Human–Centered Design of E–Health Technologies:
Concepts, Methods and Applications

Martina Ziefle
RWTH Aachen University, Germany

Carsten Röcker
RWTH Aachen University, Germany

Medical Information Science
REFERENCE

MEDICAL INFORMATION SCIENCE REFERENCE
Hershey · New York

Senior Editorial Director:	Kristin Klinger
Director of Book Publications:	Julia Mosemann
Editorial Director:	Lindsay Johnston
Acquisitions Editor:	Erika Carter
Development Editor:	Myla Harty
Production Coordinator:	Jamie Snavely
Typesetters:	Mike Brehm and Milan Vracarich, Jr.
Cover Design:	Nick Newcomer

Published in the United States of America by
Medical Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Human-centered design of e-health technologies : concepts, methods and applications / Martina Ziefle and Carsten Röcker, editors.
p. cm.
Includes bibliographical references and index.
Summary: "This book unites researchers and industry practitioners from different disciplines to share their domain-specific knowledge and contribute to a holistic introduction into the area of human-centered design for e-health applications"--Provided by publisher.
ISBN 978-1-60960-177-5 (hardcover) -- ISBN 978-1-60960-179-9 (ebook) 1. Medical informatics. 2. Human-computer interaction. 3. Human engineering. I. Ziefle, Martina, 1962- II. Röcker, Carsten, 1975-
R858.H865 2011
610.285--dc22
2010048868

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 8

Personalized Acoustic Interfaces for Human–Computer Interaction

Jan Rennies
Fraunhofer IDMT, Hearing, Speech and Audio Technology, Germany

Stefan Goetze
Fraunhofer IDMT, Hearing, Speech and Audio Technology, Germany

Jens-E. Appell
Fraunhofer IDMT, Hearing, Speech and Audio Technology, Germany

ABSTRACT

The importance of personalized and adaptable user-interfaces has been extensively discussed (European Ambient Assisted Living Innovation Alliance, 2009; Alexandersson et al., 2009). However, it often remains unclear how to specifically implement such concepts. In the field of acoustic communication, existing models and technologies offer a wide range of possibilities. Based on these technologies, this chapter presents a concrete realization of a model-based interface in the field of acoustic human-computer interaction. The core element of the implementation is a holistic approach towards a hearing perception model, which incorporates information of the acoustic environment, the context and the user himself provides relevant information for control and adjustment of adaptable and personalized acoustic user interfaces. In principle, this way of integrating state-of-the-art technologies and models into user interfaces could be applied to other sensory perceptions as e.g. vision.

INTRODUCTION

Together with vision, hearing is the most important human sense. The ability to perceive sound enables us to locate and classify sound sources and forms the basis of our orientation and communication. Both in private and at work, speech communication is of utmost importance, and has been largely influenced by advances in modern technology. A wide range of applications is available to facilitate acoustic interaction between people, ranging from mobile communication devices to video-conferencing systems. In the past years, the prevalence of computer-based

DOI: 10.4018/978-1-60960-177-5.ch008

applications has given rise to an increased importance of human-machine interaction. Acoustic information transfer between users and computers can be bidirectional, i.e. the user can both receive acoustic signals (e.g. spoken information) and interact with the system, e.g. by speech commands. Voice-controlled systems are particularly useful when human-machine communication is needed in hands-free applications or when the user is unable to use other means of input. Many other fields in modern societies could benefit from a well-working acoustic human-computer interaction. So far, however, the particular needs of the individual users have not been carefully taken into account in the design process and practical application of acoustic user interfaces. Particularly, but not exclusively, the significant part of people with hearing deficiencies could benefit from adaptable and personalized acoustic user interfaces. In modern societies, hearing impairments are widely spread. Recent figures estimate that about 16% of the population in industrialized countries suffer from hearing deficiencies (Shield, 2006). Due to age related deterioration of nerve cells in the inner ear, this percentage is much higher in older subgroups of the population. Different estimates report that between 37% and 56% of the population aged 60 to 70 years suffer from hearing loss (Uimonen et al., 1999; Sohn, 2001; Davis, 2003; Johansson and Arlinger, 2003). In the light of the demographic change, the number of hearing-impaired people is expected to increase rapidly in the next ten years and to almost double by 2030 (Shield, 2006).

Given the growing importance of human-computer interaction, acoustic interfaces should be accessible to the whole population in as many applications as possible. Particularly for hearing-impaired people, but also for people having special needs in communication, like e.g. jet pilots, the interfaces have to be adaptable to different environments and situations and personalized individually.

This chapter proposes a perceptual approach, which aims at a concrete realization of a model-based interface in the field of acoustic human-computer interaction. The core element of the implementation is a holistic approach towards the inclusion of a hearing perception model, which incorporates information of the acoustic environment (e.g. reverberation time, damping of walls and ceilings), the current acoustic context (e.g. presence of noise sources), as well as information on the individual user himself (e.g. hearing loss). The model can thereby provide relevant information for the control and adjustment of the interfaces for user interaction.

This chapter is organized as follows. Firstly, the factors influencing acoustic communication in daily life are described and the particular difficulties for hearing-impaired users of acoustic interfaces are summarized. Secondly, state-of-the-art technologies to increase the accessibility of acoustic communication systems are surveyed and their possibilities and limitations to support acoustic human-machine interaction are discussed. It is shown that existing models and technologies already offer a wide range of support, but that a combined approach based on the individual perception of sound is needed to realize adaptable and personalized acoustic user interfaces. Such an approach is presented subsequently before specific applications are illustrated. In the end, the chapter is briefly summarized and concluded.

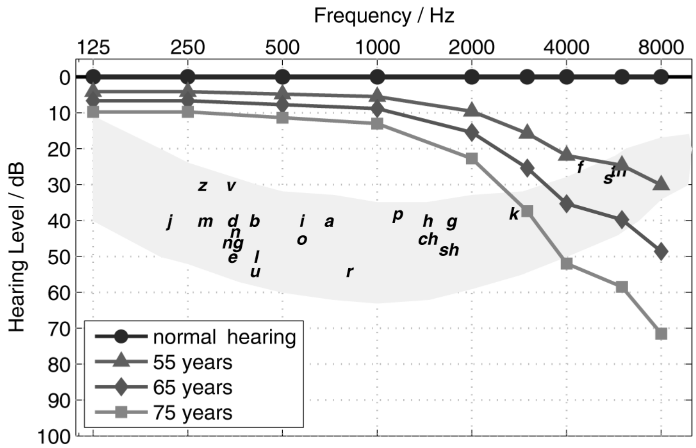
ACOUSTIC COMMUNICATION IN NORMAL AND IMPAIRED HEARING

Factors Influencing Communication Quality

Speech communication is the most natural way of information exchange. In everyday life, however, adverse acoustic factors such as background noise, competing speech or reverberation can reduce the quality of the communication or even prevent humans from understanding at all. Fortunately, the healthy hearing system is quite robust towards such difficult acoustic conditions. A well-known phenomenon is the fact that humans are able to focus on individual speakers, even when a lot of different competing sound sources in the same room superimpose with the speech signal. This ability of the hearing system is commonly referred to as the cocktail-party effect and has been subject to research for several decades now (e.g. Cherry, 1953; Bronkhorst, 2000). Several studies have shown that the auditory system can benefit from level and phase differences between the signals reaching the left and right ear. For example, a noise source from the right-hand side arrives earlier at the right than at the left ear. Additionally, the head serves as a physical barrier and, thus, the sound reaching the left ear will generally be softer than at the right ear. For comparison, a speech signal perceived from the front leads to no differences in arrival time or level between the two ears. If noise from the right and speech from the front are perceived at the same time, the normal-hearing auditory system can distinguish between the two sources since only one of them evokes differences at the two ears. This ability is very helpful in acoustic conditions with spatially distributed noise sources and supports our daily speech communication as well as our orientation. Despite the robustness of the healthy hearing system, the use of acoustic interfaces can be difficult in many situations even for normal-hearing people. For example, high levels of ambient noise in a driving car or a large amount of reverberation in a church lead to an increased effort for communication. Due to the increasing importance of human-machine interaction, technological support to ensure acceptable communication quality is highly desirable.

