). Using non-individualized HRTFs, e.g., an artificial head, will introduce disparities in the ITD and ILD. Localization performance is decreased where front-back reversals are increased (Wenzel et al., 1993, Bronkhorst, 1995; Møller et al, 1996; Hawley et al. 1999; Møller et al. 1999; Best et al., 2005). As discussed, this is due to high frequency distortions and nonexistent possibility of head movements. Although localization performance is degraded using non-individualized HRTFs,
Hawley et al. (1999) found no difference in speech intelligibility between listening with headphones and listening in the actual sound-field. In reverberant conditions perceptual factors, such as the precedence effect, must be considered (cf. Section 2.1). As previously discussed, the effect of de-reverberation is clearly heard comparing on-site listening to a room recording (Hartman, 1997). Hence, disparities in ITDs and ILDs may have a greater effect on speech intelligibility in a reverberant condition. In contrast, reverberation adds noise to ITD and ILD (Nix & Hohmann, 2006).

To preserve ongoing ITDs and ILDs is also an important topic to consider when reproducing binaural stimuli. A common reproduction approach is to use headphones, which enables an almost complete channel separation. Adequate playback equalization is required where the use of non-individualized equalization filters may cause high frequency distortions (Hammershøi & Møller, 2005). Another proposed approach is to use loudspeakers and a cross-talk cancellation system (e.g., Schroeder, 1970; Damaske, 1971; Griesinger, 1989; Møller, 1989; Gardner, 1997; Winkler et al., 2009). The principle behind cross-talk cancellation is shown in Figure 6. Using filters resembling the transfer function between listeners and the loudspeakers, the left ear stimulus to the right ear and vice versa are attenuated by inverse filtering. The cancellation of the cross signal relays on these filters to match the HRTFs of the listener. Akeroyd et al. (2007) showed cross cancellation of 25 dB with matched HRTFs and only 13 dB with mismatched HRTFs. Mismatched HRTFs also led to inaccurate reproduced ILDs and ITDs (std of 4 dB and 100 μs, cf. Figure 5). That is, binaural cues are not accurate recreated when using cross-talk cancellation if the HRTFs used in the computation do not match the individual HRTFs of the listener. An advantage of cross-talk cancellation is that the sounds are perceived outside the head (Griesinger, 1989); this can be compared to headphone listening where the rate of sound images perceived inside the head is about 30% at 0 degrees azimuth (Begault & Wenzel, 1993).

![Figure 6](image)

**Figure 6.** Principle figure of a cross-talk cancellation system, after Møller (1989). The binaural signal (X) is processed using inverse filter of the cross transfer paths (H_{left-right} and H_{right-left}) in order to recreate the binaural signal at the eardrums (Z).
3. Summary of papers

Before the five papers are summarized, considerations with regard to methodical framework are discussed.

3.1. Methodical considerations

The methodical approach was based on self-assessment, questionnaires, interviews and listening tests. In Paper I, a questionnaire addressed several aspects of hearing to assess features of sound quality relevant for classroom assistive listening devices (ALDs). The Speech, Spatial, and Qualities of Hearing Scales (SSQ) (Gatehouse & Noble, 2004) was used as a starting point. SSQ shares items with other questionnaires (cf. Section 2.2.1), but it focuses on binaural hearing. Aspects of binaural hearing were assumed important in a peer-to-peer interactive classroom setting. SSQ consists of 50 items covering a wide range of contexts. Items related to working situations in the classroom were chosen and rewritten to suit the context. The questionnaire consisted of 18 items (5 Speech items, 4 Spatial items and 9 Other Qualities items). Although a well-designed questionnaire, the range of features to assess ALD quality was predefined by the author rather than shaped by the students. However, the students rated the importance of each item with regard to its importance of how they used their hearing aid or ALD.

Because the reference is critical when students judge the quality of ALDs, the scores for ALDs were compared to scores for the hearing aid microphone mode. The questionnaire developed was based on SSQ; however, it was not validated. As a measure of reliability, items within each section of the questionnaire were intercorrelated. Furthermore, the scores were correlated against word recognition scores (collected in the experiment described in Paper III) to investigate how the different self-assessed items (e.g., speech intelligibility, localization, sound segregation, and recognition of voices) correlated with measured recognition performance.

In conjunction with the questionnaire, the students also wrote down which mode they used in a range of situations. The data were compared to their scores in the questionnaire. The approach, however, did not thoroughly investigate the students’ hearing strategies. A questionnaire that is designed to assess a broad range of communication problems is the Communication Profile for the Hearing Impaired (Demorest & Erdman, 1987). A Swedish version was developed by Hallberg et al. (1992). Three factors of communication strategies were derived: (i) Maladaptive Behaviours, e.g., avoiding conversation and pretending to understand; (ii) Verbal Strategies, e.g., asking for statements to be repeated and explaining hearing loss; (iii) Nonverbal Strategies, e.g., positioning oneself to hear and to see a speaker’s face. Such an approach would be informative evaluating ALDs in regular classes.

Self-assessment is necessary with regard to the experience of advantages and disabilities. The used questionnaire describes the students’ different judgments on hearing aid and ALD use. The diversity in preference among students is more important than defining a winner. To verify the understanding of the questions, the questionnaire was carried out in interviews. In addition to the questionnaire, open-ended questions of mode preference added data that helped evaluate the students’ answers. In total, 25 students attending classes for the hard-of-hearing participated. Their aged ranged between 10-20 years and they were fitted bilateral where the degree of hearing loss (pure tone average of 500, 1k, and 2k Hz) varied between 32-93 dB HL. Accordingly, there was a great variety among subjects with regard to several factors, such as audiological data, ALD experience, and cognitive and academic development. Data for best-ear pure-tone average (PTA) were analyzed with regard to the questionnaire score; however, the number of subjects was too low to examine fully details with regard the differences described.
To account for some of the variability (e.g., the difference in ALD experience), the auralization technique was used in listening tests for a blind-comparison between different hearing aid and ALD setups. That is, a virtual classroom was created where the students could listen and rate a wide range of sound stimuli as if they were sitting in a classroom. An advantage with the auralization approach is also that new ALD designs can be easily evaluated. The questionnaire focused on several aspects for a few alternatives (the hearing aid only and the ALD only). The listening test focused on speech intelligibility but evaluated several hearing aid and ALD configurations:

- Paper II focused on the combined ALD and hearing aid microphone mode. Mixtures of a close-up microphone (the ALD) and a binaural model (the hearing aid) were assessed.
- Paper III focused on a binaural model for an ALD by comparing an omni-directional microphone with a binaural model in the same position.

Because students move and talk in various positions in the classroom, aspects of binaural hearing (bilateral advantage) and self-monitoring of the balance between the speech source (ALD) and the surrounding (hearing aid microphone) were considered important aspects to investigate in further detail.

In Paper II and III, binaural techniques were used both with regard to synthesizing a binaural listening experience and reproducing binaural stimuli. These topics were investigated in Paper IV and V by normal-hearing subjects. In Paper IV, correlations between localization and speech intelligibility were investigated. Because cross-talk cancellation was used as in Paper II and III, the study also presents data on the localization and speech intelligibility performance of the system. Further validation of the auralization and cross-talk cancellation was carried out in Paper V. The validation was important for the interpretation of the results in previous studies because of reported errors in reproduced interaural difference using cross-talk cancellation (Akeroyd et al., 2007).

3.2. Paper I: Self-assessment of assistive listening devices

The objective of Paper I was to analyze students’ use and preference of classroom ALDs (induction loop systems).

3.2.1. Method

The dimensions of the Speech, Spatial and Qualities of Hearing Scale (SSQ) (Gatehouse & Noble, 2004) were used to assess the ALDs in dimensions related to hearing and disability. The questionnaire was composed of five Speech items, four Spatial items and nine Other Qualities items. The questionnaire was completed during interviews and took place at the students’ schools. As well as open-ended questions on mode preference at the end of the questionnaire, additional notes were taken by the interviewer. Twenty-five bilateral fitted children participated and 18 complete answers were received: nine from children attending elementary and secondary school (age 10-15 years) and nine from children attending upper secondary school (age 16-20 years). Responses were collected for both the hearing aid microphone (M) and telecoil (T) mode. The use of ALDs varies in classes for the hard-of-hearing. The T mode was discussed in relation to an ALD featuring a short microphone distance, e.g., wireless (hand-held, clip-on, or headset) and linked table microphones. The questionnaire scores were analyzed with respect to the students’ PTA.

3.2.2. Results and discussion

Speech hearing: T mode was advantageous when compared to M mode in difficult speech tasks and the difference was more prominent for subjects with severe hearing loss.

Spatial hearing: The spatial ability was rated higher in M mode when compared to T mode, both with regard to localization and to the externality of sounds. No correlations were found between the spatial hearing items and subjects’ PTA.
Other Qualities: The ability to segregate sounds was correlated to the spatial hearing items and was rated better in M mode than in T mode. A difference between modes was also found for listening effort. Less listening effort was reported using T mode. The scores for listening effort in M mode were correlated with PTA, a more severe hearing loss led to higher listening effort. No such correlations were found when using T mode.

Reviewing the results, two attributes were identified:

- Audibility: To hear, and less effort put forth to hear, what is being said.
- Awareness: To hear, locate, and segregate sounds in the environment around one self, e.g., other students not using external microphones.

T mode was better in terms of audibility. One reason for the preference of T mode was a better ability to focus on what a person says. Students with severe hearing loss benefit more using T mode when compared to the better hearing students. Better awareness was achieved in M mode. The students’ ratings of the importance of the different items indicated that Speech hearing and listening effort were more important to their choice of which mode to use than Spatial hearing and the ability to segregate sounds. A common reason to use M mode instead of T mode was also the possibility to hear what someone nearby (not using a microphone) is saying.

Whereas T mode received higher scores than M mode for items related to one-to-one and group conversation, the score for participating in a conversation that switches quickly were equal. The result indicates a loss of binaural advantages when using T mode. With regard to awareness, the result supports that ALDs decrease binaural advantages with regard to localization and sound segregation.

3.3. Paper II: Assessments of hearing aid microphone and FM/telecoil mode combinations

In Paper I it was concluded that the use of ALDs increase audibility but decrease awareness. Paper II investigates the combination of the hearing aid microphone (M) mode and the ALD (the T/FM mode). In a listening experiment, mixtures of an omni-directional microphone were compared with a binaural microphone model. To gain more insight in the results of the listening experiment, students’ hearing strategies were studied as well as ratings of their experience of different hearing aid mode combinations.

3.3.1. Method

Eleven students attending upper secondary school (age 16-20 years) participated. The listening experiment assessing speech intelligibility was carried out in a classroom at the students’ school. During the listening tests, the subjects wore their hearing aids and stimuli were reproduced using loudspeakers and cross-talk cancellation. The subjects compared an auralization of an omni-directional microphone placed 0.2 m in front of the target speech source and a binaural model at the listening position (1.5 m). Two speech masker sources were positioned at either side behind the listening position. Twelve different mixtures of M (signals generated by the binaural model) and T (signal from the close-up microphone) were compared in two tasks (liberally translated from Swedish):

a) Give 100 points to the sound that you most easily can hear what the female voice is saying. Then give the other sounds points. Lesser points the more difficult you think it is to hear what the female voice is saying.

b) Give 100 points to the sound you like the most. Then give the other sounds points. Lesser points the worse you think the sound is.

The subjects also self-assessed speech intelligibility and listening effort using ALDs in their classroom situations. The questionnaire data were collected as a part of Paper I. The items relate to
the three classroom settings: 1) listening to a student standing at the board using a wearable microphone; 2) a one-to-one conversation in a classroom with linked table microphones; and 3) a group conversation round a table where the students use wearable microphones. Each setting had three background noise conditions: a) quiet classroom; b) continuous noise, e.g., fan noise; and c) competing speech. In the same settings, the subjects also described which hearing aid mode they used at either ear.

3.3.2. Results and discussion

The scores in the listening experiment confirmed the advantage of a short microphone distance (T mode) with regard to the audibility of a single talker in presence of masking speech. The same result was shown both with regard to task (a) ability to follow speech and to task (b) sound preference. In this context, the subjects preferred the stimulus with highest intelligibility. The preference of T mode was also shown when students in the questionnaire rated the different hearing aid modes in the nine classroom situations. With regard to both speech intelligibility and listening effort, the difference between modes (M and T) is larger in the Talk distance setting than in the Talk group setting. Previous studies have confirmed the importance of listening distance. For example, Boothroyd (2004) found that FM systems were most advantageous when listening to one speaker at a long distance. The effect of masking speech was more prominent for M mode when compared to the other mode combinations. Noisy classrooms also have a more deteriorating effect on speech intelligibility and listening effort for students with a severe hearing loss. However, just a few students used both hearing aids in T mode. Instead the students preferred to use a combined mode (M+T mode or M and T mode in either ear). Figure 7 show the students' hearing aid mode use for the different conditions. The overall conclusion is that the students’ have strategies for hearing: their hearing aid mode use depends on the classroom setting and type of background noise. Students with severe hearing loss preferred to use T mode to a greater extent than the better hearing students.

![Figure 7](image-url)

**Figure 7.** Subjects hearing aid mode use for the different items and background noise conditions. Item 1 *Talk distance*, Item 2 *Talk one-to-one*, and Item 3 *Talk group*. Background noise conditions: a) quiet classroom; b) continuous noise; and c) competing speech. The y-axis shows subjects’ best ear PTA (* denotes interaural asymmetry).
3.4. Paper III: Effects on speech intelligibility using a binaural assistive listening device

Since ALDs generate a monophonic signal, the opportunity to gain binaural advantages is reduced when using the hearing aids in the telecoil (T) or FM mode. This study evaluates the effects on speech intelligibility using a binaural assistive listening device. A word recognition test was performed comparing auralizations using an omni-directional and a binaural receiver model.

3.4.1. Method

Twelve students (10-15 years old), fitted bilaterally, participated in the study. Hagerman (1984) sentences were presented together with two continuous speech maskers. The speech sounds were convolved with impulse responses generated using room acoustic modelling software (CATT-Acoustics). The room (178.5 m³) had a reverberation time of 0.3-0.5 s. The target source and the listening position were placed on opposite sides of a table at a distance of 1.3 m and at 0 degrees azimuth. The two masker sources were positioned at a distance of 2.6 m and at ±60 degrees azimuth. An omni-directional and a binaural receiver model were used, which resulted in a comparison between diotic and dichotic presented stimuli.

The subjects listened with their hearing aids, and the stimuli were presented, using cross-talk cancellation technique, by two loudspeakers in front of the listener. The listening experiment was carried out in a small sound-isolated room. After a training session, ten sentences were presented at a 5 dB signal-to-noise ratio. Percent correct answers were collected. After the last sentence, the subjects rated their effort required to recognize the words on an eleven-point scale.

3.4.2. Results and discussion

No difference between the omni-directional and the binaural receiver model was found. It has been argued that diotic summation is more relevant concerning bilateral advantage than effects due to dichotic listening (Day et al., 1988). The reason is the reduction in dichotic effects, especially for individuals with a severe hearing loss. In a study by Cox and Bisset (1984), hearing-impaired subjects who obtained a 3 dB binaural advantage could not distinguish a diotic presented stimuli from a dichotic presented stimuli, which normal hearing subjects managed to identify. Perceived listening effort was also equal for the two cases; however a correlation was found to the students’ best-ear pure-tone average. In order to draw any conclusion regarding the advantage using a binaural ALD, the following strategies are considered: (i) increase the spatial separation of the speech and the noise maskers; (ii) use an adaptive method to the word recognition test to decrease variability; and (iii) assess cognitive load or memory.

3.5. Paper IV: Self-assessment of speech intelligibility listening to binaural recordings

The use of binaural techniques to assess localization and speech intelligibility in multi talker situation were investigated in Paper IV. Artificial head recording and in-ear recordings of humans were compared in two listening tests.

3.5.1. Method

The two listening tests were performed with 20 normal hearing subjects. All recordings were done with small microphones placed either in the ears of the artificial heads or the humans used for the recordings. The recordings were reproduced in an anechoic chamber through loudspeakers using cross-talk cancellation.
Localization. The recordings for the first listening test were done with two artificial heads (Head Acoustics HMS I and HMS III) and in the ear of six humans. The first set of recordings used was made in the horizontal plane in steps of 15 degrees and the subjects were instructed to judge the direction in the horizontal plane.

Speech intelligibility. The recordings were done with the same artificial heads and in the ears of three of the humans used in the first set of recordings. The set-up is presented in Figure 8. The target speech (male voice) was played in either loudspeaker 1 or 2 and the masker (female and male voice) talkers were played in one of the loudspeakers numbered 2 to 5. The level of all talkers were adjusted to have the same A-weighted sound pressure level, resulting in the level of the maskers being approximately 3 dB higher than the target. In the listening test, the subjects were instructed to assess their ability to follow the target on a scale from 0 (not at all) to 100 (extremely difficult).

![Figure 8](image)

Figure 8: The experimental setup for the recordings (and real-life version) used for the speech intelligibility assessment in the study. The non-filled loudspeakers were placed in the anechoic chamber but were not used. Loudspeaker 6 was used for binaural reproduction but not for recording.

### 3.5.2. Results and discussion

From the localization test, it was found that there was a significant difference between the human head recordings. No difference was found between the artificial heads and the human recordings. With regard to speech intelligibility, one human head was significantly better than the Head Acoustics HMS III artificial head. A "good" head is more advantageous than artificial heads because outer-ear details are lost when using artificial heads, a finding that agrees with previous studies (Møller et al., 1986, 1999). When comparing the human head recording with best and worst localization performance, a difference in speech intelligibility was observed. No such correlation could be found for the rest of the recordings of the human and artificial heads.

Front-back reversals in the localization test were correlated with speech intelligibility for the human head with worse performance. It was concluded that the localization performance (including front-back errors) is difficult to use as a predictor for speech intelligibility in situations similar to a cocktail party. To address the errors in reproduced interaural difference using cross-talk cancellation (cf. Paper V), the study should be repeated using equalized headphones.

In Paper II, III, and IV the stimuli were presented using a two-loudspeaker cross-talk cancellation setup. Paper V investigates binaural speech intelligibility using cross-talk cancellation where headphone reproduction was used as a reference. Any differences in reproduction technique were discussed with regard to reproduced interaural differences in time (ITD) and level (ILD) and data from the talker localization test in Paper IV.

3.6.1. Method

The experiment contained two sessions: free-field and room. Twenty-three normal hearing subjects participated in the experiment, which was carried out in an anechoic chamber. In each session, both headphone and cross-talk cancellation reproduction was tested – within subject design. In the room session, subjects were either assigned the auralization or the artificial head recording – a between subject design. The room (271 m³) had a reverberation time of 0.7-0.8 s and the listening distance was 2.5 m. The speech-reception threshold in noise (SRT) was measured using the Hagerman sentence test (Hagerman, 1982; Hagerman & Kinnefors, 1994). Two setups were used: the speech and the noise were presented in front (S0N0), and the speech was presented in front and the noise was presented at the right side (S0N90). The response variable was the difference in SRT between S0N0 and S0N90. For headphone reproduction, a third monaural setup was also performed. The S0N90 setup was compared to the case where the ear attending the noise was blocked.

3.6.2. Results and discussions

The cross-talk cancellation system was measured to reproduce adequate ITDs for frequencies between 200 Hz and 1 kHz (Figure 9). Deviations of ILDs were found at low frequencies, especially below 400 Hz where higher levels were measured at the contralateral ear to the source (Figure 10). At frequencies above 4 kHz, envelope ITDs were inaccurately recreated. These errors add to the high frequency distortion using non-individualized head-related transfer functions (HRTFs), e.g., causing high rates of front-back reversals. The envelope ILDs were also delivered inaccurately at high frequencies, about 6 dB above 5 kHz. Hence, the head shadow effect was reduced.

The reported errors in delivered ITDs and ILDs by the cross-talk cancellation system reduced (compared to headphones) the advantage gained when the speech and the noise is spatially separated from 9.3 dB to 5.8 dB in the free-field condition and from 3.0 dB to 1.5 dB in the room condition. The results for headphone reproduction agreed in the free-field condition with previous studies in both virtual and actual free-field (Bronkhorst, 2000; Wagener et al., 2007). The results in the room condition can be considered to be lower when compared to listening in the actual room (Bronkhorst & Plomp, 1990; Nabelek & Pickett, 1974b) and similar to a reverberant condition synthesized and reproduced with headphones (Koehnke & Besing, 1996). No differences were found between the auralization and the artificial head recording.

The findings show that binaural cues are not fully recreated using cross-talk cancellation, limiting the benefit of binaural advantage in speech intelligibility. Localization data (Paper IV) showed performance similar to that of headphone listening. This finding emphasizes that localization performance in the horizontal plane is not a sufficient indicator for evaluating the binaural performance of a reproduction system.
Figure 9. Delivered interaural time differences (ITDs) and interaural level differences (ILDs) of the cross-talk cancellation system at 90 degrees azimuth as a function of auditory-filter frequency. (o) target stimuli and (*) artificial head recording in the listening position.

Figure 10. Delivered interaural time differences (ITDs) and interaural level differences (ILDs) carried by the signal envelope of the cross-talk cancellation system at 90 degrees azimuth as a function of 1/3 octave band frequency. (o) target stimuli and (*) artificial head recording in the listening position.
4. Discussion

4.1. Speech quality features of ALDs

Speech quality is the result of the assessment of all recognized features of a speech sample (Jekosch, 2005). In Paper I, assistive listening devices (ALDs) were assessed by a questionnaire. Based on the students’ ratings of the importance of the questionnaire items, four features of sound quality were identified: listening effort, speech intelligibility, distinction of sound, and awareness.

Listening effort

When the students rated the importance of which mode to use, listening effort was rated the most important feature to consider. A correlation between the questionnaire score for the hearing aid microphone (M) mode and best-ear pure-tone average (PTA) was also found. No such correlation was found for the hearing aid telecoil (T) mode. In the word recognition test in Paper III, perceived effort was correlated with PTA, although no correlation with PTA was found for percentage correctly repeated words. The result emphasizes the importance of assessing listening effort, a finding that agrees with Hällgren et al. (2005). In the study by Gatehouse and Noble (2006), listening effort was correlated with PTA and the experience of handicap. The correlation with PTA and handicap can be a disadvantage when assessing different ALD solutions. That is, in a group of students where hearing loss varies from mild to profound, the experience of handicap adds variability to the data when assessing ALDs (cf. Paper III).

Speech intelligibility

Another item with high importance was speech intelligibility in a group conversation setting. No significant difference between modes was found in the group setting; however, a difference was found in a one-to-one conversation setting. In Paper II, these settings were further investigated. The difference between the M and T mode is greater when listening at a long distance as when compared to a group discussion. The result agrees with a study on FM system by Boothroyd (2004), where speech intelligibility was assessed by phoneme recognition in noise.

The greatest speech intelligibility advantage using T mode was found in the presence of a speech masker and this difference was greatest for students with severe hearing loss. This result agrees with word recognition studies where the difference between a noise masker and a speech masker was attributed to a lessened ability to take advantage of spectral and temporal dips (e.g., Hygge et al. 1992; Moore, 2003). In the word recognition test (Paper III), correctly repeated words were measured in the presence of a speech masker for dichotically presented stimuli. This ability was correlated with sound segregation. No correlation was found between speech hearing items and the word recognition scores. For speech intelligibility, it is important to consider both energetic and informational masking (Brungart, 2001; Brungart et al., 2001). Informational masking corresponds to masking effects not explained by the energies of the masker (energetic masking).

Distinction of sounds

The ability to identify students by voice and to judge the mood of other speakers from their voice was also important for the students. The items are related to the factor distinction of sounds defined by Kramer et al. (1996). Although important for how the student chose to use their hearing aids, the ability was rated the same for both the M and T mode. The ability – distinction of sounds – was correlated to items related to speech intelligibility in presence of a speech masker and listening effort. Furthermore, the items were correlated with students’ PTA and to the word recognition scores (percentage correctly repeated words and perceived effort). As discussed by Gatehouse and Noble (2004), difficulty identifying other people and their mood is correlated with the experience of handicap and a reluctance to engage in conversations.
Awareness

Awareness was defined as the ability to hear, locate, and segregate sounds in the environment around one self, e.g., other students not using external microphones. The importance ratings on the questionnaire as well as additional comments by the students revealed that the importance of the ability to locate and segregate sounds was low. As discussed by Dillon (2001), hearing-impaired individuals seldom complain about poor localization ability, although they are aware of their difficulties such as to locate people talking. The relationship between interaction and localization is described by Noble et al. (1995). They found that disadvantages of localization disability could lead to confusion of sounds, loss of concentration, and a desire to escape settings where this occurred. The importance of retrieving information about the environment is also emphasized by Kramer et al. (1995) and discussed with regard to FM systems (Dillon, 2001) and directional microphones in hearing aids (Stadler & Rabinowitz, 1993). An aspect of awareness is also the feeling of presence that is discussed as quality of the reproduced sound (Blauert & Jekosch, 2003). As presented in Paper I, the ability to externalize sounds was correlated to the rating of (un)naturalness.

4.2. Hearing strategies – The combined ALD and hearing aid microphone mode

Many hearing aids feature a combined mode where the signals from the hearing aid microphone and ALD (telecoil or FM) are mixed, denoted the M+T/FM mode. For bilateral hearing aid users, a possible way to combine M mode and T/FM mode is to use different modes for each ear. In Paper I, different advantages using M mode and T mode were described. T mode was more useful with regard to speech intelligibility and listening effort (audibility). M mode was more useful with regard to being aware about things happening in the environment around one-self. The result in Paper II does support the conclusion that a combined M and T/FM mode is a feasible solution. A combined mode (also using one hearing aid in M mode and the other in T/FM mode) increases speech intelligibility when compared to using both hearing aids in M mode. Word recognition improvements using M+FM mode when compared to M mode have also previously been shown (e.g., Boothroyd & Iglehart, 1998).

The dilemma between audibility and awareness is affected by the auditory scenario in the classroom. In Paper I, some students reported a preference of M mode in noisy classrooms, whereas others preferred T mode due to better ability to ignore competing sounds. The data on students’ use of different hearing aid modes (Figure 7) revealed that increased amount of T mode (audibility) was preferred in the masking speech conditions. Students with more severe hearing loss used T mode to a greater extent than the better hearing students. A combination of M and T mode offered improved audibility with preserved awareness when compared to only using M mode. However, audibility deteriorates when adding signals from the hearing aid microphones in difficult listening situations and students with severe hearing loss only used T mode. This conclusion supports findings in previous studies (e.g., Hawkins, 1984). Students who did not use the combined mode noted that sounds were often confusing and that speech sounded unnatural (Paper I).

The compromise between M and T/FM mode has also been addressed by Lewis and Eiten (2003). They showed that the signal from the external microphone was preferred relative to the hearing aid microphone when listening to the teacher. The hearing aid microphone signal was preferred when listening to others (not the teacher) in the classroom. In Paper I, the students described two communication channels in the classroom. The ALD was used for conversations with regard to teaching and M mode was used for non-teaching conversations. The observation is of interest in a learning environment where peer-to-peer interaction is emphasized, because a clear distinction between teaching and non-teaching conversations is difficult to make.

An automatic switch in the FM system has been suggested as a solution to the dilemma between audibility and awareness (Dillon, 2001). The FM mode is selected while transmitting otherwise the
M mode is selected. These systems are referred to as Speech-operated switching (SOX), voice-operated switching (VOX), FM priority, or FM precedence. In wireless multi-microphone systems in classrooms, such priority systems also exist. The teacher’s microphone has priority over the microphones used by the students.

In Paper I, ALDs with a short microphone distance were compared to centrally placed devices, e.g., devices placed on a table between users. Moving the microphone from the speaker’s mouth to a table is intended to facilitate peer interaction (Wennergren, 2004; cf. Section 1.2). The preferences varied among subjects and the answers can be interpreted as a parallel to the discussion on audibility (short microphone distance) and awareness (long microphone distance).

From the students’ hearing mode preference (Figure 7), the following can be concluded: (i) the students frequently change their hearing aid mode use depending on the situation; and (ii) the choice of mode use varies between aspects, where degree of hearing loss is one factor. Clearly, the ability to make adjustments is important. Hence, a combined mode is useful with regard to a group-based solution (long microphone distance) and to a FM priority system.

4.3. Binaural advantages

The sections above focus on the students’ ability to hear what is happening around them in the classroom. Furthermore, in a classroom where group work and peer-to-peer interaction are stressed, spatial hearing aspects are important considerations, aspects that include improved listening with two ears instead of only one (binaural advantages). ALDs provide a monophonic signal; using the hearing aid in T/FM mode will result in that the same signal is perceived in both ears (diotic listening). In Paper III, a binaural ALD was investigated: the ALD was replaced with binaural model. No benefits of the binaural ALD were found. The result agrees with the study by Cox and Bisset (1984). There are practical problems when using FM or induction loop (IL) to transmit a separate signal to either hearing aid for bilateral-fitted students. The introduction of digital radio transmission, a solution is achievable.

The result in Paper I supported that ALDs decrease binaural advantages with regard to localization and sound segregation. However, binaural advantages with regard to speech intelligibility were not as evident. M mode and T mode were equally rated in a group conversation that switches quickly, whereas larger difference where found in a one-to-one conversation. As discussed for Paper II, the difference between M mode and T mode was more prominent listening to a person at a long distance than in a group conversation. Rather than a loss of binaural advantages using T mode, the higher ratings of M mode in group conversation can be explained by listening distance.

4.4. Validation of the auralization process

The listening tests in Paper II and III used room acoustics modelling and auralization to render a classroom listening situation. To present binaural stimuli to the subjects, two loudspeakers and cross-talk cancellation was used.

Auralization

In Paper V, an auralization and an artificial head recording were compared. No difference between the two methods was found measuring sentence word recognition scores in normal-hearing listeners. A more thorough validation of the auralization technique in classrooms (400 m³) was performed by Yang and Hodgson (2007). The authors conclude that if the room is not too absorptive or noisy, auralization is reliable. A problem in room acoustic modelling is accounting for diffusion. The scattering coefficients are often adjusted to match the model with measurements (as performed in Paper V and by Yang & Hodgson, 2007). This is a significant problem when modelling acoustic treatments or rooms not yet built. In Paper II and III, the rooms only existed as a model. The scattering coefficients were set to match the predicted reverberation time based on the
RTC (randomized cone-tracing) with reverberation time calculated according to the room dimensions and absorptions (the Eyring and Sabine formula) (see also Dalenbäck, 1995). In a room with high absorptions, the early reflections dominate over the late reflections. Early reflections are predicted using the image source model, a method that has limitations handling diffuse reflections (cf. Section 2.1.3). The predicted impulse response has less diffused reflection that will increase clarity (C50) and speech intelligibility when compared to a real room experience (Yang & Hodgson, 2007). In a room with greater amount of late reflections, the performance in the virtual room will also differ; however, the performance will be lower when compared to listening in the actual room (Yang & Hodgson, 2007). This difference is due to the lack of binaural advantage and dereverberation listening with non-individualized head-related transfer functions (HRTFs) (cf. Section 2.3). A noise source in a room can be viewed in terms of both energetic and information masking. The importance of information masking is emphasized with regard to spatial separation (Arbogast et al., 2002). In a room with short reverberation, the errors of lack of diffuse reflections would not affect spatial separation other than high frequency errors due to non-individualized HRTFs. In a reverberant room, energetic masking is assumed to be correctly predicted; however, the errors due to reduced binaural advantages will decrease the ability to separate the sources spatially. The difference between predicted and on-site listening found by Yang and Hodgson (2007) can be explained by higher informational masking using auralization. Artificial head recordings will introduce the same errors as an auralization in a noisy reverberant environment (Paper V).

**Cross-talk cancellation**

The approach used for binaural reproduction – loudspeakers and cross-talk cancellation – decreases advantages due to binaural hearing in speech intelligibility tasks when compared to headphone reproduction (Paper V). In Paper IV, self-rated speech intelligibility was also lower for the cross-talk reproduced stimuli than for the on-site setup (the talkers were reproduced in the loudspeakers in same way as they were binaurally recorded). Implications for Paper II and III are that the possibility to gain binaural advantages using the binaural model (hearing aid microphone mode) when compared to the omni-directional model (T/FM mode) is reduced using cross-talk cancellation. In Paper II, the listening test was performed in a classroom, which also decreases the performance of the cross-talk due to room reflections (Sæbø, 2001).

The ability to localize a sound source using cross-talk cancellation is similar to that of using headphones (Paper IV; Gardner, 1997). With regard to binaural speech intelligibility assessed in Paper V, the cross-talk cancellation systems used also showed similar performance with regard to binaural speech intelligibility compared to other studies. The advantage of spatial release from masking in free-field was 5.8 dB for cross-talk and 9.3 dB for headphones. Winkler et al. (2009) found an advantage of 8 dB for cross-talk and 11 dB for headphones.

The reproduced interaural difference of the cross-talk cancellation system was accurate within 300-4k Hz (Figure 9 and 10). With regard to presenting stimuli to hearing-impaired subjects wearing hearing aids, these errors, especially in high frequency, can be assumed to have less of an effect when compared to normal-hearing subjects.

**4.5. Generalization of the results**

The studies reported were limited to room induction loop (IL) systems in classes for the hard-of-hearing. The IL systems used are combined systems because FM is used to transmit the signal from the external wireless microphones to a stationary receiver and a room IL is used to couple the signal to the hearing aid.

The result discussed can be generalized regardless of using IL or FM to couple the ALD signal to the hearing aid. The dilemma of audibility and awareness is rather a question of the design and use of the external microphones than a question of the technique used to couple the signals to the
hearing aids. The dilemma has also been addressed with regard to FM systems (e.g., Lewis & Eiten, 2003). Furthermore, FM and IL systems have been shown to provide equal word recognition benefits (Nabelek et al., 1986; Noe et al., 1997). Aspects such as convenience and consistency can be discussed as to which systems to prefer (Dillon, 2001). These aspects were not addressed in this thesis; however, these aspects can be important for the overall impression and affect self-rated qualities.

There are differences between a class for the hard-of-hearing and a regular class. A class for the hard-of-hearing is smaller and the classmates also have a better understanding of hearing loss. Despite this, the studies reported showed how students have to choose between better audibility using ALD and better awareness using only the hearing aids. It can be expected that the differences in these attributes are larger for hearing-impaired students in a regular class.
5. Conclusion

Sound quality is the result of an assessment of the adequacy of sound. With regard to classroom assistive listening devices (ALDs), important features of sound quality recognized by the students were listening effort, speech intelligibility, distinction of sound, and awareness. These features must be considered when evaluating ALDs. A combined approach based on questionnaires and listening experiments is a promising tool to determine which design will best benefit the students.

When students use ALDs, the detrimental effects of late room reflections and competing sound sources are reduced by moving a microphone close to the mouth. By moving the microphone closer to the speaker’s mouth, the students improve their speech intelligibility and they use less listening effort; that is, audibility is increased. Distinction of sound is the ability to recognize additional characteristics of a speech sample, such as the ability to identify students by voice and to judge the mood of speakers from their voice. The feature was rated important as to which hearing technology students prefer, but was not found to differ between using an ALD and only using the hearing aids. Awareness – the ability to hear sounds in the environment around one self – is reduced using ALD, this strategy makes the students feel detached from their surroundings. Students also speak of two communication channels in the classroom. The first is transmitted over the induction loop or FM system and is identified as the teaching channel. The second is acoustically transmitted and is used for non-teaching conversations. Awareness is also the ability to take an active part in both these channels. The finding has implications for teaching in a learning environment where peer-to-peer interactions are emphasized, because there is no such clear distinction between teaching and non-teaching conversations.

The students are active in their use of different hearing aid modes, indicating that they have effective hearing strategies. Students prefer to use a combined mode where the signals from the ALD and hearing aid microphone are mixed or they combine them by using different modes for each ear. The combined mode facilitates both audibility and awareness. In difficult listening conditions, the signal from the hearing aid microphone masks the signal from the ALD. To maintain audibility, students with severe hearing loss prefer to use only the ALD; however, using this strategy means they may be left out of parts of the communication in the classroom. Binaural hearing was not found to be an important feature when choosing between hearing aid modes.
6. Further Research

Research on assistive listening devices (ALD) has focused on word-recognition benefits associated with close-up microphones. ALDs offer significant improvement for students with hearing loss listening at long distances. In a classroom, auditory requirements change as teachers change their pedagogical approaches. These changes can take place throughout a school day. Some auditory scenarios make using ALDs a disadvantage when compared to using only hearing aids. The problem can be assessed in a number of ways. In this thesis, for example, students completed a questionnaire that asked them to assess their experience using ALDs and hearing aids. Because the students attend classes for the hearing-impaired, a starting point for further research would be to repeat the studies for students with hearing loss in regular classes. What features of sound quality do they assign ALDs and what balance between audibility and awareness do they prefer? In such a design, it is important to consider the large varieties in classroom environments, ALD technology, and audiological data. This approach would also be beneficial if hearing strategies were investigated, such as explored in the Communication Profile for the Hearing Impaired (Demorest & Erdman, 1987; Hallberg et al., 1992). This approach is descriptive; a follow-up listening test is a reasonable next step.

Auralization techniques are a feasible tool to render audible different classroom conditions. There are many design and response variables to consider in such an experimental design. In a listening test, a topic that needs further assessment is the combined telecoil/FM and hearing aid microphone mode. In addition to speech intelligibility, interesting response variables are aspects of cognitive load and memory. Furthermore, judgements of sound quality (e.g., fullness, loudness, brightness, softness, nearness, spaciousness and clarity; Gabrielsson & Sjögren, 1979) are needed. Bilateral fitting improves sound quality (Balfour & Hawkins, 1992). Whereas, confusion of sounds and an unnatural sound were reported in Paper I for the combined telecoil and hearing aid microphone mode. Two approaches are identified for future studies: (i) an auralization variant of a questionnaire (e.g., the one described in Paper I) where the items are auralized rather than visualized and explained in text; and (ii) a test platform where students are asked to adapt aspects of room acoustics and ALD configurations (cf. Wingstedt, 2008). The approaches require accurate reproduction of binaural stimuli to subjects wearing hearing aids. Cross-talk cancellation was addressed in this thesis; however, future studies are needed before the approach can be considered effective. For example, a study could redo the evolution of cross-talk cancellation with regard to binaural speech intelligibility (Paper V) using a limited frequency range.

Although questionnaires and listening tests are useful to evaluate ALD solutions, good acoustic predictors are important for the development of new technologies. Several measures exist that have been addressed with respect to hearing aid evaluation and room acoustic quality, e.g., the Speech Intelligibility Index and the Speech Transmission Index. It would be informative to include such measures in the experimental design. Quality aspects for telecommunication speech services (Möller, 2005) also provide models useful for the evaluation of ALDs. The models include usability, which is not explored in this thesis. Because this study emphasizes how ALDs influence student interaction, future studies could benefit from both qualitative and quantitative approaches that assess collaborative aspects for different ALD designs (cf. Törlind, 2002). Hearing aids offer a number of different programs, such as combined microphone and telecoil/FM modes. An interesting topic for further research is also to explore feasible user-interface designs, e.g., when students have the opportunity to set individually the balance of the combined mode in a classroom environment.
7. References


Paper I
Self-assessment of classroom assistive listening devices

Johan Odelius, Örjan Johansson
Division of Sound and Vibration, Department of Human Work Sciences
Luleå University of Technology

Abstract
Self-assessment of classroom assistive listening devices (ALDs) based on induction loop systems was carried out in Swedish classes for hearing impaired students. A questionnaire was developed and completed by 25 students (bilateral hearing aid users, 10-20 years old). Responses for hearing aid microphone mode (M) and telecoil mode (T) were collected. Two attributes, audibility and awareness, were identified and assigned to either mode. Better audibility was achieved in T-mode. Students with severe hearing loss benefited more using T-mode when compared to the better hearing students, especially in more difficult listening situations. Better awareness was achieved in M-mode; students could better hear, locate and segregate sounds in the environment around them. Depending on the situation, students make different choices between audibility and awareness.
Self-assessment is a promising approach for determining what combination of ALD design and function that will best benefit the students.

Keywords: Assistive Technology, Hearing Aid Satisfaction, Psychoacoustics/Hearing Science, Speech Perception

1 Introduction
To assist communication in classes for hearing-impaired students, external microphone systems (assistive listening devices or ALDs) are a standard tool (Ross, 1992). The sound picked up by an ALD is transmitted to the students’ hearing aids, most often using a FM system or an induction loop (IL) system (e.g. Holmes & Saxon, 2000; Ross & Bakke, 2000). Using a switch on the hearing aid, the students can either listen to the signal from the hearing aid microphone (M) or the signal from the ALD received by a telecoil or FM receiver (T/FM). Many hearing aids also offer an additional mode where the signals from M and T/FM are mixed. In Sweden, IL systems are the dominant tool for ALD transmission in classes for hearing-impaired students.