In cases when acoustic communication is difficult already for normal-hearing people, the problems resulting from a reduced functionality of the hearing system can be considerable. Hearing impairments exists in various facets and it is beyond the scope of this chapter to detail the effects of each facet and its impact on acoustic communication (see e.g. Moore, 1985, 1995, 1996). In general, two broad classes of hearing impairments are distinguished (Yost, 2000; American Speech-Language-Hearing Association, 2009). The class of conductive hearing losses is related to the physical property of the ear to transmit sound waves. For example, infections or disorders of the outer and middle ear or an obstructed ear canal impede the transmission of sounds and increase the so-called threshold of hearing, which describes the sound pressure level that is needed for a specific acoustic event to be perceived by the human listener. The other, more prevalent class of hearing impairments is based on deficiencies in the inner ear or even higher stages of the neural sound processing. These so-called sensorineural hearing losses are typically related to a loss of nerve cells which perform the transduction from physical oscillations to neural information. These cells deteriorate over time, resulting in the well-known age-related hearing loss. A further cause of irreversible cell deterioration is the exposition to high-level sounds over long periods of time (e.g. occupational noise or loud concerts, see Felchlin and Hohmann, 1997; Bormann et al., 2005). Depending on the kind and degree of hearing impairment, the use of speech communication can be significantly impeded. In contrast to visual impairment, which can be cured by purely passive devices such as lenses or glasses in most cases, even the most recent hearing aids cannot completely restore the functionality of the complex hearing system and its nerve cells. Therefore, especially under disadvantageous acoustic conditions, hearing-impaired people inevitably have problems in communication and orientation, since part of the speech information cannot be extracted from the signal anymore. The particular problems of hearing-impaired people are addressed in the following.

Figure 1. Audiogram of normal-hearing and hearing-impaired people suffering from age-related hearing loss (ISO, 1990). The gray area indicates the typical intensity distribution of speech across frequency



Effects of an Elevated Threshold of Hearing

Figure 1 illustrates the effect of hearing loss on speech intelligibility in audiogram representation. The audiogram shows the deviation from normal hearing, expressed as hearing level in decibels (dB), as a function of frequency. By convention, increasing hearing levels are plotted downwards. The audiogram gives insight into the individual frequency dependence of the hearing threshold. The measurement of the audiogram is the standard audiometric test procedure (Katz et al., 2009) and indicates which frequency components of sounds can be heard in the absence of any other sounds. All sounds below the curve in the audiogram are loud enough to be perceived. As indicated by circles in Figure 1, the normal hearing system is able to perceive sounds with a hearing level of about 0 dB and larger. There is some controversy as to the exact definition of normal hearing and hearing impairment, but from a clinical point of view, a deviation in the hearing threshold of up to 20 to 25 dB from the reference at 0 dB hearing level is still understood as being ’normal hearing’ (Martini, 1996). Higher hearing levels indicate a hearing loss.

The curves indicated with different markers in Figure 1 show average hearing losses for different age groups that have to be expected purely due to age-related loss of nerve cells (ISO, 1990). Since the curves show the median observed hearing loss at different ages, it should be pointed out that half of the population has a stronger hearing loss compared to this average. It can be seen that, particularly at high frequencies, hearing thresholds are elevated. The gray area in Figure 1 indicates the typical level range and frequency distribution of speech. Individual speech fragments are positioned according to their principal frequency content (note that neither of these fragments consists of a single frequency only). Given the average hearing losses, an increasing part of the speech information can no longer be perceived by older people and, for example, the distinction between ’f’, ’s’ and ’th’ is not possible any more. As described above, in addition to the age-related hearing loss, diseases or exposure to occupational or recreational noise over longer periods and conductive hearing losses may further increase the overall hearing loss and, thus, the loss of speech information in conventional speech communication. Hearing impairments as depicted in Figure 1 indicate that already in a quiet environment, parts of the speech information cannot be perceived when spoken at ordinary levels. Particular problems arise in non-optimal acoustic conditions when noise and reverberation are present. In such cases, communication can be very difficult and tiring for hearing-impaired people, even though their audiogram data might indicate only small deviations from normality.

Effects of a Reduced Dynamic Range and Loudness Recruitment

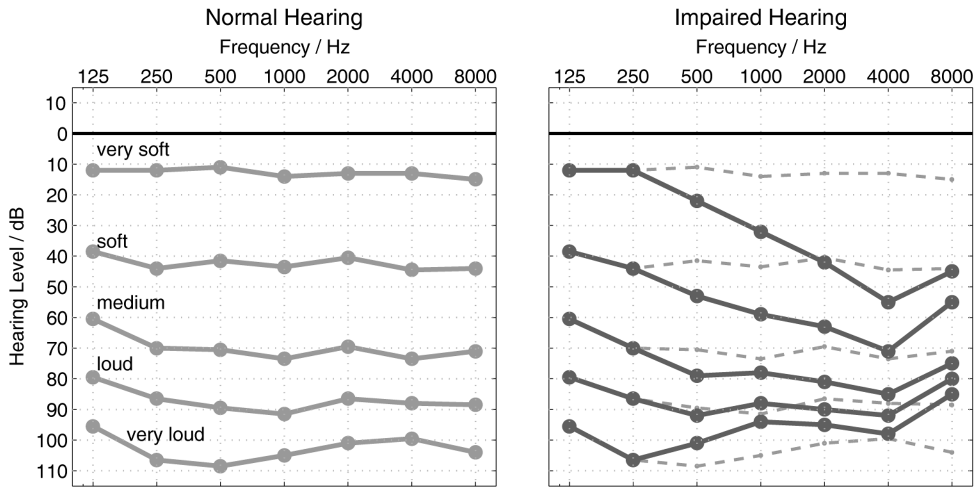
Normal-hearing people are able to perceive sound in the range of about 0 dB to 110 dB hearing level, which represents the level range between the hearing threshold and level at which sounds become uncomfortably loud. The area between the threshold of hearing and the uncomfortable level represents the dynamic range of the hearing system.

It is important to mention that even though the threshold of hearing is increased in sensorineural hearing impaired, the level of uncomfortable loudness typically is not increased accordingly. In contrast, the uncomfortable level is often similar (or even slightly lower) than for normal-hearing people. In combination with the elevated hearing threshold, this means that the level range at which sounds can be perceived is reduced. In addition to a loss of information due to the elevated threshold, this reduced dynamic range causes problems in communication for hearing-impaired people, because sounds which are clearly audible are perceived differently by hearing-impaired people.

One important reason for this is a modified perception of loudness, which is illustrated in Figure 2. The left panel shows equal-loudness contours of a normal-hearing listener, i.e. the levels at which different frequencies are perceived as equally loud. Five different curves show the levels which are perceived as 'very soft', 'soft', 'medium', 'loud' and 'very loud', respectively. When expressed in dB hearing level as in the audiogram, the equal-loudness loudness contours are fairly flat for normal-hearing listeners, i.e. different frequencies are perceived as equally loud at about the same hearing level. The equivalent data of a person suffering from sensorineural hearing loss is shown by solid lines in the right panel of Figure 2. For comparison, the normal-hearing data are shown by dashed lines. The line indicating the loudness perception 'very soft' represents sounds just above the threshold and therefore approximates the shape of the hearing threshold. It can be observed that the hearing impaired person suffers from a high frequency hearing loss, i.e. low frequencies are perceived normally while an elevated threshold is found for higher frequencies (right panel). Such a shape is typically observed in age-related or noise-induced hearing losses. Compared to the normal-hearing data, the loudness perception of sounds close to threshold is clearly modified in the high-frequency region, i.e. higher hearing levels are needed to produce the same loudness sensation. For louder sounds, however, this is not the case. For example, the contour representing a 'loud' perception is very similar for the normal-hearing and the hearing-impaired listener. The hearing levels at which sounds are perceived as 'very loud' are even lower for the hearing-impaired person. This reduced dynamic range results in the phenomenon called loudness recruitment, which describes the effect that loudness grows faster with level than normal for hearing-impaired listeners.