ALDs improve signal-to-noise ratio by decreasing microphone distance and have been shown to increase speech recognition (Hawkins, 1984; Ross, 2003; Boothroyd, 2004). Both FM and IL systems provide equal benefit (Nabelek et al., 1986; Noe et al., 1997). ALDs have been developed for the context of a single talker. The greatest advantage has also been shown to be with one speaker at a long distance from the listeners (Boothroyd, 2004). In Sweden, there has been an effort to promote a more participatory learning setting in classes for hearing-impaired. The ALD’s ability to handle peer interaction was emphasized (Wennergren, 2004). Questions as to how students use these ALDs made it apparent that there is a need for evaluation and development of these communication systems.
To measure how hearing abilities affect daily living, self-assessment is preferable when compared to pure tone averages and speech recognition scores (Schow & Gatehouse, 1990). Compared to intelligibility measures, sound quality measures have been shown to be more sensitive in distinguishing between sound-reproducing systems (Gabrielsson et al., 1988). That is, sound quality can be more important to the students as to which system they prefer as when compared to speech intelligibility. To measure hearing disabilities, Gatehouse and Noble (2004) have developed a questionnaire: Speech, Spatial and Qualities of Hearing scale (SSQ). The SSQ is divided into three sections: Speech hearing, Spatial hearing and Other Qualities. Speech hearing covers various listening situations of different difficulty levels. There are also questions that relate to binaural functions, e.g. the ability to listen to several speech streams simultaneously. The Spatial hearing section concerns the ability to locate sound events and perception of sound in space. The Other Qualities category contains issues concerning segregation of sounds, recognition, listening effort and sound quality judgments, e.g. naturalness.

SSQ has been used to assess hearing aid benefits and bilateral advantage (Noble & Gatehouse, 2006). The authors found that two hearing aids are advantageous in difficult multi speaker situations and of lesser importance in more simple listening situations. Bilateral advantage was also found with respect to localization and listening effort. Since ALDs today provide a monophonic signal, any advantages from dichotic listening cannot be expected (Dillon, 2001). Therefore, aspects of binaural hearing are important to consider when evaluating ALDs in a more complex auditory scenario, as in a peer-to-peer setting.

The objective of this study is to analyse hearing impaired students’ use and preference of classroom ALDs (IL systems). A specific interest was to compare the preference of T-mode versus M-mode in situations where bilateral advantage can be expected. A self-assessment approach based on a SSQ related questionnaire design was used. The SSQ concept was considered appropriate for the evaluation of reproduction systems for hearing impaired as well as to assess bilateral advantage. In addition to the questionnaire, students were asked which mode they preferred in different scenarios. The questionnaire scores were further analysed with respect to the students’ pure tone average, which can vary from mild to profound in Swedish classes for hearing impaired.

2 Method

For the evaluation of ALDs (IL systems), a questionnaire was designed. The questionnaire was used in two studies, Study I and Study II. Study I was carried out first and evaluated hearing aid microphone (M) mode and two different types of ALDs (T-mode). The study did not ask for responses on the M+T-mode, where the signals from the microphone and telecoil are mixed. A further possibility for bilateral hearing aid users is to use one hearing aid in M-mode and the other in T-mode. In Study II, both these alternatives (denoted MT-mode) were evaluated in addition to M-mode and T-mode. This paper will focus on the results of the evaluation of M-mode and T-mode using a type of ALD common in both studies.

2.1 Subjects

In total, 25 students from schools in Sweden that have classes for hearing impaired participated. In Study I, 15 subjects participated. Their age ranged from 10-15 years (elementary and secondary school children). Of the total, 13 students came from the same school. In Study II, 10 students from a third school participated. Their age ranged from 16-20 years (upper secondary school children). All subjects used bilateral hearing aids and their unaided best ear pure-tone average (PTA; average of the hearing levels at 0.5, 1, and 2 kHz) ranged from 32-93 dB HL (median 52 dB HL).
2.2 Questionnaire

The Speech, Spatial and Qualities of Hearing Scale (SSQ), consisting of 50 items, covers a wide range of contexts (Gatehouse & Noble 2004). For the evaluation of ALD, a new and shorter questionnaire based on the SSQ was designed. Items related to working situations in the classroom was chosen and rewritten to suit the context. The questionnaire consisted of 18 items (5 Speech items, 4 Spatial items and 9 Other Qualities items) and is described in Table 1 (liberally translated from Swedish). For each question, judgements of the qualities of M-mode and T-mode were made on a discrete eleven-point response scale ranging from 0 to 10. The different response choices are described in Table 2. The ALD was specified to be a product used with a short microphone distance, e.g. wearable and hand-held microphones or linked table microphones. The questionnaire also had a third response choice. Study I asked for responses for an alternative ALD, a device placed in between the users with an omni directional microphone, e.g. on a table similar to a teleconference phone. In Study II, the third response choice was the MT-mode. Subjects in Study II also rated the importance of each question in relation to their choice of which mode to use. The last question was open-ended; the subjects were asked to describe the situations they preferred to use for each of the three response alternatives.

The questionnaire was completed during interviews and took place at the students’ schools. In Study I, the interviews were led by the author (JO) and in Study II by the students’ audiologist. The audiologist was thoroughly instructed about the questionnaire. The interview approach was preferred and ensured that the context of each question was understood and additional notes were taken by the interviewer. A neutral interview style was used so as not to influence response choices. No data were collected regarding the subjects’ ALD experience. No measurements were further performed to verify the performance of the induction loops in the schools or the function of the participants’ hearing aids.

Word recognition test

Questionnaire items were correlated against word recognition scores. The word recognition test was performed by 12 of the subjects in Study I after they had completed the questionnaire. The listening test is here briefly described; a more detailed description can be found in Odelius et al. (2006).

Five word sentences (Hagerman, 1984) were presented together with two continuous speech maskers. The speech sounds were convolved with binaural impulse responses generated using room acoustic modelling software (CATT-Acoustics). The room measured 7x8.5x3 m and had a reverberation time of $T = 0.4-0.5$ s for $f \leq 1$ kHz and $T = 0.3-0.4$ s for $f > 1$ kHz. The target source and the listening position were placed on opposite sides of a table at a distance of 1.3 m and at 0 degrees azimuth. The two masker sources were positioned at a distance of 2.6 m and at ±60 degrees azimuth. An artificial head was used as binaural model and the directivity index of a singer (Marshall & Meyer, 1985) was applied to all sources.

The listening experiment was carried out in a sound-isolated room at the students’ school. The room measured 2x2.8x1.9 m ($T \leq 0.2$ s for $f \leq 250$ Hz and $T \leq 0.1$ s for $f \geq 500$ Hz). Subjects used their hearing aids in M-mode. A cross-talk cancellation setup with Lexicon MC-1 and two ADAM-S2A speakers were used to enable dichotic listening.

After a training session, ten sentences were presented at a 5 dB signal to noise ratio. Percent correct answers were collected. After the last sentence, the subjects rated their effort required to recognize the words on an eleven-point scale.
Table 1. Questionnaire design. Items are liberally translated from Swedish with the purpose to express the contents of each question. The wordings in Swedish were made with words intended to be understood by the students.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Rating Method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPEECH</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Can you hear what a student is saying standing at the board?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>1a</td>
<td>Can you hear what a student is saying standing at the board?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a) quiet classroom</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td></td>
<td>b) continuous noise (e.g. fan noise)</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td></td>
<td>c) competing speech</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>2</td>
<td>Can you hear what a student is saying in a one-to-one conversation?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>2a</td>
<td>Can you hear what a student is saying in a one-to-one conversation?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a) quiet classroom</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td></td>
<td>b) continuous noise</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td></td>
<td>c) competing speech</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>3</td>
<td>Can you hear what other students are saying in a three person group conversation round a table?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>3a</td>
<td>Can you hear what other students are saying in a three person group conversation round a table?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a) quiet classroom</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td></td>
<td>b) continuous noise</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td></td>
<td>c) competing speech</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>4</td>
<td>Can you listen to two students talking at the same time?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>5</td>
<td>Can you follow a group conversation without missing the start of each new speaker?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td><strong>SPATIAL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Can you decide where a sound event in a room is happening?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>7a</td>
<td>Can you decide where a sound event in a room is happening?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a) correct direction</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td></td>
<td>b) correct distance</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>8</td>
<td>When someone talks, can you without looking decide where that person is sitting?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>9</td>
<td>Can you determine the movement of someone walking in a room?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>10</td>
<td>Do you hear the sounds as if they were inside your head or as if they are in the classroom?</td>
<td>(0 In the head - 10 Out there)</td>
</tr>
<tr>
<td><strong>OTHER QUALITIES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Can you easily ignore other sounds and the sounds from others talking in the classroom?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>11</td>
<td>Can you segregate different sounds from each other?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>12</td>
<td>Can you immediately identify other students by their voices?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>13</td>
<td>Can you determine by another student’s voice if she/he is happy, sad, surly or tired?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
<tr>
<td>14</td>
<td>How much effort do you require to hear when listening to a student standing at the board?</td>
<td>(0 Extremely – 10 Not at all)</td>
</tr>
<tr>
<td>15</td>
<td>How much effort do you require to hear in a one-to-one conversation?</td>
<td>(0 Extremely – 10 Not at all)</td>
</tr>
<tr>
<td>16</td>
<td>How much effort do you require to hear in a three person group conversation round a table?</td>
<td>(0 Extremely – 10 Not at all)</td>
</tr>
<tr>
<td>17</td>
<td>Do you perceive the sound to be artificial or unnatural?</td>
<td>(0 Unnatural - 10 Natural)</td>
</tr>
<tr>
<td>18</td>
<td>Do you think the sound is pleasant to listen to?</td>
<td>(0 Not at all - 10 Perfectly)</td>
</tr>
</tbody>
</table>

*a* Not included in Study I  
*b* Not included in Study II  
*c* In Study II expanded with three noise background conditions: a) quiet classroom b) continuous noise c) competing speech.  
*d* In Study I only pertained to T-mode.
Table 2. Hearing aid mode response choices.

<table>
<thead>
<tr>
<th>Study</th>
<th>Mode</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I, II</td>
<td>M</td>
<td>Hearing aid microphone mode.</td>
</tr>
<tr>
<td>I, II</td>
<td>T</td>
<td>Hearing aid telecoil mode. ALDs used with a short microphone distance, e.g. wearable and hand-held microphones or linked table microphones.</td>
</tr>
<tr>
<td>I</td>
<td>T_{long}</td>
<td>Hearing aid telecoil mode. An ALD with a device placed in between the users with an omni directional microphone, e.g. on a table similar to a teleconference phone.</td>
</tr>
<tr>
<td>II</td>
<td>MT</td>
<td>Hearing aid M+T-mode, where the signal from the hearing aid microphone and the telecoil is mixed. or One hearing aid in M-mode and the other in T-mode.</td>
</tr>
</tbody>
</table>

2.3 Analysis

Of the total 25 subjects, 18 complete cases will be reported, 9 from each study. Missing answers were due to subjects’ lack of ALD experience. Items 1, 9, and 14 were discarded since the items were not included in both studies. In Study II, Items 15 and 16 (listening effort) had three background noise conditions: a) quiet classroom; b) continuous noise; and c) competing speech. For these two items, the scores were replaced by the average of the three background noise conditions.

The questionnaire data were analysed using descriptive statistics presented as median and interquartile range (IQR; difference between the 75th and 25th percentiles). Wilcoxon signed rank test with a significance level of $p<0.05$ was used to analyse differences between M-mode and T-mode. To see if the degree of hearing loss affected the questionnaire scores, the correlation between items and PTA was computed on ranked data (Spearman’s correlation). Higher PTA was predicted to correlate with lower questionnaire ratings; a one-tailed test ($p<0.05$) was selected. Spearman’s correlation was also used to correlate questionnaire item scores with the word recognition test. Higher word recognition scores were predicted to correlate with higher questionnaire ratings ($p<0.05$, 1-tailed).

Item intercorrelations (Spearman’s correlation) were calculated for the complete cases, collapsed across M-mode and T-mode. The questionnaire data (18 complete cases on 18 items) were analysed using a systematic observation approach because factor analysis was considered unsuitable.

3 Results

3.1 Questionnaire scores

Medians and IQR of questionnaire item scores are presented in Table 3. Response scales are shown in Table 1; rating 10 always indicates a greater ability. Significant differences between M-mode and T-mode are highlighted. The correlations between questionnaire scores and subjects’ hearing loss are shown in Table 4.

Table 3 also presents rated importance of each item in relation to the subjects’ choice of which mode to use (Study II). The importance was somewhat equal across items. Spatial items and sound segregation (Item 11) stood out with lower importance. Highest importance received listening effort (Items 15 and 16).
Table 3. Median item scores (IQR in brackets) by mode (M and T). Differences between modes are highlighted (Wilcoxon signed rank test, \( p < .05 \)). Right column shows median (IQR in brackets) of rated importance of each item in relation to subjects’ choice of which mode to use.

<table>
<thead>
<tr>
<th></th>
<th>M-mode</th>
<th>T-mode</th>
<th>Sign.</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPEECH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a) Talk one-to-one quiet</td>
<td>8 (3)</td>
<td>9.5 (1)</td>
<td></td>
<td>6 (3)</td>
</tr>
<tr>
<td>2b) Talk one-to-one noise</td>
<td>7 (3)</td>
<td>8 (3)</td>
<td></td>
<td>6 (3)</td>
</tr>
<tr>
<td>2c) Talk one-to-one speech</td>
<td>6 (4)</td>
<td>8 (2)</td>
<td>&lt;.05</td>
<td>6 (3)</td>
</tr>
<tr>
<td>3a) Talk group quiet</td>
<td>8.5 (2)</td>
<td>9 (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3b) Talk group noise</td>
<td>7.5 (4)</td>
<td>8 (2)</td>
<td></td>
<td>7 (3.5)</td>
</tr>
<tr>
<td>3c) Talk group speech</td>
<td>5 (4)</td>
<td>7 (4.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Follow two speech streams</td>
<td>4 (3)</td>
<td>6 (2.25)</td>
<td></td>
<td>5 (3.5)</td>
</tr>
<tr>
<td>5) Switch quickly conversation</td>
<td>8 (3)</td>
<td>8 (3.25)</td>
<td></td>
<td>6 (4.25)</td>
</tr>
<tr>
<td><strong>SPATIAL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7a) Locate direction</td>
<td>9 (2)</td>
<td>4 (3.25)</td>
<td>&lt;.05</td>
<td>6 (5.75)</td>
</tr>
<tr>
<td>7b) Locate distance</td>
<td>6 (3)</td>
<td>2 (3.25)</td>
<td>&lt;.05</td>
<td>5 (4)</td>
</tr>
<tr>
<td>8) Locate speaker</td>
<td>8 (2)</td>
<td>5 (4.25)</td>
<td>&lt;.05</td>
<td>4.5 (3.5)</td>
</tr>
<tr>
<td>10) Externality of sounds</td>
<td>9 (3)</td>
<td>5 (5.25)</td>
<td>&lt;.05</td>
<td>5 (1.5)</td>
</tr>
<tr>
<td><strong>OTHER QUALITIES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) Ignore interfering sounds</td>
<td>7 (4)</td>
<td>9 (4)</td>
<td></td>
<td>7 (4)</td>
</tr>
<tr>
<td>11) Segregate sounds</td>
<td>8 (2)</td>
<td>5 (4.5)</td>
<td>&lt;.05</td>
<td>5 (3)</td>
</tr>
<tr>
<td>12) Identify students by voice</td>
<td>9 (2)</td>
<td>9.5 (2)</td>
<td></td>
<td>7 (3.5)</td>
</tr>
<tr>
<td>13) Judge mood from voice</td>
<td>8.5 (3)</td>
<td>8 (2)</td>
<td>&lt;.05</td>
<td>7 (3.75)</td>
</tr>
<tr>
<td>15) Listening effort one-to-one</td>
<td>5 (4.25)</td>
<td>8 (4.25)</td>
<td>&lt;.05</td>
<td>8 (4.5)</td>
</tr>
<tr>
<td>16) Listening effort group</td>
<td>6 (4)</td>
<td>7 (2)</td>
<td></td>
<td>8 (3.5)</td>
</tr>
<tr>
<td>17) Unnatural</td>
<td>10 (1.5)</td>
<td>7 (2)</td>
<td></td>
<td>6 (1.25)</td>
</tr>
<tr>
<td>18) Pleasantness</td>
<td>7.5 (1.5)</td>
<td>9 (2.5)</td>
<td></td>
<td>6.5 (5)</td>
</tr>
</tbody>
</table>

**Speech hearing**

In the Speech section, T-mode was in general rated higher than M-mode. A significant difference was found for Item 2c *Talk one-to-one speech* (\( T=18.5, p < .05 \)). Comparing the results for the different noise backgrounds in Items 2 and 3 (Table 3), one can observe an increasing difference between M-mode and T-mode when the masker is speech (alternative c). As presented in Table 4, difficult speech tasks in M-mode were negatively correlated with PTA. No correlations between speech scores and PTA were found in T-mode. Figure 1 illustrates the result using a scatter plot for Item 3c *Talk group speech*. Interpretation of the result indicates that T-mode was advantageous in difficult speech tasks and that the difference was more prominent for subjects with severe hearing loss.
Table 4. The two columns show spearman rank correlations ($r_s$) between best ear PTA and questionnaire items for M-mode and T-mode, respectively. * indicates significant correlations ($p<.05$); correlations less than 0.3 are ignored.

<table>
<thead>
<tr>
<th></th>
<th>M-mode</th>
<th>T-mode</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPEECH</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a) Talk one-to-one quiet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2b) Talk one-to-one noise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2c) Talk one-to-one speech</td>
<td>-0.37</td>
<td></td>
</tr>
<tr>
<td>3a) Talk group quiet</td>
<td>-0.46*</td>
<td></td>
</tr>
<tr>
<td>3b) Talk group noise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3c) Talk group speech</td>
<td>-0.43*</td>
<td></td>
</tr>
<tr>
<td>4) Follow two speech streams</td>
<td>-0.72*</td>
<td></td>
</tr>
<tr>
<td>5) Switch quickly conversation</td>
<td>-0.56*</td>
<td></td>
</tr>
<tr>
<td><strong>SPATIAL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7a) Locate direction</td>
<td>-0.33</td>
<td>-0.30</td>
</tr>
<tr>
<td>7b) Locate distance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8) Locate speaker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10) Externality of sounds</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OTHER QUALITIES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) Ignore interfering sounds</td>
<td>-0.58*</td>
<td></td>
</tr>
<tr>
<td>11) Segregate sounds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12) Identify students by voice</td>
<td>-0.39</td>
<td></td>
</tr>
<tr>
<td>13) Judge mood from voice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15) Listening effort one-to-one</td>
<td>-0.44*</td>
<td></td>
</tr>
<tr>
<td>16) Listening effort group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17) Unnatural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18) Pleasantness</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Speech Item 3c Talk group speech (* Study I and * Study II) plotted against PTA for M-mode (left) and T-mode (right). The score range is 0 Not at all to 10 Perfectly. Regression lines and Spearman’s correlations ($r_s$) are presented.
Spatial hearing

The spatial ability was significantly rated higher in M-mode when compared to T-mode: Item 7a Locate direction ($T=0, p<.05$), Item 7b Locate distance ($T=3, p<.05$), Item 8 Locate speaker ($T=6, p<.05$), and Item 10 Externality of sounds ($T=11.5, p<.05$). No significant correlations were found between the spatial hearing items and subjects' PTA.

Other Qualities

M-mode was significantly rated better than T-mode ($T=1.5, p<.05$) for Item 11 Segregate sounds. Differences between modes were also found for Item 15 Listening effort one-to-one ($T=13.5, p<.05$). Less listening effort was reported using T-mode. The scores for listening effort (Items 15 and 16) in M-mode were correlated with PTA. A more severe hearing loss led to higher listening effort. The correlation between Item 16 Listening effort group and PTA is shown in Figure 2. Item 12 Identify students by voice also showed a negative correlation with PTA in M-mode. No correlations between Other Qualities items and PTA were found when subjects used T-mode. Item 17 Unnatural and Item 18 Pleasantness were analysed for the data in Study II because the items only pertained to the T-mode in Study I. No significant differences between M-mode and T-mode were found.

![Figure 2](image)

**Figure 2.** Other Qualities Item 16 Listening effort group (* Study I and * Study II) plotted against PTA for M-mode (left) and T-mode (right). The score range is 0 Extremely to 10 Not at all. Regression lines and Spearman’s correlations ($r_s$) are presented.

Intercorrelation

Prominent questionnaire item intercorrelations ($p<.01$, 2-tailed) are presented in Table 5. The items were found to intercorrelate, especially for Speech hearing and Spatial hearing. Speech items were correlated with the ability to ignore interfering sounds (Item 6) and listening effort (Items 15 and 16). Items 2c and 3c were further found to correlate with Item 12 Identify students by voice. In a conversation with competing speech, it is important to have the ability to identify students by their voice. Item 12 was also correlated with listening effort (Items 15 and 16). Scores on Spatial items were correlated with the ability to segregate sounds (Item 11).

Items 17 Unnatural and Item 18 Pleasantness were only analysed for the data in Study II. Correlations ($p<.05$, 2-tailed) are presented in Table 5. The unnatural quality of the sound correlated with the ability to externalize sounds (Item 10). Listening pleasantness was found to correlate with difficult conversation tasks (Items 3a, 3b, 3c, 4 and 5), i.e. high listening pleasantness corresponded to high speech intelligibility.

8
Table 5. Spearman rank correlations ($r_s$) between items. Plain font shows correlations of $r_s=0.40$-0.69 and underlined font shows correlations $r_s>0.70$.

<table>
<thead>
<tr>
<th>SPEECH</th>
<th>SPATIAL</th>
<th>OTHER QUALITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a) Talk one-to-one quiet</td>
<td>2bc, 3abc</td>
<td>6, 15, 16</td>
</tr>
<tr>
<td>2b) Talk one-to-one noise</td>
<td>2ac, 3abc, 4</td>
<td>6, 15, 16</td>
</tr>
<tr>
<td>2c) Talk one-to-one speech</td>
<td>2ab, 3abc, 5</td>
<td>6, 12, 15, 16</td>
</tr>
<tr>
<td>3a) Talk group quiet</td>
<td>2abc, 3bc, 5</td>
<td>6, 15, 18</td>
</tr>
<tr>
<td>3b) Talk group noise</td>
<td>2abc, 3ac, 5</td>
<td>15, 16, 18</td>
</tr>
<tr>
<td>3c) Talk group speech</td>
<td>2abc, 3ab, 4, 5</td>
<td>6, 12, 15, 16, 18</td>
</tr>
<tr>
<td>4) Follow two speech streams</td>
<td>2b, 3c, 5</td>
<td>15, 16, 18</td>
</tr>
<tr>
<td>5) Switch quickly conversation</td>
<td>2c, 3abc, 4</td>
<td>6, 15, 16, 18</td>
</tr>
<tr>
<td>SPATIAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7a) Locate direction</td>
<td>7b, 8, 10</td>
<td>11,</td>
</tr>
<tr>
<td>7b) Locate distance</td>
<td>7b, 8</td>
<td>11, 13</td>
</tr>
<tr>
<td>8) Locate speaker</td>
<td>7ab, 10</td>
<td>11</td>
</tr>
<tr>
<td>10) Externality of sounds</td>
<td>7a, 8</td>
<td>17</td>
</tr>
<tr>
<td>OTHER QUALITIES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) Ignore interfering sounds</td>
<td>2abc, 3ac, 5</td>
<td>15, 16</td>
</tr>
<tr>
<td>11) Segregate sounds</td>
<td>7ab, 8</td>
<td></td>
</tr>
<tr>
<td>12) Identify students by voice</td>
<td>2c, 3c</td>
<td>13, 15, 16</td>
</tr>
<tr>
<td>13) Judge mood from voice</td>
<td>7b</td>
<td>12, 16</td>
</tr>
<tr>
<td>15) Listening effort one-to-one</td>
<td>2abc, 3abc, 4, 5</td>
<td>6, 12, 16</td>
</tr>
<tr>
<td>16) Listening effort group</td>
<td>2abc, 3bc, 4, 5</td>
<td>6, 12, 13, 15</td>
</tr>
<tr>
<td>STUDY II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17) Unnatural</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>18) Pleasantness</td>
<td>3abc, 4</td>
<td></td>
</tr>
</tbody>
</table>

3.2 Questionnaire correlation with word recognition

The result of the word recognition test is plotted with respect to subjects’ PTA in Figure 3. The speech recognition threshold at the 5 dB SNR was high, and the median score was 90% (IQR=12%). No correlation was found between the word recognition scores and subjects’ PTA ($r_s=-0.41$, $p$=n.s.; the deviating threshold of 44% was discarded). However, more effort was put forth by subjects with more severe hearing loss ($r_s=-0.67$, $p<.05$). The median effort rating was 3 (IQR=2.75) on the eleven-point scale ranging from 0 Extremely to 10 Not at all.

Spearman’s rank correlations between questionnaire items (M-mode) and the word recognition test are presented in Table 6. Word recognition was not found to correlate with Speech hearing items; however, high correlation was found between recognition scores and Item 11 Segregate sounds ($r_s=0.82$, $p<.01$). Speech recognition in the presence of two speech maskers was correlated to sound segregation rather than to self-assessed speech intelligibility. Word recognition scores were also correlated with Item 13 Judge mood from voice. Perceived effort in the word recognition test was correlated with listening effort in a group conversation (Item 15; $r_s=0.89$, $p<.01$). A high correlation was also found between the effort ratings and Item 13 Judge mood from voice ($r_s=0.82$, $p<.01$). Furthermore, questionnaire scores for Item 2c Talk one-to-one speech, Item 11 Segregate sounds, and Item 12 Identify students by voice were correlated with perceived effort.
Figure 3. Word recognition scores (left) and Perceived effort (right) are plotted against PTA. Regression lines and Spearman’s correlations ($r_s$) are presented.

Table 6. Spearman rank correlations between questionnaire items (M-mode) and word recognition test: Recognition (percentage correctly repeated words) and Effort ratings (0 Extremely - 10 Not at all). * indicates significant correlations ($p<.05$); correlations below 0.3 are ignored.

<table>
<thead>
<tr>
<th>Item</th>
<th>Recogn.</th>
<th>Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPEECH</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a) Talk one-to-one quiet</td>
<td>0.55*</td>
<td></td>
</tr>
<tr>
<td>2b) Talk one-to-one noise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2c) Talk one-to-one speech</td>
<td></td>
<td>0.37</td>
</tr>
<tr>
<td>3a) Talk group quiet</td>
<td></td>
<td>0.42</td>
</tr>
<tr>
<td>3b) Talk group noise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3c) Talk group speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Follow two speech streams</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>5) Switch quickly conversation</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SPATIAL</strong></td>
<td>0.35</td>
<td>0.49</td>
</tr>
<tr>
<td>7a) Locate direction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7b) Locate distance</td>
<td>0.34</td>
<td>0.44</td>
</tr>
<tr>
<td>8) Locate speaker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10) Externality of sounds</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td><strong>OTHER QUALITIES</strong></td>
<td>0.35</td>
<td>0.65*</td>
</tr>
<tr>
<td>6) Ignore interfering sounds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11) Segregate sounds</td>
<td>0.82*</td>
<td>0.82*</td>
</tr>
<tr>
<td>12) Identify students by voice</td>
<td>0.44</td>
<td>0.58*</td>
</tr>
<tr>
<td>13) Judge mood from voice</td>
<td>0.56*</td>
<td>0.82*</td>
</tr>
<tr>
<td>15) Listening effort one-to-one</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>16) Listening effort group</td>
<td>0.41</td>
<td>0.89*</td>
</tr>
</tbody>
</table>

3.3 Strategies for hearing

At the end of the questionnaire, subjects were asked to describe in which situations they preferred to use any of the modes (Table 2). Subjects preferred to use T-mode if listening to a student/teacher talking in front of the class, i.e. listening to someone speaking at a long distance. The choice of M-mode or T-mode depended on speech level or quality. One subject expressed it this way: Most often I use T-mode, since more than half of the other students in the class speak with a quiet voice. Similarly, another subject preferred to use T-mode when other students speak and M-mode: When
The teachers speak, since they speak with a clear voice. A common reason to use M-mode instead of T-mode was the possibility to hear what someone is saying close to you without using a microphone. The subjects often also distinguished between conversation that is a part of the teaching and non-teaching conversations. This distinction was expressed by the subjects as follows: When one talks with classmates about things that do not concern the teaching or as I would like to hear when classmates tell jokes or similar (M-mode); and At teaching and discussion or as When the teacher says something important (T-mode).

One reason for the preference of T-mode was a better ability to focus on what a person says. One student also answered that one hears best with T-mode and that M-mode was preferred in noisy classroom situations. The two examples illustrate a difference in answers regarding preferable mode in noisy classrooms. Some students prefer T-mode when the classroom is noisy while others prefer M-mode. One can here observe a conflict between the better ability to ignore competing sounds using T-mode and the better ability to catch what is happening around oneself using M-mode. Subjects answered that a solution to the conflict was to use MT-mode (a combination of M-mode and T-mode, see Table 2) since this allows them to hear both what is said and what is happening around them. One student also explained that the use of MT-mode was preferred since it helps them hear what someone is saying even if the speaker does not use a microphone. Students who did not use MT-mode noted that sounds were often confusing and that speech sounded unnatural.

The answers for mode preference depended on the degree of the subjects' hearing loss: the better hearing students tended to prefer M-mode as compared to students with severe hearing loss who to a greater extent preferred T-mode.

One student noted that M-mode gives a more pleasant sound than T-mode. An unnatural or artificial character in T-mode was also reported. One student thought T-mode sounded mechanical and one student said T-mode sounded robot-like. T-mode sometimes sounded rough was another observation. In Study I, some students reported that T-mode produced a buzzing sound and a crackling sound.

The answers were diverse comparing the two different types of ALDs assessed in Study I (Table 2). Some subjects preferred to use the centrally placed ALD while others preferred using an ALD where the speaker is close to the microphone. One student emphasized a preference for the centrally placed ALD; the student felt safer because it was easier to hear events in the surrounding area. As previously noted, the advantage with a short microphone distance was a better ability to hear what is being said.

4 Discussion

The questionnaire developed is based on SSQ; however, the present questionnaire is not validated. As a measure of reliability, items within each section of the questionnaire were intercorrelated. Furthermore, when the subjects only used their hearing aids (M-mode), their ability to follow a conversation decreased in the presence of ambient noise. Competing speech was the most difficult condition. The deteriorating effect of masking speech was most prominent for subjects with more severe hearing loss. This result agrees with word recognition studies, where the difference between a noise masker and a speech masker was attributed to a lessened ability to take advantage of spectral and temporal dips (e.g. Hygge et al. 1992; Moore, 2003). Ratings of listening effort in the questionnaire were also consistent with effort ratings of a comparable auditory scenario in the word recognition test. The results verify students' ability to imagine a listening situation while answering a questionnaire.
The difference between a noise masker and a speech masker can also be explained by informational masking (Brungart, 2001; Brungart et al., 2001), which corresponds to masking effects not explained by the energies of the masker (energetic masking). It has been shown that the advantage due to spatial separation of sources is greater for informational masking than for energetic masking (Arbogast et al., 2002). In this study, no correlation was found between the word recognition scores and the Speech hearing items, but a correlation was found in the ability to segregate sounds. Accordingly, informational masking was considerable for word recognition in presence of masking speech and this ability was related to sound segregation.

There was high variability in the questionnaire scores. Audiological data can explain this variability: only data for best ear PTA were analysed. There were also differences between Study I and Study II. Subjects in Study I attended elementary/secondary school. In Study II, subjects were older and attended upper secondary school. This difference means there was a variety in ALD experience as well as a variety in cognitive and academic development. Differences between Study I and II are illustrated in Figure 1 and 2. The number of subjects was too low to examine fully these differences. Instead, the data describe the diversity in ALD preference among students attending classes for hearing impaired.

The induction loop (IL) installation and the hearing aid telecoil adjustment affect the assessment of the hearing aid telecoil (T) mode. As previously discussed, noise was noted when using T-mode. The quality of the IL systems in the schools in the study is controlled on a regular basis. However, no measurements were performed to verify the function of these systems.

4.1 Mode comparison

The focus of this paper lies in the comparison between the use of only hearing aids (M-mode) and the use of an ALD (T-mode). Reviewing the results, two attributes were identified:

- **Audibility**: To hear, and less effort put forth to hear, what is being said.
- **Awareness**: To hear, locate, and segregate sounds in the environment around one self, e.g. other students not using external microphones.

T-mode was better in terms of **audibility**, especially when listening to a person at a long distance and in the presence of speech. Better **awareness** was achieved in M-mode.

Listening effort is important to consider with regard to **audibility**. Speech recognition is not necessarily reduced in difficult listening situations, i.e. the audition performance is not decreased. However, speech understanding will require more cognitive processing (top-down), e.g. working memory capacity and semantic and lexical knowledge. Measuring listening effort is thus important to capture the cognitive load (Hällgren et al., 2005). In the presented word recognition test, high speech recognition was measured regardless of subjects’ PTA; however, subjects with severe hearing loss needed to use more effort.

Bilateral advantage in terms of **audition** was not as evident in the data as when compared to Noble and Gatehouse’s (2006) findings. Whereas T-mode received higher scores than M-mode for items related to one-to-one and group conversation, the score for participating in a conversation that switches quickly were equal. The result indicates a loss of bilateral advantage when using T-mode. With regard to **awareness**, the result supports that ALDs decrease bilateral advantage with regard to localization and sound segregation.

The students’ ratings of the importance of the different items indicated that Speech hearing and listening effort were more important to their choice of which mode to use than Spatial hearing and
the ability to segregate sounds. An explanation is that *awareness* lies more in the possibility and necessity to hear what is happening nearby and the comments by students not using a microphone rather than in localization and segregation of sounds. The ability to externalize sounds was correlated with the ratings of (un)naturalness. The finding relates to the feeling of presences, which is discussed as a measure of plausibility of reproduced sound (Blauert and Jekosch, 2003). This finding suggests that the ability to hear and retrieve information from the surroundings is a quality parameter to consider in sound-transmission systems, such as classroom ALDs. An aspect of *awareness* is also the ability to control one’s own perception, which is reduced using external microphones.

The sound character when using T-mode was reported different from M-mode. To some students, the T-mode sounded *robot-like*, an unnatural sound. A robot-like sound can also be a result of a more narrow frequency bandwidth. A difference between the modes was not found for the item related to an unnatural quality of the sound. A feasible explanation lies in the different sound quality preference among subjects.

For Other Quality items, such as the ability to identify students by voice and judge a student’s mood by her/his voice, no differences between M-mode and T-mode were found. However, students did rate these items to be important for their choice of which mode to use. The ability to identify students by voice and the ability to judge mood from voice were correlated with listening effort. To identify student by voice was also correlated with the ability to participate in conversations in the presence of competing speech. As discussed by Gatehouse and Noble (2004), difficulty identifying other persons and their mood is correlated with the experience of handicap and a reluctance to engage in conversations.

**Mode combinations**

For bilateral hearing aid users, a possible way to combine the advantage of M-mode and T-mode is to use one hearing aid in M-mode and the other in T-mode. Many hearing aids also feature a M+T-mode, where the signals from the hearing aid microphone and telecoil are mixed. Word recognition improvements using M-FM-mode when compared to M-mode have been shown (Boothroyd & Iglehart, 1998), even if Hawkins (1984) found the improvements to be smaller when compared to only FM-mode. In the present study, subjects expressed that MT-mode combines the qualities of *audibility* and *awareness*. Comments such as confusion of sounds and an unnatural sound call for new studies where the different mode combinations are examined.

In Study I, ALDs with short microphone distance were compared to a centrally placed device, e.g. on a table between users. Moving the microphone from the mouth to a table is intended to facilitate peer interaction. The preferences varied among subjects and the answers can be interpreted as a parallel to the discussion on *audibility* and *awareness* between M-mode and T-mode. An advantage with MT-mode, in contrast to a centrally placed ALD, is the possibility to do individual adjustments.

**Generalization**

The scope of the studies reported was limited to IL systems. The discussion on *audibility* and *awareness* can be generalized to ALDs regardless of the use of IL or FM systems. Lewis and Eiten (2003) evaluated audibility of FM level settings relative to the hearing aid microphone input. Maximum level (+24 dB) was preferred when listening to the teacher and minimum level (-6 dB) was preferred when listening to others in the classroom. FM and IL systems have also been shown to provide equal word recognition benefits (Nabelek et al., 1986; Noe et al., 1997).
The relation between the importance of **audibility** and **awareness** to the pedagogical paradigm is important. The auditory scenario is different in a peer-to-peer setting than in a classroom where the teacher lectures; students make different choices between M-mode and T-mode. Students also described T-mode as a teaching communication channel and M-mode as a channel for other conversations. Subjects in present study attend classes for hearing impaired. These classes have better room acoustic conditions and lower number of students when compared to a regular class. Hence, it can be expected that the difference in attributes between T-mode (**audibility**) and M-mode (**awareness**) is larger for hearing impaired students in a regular class.

### 5 Conclusions

This paper investigates the preference of classroom assistive listening devices (ALDs) based on induction loop systems. By comparing the assessments of T-mode (hearing aid telecoil mode) and M-mode (hearing aid microphone mode), the following conclusions were made:

- M-mode and T-mode give two advantages to choose from: better **audibility** or better **awareness**. Less effort is required to hear using T-mode. M-mode gives a better recognition of the environment around students and a possibility to hear other students not using an ALD.
- Students with severe hearing loss benefit more using T-mode when compared to the better hearing students.
- The loss of bilateral advantage using an ALD results in a decrease in localization and segregation of sounds.
- Students have strategies for hearing: different choices between M-mode and T-mode are made depending on the situation.
- Students noted that they solve the choice between **audibility** and **awareness** by using one hearing aid in M-mode and the other in T-mode, or by using one or both hearing aids in a M+T-mode.

The pedagogical paradigm changes the auditory scenario and consequently also the quality specifications of ALDs. Evaluation of these systems in terms of both **audibility** and **awareness** requires an approach where several attributes to hearing and disability are assessed. A questionnaire in the dimensions of Speech hearing, Spatial hearing and Other Qualities is a promising tool for this evaluation. Open-ended questions as well as additional notes add to the understanding of the result. Further evaluations of ALDs and listening conditions in classrooms are needed. Specifically, it may be informative to study the different solutions where the signal from the ALD is combined with the hearing aid microphone signal.

### 6 Acknowledgements

The authors thank the Swedish Association of Hard of Hearing People (HRF) and the Swedish Inheritance Fund for their generous support. The collaboration and contributions of the staff and students of the Dialogue project (http://www.dialogprojektet.se), Sweden’s National Upper Secondary Schools for the Deaf and for the Hard of Hearing (RGH/RGD), and the National Agency for Special Needs Education and Schools (SPSM) are gratefully acknowledged. We would also like to thank Laila Sandmon for her assistance gathering the data.

An earlier version of this manuscript related to Study I was presented as a part of a Licentiate thesis (Odelius, 2007).
7 References


Assessments of hearing aid microphone and FM/telecoil mode combinations

Johan Odelius, Örjan Johansson, and Roger Johnsson
Division of Sound and Vibration, Department of Human Work Sciences
Luleå University of Technology

Abstract
This study investigates how assistive listening devices (ALD) and hearing aids are combined with respect to individual preference and acoustic setting. A listening experiment and a questionnaire were performed with eleven hearing-impaired students. The listening experiment compared signal combinations of a close-up microphone and an artificial head in the listening position at a distance of 1.5 m. Speech intelligibility and preference were self-assessed. The target was female speech and two male voices were used as a masker. Stimuli were binaurally reproduced using cross-talk cancellation. In the questionnaire, speech intelligibility and listening effort were assessed in different classroom settings. The hearing aid microphone mode, telecoil mode, and combinations of the two modes were compared. The subjects also answered which mode combinations they preferred to use in each setting. The results of the listening experiment confirmed the advantage of a short microphone distance. The results were validated by the questionnaire. The students were active in their use of the different hearing aid modes, where a combination of the microphone and telecoil mode was preferred.