The perception of loudness in normal-hearing listeners and especially the loudness recruitment in hearing impaired have substantial consequences on acoustic communication. On the one hand, too loud sounds are uncomfortable to listen to and therefore have to be avoided. On the other hand, the effort to focus on audible but too soft sounds can be very tiring over longer periods. The general goal of sound presentation is, therefore, to adjust the level to a reasonable, in most scenarios comfortable, listening level for both, normal and hearing-impaired listeners. Due to the steep growth of loudness in hearing-impaired listeners, already a minor deviation in level can lead to too loud or

Figure 2. Equal-loudness contours of a normal-hearing (left) and a hearing-impaired listener (right). The lines indicate the levels of sounds perceived as equally loud as a function of frequency. In the right panel, the normal-hearing data are re-plotted as dashed lines for comparison



too soft sounds. Communication situations can, therefore, be very annoying for people suffering from a sensorineural hearing loss, e.g. because constant manual adjustments of volume are needed. Naturally, this reduces speech intelligibility and quality in speech communication. As a result, increased hearing effort is required since parts of the speech are difficult to hear. The consequences of a reduced communication quality can be considerable. During acoustic communication at work or in private, misunderstandings and misinterpretations may lead to reduced self-assurance, frustration, decreased productivity, reduced social activity, and even solitude or loss of working place (for an overview, see Shield, 2006, p. 61ff.).

Technical support for hearing-impaired people is therefore highly desirable, in particular in the field of acoustic human-computer interaction. Since the effects of hearing impairment are highly individual, an increased accessibility of acoustic interfaces requires personalized technology. The following section describes different technical approaches which can support acoustic human-machine interaction for both normal-hearing and hearing-impaired people.

STATE-OF-THE-ART TECHNOLOGIES TO IMPROVE AND EVALUATE ACOUSTIC USER INTERFACES

A number of technologies exist to support acoustic communication, which are also applicable to human-machine interaction. This section describes different signal-processing approaches comprising enhancement of communication signals, application of personalized algorithms and technical evaluation of communication quality. Despite the fact that all of these approaches are still subject to current research, the technologies that are already available offer a wide range of solutions for improved human-machine interaction. Basic principles, technical limitations and examples of ongoing research are presented for each field.

Digital Signal Enhancement

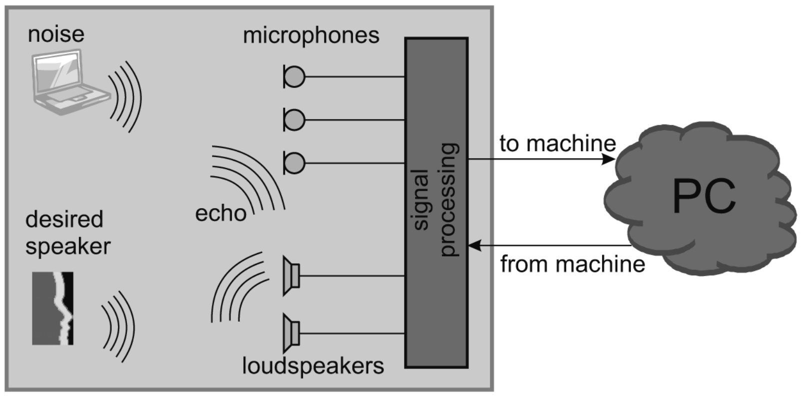
As outlined in the introduction, the use of acoustic interfaces may be very difficult under adverse acoustical conditions due to background noise, acoustic echoes or reverberation for both normal-hearing and hearing-impaired users. Part of the solution to such problems can be offered by signal processing strategies, which enhance the quality of the communication signal by acoustic echo cancellation (Hänsler and Schmidt, 2004; Goetze et al., 2005), de-noising (Ephraim and Malah, 1985; Huang et al., 2006) or dereverberation (Goetze et al., 2006; Habets, 2007; Goetze et al., 2008).

Figure 3 illustrates the principle challenges for human-machine communication, in which acoustic information is exchanged between the user and a computer system in a realistic, hands-free communication situation. The signal containing the information is presented to the user via loudspeakers, and the user can give spoken commands to the machine which are recorded by microphones. The signal processing unit removes disturbances from the signal and performs an analysis of the information contained, e.g. by automatic speech recognition (ASR). Note that the sound is picked up at a spatial distance from the speaker (non-close-talk situation) by the microphones. On the one hand, this is convenient for the user since he or she does not need to wear the microphone. On the other hand, this microphone positioning leads to problems for the signal processing unit since the microphones are not directly located at the source of the sound signal. This leads to a substantially decreased signal-to-noise-ratio (SNR) compared to the use of a microphone that is placed e.g. close to the mouth of the speaker (close-talk scenario) (Huang et al., 2006). In modern work processes and communication systems, hands-free sound pick-up is a desirable way of human-machine communication, since it allows for more flexibility and multi-user input. The users can move freely and use both hands for other tasks. However, the signal is, furthermore, superimposed by acoustic echoes and reverberation caused by repeated reflections of the signal itself at the room boundaries like walls, floor and ceiling. The input signal for the human-computer interface coming from the human speaker is, thus, a combined signal comprising the desired speech signal, acoustic echoes, reverberation and ambient noise (see Figure 3), which may heavily reduce the quality of the received signal. Under such adverse acoustic conditions, automatic speech recognition systems have a considerably lower performance (see e.g. Mildner et al., 2006; Benesty et al., 2008). Not only ASR-systems suffer from reduced speech quality. In telecommunication, the same challenges occur when information is sent from one place to another via acoustic user interfaces (Hänsler and Schmidt, 2004; Rohdenburg et al., 2005). Without further processing, unwanted echoes, reverberation and ambient noise are transmitted together with the desired speech signal. This is not only annoying and tiring for a human listener and disturbing for both, an ASR system and a human listener, it can also lead to a closed electro-acoustic feedback loop and, thereby, to an instable system producing howling. As outlined in the introduction, the resulting communication problems can be substantial for hearing-impaired users of such acoustic interfaces. Therefore, it is highly desirable to enhance the signal quality for human-machine interaction as well as for telecommunication systems.

Reduction of Ambient Noise

The removal of noise from the microphone signal is a particularly powerful technique when the spectral or statistical content of noise and speech signal differ or if the desired sound source and the undesired disturbance arrive from different directions. In many situations like factory workplaces (machinery noise), cars (noise from engine and tires) or open-office areas (ventilation, typing, printer noise, etc.), this condition is at least partly fulfilled, which offers a large potential for signal

Figure 3. Illustration of the acoustic difficulties in human-machine interaction. The speech signal of the machine is transmitted to the user and is disturbed by noise and reverberation, leading to decreased speech intelligibility. In addition, the signal of the loudspeaker feeds back into the microphone and superimposes the desired signal of the human speaker. Digital signal processing strategies, such as noise reduction and acoustic echo cancellation can help to considerably improve the signal quality and to avoid feedback in the system



enhancement using noise reduction schemes. Depending on the number of available microphones, single-channel (Hänsler and Schmidt, 2004; Ephraim and Malah, 1985) or multi-channel (Goetze et al., 2006; Bitzer and Simmer, 2001) noise reduction schemes can be applied.