Keywords: Assistive Technology, Hearing Aids, Psychoacoustics/Hearing Science, Hearing Aid Satisfaction, Speech Perception

1 Introduction
External microphone systems, referred to as assistive listening device (ALD), are used in classrooms to support hearing-impaired students with communication when hearing aids are insufficient (Ross, 1992; Holmes and Saxon, 2000). The speech to noise ratio is increased by moving a microphone closer to the speaker and speech recognition improvements have been shown (Hawkins, 1984; Lewis et al., 1991; Ross, 2003; Boothroyd, 2004). The signal from the external ALD microphone is transmitted to the hearing aids, most often, using a FM system or an induction loop (IL) system. Using a switch on the hearing aid, the students can either listen to the signal from the hearing aid microphone or the signal from the ALD received by a FM receiver or a telecoil, respectively. In Sweden, IL systems are the dominant tool for ALD transmission in classes for hearing-impaired students. In Europe, hearing aids most often have a built-in telecoil (some 85-90%); i.e., no external receiver is necessary. In the USA, only about 30-40% of the hearing aids include a telecoil, and FM systems are used to a greater extent (Ross, 2002). FM and IL systems have been shown to provide equal benefit (Nabelek et al., 1986; Noe et al., 1997).
Following notations will be used to describe different hearing aid modes: (M) microphone mode, (T) telecoil mode, (M+T) combined mode where the signals from the internal microphone and telecoil are mixed, and (M/T) one hearing aid in M mode and the other in T mode.

In a previous study on IL systems, hearing-impaired students’ questionnaire scores from mode M and T were compared (Odelius & Johansson, 2010). It was concluded that T mode gives better audibility and that M mode gives better awareness. Audibility was defined as the ability to hear with less effort put forth to hear what is being said. Awareness was defined as the ability to hear, locate, and segregate sounds in the environment around one self. The students noted that they solved the choice between audibility and awareness by using one hearing aid in M mode and the other in T mode or by using one or both hearing aids in the M+T mode. Higher word recognition scores for M+FM mode when compared to M mode have been shown (Boothroyd & Iglehart, 1998). The word recognition improvements using M+FM are, however, smaller than when only using FM (Hawkins, 1984). The compromise between good audibility of a single talker and to hear other persons in the room was also highlighted by Lewis and Eiten (2003). In their experiment, the level of the FM signal was varied between -6 to +24 dB relative to the hearing aid microphone signal. The setting with +24 dB was in favour when listening to the teacher and -6 dB was in favour when the task was to listen to others in the room. As the authors also conclude, there is still more to be learned about the interaction between hearing aids and FM systems.

This study investigates the effect on speech intelligibility and preference when combining M mode and T/FM mode. A listening experiment was carried out where the subjects listened to female speech in the presence of male speech. The subjects compared mixtures of an omnidirectional microphone close to the female speaker with a binaural microphone model in the listening position. In a questionnaire and in interviews, the subjects also rated their experience of different hearing aid mode combinations as well as which mode combinations they use. The questionnaire focused on three classroom settings with regard to three different background noise conditions in each case.

2 Subjects

Eleven hearing-impaired students, nine girls and two boys, participated in the study. The students attended one of Sweden’s National Upper Secondary Schools for the Deaf and for the Hard of Hearing. The school was equipped with IL systems. The age of the students ranged from 16 to 20 years. The students were fitted bilaterally and their unaided best ear pure-tone average (PTA, average of the hearing levels at 0.5, 1, and 2 kHz) ranged from 38-83 dB HL (median 67 dB HL). One subject had an interaural asymmetry equal or more than 15 dB. One of the subjects had also previously participated in a questionnaire and listening experiment study carried out by the authors. Because the experience was not considered to influence the results, the student was included in the analyses.

3 Listening experiment

3.1 Method

Speech intelligibility was assessed in a listening experiment based on Multi Stimulus test with Hidden Reference and Anchor (MUSHRA) (ITU-R BS 1534-1). Stimuli were generated using the room acoustic modelling software CATT Acoustics, which features a variant of the ray
trace technique. During the listening tests, the subjects wore their hearing aids and stimuli were reproduced using loudspeakers and cross-talk cancellation.

**Room acoustic model**

A rectangular classroom (9.4 x 8.7 x 3.1 m) was modelled (Figure 1). Target speech source and listening position were placed on opposite sides of a table at a distance of 1.5 m and 0 degrees azimuth. Two speech masker sources were positioned at either side and behind the listening position. A noise masker source was also placed in one of the top corners in the room. The room was furnished and absorption and diffusion coefficients were set to create reasonable room acoustic characteristics. Reverberation times are presented in Table 1. Two built-in receiver models were used in the model: An omnidirectional microphone was placed 0.2 m in front of the target speech source at -40 degrees elevation and an artificial head (ITA kunstkopf) was placed at the listening position. The directivity of a singer (Marshall & Meyer, 1985) was applied to all speech sources.

**Table 1.** Reverberation time in the room acoustic model. RT₆₀ is derived using least-square fits to the decay in the interval -5 to -35 dB.

<table>
<thead>
<tr>
<th>Octave (Hz)</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1k</th>
<th>2k</th>
<th>4k</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT₆₀(s)</td>
<td>0.46</td>
<td>0.43</td>
<td>0.36</td>
<td>0.32</td>
<td>0.26</td>
<td>0.24</td>
</tr>
</tbody>
</table>

**Figure 1.** The classroom situation modelled in CATT Acoustics.

**Stimuli**

Anechoic recordings of a female voice and a male voice were used as target and masker, respectively. The recordings were convolved in CATT Acoustics. The target female speech was cut in sections of 12-15 s. For the male speech masker, random cuttings were used and randomly assigned to one of the two masker speech positions. The equivalent sound pressure level of the male speech masker was set 3 dBA below the level of the female speech target at the listening position. To render an ecological classroom sound as well as to establish equal background noise criteria throughout the test, Brownian noise (-6 dB/octave) was added. The noise was supposed to act as a ventilation sound and was convolved in CATT Acoustics. The level of the background noise was set 25 dBA below the female speech level. Speech and noise sounds were limited in frequency range, in octave bands, from 125 Hz to 4 kHz. The presentation level was set to a normal speech level adjusted by each subject before the listening test.
Twelve different mixtures of M (signals generated by the binaural model) and T (signal from the close-up microphone) were created (Table 2). Stimuli 6, 7, and 8 were presented in two variants – denoted left and right – and the left and right channel were altered. The two signals were equalized to the same A-weighted sound pressure level and the level was kept equal across the different mixtures. The levels were adjusted before summation as follows: for (M+T), the M and T signals were decreased 3.0 dB; for (M_{10}+T), the M signal was decreased 10.0 dB and the T signal was decreased 0.5 dB; for (M+T_{10}), the M signal was decreased 0.5 dB and the T signal was decreased 10 dB.

Table 2. The twelve stimuli in the experimental design. Each session was separately assessed using multiple stimulus test where stimulus 5 and 9 were hidden anchors.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Left ear</th>
<th>Right ear</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Session I</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>2</td>
<td>M_{10}+T</td>
<td>M_{10}+T</td>
</tr>
<tr>
<td>3</td>
<td>M+T</td>
<td>M+T</td>
</tr>
<tr>
<td>4</td>
<td>M+T_{10}</td>
<td>M+T_{10}</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>Session II</td>
<td></td>
</tr>
<tr>
<td>6_{left}</td>
<td>T</td>
<td>M+T</td>
</tr>
<tr>
<td>6_{right}</td>
<td>M+T</td>
<td>T</td>
</tr>
<tr>
<td>7_{left}</td>
<td>M+T</td>
<td>M</td>
</tr>
<tr>
<td>7_{right}</td>
<td>M</td>
<td>M+T</td>
</tr>
<tr>
<td>8_{left}</td>
<td>T</td>
<td>M</td>
</tr>
<tr>
<td>8_{right}</td>
<td>M</td>
<td>T</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>M</td>
</tr>
</tbody>
</table>

Procedure
The listening experiment was carried out in a classroom at the students’ school. The room had a reverberation time of $0.6\text{ s}$ and a background noise level of 27 dBA. Dichotic listening was enabled using the cross-talk cancellation technique. Cross-talk cancellation was achieved using a Lexicon MC-1 and two ADAM-S2A studio monitors placed in front of the listener (Figure 2). The listening position was situated in the centre of the room to avoid early reflections from the sides. During the experiment, subjects wore their hearing aids set in M mode.

The experiment was run in conjunction with another listening experiment evaluating different ALD solutions. The same user interface and stimuli were used in both experiments. The order of the two experiments was randomized. A training session introduced the subjects with user interface and type of stimuli. During the training session, the subjects also had the opportunity to adjust the output level.

The listening test was divided into two sessions (Table 2). Session I assessed the relation between M and T. Session II assessed different mixtures of M, T, and M+T. Stimulus 5 and 9 (M) were the same and were used as an anchor for both sessions. The two sessions were run twice, with two different tasks, i.e. four random-ordered runs. The different tasks were defined as follows (liberally translated from Swedish):
a) Give 100 points to the sound where you most easily can hear what the female voice is saying. Then give the other sounds points. Lesser points the more difficult you think it is to hear what the female voice is saying.

b) Give 100 points to the sound you like the most. Then give the other sounds points. Lesser points the worse you think the sound is.

To each sound stimulus, there was a slider that allowed the score to be set from 0 to 100 points. The subjects could freely decide the order and how many times to play each sound.

Figure 2. Experimental setup. Two speakers (ADAM-S2A) were placed in front of the listener and stimuli were played back using cross-talk cancellation (Lexicon MC-1).

3.2 Results

One student was excluded from the analyses due to a problem with the hearing aids at the time of the listening experiment. Figure 3 shows the scores of the two tasks a) ability to follow speech and b) sound preference. In session II, the scores of stimuli 6, 7, and 8 were represented by each subject’s maximum score of the left and right alternative (see Table 2). That is, stimulus 6 corresponds to each individual’s maximum score of $\delta_{\text{left}}$ T/M+T and $\delta_{\text{right}}$ M+T/T. The scores for task (a) and (b) were similar and were also found to correlate (Spearman’s rank correlation, $p<.01$). Due to this correlation, further analyses of the results will focus only on task (a).

The spread in data varied between stimuli, e.g. session I in Figure 3. Stimulus 1 (T) received the highest value, whereas the decrease in scores for stimuli 2-5 varied between subjects. An explanation of the variety due to the degree of hearing loss was not found: the scores and subjects’ PTA were not correlated.

Statistical analyses based on ranked data were chosen. Friedman’s ANOVA with a significance level of $p<.05$ was used. Wilcoxon signed rank test was used for planned comparisons where the significance level was adjusted according to the Bonferroni correction. The results are presented in Table 3. In session I, the ratings decreased as the magnitude of the M mode increased. In session II, M mode was rated lower than M+T/T (M+T and T presented to either ear) and M/T (M and T mode presented to either ear).
Table 3. The results of Friedman’s ANOVA and the planned comparisons.

<table>
<thead>
<tr>
<th>Session</th>
<th>Friedman test</th>
<th>Stimulus</th>
<th>Planned comparisons</th>
<th>Wilcoxon test</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$\chi^2(4)=27.7$</td>
<td>1 T &gt; 2 M,10+T</td>
<td>T=12.5 p=n.s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 T &gt; 3 M+T</td>
<td>T=0 p&lt;.007</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 T &gt; 4 M+T,10</td>
<td>T=0 p&lt;.007</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 T &gt; 5 M</td>
<td>T=0 p&lt;.007</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 M,10+T &gt; 5 M</td>
<td>T=0 p&lt;.007</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 M+T &gt; 5 M</td>
<td>T=5 p=n.s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 M+T,10 &gt; 5 M</td>
<td>T=5.5 p=n.s.</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>$\chi^2(3)=14.1$</td>
<td>6 M+T / T &gt; 9 M</td>
<td>T=1 p&lt;.0167</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 M / M+T &gt; 9 M</td>
<td>T=9 p=n.s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 M / T &gt; 9 M</td>
<td>T=3 p&lt;.0167</td>
<td></td>
</tr>
</tbody>
</table>

4 Questionnaire
To gain more insight in the results of the listening experiment, the subjects self-assessed their use and speech intelligibility judgments of ALDs in real situations. The questionnaire and interview results are a part of a larger questionnaire study on IL systems (Odelius & Johansson, 2009).

4.1 Method
The questionnaire was completed during interviews and took place at the students’ schools. The interviews were led by the students’ audiologist. A neutral interview style was used so as not to influence response choices.

Subjects assessed speech intelligibility and listening effort when using mode M, T, and MT in three classroom settings and in three noise conditions. The notation MT corresponds to either the M+T mode or the case when M mode and T mode are used at either ear (M/T). The
questionnaire items are presented in Table 4. The items relate to three classroom settings (Figure 4): 1) listening to a student standing at the board using a wearable microphone; 2) a one-to-one conversation in a classroom with linked table microphones; and 3) a group conversation around a table where the students used wearable microphones. Each setting had three background noise conditions: a) quiet classroom; b) continuous noise, e.g. fan noise; c) competing speech. The students also explicitly wrote down which mode they preferred to use at either ear for each setting and noise condition.

**Table 4.** Questionnaire items liberally translated from Swedish. Each item had three background noise conditions: a) quiet classroom; b) continuous noise, e.g. fan noise; and c) competing speech.

<table>
<thead>
<tr>
<th>SPEECH INTELLIGIBILITY (0 Not at all - 10 Perfectly)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Can you hear what a student is saying standing at the board?</td>
</tr>
<tr>
<td>2 Can you hear what a student is saying in a one-to-one conversation?</td>
</tr>
<tr>
<td>3 Can you hear what other students are saying in a three person group conversation around a table?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LISTENING EFFORT (0 Extremely - 10 Not at all)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 How much effort do you require to hear when listening to a student standing at the board?</td>
</tr>
<tr>
<td>5 How much effort do you require to hear in a one-to-one conversation?</td>
</tr>
<tr>
<td>6 How much effort do you require to hear in a three person group conversation around a table?</td>
</tr>
</tbody>
</table>

![Figure 4](https://example.com/figure4.png)

Figure 4. The three classroom settings: (left) listening to a student standing at the board using a wearable microphone; (middle) participating in a one-to-one conversation in a classroom with linked table microphones; and (right) a group conversation around a table with the students using wearable microphones.

### 4.2 Results

The ratings of speech intelligibility (Item 1-3) and listening effort (Item 4-6) are presented for ten complete cases as median and interquartile range (IQR) (Table 5). There was a significant difference between the three modes (M, MT, and T) for all items (Friedman’s ANOVA, \( p<.05 \)). The results of the paired comparisons, Wilcoxon signed rank test (Bonferroni correction), is depicted in Table 5. T mode was rated better than M mode both with regard to speech intelligibility and listening effort. The result for M and T mode combinations was not as decisive. MT was rated in between T mode and M mode, where significant differences with regard to speech intelligibility were found in difficult speech tasks (Item 1c, 3b, and 3c). For listening effort, MT mode was rated higher than M mode in the masking speech conditions. A difference between mode T and MT was instead found in the quiet classroom condition.
Table 5. Median questionnaire scores (IQR in brackets) by mode (T, MT, and M) for speech intelligibility and listening effort. Differences between modes are highlighted (Wilcoxon signed rank test, \( p < .0167 \)).

<table>
<thead>
<tr>
<th>1. Talk distance</th>
<th>Mode</th>
<th>Comparison</th>
<th>T</th>
<th>MT</th>
<th>M</th>
<th>T&gt;MT</th>
<th>MT&gt;M</th>
<th>T&gt;M</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) quiet classroom</td>
<td>T</td>
<td>10 (1)</td>
<td>9 (2)</td>
<td>7 (1)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) continuous noise</td>
<td>MT</td>
<td>9 (3)</td>
<td>7 (3.25)</td>
<td>5 (2)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) competing speech</td>
<td>M</td>
<td>9 (1)</td>
<td>7.5 (2)</td>
<td>5 (2)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>2. Talk one-to-one</td>
<td>T</td>
<td>10 (1)</td>
<td>9 (2)</td>
<td>7.5 (3)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) quiet classroom</td>
<td>MT</td>
<td>8.5 (2)</td>
<td>7 (3.25)</td>
<td>5.5 (2)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) continuous noise</td>
<td>M</td>
<td>8.5 (2)</td>
<td>7 (2)</td>
<td>4 (3)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Talk group</td>
<td>T</td>
<td>9 (2)</td>
<td>7.5 (3)</td>
<td>7 (2)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) quiet classroom</td>
<td>MT</td>
<td>8 (3)</td>
<td>7 (3)</td>
<td>5.5 (3)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) continuous noise</td>
<td>M</td>
<td>7.5 (2)</td>
<td>6 (2)</td>
<td>4.5 (3)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Effort distance</td>
<td>T</td>
<td>9 (2)</td>
<td>7.5 (3)</td>
<td>6.5 (2)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) quiet classroom</td>
<td>MT</td>
<td>7 (3)</td>
<td>6 (3)</td>
<td>5.5 (2)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) continuous noise</td>
<td>M</td>
<td>7 (2)</td>
<td>6 (3)</td>
<td>4 (3)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Effort one-to-one</td>
<td>T</td>
<td>9 (2)</td>
<td>7.5 (3)</td>
<td>7.5 (3)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) quiet classroom</td>
<td>MT</td>
<td>8 (1)</td>
<td>7.5 (3)</td>
<td>6 (3)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) continuous noise</td>
<td>M</td>
<td>7 (2)</td>
<td>6 (3)</td>
<td>3.5 (1)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Effort group</td>
<td>T</td>
<td>8 (1)</td>
<td>7 (2)</td>
<td>7 (3)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) quiet classroom</td>
<td>MT</td>
<td>7.5 (2)</td>
<td>6.5 (3)</td>
<td>6 (2)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) continuous noise</td>
<td>M</td>
<td>6.5 (2)</td>
<td>6 (3)</td>
<td>4 (3)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Spearman rank correlations were also computed between the scores and subjects’ best ear PTA. Significant (\( p < .05 \), 1-tailed) correlations were found for Item 3c, 6a, and 6c. That is, speech intelligibility and listening effort in a group conversation with masking speech were correlated with the degree of hearing loss. The interpretation was that the deteriorating effect of competing speech is larger for students with a severe hearing loss.

Figure 5 illustrates the questionnaire scores collapsed across noise conditions and classroom settings. With regard to both speech intelligibility and listening effort, the difference between mode M and T is larger in the Talk distance setup than in the Talk group setup. Furthermore, the difference between mode M and T is larger in the case of competing speech than in a quiet classroom. The same trends were shown when comparing MT and M mode. The difference between mode MT and T was independent of the situations.

Hear aid mode usage

Figure 6 presents which mode the students preferred to use in the different situations. The subjects’ strategies for hearing depended on the classroom situation. Mode T was more frequently used when listening to a person at a large distance standing at the board in the classroom in comparison to participating in a one-to-one or group conversation. The use of T mode also increased if the listening task was difficult, as in the presence of masking speech.
Three hearing strategies were identified (Figure 6):

I. Subjects prefer M mode and switch one or both hearing aids to T mode if hearing becomes difficult (subject 1, 3, 4 and 6).

II. Subjects use T mode and switch to a combination of modes (M, M+T, and T) in easy listening situations (subject 7, 8, and 9).

III. Subjects prefer a combination of M and T mode and are static in their hearing aid mode usage (subject 2, 5, 10, and 11).

Subjects with mild to moderate hearing loss use hearing strategy I (M mode preference), and subjects with severe hearing loss use hearing strategy II (T mode preference).

5 Discussion

The results in the listening experiment showed the advantage of a short microphone distance with regard to the audibility of a single talker in the presence of masking speech, a result that agrees with previous reported studies on both IL and FM systems. The preference of T mode was also shown when students rated the different hearing aid modes in nine different classroom situations. However, just a few students used both hearing aids in T mode. Instead, the students preferred to use the M+T mode or they used M and T mode in either ear. When using an ALD, information from the classroom is lost. As Lewis and Elten (2003) showed, the signal from the hearing aid microphone was preferred relative to the external microphone signal when listening to people other than the teacher.
The present study demonstrates three important aspects as when the threshold is reached and the students must gain audibility by increasing the amount of the external microphones. The first aspect is listening distance, as in accordance to previous findings of FM system advantage listening to one speaker at a long distance (Boothroyd, 2004). The difference in ratings between mode M and T was larger when listening at distance as when compared to a group discussion setting around a table. In the listening test, the distance to the target speech was 1.5 m; however, T mode was rated higher than both M mode and M+T mode. The second aspect is noise background, where competing speech is the most difficult situation. The effect of masking speech is more prominent for M mode when compared to the other mode combinations. The last aspect is the degree of hearing loss. Students with severe hearing loss use T mode to a greater extent than the better hearing students. Noisy classrooms also have a more deteriorating effect on speech intelligibility and listening effort for students with severe hearing loss.