Single-channel noise-reduction schemes estimate the current SNR in several frequency bands within short time intervals of about 10 to 30 ms and calculate a suppression rule depending on that estimate. The suppression rule defines the amount of attenuation within each of the frequency bands of a given time frame. Thus, single-channel noise-reduction schemes perform adaptive filtering of the signal aiming at suppressing the noise part while leaving the desired speech part unaffected. This leads to a high amount of noise reduction for low SNR, i.e. when the noise power is higher than the power of the desired signal. If the desired signal dominates the noise (high SNR) the noise reduction only slightly suppresses the signal. However, single-channel noise-reduction schemes always suppress both signals, the desired signal and the noise signal to a certain extend since both signals are picked up simultaneously by the microphone and none of the signals is available separately. Although the filter can be designed to mathematically perform the best trade-off in terms of noise reduction versus not affecting the desired signal, an unwanted side-effect of a good noise reduction is always a certain amount of cancellation of the desired signal component, which reduces the signal quality. Furthermore, state-of-the-art single-channel noise-reduction schemes still suffer from the so-called *musical noise* problem. Musical noise is caused by residual noise that is small in amplitude but clearly perceivable by a human listener since it sounds unnatural due to its non-stationary nature (Cappe, 1994; Rohdenburg, 2008). Noise-reduction schemes incorporating models of the human auditory system (Gustafsson, 1999; Goetze et al., 2006) partly avoid the musical noise problem and, therefore, lead to perceptually better results. This is achieved on one hand by exploiting the fact that noise parts that are below the hearing threshold are not perceived be the human listener and, thus, do not have to be suppressed. This provides more degrees of freedom to the noise suppression filter. On the other hand, distortions of the signals additionally can be hidden below the threshold of hearing which leads to better sounding signals (Goetze et al., 2006) as well as increased performance of ASR systems (Mildner et al., 2006). Thus, for communication systems as well as for human-machine interfaces in general the use of models of the human auditory system increases the performance of technical systems.

If more than one microphone is available, so-called beamformers can be applied that exploit spatial information about desired sound source and ambient noise. Beamforming microphone arrays (Monzingo and Miller, 1980; Bitzer and Simmer, 2001) and their extensions by multi-channel post-filters (e.g. Simmer et al., 2001; Goetze et al., 2006) work similar to the binaural hearing system. They exploit information about the position of the desired sound source by spatially sampling a given sound field at multiple positions using multiple microphones. By this, the digital filter is capable to enhance the signal using the level and phase differences. Ambient noise can in principle be suppressed without affecting the desired signal part (Bitzer and Simmer, 2001). Since beamforming microphone arrays rely on knowledge about the position of the desired source, this information has to be obtained first by approaches for direction-of-arrival estimation (Knapp and Carter, 1976; Doblinger, 2006; Rohdenburg et al., 2008).

Acoustic Echo Cancelation

As depicted in Figure 3, in addition to ambient noise, also acoustic echoes stemming from the signal uttered by the loudspeakers are picked up by the microphones. The cancellation of acoustic echoes seems to be much easier at a first glance, since knowledge about the signal to be canceled out (i.e. the loudspeaker signal) is available. However, not exactly the loudspeaker signal is picked-up by the microphones again, but a signal changed by numerous reflections within the room. Thus, the influence of the room, which can be characterized by the so-called room impulse response, has to be estimated for an exact compensation. Unfortunately, this compensation leads to a high computational load that may be too high even for nowadays computers for a real-time processing (Hänsler and Schmidt, 2004). Here, two step approaches that cancel only parts of the acoustic echoes and suppress the remaining parts by additional filtering similar to noise-reduction filters can be applied (Goetze et al., 2005). Again, approaches that exploit models of the human auditory system lead to perceptually better results for both human listeners (Gustafsson, 1999; Goetze et al., 2006) and human-machine interfaces (Mildner et al., 2006).

Dereverberation

As obvious from preaches in churches, high amounts of reverberation decrease speech intelligibility (see e.g. Duquesnoy and Plomp, 1980). Also the performance of acoustic signal processing schemes like multi-channel position estimation, noise reduction or automatic speech recognition (Monzingo and Miller, 1980; Huang et al., 2006; Doblinger, 2006; Rohdenburg et al., 2008) is significantly decreased by reverberation. Hence, it is desirable to reduce reverberation from acoustic signals. Since reverberation is caused by the influence of the acoustic environment (the room) that is characterized by the room impulse response. the most straightforward approach is to invert this impulse response (Neely and Allen, 1979). However, due to the signal-theoretic properties of common room impulse responses this is not easy to achieve technically and such system have difficulties working in real-world systems (Goetze et al., 2008). Another approach is to estimate the part of the signal that does not contain reverberation and separate it from the reverberant part (Habets, 2007). This approach leads to digital filters that try to suppress the reverberant part similarly to the noise reduction filters described above. It should be noted that more than the previously discussed technologies, dereverberation schemes are topic of currently ongoing research (Habets, 2007; Hänsler and Schmidt, 2008; Naylor and Gaubitch, 2010) since a perceptually satisfying solution is hard to find. However, first practically feasible approaches exist that enhance the quality of a reverberant signal as a pre-processing for either a technical system or for the human listener (Naylor and Gaubitch, 2010). First approaches to incorporate models of the human auditory system for dereverberation exist that, again, lead to perceptually much better results (Mertins et al., 2010).

Acoustic Event Detection and Classification

Humans have an astonishing ability to automatically detect and classify single acoustic events that are important in a specific situation and to associate them with certain phenomena. Hence, our auditory system plays a major role in identifying e.g. critical situations and in providing orientation in everyday life. Similarly, technologies for acoustic event recognition aim at monitoring the environment, e.g. for intrusion detection by identifying breakage of glass, or at monitoring systems, e.g. by recognizing machine failures. Acoustic event recognizers typically separate acoustic events from background noises before classification. For that purpose, one or more microphones continuously record the surrounding sound. Current approaches again are based on mimicking the sound pre-processing of the human hearing system (van Hengel et al., 2009). The result of this pre-processing is a so-called cochleogramm which describes the spectral and temporal energy distribution of the signal similarly to the processing within the inner ear of humans. On the basis of threshold transitions relative to the background sound energy, the detector identifies events which are further analyzed by the classifier. In a training phase, classes of representative cochleogramms are derived from a low number of training data. These representative cochleogramms are compared by the classifier to the incoming events and a report is given when a specific event is detected. Acoustic event recognition can also be used to support human machine interaction by providing additional information on the recent situation underlying the interaction or by invoking an alarm when certain events are recognized (van Hengel and Anemüller, 2009). Despite improved acoustic conditions and signal quality by means of the technologies described above, there may be situations for hearing-impaired (and also normal-hearing) people, in which acoustic communication is extremely difficult. Such problems have also been identified in non-occupational contexts and technical solutions have been developed, in particular in the field of ambient-assisted living (van Hengel and Anemüller, 2009). These solutions are not restricted to their original applications, and the developed concepts like robust speech recognition for human-machine interfaces and acoustic event detection may find applications also in modern workplaces (Rennies et al., 2009a) and in domestic and health sectors (van Hengel and Anemüller, 2009). Acoustic events indicating alarms, incoming messages, or operational elements of machinery may be hard to hear, particularly for hearing-impaired persons. But even if these events are audible, the localization or classification of the events may not be possible or error-prone. Automatic acoustic event detection can support the hearing-impaired person, giving indications about the type of signal, its direction and intensity. Examples range from indicating critical alarms at work places, detection of a ringing door bell, an over boiling in the kitchen to automatic detection of critical situations like yelling of patients, e.g. in nursing homes.

Personalized Signal Processing

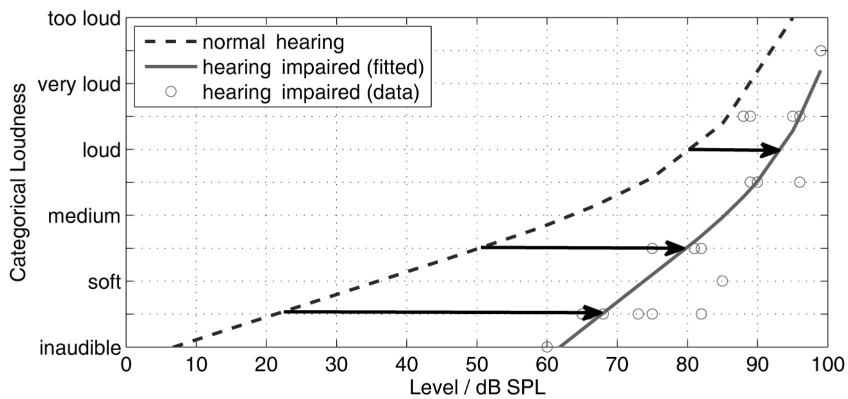
Additional benefit for hearing-impaired users, as well as for normal-hearing people having special listening requirements, can be expected if communication devices are fitted to their individual needs, i.e. adjusted to the hearing loss, requirements and preferences of the individual user (Appell et al., 2007; Baumgartner et al., 2009; Rohdenburg et al., 2009). Such personalized hearing systems partly cover the functionality of a hearing aid, which may not be available for the user, or which may be impossible to wear at certain conditions (e.g. in combination with hearing protectors or helmets). Personalized algorithms account for the reduced dynamic range (cf. Figure 2) by nonlinearly mapping the sound energy into the remaining audible range. The modified loudness perception can, thus, be partly compensated, which facilitates communication for hearing-impaired users. As already mentioned, hearing-impaired people understand speech only if the level of soft sounds is increased. Thus, these sounds have to be amplified by the supporting system. However, since a higher sound level leads to an unpleasantly loud perception for sounds that already have a loud level, these sound must not be amplified further, because hearing-impaired people typically perceive loud sounds similarly as normal-hearing people (see Figure 2) and a level increase at high sound input levels leads to a uncomfortably loud sensation. Thus, the dynamic range of a normal-hearing person has to be compressed to the dynamic range of the hearing-impaired person as depicted in Figure 2 (Appell et al., 2002). Figure 4 illustrates the concept of dynamic compression. In this example, loudness was measured for a sound at a frequency of 3 kHz by means of categorical loudness scaling. Using this method, the sound is presented at different sound pressure levels (dB SPL) and the listeners have to indicate the loudness on a categorical scale ranging from inaudible, soft, medium and loud to too loud (for details, see Brand and Hohmann, 2002). Circles in Figure 4 indicate the individual data points measured with the hearing-impaired listener, i.e. the categories indicated by the listener for the sound presented at a certain level. The solid line is a curve fitted to the data. The dashed line represents a reference curve of normal-hearing listeners. It can be seen that the curve for the hearing-impaired listener is steeper than for the normal-hearing reference, which reflects loudness recruitment. The horizontal distance between the two curves represents the level difference that is needed to produce the same loudness sensation for the hearing-impaired as for the normal-hearing person. The arrows indicate that this level difference is much higher for softer sounds than for louder sounds. This means that in order to compensate for the modified loudness perception of the hearing-impaired person, a larger amplification is needed for softer sounds than for louder sounds, i.e. that a compressive amplification is needed. The example in Figure 4 illustrates the need for compressive amplification only for a single frequency, while the effect of hearing impairment on loudness perception generally depends on frequency (see Figure 1). Therefore, different amplification rules apply for different frequency regions. Provided with this information, personalized signal-processing strategies can compensate for the individual loudness perception of hearing impaired persons and thereby substantially improve acoustic communication.

Technical Evaluation of Communication Quality

Although speech communication is the most natural and convenient way of information transmission, the use of multi-modal information may be desirable in certain situations, e.g. when acoustic conditions are difficult. In such cases, additional visual, tactile or text-based information could be given to the user. In order to detect such difficult acoustic situations, an automatic monitoring of the acoustic conditions is required.

The most important indicator of the quality in acoustic human-machine interaction is whether speech can be understood properly or not since the primary goal is the transmission of information. To evaluate communication with respect to speech intelligibility, it is - ideally - measured experimentally using the same people normally

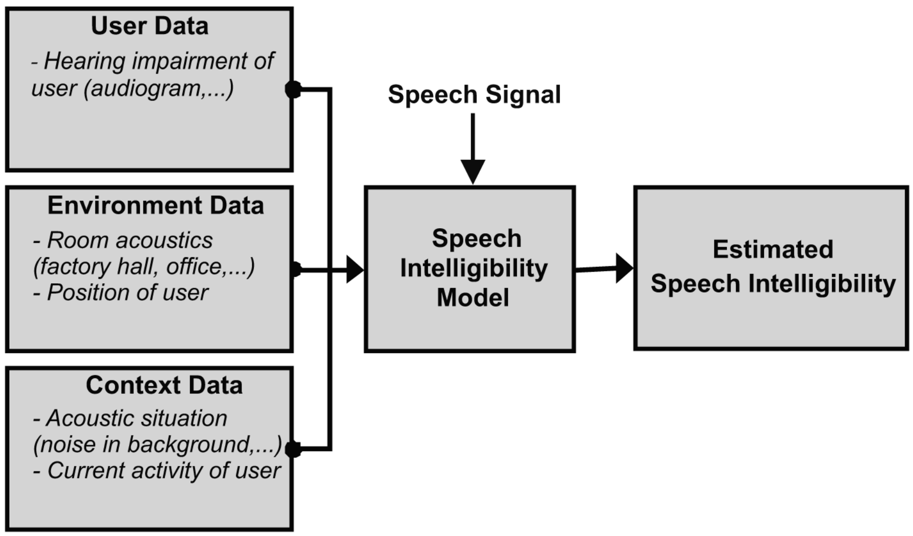
Figure 4. Derivation of amplification rules based on loudness scaling data of a hearing-impaired person. Fitted loudness scaling data are shown for a normal-hearing (dashed) and a hearing-impaired person (solid). Circles indicate the measured data of the hearing impaired. Arrows represent amplifications that would lead to the same loudness perception for the hearing-impaired listener as for normal-hearing person



using the interface as test listeners. To ensure comparability, exactly the same conditions as typically encountered should be ensured. Several tests were developed, which are in principle able to yield exact estimates of speech intelligibility (see e.g. Wagener et al., 1999a,b,c; Sukowski et al., 2009). Naturally, such measurements involving real subjects are impracticable due to their high costs and long durations, and cannot be used for continuous monitoring. Therefore, several technical methods were developed to predict speech intelligibility for given physical boundary conditions. Standardized models such as the Speech Intelligibility Index (SII, ANSI, 1997) or the Speech Transmission Index (STI, IEC, 1998) are widely used in research and applications. The general principle of these models is to compute a ratio between the desired speech energy and the unwanted noise energy (signal-to-noise ratio). As shown in a simplified way in Figure 1, individual speech fragments mainly contain different frequency components. Accordingly, some frequency regions contribute more to speech intelligibility than others. The models account for this dependence by weighting the energy ratios in different frequency regions according to their importance for speech recognition. Additionally, other detrimental effects can be included in the predictions, like e.g. the influence of reverberation or band-limiting transmission systems. Figure 5 illustrates the basic principle of speech intelligibility prediction. The speech intelligibility model requires the speech signal to be transmitted and information about the user (e.g. type and degree of hearing impairment), the environment (e.g. room acoustics) and the context (e.g. current noise sources) to compute an estimate of the current speech intelligibility. In general, speech intelligibility models such as SII or STI calculate an index with a value ranging from 0 (completely unintelligible) to 1 (fully intelligible). Intermediate values can be transformed into other measures of speech intelligibility, e.g. the percentage of correctly understood words or sentences.