The compromise discussed by Odelius and Johansson (2010) between good audibility and retrieving information from the surroundings remains. The results in this study do support that a combination of M and T mode is a feasible solution. From the listening experiment, it was concluded that using one hearing aid in M mode and the other in T mode (M/T) significantly increased speech intelligibility when compared to using both hearing aids in M mode. In noisy classrooms, the dilemma between audibility and awareness increases: Because in noisy classrooms there is a greater need to reduce ambient noise but increased classroom activity also increases the need to retrieve information from the surroundings, a combination of M and T mode is of greater use. M and T mode combinations received higher questionnaire scores than M mode in listening situations with competing speech. The data on students’ usage of different hearing aid modes (Figure 6) revealed that audibility, i.e., increased amount of T mode, was preferred in the masking speech conditions. Consequently, the results from the questionnaire verify the results from the listening experiment that combinations of M and T mode offer improved audibility with preserved awareness when compared to only using M mode. However, audibility deteriorates when adding signals from the hearing aid.
microphones in difficult listening situations and students with a severe hearing loss chose to switch to only using T mode. The conclusion agrees with Hawkins’ study (1984) where word recognition improvements using M+FM were smaller than when only using FM.

Multiple stimuli test is a powerful tool for the assessment of different ALD solutions and hearing aid mode combinations. Reproducing binaural stimuli using cross-talk cancellation distorts on-going interaural time differences (ITD) and interaural level differences (ILD) (Akeroyd, 2007). Odelius et al. (2010) showed that ITD and ILD were adequately recreated for frequencies between 300 Hz and 4 kHz, the frequency range evaluated in the listening experiment.

In the listening experiment, there was no difference in scores for speech intelligibility and preference. In this context, the subjects preferred the stimulus with the highest intelligibility. Future studies should assess sound quality parameters in more detail. It has been shown that hearing aid users may prefer aids with good speech quality rather than with high speech intelligibility (Gabrielsson et al., 1988). Since the preferable mixture of T/FM and M mode depend on the auditory scenario, further studies should aim to better replicate classroom situations with multiple talkers.

6 Acknowledgements

The collaboration and contributions of Laila Sandmon at Sweden’s National Upper Secondary Schools for the Deaf and for the Hard of Hearing (RGH/RGBD) and Håkan Bergkvist at the National Agency for Special Needs Education and Schools (SPSM) are gratefully acknowledged.

7 References


CATT-Acoustic v8. Room Acoustic Prediction and Desktop Auralization. URL: http://www.catt.se


Odelius, J. and Johansson, Ö. 2010. Self-assessment of assistive listening devices. *Accepted for publication in Int J Audiol*.


Paper III
Effects on speech intelligibility using a binaural assistive listening device

Johan Odelius, Örjan Johansson, Anders Ågren

Luleå University of Technology
Division of Sound and Vibration
971 87 Luleå, Sweden
johan.odelius@ltu.se

ABSTRACT

The objective of the study was to evaluate effects on speech intelligibility using a binaural assistive listening device. A word recognition test, with two continuous speech maskers, was performed comparing an omni directional and a binaural recording. The stimuli were presented at 5dB signal to noise ratio. In conjunction, the degree of perceived effort was rated on an eleven-point scale. Twelve hard of hearing students, fitted bilaterally, participated in the study. The subjects listened with their hearing aids and the stimuli were presented, utilizing cross-talk cancellation technique, by two loudspeakers in front of the listener. The stimuli were generated using the room acoustic modelling software CATT. A rectangular classroom was modelled, where the source of the target speech and the receiver were placed on opposite sides of a table, at a distance of 1.3 m and 0 degrees azimuth. The two speech maskers were positioned at a distance of 2.6 m and at 60 and 300 degrees azimuth, respectively. An omni directional and a binaural receiver model were used, which resulted in a comparison between diotic and dichotic presented stimuli, respectively. No significant dichotic advantage in word recognition and perceived effort was found.

1. INTRODUCTION

The study origin from a project concerning assistive listening device (ALD) in Swedish hearing impairment classes. The signal from an ALD is transmitted to the hearing aids using, most commonly, an induction loop or a FM system. The students have the opportunity to listen using their hearing aid microphone (M-mode) or using the external ALD microphone (T-mode). Assume a student is bilaterally fitted and uses the same mode on both aids. Then M-mode will result in dichotic listening, i.e. a different sound is presented to each ear. Using T-mode identical sounds are presented to both ears, which is termed diotic listening.

Binaural advantage, defined as the advantage of listening with two ears instead of one, can be viewed in terms of recognition and localization [1]. Recognition can be expressed as a signal to noise ratio (SNR) difference in a speech intelligibility task. There are three reasons for this difference [1]; head diffraction, binaural squelch and binaural redundancy. Head diffraction benefit is due to the possibility to attend to the one ear with the better SNR. Binaural squelch relies on the brain taking advantage of differences between the different signals to the ears. The total dichotic effect of head diffraction and binaural squelch compared to listening with one ear is about 5 dB, averaged over all directions [1]. The brain can also take advantage combining two identical signals, i.e. diotic listening, and this is referred to as binaural redundancy or diotic summation. Improvement in SNR of 1-2 dB has been reported [1,2].

A SNR difference due to binaural squelch is also referred to as binaural intelligibility level difference (BILD). BILD increases with increasing difference in azimuth between target speech and noise masker [3]. Type of noise masker is of importance, where two speaking voices increase BILD about 2 and 5 dB compared to white noise and one speaking voice, respectively [4]. BILD decrease for people with hearing loss, where the greatest reduction occurs with a severe hearing loss and with the greatest asymmetry of hearing loss [1]. In a study [2], the BILD was 9.8 dB for normal hearing, 7.1 dB for symmetrical hard of hearing, and 7.2 or 4.2 dB for asymmetrical hard of hearing, depending whether the noise masker was moved to their good or bad side, respectively.
Binaural advantages have been an issue regarding bilateral versus unilateral hearing aid fitting, so-called bilateral advantage. Advantages have been shown in several studies, e.g. [5-8]. It has also been shown that the bilateral advantage is considerable in difficult multi speaker situations (cocktail-party) and of less importance in more simple speech situations [9]. This result highlights the role of localization regarding speech intelligibility in real life situations. The same study also showed bilateral advantage with respect to listening effort.

Since an ALD system (including induction loop or FM system) today provide a monophonic signal, advantages from dichotic effects can not be expected [1]. The only binaural advantage using ALD is thus diotic summation. Models of binaural processing have been developed and utilized in noise reduction filters [3]. Advanced signal processing can accordingly account for some aspects of binaural advantage. Another approach would be to develop a binaural ALD. The approach requires a different transmission system between ALD and hearing aids than those in use today. Overlooking that problem and other practical problems, binaural ALD can be evaluated in listening tests utilizing modelling software and auralization.

The objective of this study is to evaluate the effect on speech intelligibility using a binaural ALD. This was carried out doing a speech intelligibility comparison between an omni directional and a binaural receiver model in the room acoustic modelling software CATT [10].

2. METHOD

Speech intelligibility was assessed with a word recognition test using Hagerman sentences [11] with continues speech as background sound. In conjunction the perceived effort was rated on an eleven-point scale. Stimuli were generated in a classroom model using CATT.

2.1. Subjects

Twelve hard of hearing students, six girls and six boys, participated in the study. The age of the subjects ranged from 10 to 15 years (mean 12.3 years). The participating students were fitted bilaterally and their unaided best ear pure-tone average (PTA, average of the hearing levels at 0.5, 1, and 2 kHz) ranged from 31-75 dB (mean 52 dB, SD 14 dB). Three subjects had an interaural asymmetry equal or more than 10 dB (average of the hearing levels at 0.5, 1, 2, and 4 kHz [12]).

2.2. Room acoustic model

A rectangular classroom 7x8.5x3 m was modelled in CATT, see Figure 1. The target source and the receiver were placed on opposite sides of a table, at a distance of 1.3 m and 0 degrees azimuth. Two masker sources were positioned at a distance of 2.6 m and at 60 and 300 degrees azimuth, respectively. The room was furnished and absorption and diffusion coefficients were set to create normal room acoustic characteristics see reverberation times in Table 1. T-30 is derived from ray tracing using least-square fits to the decay in the interval -5 to -35 dB. T-30 was computed as an averaged over three microphone positions spread through out the room using an omni directional source in the middle. The Eyring formula based reverberation time, EyrT, is computed from the mean free path, calculated from all trays, and the mean of all absorption values encountered [10]. A large difference between T-30 and EyrT indicate a lack of diffuse reflection.

Two receiver models were utilized in CATT; one omni directional and one binaural. ITA kunstkopf artificial head was used as binaural model [10]. The directivity index of a singer [13] was applied to all sources, see Table 2.
Table 1: Reverberation time in the CATT modelled classroom. T-30 is derived from ray tracing and EyrT is based on Eyring formula.

<table>
<thead>
<tr>
<th>Octave (Hz)</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1k</th>
<th>2k</th>
<th>4k</th>
<th>8k</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-30 [s]</td>
<td>0.41</td>
<td>0.49</td>
<td>0.45</td>
<td>0.42</td>
<td>0.40</td>
<td>0.34</td>
<td>0.31</td>
</tr>
<tr>
<td>EyrT [s]</td>
<td>0.33</td>
<td>0.35</td>
<td>0.33</td>
<td>0.28</td>
<td>0.23</td>
<td>0.19</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 2: Directivity index of the sound sources by octave bands.

<table>
<thead>
<tr>
<th>Octave (Hz)</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1k</th>
<th>2k</th>
<th>4k</th>
</tr>
</thead>
<tbody>
<tr>
<td>DI (dB)</td>
<td>2.2</td>
<td>3.0</td>
<td>5.3</td>
<td>0.6</td>
<td>2.1</td>
<td>5.1</td>
</tr>
</tbody>
</table>

2.3. Stimuli

A Hagerman sentence consists of five words with the structure <name> <verb> <number> <adjective> <substantive> [11]. The target sentences were presented together with two continuous speech maskers. The continuous speech was a recording of a woman reading a continuous story from a fiction book. The Hagerman sentences and the speech masker were convoluted in CATT using the two receiver models. Three different parts of the recording were generated at both masker positions. Hence, six different pairs of speech maskers were used. The masker pairs were then added to the target sentences and, for both receiver models, the sound pressure levels were adjusted to create a 5 dB SNR. The sound pressure level of the binaural signal was
computed as the mean between left and right channel. For the omni directional receiver model the same signal was used in both left and right channel.

The two test conditions, the omni directional and the binaural receiver model, are denoted as diotic and dichotic, respectively.

2.4. Procedure

The listening experiment was carried out in a room at Hörcentralen at Alvikskolan in Stockholm. The room measured 2x2.8x1.9 m and data is specified in Table 2. The background noise level, 23 dB(A) and 57 dB(C), imply a presence of low frequency noise.

Table 2. Specifications of the room where the listening experiment was performed.

<table>
<thead>
<tr>
<th>Octave [Hz]</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1k</th>
<th>2k</th>
<th>4k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverberation time [s]</td>
<td>0.20</td>
<td>0.15</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>Background noise: 23 dB(A) and 57 dB(C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Instead of using T-mode, M-mode and cross-talk cancellation technique was used in order to enable dichotic listening. The two channel stimuli were presented, utilizing Lexicon MC-1, by two ADAM-S2A speakers in front of the listener. Subjects with the opportunity to differ the directionality of their hearing aid microphones were asked to use the omni directional feature.

Before the word recognition test, Hagerman sentences were presented to the subjects at 65 dB SPL. The purpose was to familiarize the subject with the testing material and a possibility to adjust the amplification. It should also be mentioned that before the listening experiment, the subjects participated in an interview study regarding their use and experience of ALD.

Three word recognition tests were conducted, each containing 10 Hagerman sentences. As previously described, this paper focuses on two of these tests; the diotic and dichotic presented stimuli. The third test was a Hagerman test with a noise masker and will not be discussed in this paper. The order of the three tests was randomly assigned to each subject. The six different pairs of speech maskers that were generated were used in random order. Two training sessions precede the three tests, where the stimuli consisted of five sentences together with speech masker and noise masker, respectively, at random order.

Each word recognition test was then carried out as follows. The masker was present at all time. Using a computer interface, the subject received a visual indication before each Hagerman sentence. After the last sentence, the subject rated the perceived effort on an eleven point scale.

3. RESULT

The results of the diotic and dichotic word recognition tests, respectively, are presented in Table 3. Response variables are percent corrected repeated words for all ten Hagerman sentences and rated perceived effort (eleven point scale). For each subject the paired difference was defined as the diotic test minus the dichotic test. One of the subjects with a moderate hearing loss was excluded in the analysis. The subject’s result was considered insufficient due to low word recognition.

The null hypothesis, i.e. no difference between diotic and dichotic test condition, could not be rejected with regard to both word recognition and perceived effort.
Table 3. Word recognition and perceived effort for the diotic and dichotic Hagerman test.

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
<th>[Min,Max]</th>
<th>t-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word recognition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diotic</td>
<td>0.86</td>
<td>0.80</td>
<td>0.17</td>
<td>[0.44,0.96]</td>
<td></td>
</tr>
<tr>
<td>Dichotic</td>
<td>0.90</td>
<td>0.84</td>
<td>0.15</td>
<td>[0.44,0.98]</td>
<td></td>
</tr>
<tr>
<td><strong>Perceived effort</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diotic</td>
<td>6</td>
<td>6.4</td>
<td>2.3</td>
<td>[3,10]</td>
<td>p=0.16</td>
</tr>
<tr>
<td>Dichotic</td>
<td>7</td>
<td>6.6</td>
<td>2.4</td>
<td>[3,10]</td>
<td>p=0.43</td>
</tr>
</tbody>
</table>

A one-way ANOVA was performed to see whether there were any significant effect due to PTA and interaural asymmetry. Concerning best ear PTA, the subjects were arranged in three groups; mild hearing loss PTA less than 40 dB, moderate hearing loss PTA between 40 and 70 dB, and severe hearing loss PTA greater than 70 dB. As previously described, interaural asymmetry was defined as PTA difference, average of the hearing levels at 0.5, 1, 2, and 4 kHz, equal or greater than 10 dB. The mean results are shown in Table 4.

Table 4. Table of means for word recognition and perceived effort due to different hearing loss characteristics.

<table>
<thead>
<tr>
<th>Hearing loss</th>
<th>Mild (N=2)</th>
<th>Moderate (N=6)</th>
<th>Severe (N=3)</th>
<th>Symmetrical (N=8)</th>
<th>Asymmetrical (N=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word recognition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diotic</td>
<td>0.94</td>
<td>0.74</td>
<td>0.82</td>
<td>0.83</td>
<td>0.71</td>
</tr>
<tr>
<td>Dichotic</td>
<td>0.90</td>
<td>0.83</td>
<td>0.83</td>
<td>0.84</td>
<td>0.85</td>
</tr>
<tr>
<td><strong>Perceived effort</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diotic</td>
<td>4.0</td>
<td>6.5</td>
<td>7.7</td>
<td>6.3</td>
<td>6.7</td>
</tr>
<tr>
<td>Dichotic</td>
<td>4.5</td>
<td>6.3</td>
<td>8.7</td>
<td>6.5</td>
<td>7.0</td>
</tr>
</tbody>
</table>

There were no significant differences in degrees of hearing loss. Though, the results indicate a higher rated perceived effort with increasing hearing loss. Concerning degrees of interaural asymmetry, there is a significant (p<0.05) difference in dichotic advantage between the groups. Further, the dichotic scores are significantly (p<0.05) better compared to diotic test condition for subjects with an asymmetrical hearing loss. There were no significant differences in perceived effort between the groups.

4. DISCUSSION AND CONCLUSION

The loss of dichotic bilateral advantage, i.e. head diffraction effects and binaural squelch, must be put in to respect discussing signal to noise ratio (SNR) advantages using an assistive listening device (ALD). However, it has been argued that diotic summation is more relevant concerning bilateral advantage than effects due to dichotic listening [7]. The reason is the reduction in dichotic effects, especially for individuals with a severe hearing impairment. On the contrary, bilateral advantage has been shown irrespective of hearing level [7] and, according to [1], bilateral advantage is greatest for those individuals with severe hearing impairment. Defining binaural squelch in respect to monaural listening, diotic summation is one component of binaural squelch [1]. In one study [14], hard of hearing subjects who obtained a 3 dB binaural advantage could not distinguish a
Doctic presented stimuli from a dichotic presented stimuli, which normal hearing subjects managed. The results, concerning the word recognition test, could be seen as a verification of the null hypothesis, i.e. no difference exists between the dotic and dichotic test condition. The result of bilateral advantage irrespective of hearing level was also supported. In order to draw any conclusion regarding the advantage using a binaural ALD, it would be preferable to test the effects of binaural squeal and dotic summation separately.

Degree of asymmetry has been shown not to affect the bilateral advantage [7]. Concerning word recognition scores, this study showed a significant difference in dichotic advantage between the groups of different degrees of interaural asymmetry. Subjects with an asymmetrical hearing loss had significant higher score when the stimuli were presented dichotically than when it was presented doticly.

Due to the reported bilateral advantage with respect to listening effort [9], one could assume a lower rating in perceived effort in the dichotic test compared to the dotic test. This advantage could not be seen in this study. There were no significant difference in perceived effort due to degree of hearing loss and asymmetry. The results indicate a higher rated perceived effort with increasing hearing loss both in the dotic and dichotic listening condition.

One must keep in mind the low number of subjects that participated. The, to the listeners, unfamiliar binaural processing in CATT must also be considered together with the use of cross-talk cancellation technique. Further, the modelled situation with a fix speaker position and two fix masker position could be discussed. Alternatively, the speech source and masker will alter more similar to a multi speaker situation where bilateral advantages are considerable.

5. REFERENCES

Paper IV
Self assessment of speech intelligibility listening to binaural recordings

Roger Johnsson\(^a\)
Johan Odelius\(^b\)
Arne Nykänen\(^c\)
Division of Sound and Vibration
Luleå University of Technology
971 87 Luleå
Sweden

ABSTRACT
Our ability to focus on a talker in an environment with several talkers is improved by our binaural hearing. The ability to listen to one talker is improved if the talkers are spatially well separated. A common problem with binaural recordings is front-back confusion, which can make it more difficult to focus on the target talker among several talkers. In this study the ability to localize a talker is compared to the ability to follow a talker in situations with several talkers. In the first part the subjects were asked to localize a talker in the horizontal plane listening to binaural recordings. In the second part of the study the subjects were instructed to assess their ability to follow a target male voice in presence of a masker consisting of two voices (male and female). Target and maskers were presented through separate loudspeakers in various positions. The same listening test was also performed with binaural recordings of the test environment using different artificial heads and in-ear recordings of humans. In both parts the binaural recordings were presented to the subjects through loudspeakers using cross-talk cancellation. Correlation between localization performance and self assessed speech intelligibility was analyzed.

1. INTRODUCTION
Binaural techniques have been used for recording and reproducing sounds for over two decades. It is today an established tool for product sound quality evaluation. The advantage with binaural recordings and reproduction is the capability to store and reproduce 3-dimensional sound fields with high realism with only two channels.

The auditory system benefits from having two ears (binaural) in tasks such as localization, detection (detect a sound signal within noise) and speech intelligibility (not only hear that some one speaks, but actually hear what the person says). This is due to the binaural hearing systems ability to suppress noise, reverberance and colouration to a certain extent\(^1\). What commonly is known as the cocktail-party effect is possible thanks to binaural hearing. Cocktail-party effect is our ability to concentrate on one talker in an environment with several talkers and discriminate the speech of this talker from the rest of talkers.

\(^a\) Email address: rojn@ltu.se
\(^b\) Email address: johode@ltu.se
\(^c\) Email address: arny@ltu.se
The idea behind the binaural technique is as follows: If signals can be recorded in the ears of the listener and later reproduced in such way that the reproduced signals at the eardrums are the same as they were, than the auditory impression (including spatial aspects such as distance and direction) is recreated. The reproduction is normally done through headphones as this gives a complete channel separation. The signals can also be played back through loudspeakers but then cross-talk cancellation is needed to cancel the sound from both loudspeakers reaching both ears. This cancellation is done by adding an artificial cross-talk to the binaural signals. The recording of binaural signals can be done with small microphones placed in the ears of the subject. However, it is often not a practical solution to make the recordings in the ears of the listeners. Instead the recordings are made with an artificial head (also called dummy head), which is a simplified model of an average human person. If the recordings are made with artificial heads the binaural signals are passed through equalizers that make the signals spectrally compatible with sounds recorded with conventional microphones. During the reproduction in headphones the signals are played back through equalized headphones to create the same signal in the listeners’ ear canals as would be presented in a real situation.