In principle, these models are able to predict the influence of external acoustic conditions and can, therefore, be used to technically evaluate the quality of acoustic human-computer interaction. Naturally, this concept relies on a trustworthy estimate of speech intelligibility based on the available information. While the standardized models (SII, STI) successfully account for the effects of background noise, reverberation and an elevated hearing threshold, other important factors are not included in their calculations. Under cer-

Figure 5. A model of speech intelligibility can combine information about the user, the acoustic environment and the context to estimate how well the user is able to understand acoustic information in the given situation. This estimate can be used in monitoring, planning and improvement of human-machine interaction



tain conditions, this may lead to wrong predictions, which limits the models' applicability. Current research focuses on the development of more generally applicable models taking account of other relevant aspect like temporally fluctuating noises (e.g., Rhebergen and Versfeld, 2005; Meyer et al., 2007), spatial distributions of different speech and noise sources (e.g., Beutelmann and Brand, 2006; Rennies et al., 2010a) or other relevant aspects of hearing impairment beyond an elevated threshold (e.g., Jürgens et al., 2009).

Apart from speech intelligibility, other factors also affect the quality of acoustic signals and a technical assessment of these aspects is very helpful to evaluate acoustic communication. It was discussed in the previous section how loudness perception contributes to communication in normal and impaired hearing. As for speech intelligibility, a number of models already exist, which transform a physical sound signal to a predicted loudness quantity (e.g. Moore, 1996; DIN, 1991; Moore et al., 1997; Zwicker and Fastl, 1999; Chalupper and Fastl, 2002; ANSI, 2007; Moore and Glasberg, 1997, 2007). For many physical signals, this quantity relates very well to the loudness perception of real listeners, and even the modified loudness perception resulting from hearing impairment is taken into account by some models (Moore et al., 1997; Chalupper and Fastl, 2002; Appell, 2002). However, so far some aspects remain unaccounted for. For example, measurements have shown that very short sounds highly contribute to the overall loudness perception even though their physical energy is small. Therefore, the technical prediction of the currently standardized models (DIN, 1991; ANSI, 2007) will deviate from the real loudness perception when such signals are considered, which may be the case e.g. in machinery noise. Therefore, ongoing research aims at the improvement of the existing loudness models, focusing on the special influence of short and quickly fluctuating sounds (e.g., Rennies et al., 2009b, 2010c; Rennies and Verhey, 2009) or the effect of binaural listening (e.g., Moore and Glasberg, 2007).

In conclusion, current models of speech intelligibility and loudness provide a way to detect and quantify difficulties in acoustic communication. While some relevant factors influencing speech quality are still investigated and current models do not include all of these factors, the available models account for the most important aspects such as background noise, reverberation and hearing threshold. In the context of acoustic human-machine interaction, the benefit of using such models can be twofold: on the one hand, they can be used in planning and design phases of interfaces to ensure good communication quality based on simulated acoustic conditions. On the other hand, they can be used to continuously monitor acoustic conditions in the environment in which the interface is used. In case of bad acoustic conditions, further processing strategies or other modalities can be used to ensure that the desired information is perceptible.

Such an integrative approach of combined acoustic monitoring and signal enhancement strategies is outlined in the following section. A concrete example of continuous communication quality monitoring is given subsequently.

TOWARDS A HEARING PERCEPTION MODEL FOR ACOUSTIC USER-INTERFACES

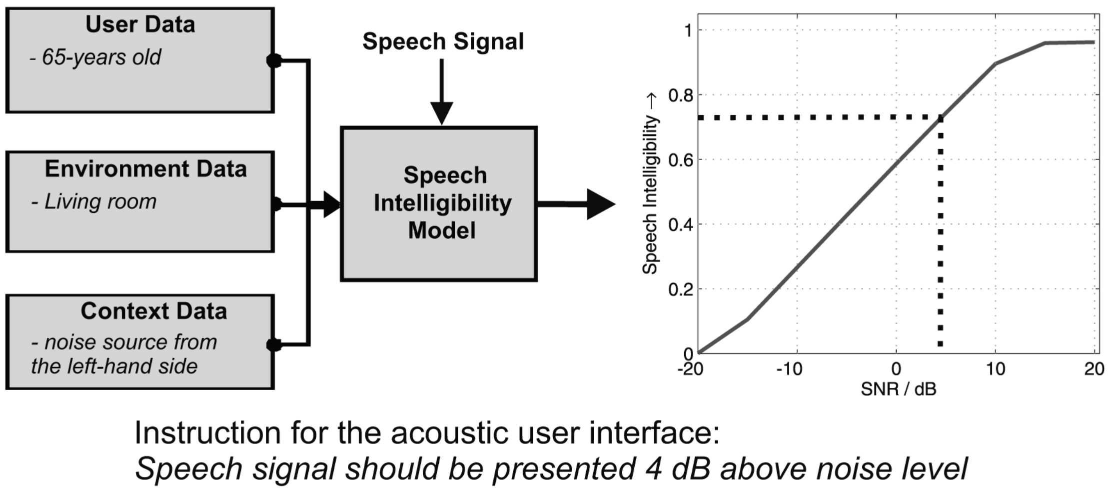
In the previous section, several separated techniques for the enhancement and evaluation of acoustic human-machine interaction have been described. These technologies focused on different aspects ranging from speech communication to acoustic event detection. This section extends these individual approaches towards an integrative and comprehensive model-based concept for design and application of human-machine communication. Thereby, the central question is how acoustic interfaces can be enabled to support the individual requirements of different users, i.e. individuals having hearing deficiencies, or normal-hearing users asking for a reduced listening effort in acoustically adverse conditions.

In principle, each of the methods described in the previous section receives and processes input signals and parameters to compute its output. For models predicting speech intelligibility, this concept is exemplarily shown in Figure 5. The model combines information on the user, the recent context, and the environment to calculate an estimate of the speech intelligibility. Noise reduction schemes partly use similar information. They calculate an enhanced speech signal and, to do so, require knowledge about the relevant information to be transmitted and the acoustic context (e.g. signal statistics of background noise, position of desired speaker, etc.). Methods for echo cancellation and dereverberation focus on sound quality improvements by compensating for adverse environmental conditions (e.g. reverberation, closed feedback loops) and lead to better results if they are based on models of the human auditory system. Personalized signal processing strategies require data of the user himself as input, e.g. of the individual hearing loss or listening preferences. It was discussed above that such an individualized signal processing would be very beneficial for people suffering from hearing deficiencies. In practice, however, it is not possible to measure each aspect of the individual hearing loss and preferences relevant for speech communication and acoustic interaction before the user can handle the respective communication and interaction systems. Therefore, information on the individual user has to be acquired and provided in a simplified way. This can be achieved by computational models, which mimic the perceptual characteristics of an individual user. Naturally, improvements of acoustic user interfaces based on models can only be achieved if the models take the relevant aspects of hearing perception into account. For example, the age of a user can be a first approach to describe his hearing abilities since, in general, hearing deficiencies increase with age. As shown in Figure 1, an average hearing loss can be computed on the basis of the age (ISO, 1990). However, the individual vulnerability to hearing loss varies substantially. The data shown in Figure 1 are median values, i.e. half of the people included in the studies on which the calculations are based have a larger hearing loss. The spread of the hearing losses can be demonstrated by looking at percentiles other than the median value. For the group of 65-year olds (curve indicated by diamonds in Figure 1), the median hearing loss is about 49 dB at a frequency of 8 kHz. For comparison, the 80th and 20th percentiles are at about 27 and 76 dB HL. This means that, for the same age, a large range of hearing losses has been observed. Therefore, in most of the cases, age alone hardly represents the individual perception and would lead to wrong conclusions if prerequisites of acoustic user interfaces were derived from it. Given the difficulties to generalize hearing losses over larger groups of individuals, a better source of information about hearing perception is the audiogram of an individual user. An acoustic user interface provided with this information could predict audibility of an acoustic output signal over frequency and, hence, could react accordingly by providing additional amplification to the output signal. Although the incorporation of individual audiogram information is already a major step towards individualization of acoustic user interfaces, it has to be noted that the pure use of audiogram information can lead to wrong conclusions about the requirements of acoustic user interfaces. As mentioned above, the individual perception of loudness can also be considerably influenced by the individual kind of hearing loss. Therefore, a 'blind' amplification of sound signals without knowledge of the individual level at which sounds become too loud may lead to uncomfortably loud levels. Thus, measuring this uncomfortable level in addition to the hearing threshold provides an estimate for the entire dynamic range at which sounds can be presented to a user of an acoustic user interface. In audiological practice, the individual dynamic range is measured using sounds of very limited frequency content, e.g. pure tones (containing only a single frequency) or narrow bands of noise (Kollmeier, 1992). However, most natural sounds including speech are highly complex and consist of many time-varying frequency components. This means that even when the narrowband dynamic range of a user is known, many other aspects may affect his acoustic communication, e.g. his ability to understand speech in noise or to process temporal fluctuations, or his attention, motivation or lexical skills. As a consequence, people with a similar hearing threshold and uncomfortable level can have significantly different performances in speech recognition tasks.

These various factors contributing to speech intelligibility are still subject to research and are not fully understood yet. Consequently, no existing model can fully describe the individual perception of sounds. However, in the context of acoustic user interfaces, models incorporating the most important aspects of hearing can significantly improve the accessibility, not only for people with hearing deficiencies. Such hearing perception models need to be embedded in a general framework combining information about the user (as described above) and the acoustic context and environment. By estimating the relevance of this information for hearing perception, the model can provide important input to develop and control personalized and adaptable acoustic user interfaces.

An exemplary realization of how such a model-based approach could be used to personalize and adapt user interface is shown in Figure 6. The task of the interface is to transmit acoustic information to a user. A model estimates speech intelligibility on the basis of data about the acoustic environment and context, and about the user. In this example, only very general information is available. The acoustic environment is a living room with a 'typical' reverberation time. The current acoustic context consists of a single noise source placed at the left side of the user, while the user is assumed

Figure 6. A model of speech intelligibility used to derive instructions for an acoustic user interface. The model estimates speech intelligibility based on the given acoustic environment and data as well as on information about the user

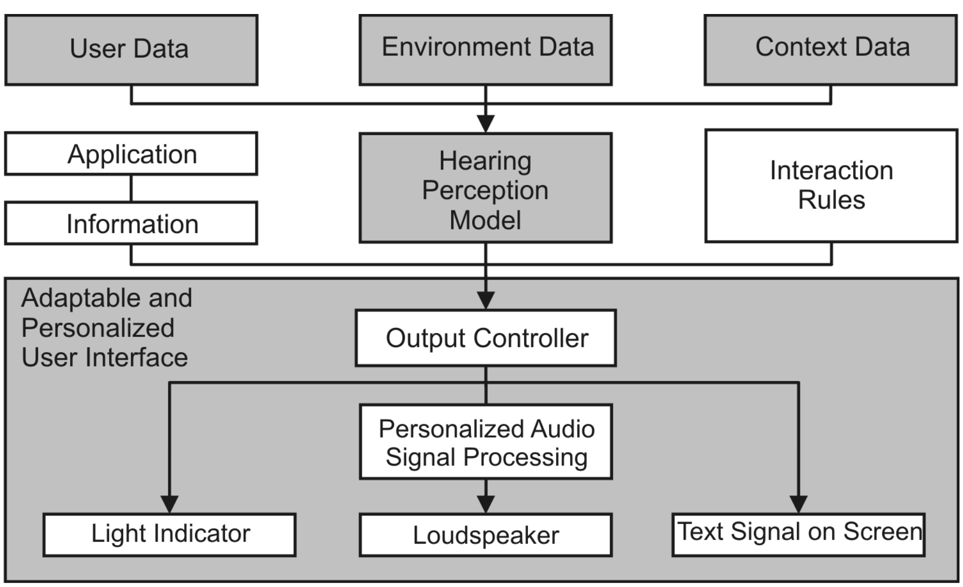


to receive the desired information from the front. As an example, the disturbing source at the left of the user could be an active loudspeaker (e.g. of a hi-fi system) acting as a disturbance for the acoustical sound information that is played back e.g. by the television in front of the user. The hearing capability of the user is modeled based on the mean expected hearing loss of a 65-years old person that can be taken from Figure 1. Provided with this information the model calculates the expected speech intelligibility for different SNR configurations. In the example shown in Figure 6, the model presented by Beutelmann and Brand (2006) was used to predict speech intelligibility. The result is shown as a solid line in the right panel of Figure 6. As expected, the predicted speech intelligibility increases with increasing SNR. Since the level of external noise sources generally cannot be modified by the acoustic interface, better SNR can only be realized by a level increase of the speech to be presented by the acoustic interface. Based on a predefined threshold of intelligibility and a corresponding SNR value, the interface can now calculate the amplification of the speech signal required to achieve that SNR.

In the example shown in Figure 6, the threshold for good intelligibility was set to an index value of 0.75. This corresponds to an SNR of about 4 dB, as indicated by the dotted line in Figure 6. The model-based instruction for the interface would therefore be to present the acoustic information at least 4 dB above the noise level in order to ensure good speech intelligibility. This example is meant to be an illustration of the possible applicability of speech intelligibility models in acoustic user interfaces. It shows that, in general, current models can be used to derive technical instructions for acoustic interaction between humans and machines. As outlined above, age alone is hardly a representative indicator of the user's hearing deficiencies. Individual data as e.g. provided by the audiogram or loudness perception measurements should be used instead to enable a higher degree in the personalization of acoustic user interfaces. Depending on the desired field of application, acoustic data about environment and context can often already be estimated in the planning phase of acoustic interfaces by commonly used room acoustic simulation software (Allen and Berkley, 1979), or can be measured at existing environments with relatively low effort (MacWilliams and Sloane, 1976; Rife and Vanderkooy, 1989; Vanderkooy, 1994; Müller and Massarani, 2001; Pintelon and Schoukens, 2001). Similarly, the information about the individual user can be estimated, for example based on age (ISO, 1990), or it can be measured with standard audiological techniques. For a given acoustic environment and context, the entire range of estimated (or real) hearing losses can be used to compute predicted speech intelligibility. This way, an easy detection of acoustically problematic conditions can be achieved already in planning phases of acoustic human-machine interaction. By variations of the environment and context data, the benefit from acoustic modernizations like additional damping, less noisy machinery, or signal enhancement strategies can be estimated.