Even though the binaural techniques produce recordings with surprisingly natural reproduction they often suffer from localization errors. Common imperfections in the reproduced sound field are:
- front-back confusions – the predicted position of the sound is a mirror of the actual sound in the frontal plane, Figure 1
- distance errors – the sound is perceived too close (or even inside the head) or too far away
- elevation errors – the elevation of the sound source is perceived wrong

The localization performance of both artificial heads and recordings in humans’ ears have thoroughly been investigated and presented in the literature. The conclusion of these studies is that the localization performance for reproduced binaural recordings (independent if these recordings were made with artificial heads or done in the ears of humans) are worse than in a real life situation.

Speech communication is seldom done in quiet environments; often the communication is disturbed by other speech signals, much like a cocktail party. The speech intelligibility is affected by several factors such as; spatial separation, speech (and noise) spectra. In a study by Ericson and McKinley, using headphone listening, it was found that directional (the talkers were spatially separated) presentations of speech are more intelligible than diotic presentations. But an even larger effect was the gender of the talker and masker. If the talker and masker were of
different gender the speech intelligibility in general was higher than if the talker and masker were of same gender, Figure 2.

![Figure 2: Speech intelligibility for diotic, dichotic and directional presentations of two talkers in silence. The talkers were male and male (MM), female and female (FF) or male and female (MF).](image)

The aim of this study is to compare the subject’s ability to correctly localize the position of the source in a binaural recording with their ability to follow a talker in a binaural recording with several other talkers. The hypothesis behind the study is that subjects that are good at correctly localize sounds will be able to use this skill to spatially separate the talkers in a cocktail party situation and thereby improve their speech intelligibility.

## 2. EXPERIMENTAL SETUP

The study presented in this paper is based on two listening tests. In the first test the subjects’ ability to make a correct localization in the horizontal plane was investigated and in the second test the subjects were instructed to assess their ability to follow a target male voice in presence of a masker consisting of a male and a female voice. Both listening tests were based on binaural recordings using both artificial heads and in-ear recordings of humans. The recordings were reproduced in an anechoic chamber through loudspeakers using cross-talk cancellation. In the second listening test also real-life versions of the recorded signals were presented to the subjects.

### A. Recordings

All recordings used in this study were done with small microphones (Sennheiser KE-4-211-2) placed either in the ears of the artificial heads or the humans used for the recordings. The miniature microphones were mounted in the ears flush with the entrance of the blocked ear canal. Due to the size of the microphones and the cables it was not possible to use an earplug to get a perfectly sealed ear canal. Subjects with small ear canals could not be used due to the microphone size. The subjects had to be carefully chosen to obtain a fit that was as tight as possible. The recordings for the first listening test were done with two artificial heads (Head Acoustics HMS I and HMS III) and in the ear of 6 humans (HH 1-6). For the second listening test recordings were done with same artificial heads and in the ears of 3 of the humans (HH 2, HH 4 and HH 6) used in the first set of recordings.

The first set of recordings used for the localization test was made in the horizontal plane in steps of 15 degrees in a total of 24 positions. As source a small coaxial loudspeaker was used to
get the characteristics of a point source. The recorded signal was a male voice (the speech was created using a text-to-speech program) that said (in Swedish) *Hi, my name is Erik. Your task is to determine from what direction I’m speaking.* During the recordings the subject was standing on a turn table that was rotated in steps of 15 degrees. In 8 positions (in front and then in steps of 45 degrees in the horizontal plane) recordings were also made, later used in the first listening test for a short training session. In these angles the male voice was saying in what direction he was speaking from.

The second set of recordings was made using the set-up presented in Figure 3. The target talker was played in either loudspeaker 1 or 2 and the masker talkers were played in one of the loudspeakers numbered 2 to 5. The combinations of target and maskers are summarized in table 1. The level of all talkers were adjusted to have the same A-weighted sound pressure level, resulting in the level of the maskers being approximately 3 dB(A) higher than the target.

![Figure 3: The experimental setup for the recordings (and real-life version) used for the speech intelligibility assessment in the study. The grey non-filled loudspeakers were placed in the anechoic chamber but were not used neither for recordings nor for reproduction. Loudspeaker 6 was used for binaural reproduction but not for recording.](image)

**Table 1**: Combinations of target and masker talkers used for the speech intelligibility assessment. The numbering of the loudspeakers is presented in Figure 3.

<table>
<thead>
<tr>
<th>Test nr</th>
<th>Target talker played in loudspeaker</th>
<th>Masker talkers played in loudspeaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

**B. Subjects**

20 subjects (11 males and 9 females) participated in the listening tests. The same 20 subjects participated in both listening tests. All subjects are staff at Luleå University of Technology. The average age for male subjects was 37 years (standard deviation 9 years) and for female subjects 35 years (standard deviation 5 years). 16 of the subjects had earlier participated in other listening tests. None of the subjects had any known hearing loss and considered themselves to be normal hearing.
C. Experimental facilities
Both recordings and listening tests were carried out in an anechoic chamber, at Luleå University of Technology. The anechoic chamber was chosen because the reproduction was made through loudspeakers using cross-talk cancellation. On top of this the anechoic chamber offers an environment that is not disturbed by outside sounds and other stimuli that may influence performance.

D. Localization test
In the first listening test the subjects were instructed to judge the direction of the talker on a diagram, shown in Figure 4. The diagram had marks in the directions front, back, left and right. No information was given to the subjects about where the source could be in the horizontal plane. They were instructed to mark each talker position with a cross on the circle in the diagram. Each subject listened to only 12 of the 24 positions for each head in order to reduce the number of stimuli each subject listened to. This was done by randomly selecting for each angle from front to back either the recording with the talker to the left or right. If the recording with the talker in front or back of the subject should be used was randomized. 8 different heads and 12 (of 24 possible) angles give a total of 96 stimuli for each subject. The listening test began with a short introduction/training session with 8 stimuli in angle steps of 45 degree around the horizontal plane (the selected head in each angle was randomized).

E. Speech intelligibility test
The subjects were instructed to assess their ability to follow a target male voice in presence of a masker consisting of a male and female voice. Both masking talkers were reproduced from the same position. The masker position changed for the different tests according to table 1. The target talker began 5 seconds before the masker in order to make it easier for the subjects to identify the target voice. The total length of each stimulus was 20 seconds. A total of 5 heads and 6 different target/masker combinations plus one real life version (the talkers were reproduced in the loudspeakers in same way as they were binaurally recorded) of the target/masker combinations gave a total of 36 stimuli. The listening test began with a short session with 3 examples to let the subjects get used to the setup. The speech intelligibility was assessed on a scale from 0, not at all to 100, extremely difficult to follow the target talker, shown in Figure 4. The length of the scale was 100mm giving a rating between 0 and 100.

![Figure 4: The diagram used for the localization test of the talker (left) and the scale used to self assess the subject’s ability to follow the target talker (right).](image)
3. RESULTS

It was considered to be a correct localization if the distance between the stimuli angle and responded angle was less than 7.5 degrees (the angle distance between the stimuli was 15 degrees). The results for two of the heads are shown in Figure 5. The abscissa gives the stimulus position and the ordinate gives the responded position. The dots represent the answers and the area of a dot is proportional to the number of answers for the combination of stimulus and response. Correct answers are found on the positive diagonal (solid line). Front-back confusions are found on the diagonals indicated by the dashed lines. Due to the randomization of the selected angles for the listening test not all angles has the same number of stimulus.

![Figure 5: Result of the localization test for all subjects for two of the heads used in the test (240 stimuli in each diagram). To the left are results for in-ear measurements of human 2 and to the right for artificial head HMS I.](image)

Horizontal angular errors for the six in-ear recordings of human heads (HH 1-6) and the two artificial heads (HMS I and HMS III) are showed in Figure 6. Due to the skewed distribution Friedman’s ANOVA was used to analyze the data. A significant difference was found between the eight recordings (χ²(7)=24.08, p<.05). Post-hoc tests were performed for the three human heads; HH 2 (Mdn=16), HH 4 (Mdn=24.5) and HH 6 (Mdn=30), and the two artificial heads; HMS I (Mdn=18) and HMS III (Mdn=22), used in the speech intelligibility test. Wilcoxon signed rank test was used for the post-hoc tests where the significance level was adjusted using the Bonferroni correction, i.e. p<.005. The angular error was significantly lower for HH 2 than for HH 6 (z=-3.17, p<.005, r=-.20). The angular error for HH 4 was neither found larger nor smaller when compared to HH 2 and HH 6, respectively. There were further no differences found for the two artificial heads (HMS I and HMS III), neither compared to each other nor to HH 2, 4 and 6.

Front-back errors, showed in Figure 7, were also analyzed. However, no differences could be found between the eight recordings (χ²(7)=7.91, p=n.s.).
Figure 6: Boxplots of angular error for the six in-ear recordings of human heads (HH 1-6) and for the two artificial head recordings (HMS I and HMS III).

Figure 7: Boxplots of front-back error (percentage) for the six in-ear recordings of human heads (HH 1-6) and for the two artificial head recordings (HMS I and HMS III).

Ratings in the self-assessed speech intelligibility test are shown in Figure 8. The ratings were normalized with respect to each subject mean and standard deviation:

\[
z_i = \frac{(x_i - m_i)}{s_i} - m + s
\]

where
- \(z_i\): normalized result
- \(x_i\): ratings for subject \(i\)
- \(m_i\): mean rating of subject \(i\)
- \(s_i\): standard deviation for subject \(i\)
- \(m\): mean rating of all subjects
- \(s\): standard deviation for all subjects
As for angular errors, Friedman’s ANOVA was performed using Wilcoxon signed rank test as post-hoc tests where the significance level was adjusted using the Bonferroni correction. A significant difference among the six cases was found ($\chi^2(5)=45.78, p<.05$). Rated difficulty to follow the target was significantly lower for the real life case when compared to the five human head recordings and the artificial head recordings, respectively. It was further found easier to follow the target for HH 2 when compared to HH 4 ($z=-4.07, p<.0033, r=-.37$) and HH 6 ($z=-3.62, p<.0033, r=-.33$). The artificial head HMS III was also rated easier than HH 4 ($z=-4.19, p<.0033, r=-.38$) and HH 6 ($z=-3.14, p<.0033, r=-.28$).

![Figure 8: Boxplots of normalized ratings of the difficulty to follow a target talker for the three in-ear recordings of human heads (HH 2, HH 4 and HH 6), the two artificial head recordings (HMS I and HMS III) and for the real life case.](image)

Spearman rank correlations ($\rho$) were computed between front-back errors and self-assessed speech intelligibility (not normalized ratings) in test 3 (target at 0º and masker at 180º) and test 5 (target at 30º and masker at 150º). The correlations were computed to test the hypothesis that front-back errors make a listening situation in a cocktail party scenario more difficult. A significant correlation was only found for HH 4 ($\rho=0.49, p<.05$) and HH 6 ($\rho=0.52, p<.05$). For HH 4 and HH 6 subjects with many front-back errors also rated their ability to follow the target as more difficult. The speech intelligibility ratings were further correlated with angular errors. However, no significant correlation could be found for any of the three in-ear recordings of human heads and the artificial head recordings.

### 4. DISCUSSION AND CONCLUSIONS

From the localization test it was found that the angular error for HH 2 was less than for HH 6. It was also found that the assessed speech intelligibility was rated higher for HH 2 than HH 6. No such correlation could be found for the rest of the in-ear recordings of the humans or the artificial heads. The speech intelligibility was rated higher for both HH 2 and HMS III compared to HH 4 and HH6. For HH 4 and HH 6 there was also a correlation found between the number of front-back errors for the subjects and their assessed speech intelligibility. This together indicates that the localization performance (including front-back errors) has an influence on the speech intelligibility in situations similar to a cocktail party.
However, the lack of correlation between localization and speech intelligibility for both in-
ear and artificial head recordings imply that the speech intelligibility is not that sensitive to the
ability to correctly localize the target. Instead factors that cannot be explained by position, such
as the spectra of the talkers, have a large influence on the speech intelligibility.

It has been shown that there are differences in the subjects’ ability to follow a target talker
between the binaural recordings done with artificial heads and in-ear recordings of humans. This
even though there was no difference in the number of front-back errors between the artificial
heads and in-ear recordings of the humans. Even for the best binaural recording (HH 2) the self
assessed speech intelligibility was lower than for the real life case.

To investigate if the cross-talk cancellation used was insufficient in presenting correct
interaural level differences (ILDs) and interaural time differences (ITDs), which could explain
the low correlation between localization performance and assessed speech intelligibility, the
study should be repeated using equalized headphones.

REFERENCES

3. Jens Blauert, Klaus Genuit “Evaluating sound environments with binaural technology – some basic considerations,”
83-100 (1999)
Chap. 32 in Binaural and spatial hearing in real and virtual environments, edited by Robert H. Gilkey and
Timothy R. Anderson (Lawrence Erlbaum Associates Inc, Mahwah, 1997)
8. ITU-R BS.1116-1, “Methods for the subjective assessment of small impairments in audio systems including
multichannel sound systems,” (1994)
Paper V
Binaural speech intelligibility in free-field and reverberation: A comparison of headphone and cross-talk cancellation reproduction

Johan Odelius, Roger Johnsson, and Örjan Johansson
Division of Sound and Vibration, Department of Human Work Sciences
Luleå University of Technology

Abstract
Binaural speech intelligibility was compared for headphone and cross-talk cancellation reproduction. Speech reception threshold (SRT) in noise was measured in normal hearing subjects. Spatial release from masking (SRM) was computed as the difference in SRT when both the speech and noise were presented in front and when the speech was presented from the front and the noise was presented from the right side. The comparison was performed separately for a virtual free-field and a room condition. The room stimulus was either an auralization using room acoustic modelling software or an artificial head recording in the actual room. The room (271 m$^3$) had a reverberation time of 0.7-0.8 s and the listening distance was 2.5 m. Cross-talk cancellation reduced SRM when compared to headphone reproduction in the free-field condition from 9.3 dB to 5.8 dB and in the room condition from 3.0 dB to 1.5 dB. No difference was found between the auralization and the artificial head recording. It was concluded that binaural cues are distorted using cross-talk cancellation and, in spite of good horizontal localization performance, binaural advantages in speech intelligibility are limited.

1 Introduction
The spatial separation of speech and noise improves speech intelligibility [1]. The improvement is attributed to two factors: head shadow and binaural processing. The head shadow effect is due to the possibility to attend to the ear which has the better signal-to-noise ratio (SNR). Binaural processing relies on the brain taking advantage of interaural differences. The spatial release from masking (SRM) can be measured as a SNR threshold difference in a word recognition task, the speech reception threshold in noise (SRT). The difference in SRT when speech and noise are presented from the front (S0N0) and when noise is presented from the side (S0N90) is in free-field about 8-10 dB (see reviews by [2, 3]). Different test procedures, type of speech material (numbers, monosyllables or sentences), and type of noise masker affect the result. SRM will hereafter be used as abbreviation for the intelligibility level difference between the S0N0 and S0N90 conditions. SRM$^{90}$ will be used if the 90-degree setup refers to the left rather than to the right side. To separate the contribution of head shadow and binaural processing, the SRT can be measured in the setup S0N90 by blocking the ear directed towards the noise source (S0N90B). The difference between the blocked monaural case and its binaural equivalence quantifies the binaural
processing and is denoted as binaural intelligibility level difference (BILD). The BILD is typically 3-4 dB [3]. The effects of binaural processing and head shadow are not additive. Bronkhorst and Plomp [4] studied the individual contribution of interaural time difference (ITD) and interaural level difference (ILD). They found a SRM of 5.0 dB and 7.8 dB for stimulus with non-zero ITDs only and ILDs only, respectively. When stimulus contained both ITDs and ILDs, the SRM was 10.1 dB.

In more realistic reverberant environments, listener performance will deteriorate and the binaural gain will decrease; in particularly, the contribution of head shadow decreases [2]. The reverberant/room condition is most often defined by its reverberation time. For binaural hearing, however, the placement and especially the distance between source and listening positions are more important [5, 6]; this can be quantified by the direct to reverberant ratio. The effect of head shadow relies on the listener to be in the direct field of the source, i.e. within the critical distance \( r_c = 0.1 \sqrt{\frac{V}{T}} \), where \( V \) is the volume, \( T \) is the reverberation time, and \( \Gamma \) is the directivity gain. Peutz [7] defined a distance where speech reception (articulation loss of consonants) no longer depends on the distance between source and listener. The distance, defined as \( r_{al} = 0.2 \sqrt{\frac{V}{T}} \), is approximately 2.5 times the critical distance for \( \Gamma = 2 \) (half spherical source). Koehnke and Besing [8] found a decrease in SRM from 11.7 dB to 3.5 dB between a free-field condition and a room condition (SRM decreased from 15.7 dB to 4.6 dB). The conditions were simulated using artificial head measurements of source-to-eardrum transfer functions (source to microphone distance not specified). The reverberation time in the room (\( V = 84.1 \text{ m}^3 \)) was 0.4 s below 800 Hz and 0.25 s above 800 Hz. In a listening experiment on-site, Bronkhorst and Plomp [9] obtained SRM of 5.4 dB and SRM* of 6.1 dB. The listening position \( (r=0.8 \text{ m}) \) was inside the critical distance in a room \( (V=196.4 \text{ m}^3 \) with a reverberation time of 0.9 s. As previously discussed, the authors found a SRM of 10.1 dB in free-field [4]. Nabalek and Pickett [10] reported a BILD of 5 and 3 dB for reverberation times of 0.3 and 0.6 s, respectively. In their study, the speech and the noise sources were at ±30 deg azimuth and the listener sat in the actual room at a distance of 3.4 m (11 ft), which was just inside the distance defined by Peutz for \( T=0.3 \) s and outside for \( T=0.6 \) s.

In binaural synthesis, measured head related transfer functions (HRTFs) are used to add binaural cues as well as monaural cues to the sound [11, 12]. Using non-individualized HRTFs, such as an artificial head, will introduce disparities in ITD and ILD. Localization performance is decreased, and front-back reversals is increased [13, 14, 15, 16]. For front-back reversals, the importance of high frequencies (above 7-8 kHz) is accentuated [17, 18]. In addition, the possibility of head movements is lost in binaural synthesis. Head movements are often unconsciously used by listeners to resolve front-back confusion [11]. While localization performance is degraded using non-individualized HRTFs, Hawley et al. [16] found no difference in speech intelligibility between listening in the actual sound field and listening to artificial head recordings presented with headphones. The study was performed in reverberation \( (T\geq0.3 \text{ s}) \). The listener position was at a distance of 1.5 m (5 ft) from the speakers, a position that indicates the listener sat in the direct-field (complete acoustical data not presented). Besides the direct to reverberation ratio, other factors such as the precedence effect must also be considered in a reverberant condition. The precedence effect contributes both to localization and speech intelligibility, e.g. the Haas effect and dereverberation [19]. The Haas effect refers to the advantage rather than the interference of early reflections to speech perception. The speech is also very sensitivity to
reverberation characteristics, especially when compared to music perception. The effect of dereverberation (decolouration) is clearly heard comparing on-site listening to a room recording [19]. Hence, disparities in ITDs and ILDs may have a greater affect on speech intelligibility in a reverberant condition. In contrast, reverberation adds noise to ITD and ILD [20], a disparity that would not affect performance because of the higher variability of the ITDs and ILDs.

The preservation of ongoing ITDs and ILDs is also a topic when reproducing binaural stimuli. A common reproduction approach is to use headphones, which enables an almost complete channel separation. Adequate playback equalization is required as the use non-individualized equalization filters may cause high frequency distortions [12]. Another proposed approach is to use loudspeakers and a cross-talk cancellation system (see e.g. [21, 22]). The approach can be beneficial, for instance, when presenting stimuli to subjects wearing hearing aids or in a blind comparison between a binaural synthesis and a "real case". Sound images are also perceived outside the head [21], which can be compared to headphone listening where the proportion of sound images perceived inside the head is some 30% at 0 deg azimuth for HRTF stimuli [23]. The amount decreases when listening to reverberant stimuli and can also be reduced after extensive training [24]. The performance of the cross-talk cancellation – i.e. the cancellation of left ear stimulus to the right ear and vice versa – relies on accurate HRTFs. Akeroyd et al. [25] showed a decrease in cancellation from 25 dB with matched HRTFs (setup and playback) to 13 dB with mismatched HRTFs. Mismatched HRTFs also led to inaccurate reproduction of ILD and ITD (std of 4 dB and 100 μs, respectively). That is, binaural cues are not accurately recreated using cross-talk cancellation if the HRTFs used in the computation do not match the individual HRTFs of the listener.

This study quantifies differences in binaural speech intelligibility between cross-talk cancellation and headphone reproduction. The amount of SRM was measured in a virtual free-field and a virtual room condition. HRTFs of an artificial head were used to render the free-field condition, and room acoustic simulation (auralization) was used to render the room condition. In addition to the auralization, an artificial head recording was taken in the actual room and used to define a further condition. The purpose was to examine the interaction between Method (auralization vs. recording) and Reproduction (cross-talk cancellation vs. headphones). To further compare the auralization with the artificial head recording, the BILD was measured when using headphones.

The binaural performance of the cross-talk cancellation system was defined by measuring the values of ITD and ILD that it delivered. Previous data on talker localization [26], using the same experimental setup, is also presented to uncover any difference in reproduction technique to the binaural and localization performance of the cross-talk cancellation system.

2 Method

The experimental design is presented in Table 1. The experiment contained two sessions: free-field and room. In each session, both headphone and cross-talk cancellation reproduction were tested – a within subject design. In the room, session subjects were either assigned the auralization or the artificial head recording – a between subject design. For the different setups, the speech reception threshold in noise (SRT) was measured using the Hagerman sentence test (40% threshold detection) [27, 28].
Table 1. Experimental design. The blocks HP and CTC refer to headphone and cross-talk cancellation reproduction, respectively. The spatial release from masking (SRM) is the difference in the speech reception threshold in noise (SRT) between S0N0 (the speech and the noise are presented from the front) and S0N90 (the speech is presented from the front and the noise is presented for the side). The binaural intelligibility level difference (BILD) is the SRT difference between S0N90B (the channel attending to the noise source is blocked) and S0N90.

<table>
<thead>
<tr>
<th>Session</th>
<th>Free-field</th>
<th>Auralization</th>
<th>Room</th>
<th>Recording</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>HP</td>
<td>CTC</td>
<td>HP</td>
<td>CTC</td>
</tr>
<tr>
<td>Measure</td>
<td>SRM BILD</td>
<td>SRM</td>
<td>SRM</td>
<td>BILD</td>
</tr>
<tr>
<td>Setup</td>
<td>S0N0</td>
<td>S0N90</td>
<td>S0N0</td>
<td>S0N90B</td>
</tr>
</tbody>
</table>

2.1 Subjects

In total, 23 normal hearing subjects aged 22-49 year (M=32) participated in the experiment. All subjects performed the free-field session. Within one week and with a minimum pause of 24 h, the subjects returned to perform the room session. Half of the subjects were assigned (randomly selected) the auralization and the other half the artificial head recording. There was a drop out of five subjects in the second session, so nine subjects performed the test for recording and nine for auralization.

2.2 Procedure

The listening experiment was carried out in an anechoic chamber. As previously described, subjects performed the experiment in two sessions (free-field and room). Each session consisted of two test blocks (Table 1): (1) The headphone test block where both the SRM and the BILD were measured (the SRT tests were performed for the setups S0N0, S0N90, and S0N90B); and (2) the cross-talk cancellation test block where only the SRM was measured (the SRT tests were performed for the setups S0N0 and S0N90). The order of the two blocks was random as was the tests within each block.

Each session started with a training exercise to introduce the subjects to the test material and procedure. The training exercise consisted of 10 sentences. Speech and noise was presented in front (S0N0) by the same reproduction technique as were to be used in the first test block. The first four sentences, where the first two sentences did not contain any noise, were intended to familiarize the subjects with the character of the speech and noise material. For the last six sentences, the subjects were asked to repeat each five-word sentence. The SNRs were fixed according to Table 2.

Table 2. SNRs (dB) in the training exercise. The first two sentences did not contain noise.

<table>
<thead>
<tr>
<th>Sentence No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR (dB)</td>
<td></td>
<td></td>
<td>20</td>
<td>20</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>-5</td>
<td>-8</td>
<td></td>
</tr>
</tbody>
</table>
After the training exercise, the word recognition tests were initiated. Each test consisted of 14 sentences. For the first four sentences fix SNRs was used as presented in Table 3. After the fourth sentence, or earlier, as soon as two correct answers or less were obtained, an adaptive stepping method was used (Table 4; see [28] for procedures). The SNR threshold when 40% recognition was reached (the SRT in noise) was computed as the average of the adjusted SNR for the last ten sentences (no 5-14).

<table>
<thead>
<tr>
<th>Table 3. Fix SNRs (dB) for the first four sentences preceding the adaptive procedure.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence No.</td>
</tr>
<tr>
<td>Free-field</td>
</tr>
<tr>
<td>SnN0</td>
</tr>
<tr>
<td>SnN90</td>
</tr>
<tr>
<td>SnN0B</td>
</tr>
<tr>
<td>Room</td>
</tr>
<tr>
<td>SnN0</td>
</tr>
<tr>
<td>SnN90</td>
</tr>
<tr>
<td>SnN0B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4. Change of noise level according to the adaptive procedure [28].</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of correct words</td>
</tr>
<tr>
<td>Change of noise level (dB)</td>
</tr>
</tbody>
</table>

2.3 Stimuli

The Hagerman test material contains lists of female spoken sentences and slightly amplitude modulated (2.1 Hz) speech spectrum noise [27]. The sentences are low redundant and consist of five words and have the syntactical form: Name, verb, number, adjective, and substantive.

In the free-field condition, the test material was convolved with measured HRIRs (head related impulse responses, 512 bins with sample rate 44.1 kHz) of Head Acoustics HMS III artificial head. The measurement was performed in an anechoic chamber at 314 positions1.

The room condition was a classroom (V=271 m³) with a reverberation time of 0.7-0.8 s (Table 5). Room dimensions together with listening position as well as speech and noise source positions are presented in Figure 1. The distance between sources and listening position was 2.5 m. The critical distance and the direct-to-reverberant ratio ($C_{50}$, at the listening position) are shown in Table 5. The quantities were computed with respect to the directivity of the source (Paradigm Titan v.2 loudspeaker); the directivity index is shown in Table 5. The listener position was inside Peutz’s distance ($r_{w}=3.7$ m) and also inside the critical distance for frequency above 1 kHz.

1 Elevation in step of 15 deg from the top down to 30 deg below the horizontal plane and azimuth in step of 15 deg decreasing as the elevation increases 15 deg above the horizontal plane.
Table 5. Acoustic properties of the room.

<table>
<thead>
<tr>
<th>Octave (Hz)</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1k</th>
<th>2k</th>
<th>4k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverberation time (s)</td>
<td>0.8</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>Critical distance (m)</td>
<td>1.3</td>
<td>1.6</td>
<td>1.9</td>
<td>2.5</td>
<td>2.9</td>
<td>2.7</td>
</tr>
<tr>
<td>Clarity, $C_{50}$ (dB)</td>
<td>3.3</td>
<td>4.1</td>
<td>5.0</td>
<td>6.6</td>
<td>6.3</td>
<td>6.6</td>
</tr>
<tr>
<td>Source directivity, DI (dB)</td>
<td>2.1</td>
<td>3.5</td>
<td>4.7</td>
<td>6.9</td>
<td>8.5</td>
<td>7.6</td>
</tr>
</tbody>
</table>

Figure 1. Source and receiver positions in the virtual room as well as the positions used in the artificial head recordings.

The room stimuli were created as follows. **Auralization:** The test material was convolved with binaural room impulse responses generated in a room acoustic model (CATT Acoustics [29]). The room simulation is based on the image source model for early reflections and a randomized cone-tracing approach for late reflections. The room model was designed to match geometrical and acoustical characteristics of the classroom. Source and receiver data were included: measured directivity of the Paradigm Titan v.2 loudspeaker and the measured HRIRs of the HMS III artificial head. **Recording:** The test material was played back using the Paradigm Titan v.2 loudspeaker in the classroom and binaurally recorded with the HMS III artificial head.
Reproduction
For headphone reproduction, the Head Acoustics HPS IV system was used. Built-in free-field and
diffuse-field headphone equalization filters were used for the free-field and room condition,
respectively.

The cross-talk cancellation system consisted of Lexicon MC-1 and two loudspeakers (flat ±2 dB
response within 100 Hz - 10 kHz). The loudspeakers and the listening position were placed in an
equilateral triangle (length 2.5 m).

A speech presentation level of 60 dB SPL at the listening position was chosen for both the free-
field and room condition. An artificial head recording of the noise stimulus (speech spectrum
noise) in the listening position was performed to ensure the same presentation level for the cross-
talk cancellation system as when using headphones.

3 Binaural performance of the cross-talk cancellation
system
The binaural performance of the cross-talk cancellation system was measured in the same setup
used in the listening experiment. White noise was convolved with the measured HRIRs of the
HMS III artificial head at azimuth of 90 deg (0 deg elevation). Interaural differences (ITDs and
ILDs) of the test signal (target) were compared to an artificial head recording in the listening
position of the cross-talk cancellation setup (delivered). The ITD was computed as the maximum
of the cross-correlation function between the left and right channel. The sound pressure level of
left and right channel was taken as the ILD. The sensitivity for disparities in ongoing ITDs is
larger for low frequencies. For high frequencies, the sensitivity of ongoing ITDs is carried by the
envelope [30]. Accordingly, for high frequencies, envelope ITDs and envelope ILDs were
computed. The signal envelope was computed as the complex modules of the Hilbert function.
The performance for low frequencies (≤1 kHz) was compared per auditory filter frequency and is
shown in Figure 2. The result for high frequencies (≥1 kHz) is presented per 1/3 octave band in
Figure 3.

The cross-talk cancellation system did not deliver correct ITDs and ILDs for frequencies below
500 Hz. Even negative ILDs were delivered: higher levels were measured at the contralateral ear
to the source. For frequencies between 500 Hz and 1 kHz, the system was accurate, especially for
ITDs. For high frequencies, envelope ILDs were delivered fairly accurately for frequencies up to
5 kHz. Above 5 kHz, where the ILD increases above 12 dB, the system delivered ILDs of some 6
dB. Envelope ITDs were not delivered accurate for high frequencies, especially above 3 kHz.

3.1 Localization
In a previous study, horizontal localization was evaluated for the same cross-talk cancellation
setup used in present study [26]. A sentence was recorded from 24 angles, azimuths 0 to 345 deg
in steps of 15 deg. The recordings were then reproduced using cross-talk cancellation and 20
normal hearing subjects were asked to arbitrarily localize the sounds in the horizontal plane. The
experiment included several artificial heads and in-ear measurements of humans. Only data for
the HMS III artificial head are presented below.
Figure 2. Delivered ITDs and ILDs of the cross-talk cancellation system at 90 deg azimuth as a function of auditory-filter frequency. (o) target stimuli and (*) artificial head recording in the listening position.

Figure 3. Delivered envelope ITDs and envelope ILDs of the cross-talk cancellation system at 90 deg azimuth as a function of 1/3 octave band frequency. (o) target stimuli and (*) artificial head recording in the listening position.
Mean absolute error (MAE) together with spherical statistics – judgment centroid and inverse kappa ($\kappa^{-1}$) – were used to analyze the data. The judgement centroid is the average direction of the localization judgments. It is the angle of the resultant, the vector sum of all unit length judgments vectors. $\kappa$ is estimated from the length of the resultant and is used as a measure of dispersion. Since scattered judgments produce small $\kappa$ (short resultant), $\kappa^{-1}$ is reported. For further details on these quantities, see [31]. Front-back reversals were resolved from the data and presented separately since these unfairly inflate the results. Front-back reversals were defined as when judgments lay on the opposite side of the vertical interaural plane from the target. Targets of $\pm 90$ deg were excluded and judgements of $\pm 90$ deg were considered to be in the same interaural plane as the target. Table 6 shows the descriptive statistics for a selection of target angles.

**Table 6.** The localization judgments presenting stimuli using the cross-talk cancellation system.

Means of percentage front-back reversals and reversals-corrected judgment of centroids, dispersion ($\kappa^{-1}$), and mean absolute error (MAE). Values are collapsed across subjects and shown for a selection of target angles [26].

<table>
<thead>
<tr>
<th>Target (deg)</th>
<th>Front-back (%)</th>
<th>Centroid (deg)</th>
<th>Reversals-corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\kappa^{-1}$</td>
</tr>
<tr>
<td>0</td>
<td>20%</td>
<td>-1</td>
<td>0.017</td>
</tr>
<tr>
<td>30</td>
<td>25%</td>
<td>43</td>
<td>0.074</td>
</tr>
<tr>
<td>60</td>
<td>10%</td>
<td>72</td>
<td>0.116</td>
</tr>
<tr>
<td>90</td>
<td>n/a</td>
<td>81</td>
<td>0.088</td>
</tr>
<tr>
<td>120</td>
<td>82%</td>
<td>120</td>
<td>0.069</td>
</tr>
<tr>
<td>150</td>
<td>38%</td>
<td>139</td>
<td>0.044</td>
</tr>
<tr>
<td>180</td>
<td>90%</td>
<td>179</td>
<td>0.017</td>
</tr>
<tr>
<td>-150</td>
<td>58%</td>
<td>-142</td>
<td>0.093</td>
</tr>
<tr>
<td>-120</td>
<td>33%</td>
<td>-106</td>
<td>0.083</td>
</tr>
<tr>
<td>-90</td>
<td>n/a</td>
<td>-78</td>
<td>0.093</td>
</tr>
<tr>
<td>-60</td>
<td>0%</td>
<td>-77</td>
<td>0.093</td>
</tr>
<tr>
<td>-30</td>
<td>8%</td>
<td>-43</td>
<td>0.059</td>
</tr>
<tr>
<td>Mean</td>
<td>36.4%</td>
<td>-</td>
<td><strong>0.070</strong></td>
</tr>
</tbody>
</table>

**Front-back reversals**

Mean percentage of front-back reversals was 36.4%. This result is similar to previous studies on non-individualize HRTFs filtered stimuli reproduced with headphones: speech target 29% [23] and wideband noise burst 31% [13]. A noticeable difference, however, lies in type of front-back reversals. For current cross-talk cancellation system, front to back reversals (13%) were less frequent than back to front reversals (62%), as compared to reported headphone studies where front to back reversals were more frequent than back to front reversals (47% and 11% [23], and 50% and 12% [13], respectively). In addition to headphone reproduction, Wenzel et al. [13] assessed localization in on-site free-field, where the front-back reversal rate was 19% (34% front to back, 4% back to front). However, large variations existed among the subjects. Wightman and Kistler [31] used individualized HRTFs in virtual localization (with headphones) and in free-field listening. They found reversal rates of some 3% in free-field and some 6% using headphones (for stimuli with elevation 0 to 18 deg up). The results are also supported in studies by Møller et al.
their study found that localization errors – e.g. median plane errors – increased when using non-individualized HRTFs or artificial heads when compared to using individualized HRTFs or in a real-life situation.

Angle error
The size of the MAE and dispersion ($\kappa^{-1}$) were smaller than localization performance using headphones [13, 23, 31], but can be considered comparable since present data are only in the horizontal plane. The localization performance was best at 0 deg and 180 deg. Judgements of ±90 deg were in front of the listener, centroids of 81 deg and -78 deg, respectively.

4 Results
The results of the Hagerman SRT test, the SNR threshold for 40% word recognition are presented in Figure 4. Due to the threshold detection of 40%, 0.6 dB lower values are obtained when compared to the standard detection of 50% [28]. For all conditions, the SRT was approximately -8 dB when the speech and noise was presented in front (S0N0). When the speech and noise was spatially separated (S0N0 and S0N90B), the performance in the Room Auralization and Room Recording decreased (Figure 4). The experiment in the room condition was carried out after the free-field condition. Because the subjects had more experience at this point in the experiment, their word recognition performance may have improved.

4.1 Spatial release from masking (SRM)
The SRM was computed as the difference in SRTs between the setups S0N0 and S0N90. Data for the 18 complete cases were considered: free-field (N=18), auralization (N=9), and artificial head recording (N=9). Means and 95% confidence intervals are shown in Figure 5.
In the free-field condition, a repeated-measures ANOVA was used to test for SRM differences between the two reproduction approaches: headphones and cross-talk cancellation. The SRM was reduced from 9.3 dB using headphones to 5.8 dB using cross-talk cancellation, $F(1,17)=129.80$, $p<.05$.

In the room condition, a mixed model ANOVA was performed on the SRM with regard to repeated-measures variable Reproduction (headphones and cross-talk cancellation) and between-group variable Method (auralization and artificial head recording). An effect of Reproduction was found; the SRM was reduced from 3.0 dB using headphones to 1.5 dB using cross-talk cancellation, $F(1,16)=34.35$, $p<.05$. There was no effect of Method $F(1,16)=0.86$, $p=\text{n.s.}$, or any interaction effect between Method and Reproduction, $F(1,16)=0.02$, $p=\text{n.s.}$.

![Figure 5](image.png)

**Figure 5.** Means and 95% confidence intervals (individual $s$) of SRM for headphone (HP) and cross-talk cancellation (CTC) reproduction.

### 4.2 Binaural intelligibility level difference (BILD)

In addition to the SRM, the BILD was measured when stimuli were reproduced by headphones. The BILD was computed as the difference in SRTs between the setups $S_{0N90B}$ and $S_{0N90}$. Means and 95% confidence intervals are shown in Figure 6. The BILD was 5.1 dB in the free-field condition. In the room condition, the BILD decreased to 1.5 dB and 1.7 dB for the auralization and the artificial head recording, respectively. An approximation of the intelligibility advantage due to head shadow was derived by subtracting the BILD from the SRM. As previously discussed, the effects due to ITD and ILD are not additive [4]. The head shadow effect decreased from 4.2 dB in free-field to 1.4 dB for the auralization and to 1.3 dB for the artificial head recording. Both BILD and head shadow decreased in reverberation and the reduction was equal regardless of method used to render the room condition.
Figure 6. Means and 95% confidence intervals (individual $s$) of BILD for headphone reproduction in the conditions: Free-field (FF), Room Auralization (RA), and Room Recording (RR).

5 Discussion and conclusions

In the free-field condition, the results (SRT, SRM, and BILD) for headphone reproduction agreed with previous studies in both virtual and actual free-field [2, 3]. The results in the room condition can be considered to be lower when compared to on-site listening, both with regard to the SRM [9] and to the BILD [10]. The result for the SRM was similar to the study by Koehnke and Besing [8], where a reverberant condition was synthesized and reproduced with headphones. The results suggest that use of non-individualized HRTFs (artificial head) have a deteriorating affect on binaural speech intelligibility in reverberation, but not in free-field. No differences between the auralization and the artificial head recording were found. Studies on auralization based on room acoustic modelling have shown some inconsistent results in terms of speech intelligibility. Kleiner [32] found lower speech intelligibility scores listening to an auralization when compared to on site listening, whereas no differences were found between an artificial head recording and on site listening. Yang and Hodgson [33] conclude that if the room is not too absorptive or noisy, speech intelligibility tests using auralization are reliable. The affect of non-individualized HRTFs to the precedence effect and speech intelligibility in reverberation is an interesting topic for future studies.

There is a training effect in SRT measurements, some 0.1 dB [28]. Reduced amount of sound perceived inside the head due to extensive training of listening to binaural synthesis has also been reported although the effect was more prominent for anechoic stimuli than for reverberant stimuli [24]. Any considerable training effect was not observed in present study; the SRT for the different setups were consistent across subjects (Figure 4).

The performance for virtual free-field using headphones is similar to the performance in the actual free-field [16]. Using cross-talk, cancellation binaural cues are distorted [25]. Present cross-talk cancellation system reproduced adequate ITDs for frequencies between 200 Hz and 1
kHz (Figure 2). Localization data [26] also showed performance similar to that of headphone listening (Table 6). Deviations of ILDs were found at low frequencies, especially below 400 Hz where higher levels were measured at the contralateral ear to the source. The deviations had a minor effect on the localization performance, but negative effects on speech intelligibility due to upward spread of masking can be expected. At frequencies above 4 kHz, envelope ITDs were inaccurately recreated (Figure 2). These errors add to the high frequency distortion using non-individualized HRTFs that, for example, causes high rates of front-back reversals as was observed in the localization experiment. The envelope ILDs were also delivered inaccurately at high frequencies, some 6 dB above 5 kHz. Hence, the head shadow effect was reduced. The reported errors in delivered ITDs and ILDs by the cross-talk cancellation system reduced the SRM from 9.3 dB to 5.8 dB in free-field and from 3.0 dB to 1.5 dB in reverberation. The findings show that binaural cues are not fully recreated using cross-talk cancellation, and by that the benefit from binaural advantage in speech intelligibility is limited. This finding emphasises that the localization performance in the horizontal plane is not a sufficient indicator for evaluating the binaural performance of a reproduction system.

6 References