The incorporation of hearing perception models for better individualization of sound output is obviously applicable in many scenarios, like phone and teleconferencing conversation, communication in cockpit or in technical systems providing information by speech output. Also multi-modal user interfaces are enabled for more individualization and adaptability when hearing perception models are embedded in respective software architecture for multi-modal user interfaces. Such a concept is schematically shown in Figure 7. The interface is used to transmit information to a user. This information may be of general kind or depend on the current application. A model of hearing perception is the core element in the schematic design. It predicts individual hearing perception based on the individual user data, as well as on information of the acoustic environment and context. Thereby, it can provide an estimate of how personalized audio signal-processing strategies as described in the previous section can increase the quality of the acoustic information for the user. The prediction of the model is transferred to an output controller, which manages the actual multi-modal presentation of information to the user. Depending on the achievable quality of

Figure 7. Schematic structure of a personalized and adaptable acoustic user interface. The core element is a hearing perception model, which predicts the potential difficulties in acoustic communication and provides input for signal-processing strategies and information output, which is managed by an output controller also considering other modalities to present information to the user



the acoustic presentation, as estimated by the hearing perception model, the output controller initializes the information presentation in an acoustic way or, if the potential acoustic presentation is not suitable for the user in his or her current environment and context, by means of other modalities like for example light or text output. This interaction between user and machine can be adjusted by defining interaction rules and providing them to the output controller to optimize accessibility. A schematic as shown in Figure 7 represents a concrete realization of an adaptable and personalized architecture of acoustic user interface. The general concept of incorporating individualized human perception modeling is principally also applicable to other senses and types of impairments such as e.g. visual deficiencies or a reduced mobility, provided adequate models can account for the relevant aspects in the context of human-computer interfaces. The combination of human perception models for different senses and disabilities would lead to an even more holistic approach for personalization and adaptability of human-machine interaction. The hearing perception modeling approach described here already enables personalization and adaptability in acoustic user interaction. This is particularly relevant since the demand of human-machine interaction increases in many fields of modern societies, while, at the same time, the portion of hearing-impaired persons will grow dramatically (Shield, 2006).

The concept of personalization and adaptability addressed in theory in this section is illustrated by potential applications and scenarios incorporating hearing perception models in the following section.

APPLICATIONS

This section aims at illustrating the concepts developed in the previous sections. As mentioned in the introduction, adaptable and personalized user interfaces are not widely spread yet. The following examples can be considered as first steps towards achieving this goal in the field of acoustic communication.

Automatic Monitoring of Communication Quality

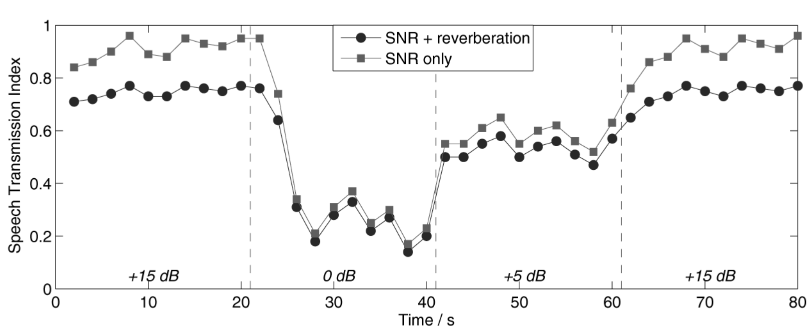
As described above, a model of speech intelligibility can be used to adjust acoustic user interfaces according to the current acoustic conditions and needs of the user (see Figure 6). Today, predictions of hearing perception models are typically calculated offline. However, in practical applications like telecommunication systems or human-machine interfaces it is important that estimates of the quality of acoustic communication are calculated during runtime. Particularly in teleconferencing situations the speaker is not aware of low intelligibility at the distant listener. Without technical support the listener needs to make constant inquiries which can considerably disturb the communication. Therefore, we developed a hearing-perception model based method that continuously monitors the acoustic conditions with respect to intelligibility (Rennies et al., 2010b). In case low intelligibility is observed by the model, the speaker can be informed enabling him to take respective actions (e.g. to move the microphone to a better position). The same approach is applicable to human-machine interfaces to estimate the benefits of different sound processing schemes or to select amongst different output modalities.

An example of such a monitoring process is illustrated in Figure 8. Speech intelligibility was estimated using the STI (IEC, 1998) as a function of time. The data were simulated using continuous speech from the Oldenburg sentence test (Wagener et al., 1999a) as the desired signal, superimposed with an undesired continuous noise. The noise had the same spectrum as speech, mimicking a disturbing conversation in the background. A reverberation typical for office rooms was included in the simulations. Initially, the speech was superimposed by a background noise of relatively low level at an SNR of 15 dB, resulting in acceptable speech intelligibility. After 20 s, the SNR was lowered to 0 dB. In practice, such a reduction in SNR could occur when an additional noise source is switched on. After 40 and 60 s, the SNR was changed to 5 and 15 dB, respectively, corresponding to a stepwise reduction of the effect of the additional noise source. The time instances when the changes in SNR were made are indicated by vertical dashed lines in Figure 8. As indicated by the symbols, the STI was calculated at intervals of 2 s, based on (i) the estimated SNR alone (squares), and (ii) on estimations of both SNR and reverberation time (circles). For both calculation methods, the effect of SNR can be clearly observed: the predicted speech intelligibility is highest in the beginning and the end, when the SNR is 15 dB. Accordingly, it is lowest for an SNR of 0 dB and intermediate for an SNR of 5 dB. The data also indicates slight variations in the predicted STI, even when the SNR is constant. This can be explained by two effects. On the one hand, the energy distribution of running speech varies over time, which leads to a time-varying SNR. On the other hand, the estimation of both SNR and reverberation time is based on statistical methods (Ephraim and Malah, 1985; Marzinzik and Kollmeier, 2002; Löllmann and Vary, 2008), which introduce fluctuations in the prediction. In particular at higher SNR, the estimated STI is lower, when the reverberation time is included in the calculation. This reflects the detrimental effect of reverberation on speech intelligibility. In general, the estimation of reverberation characteristics becomes particularly important with increasing SNR and increasing reverberation times. Since reverberation may substantially influence speech intelligibility, it has to be accounted for within hearing perception models.

The data shown in Figure 8 have been collected at intermediate sound pressure levels and under the assumption of normal hearing. It could be seen that even for normal-hearing people, the predicted speech intelligibility was quite low in adverse acoustic environment (i.e., at low SNR). For even lower speech levels and users with a hearing loss, the elevated hearing threshold would lead to an additional loss of speech information.

In conclusion, the monitoring system can account for the external acoustic conditions, but when such an estimation of speech intelligibility is used in real applications, the uncertainties associated with the estimation procedures have to be kept in mind. In general, the estimation of signal parameters such as SNR or reverberation time is more accurate when averages are made over longer periods of time. However, to allow for an

Figure 8. Continuous estimation of speech intelligibility. The estimated Speech Transmission Index (STI) is shown as a function of time. The signal-to noise ratio (SNR) was 15, 5 or 0 dB as indicated by vertical lines for four periods of time. Circles represent data including an estimation of the reverberation time; squares indicate estimations based on the SNR alone



acceptably fast detection of changes in the acoustic environment, the time frames used for averaging must not be too long. The simulations shown were based on intervals of 2 s, which might represent a reasonable compromise between estimation accuracy and short update intervals. Further research is particularly needed to reduce estimation errors by balancing the effect of background noise and reverberation depending on the SNR. In practice, reverberation time could be estimated over longer periods since room acoustic parameters generally do not change quickly, while the SNR can vary substantially from one estimation interval to another, e.g. when a noise source is switched on. In combination with an acoustic user interface, the information provided by such an online monitoring of speech intelligibility can be used to adapt the output modality or to control other processing strategies such as noise reduction (see previous section).

The application of perception models in human-machine interaction requires reliable and generally applicable model predictions. As discussed above, current research aims at improving existing models. For this purpose, we have developed a speech intelligibility prediction (SIP) toolbox, which comprises all standardized and several state-of-the-art models for speech intelligibility and loudness prediction (see Kühler et al., 2010). The SIP Toolbox allows for an easy comparison of the different models and can support the choice of a particular model for a certain application, e.g. in the context of the monitoring of communication quality.

Personalized Hearing-Loss Compensation in Consumer Electronics

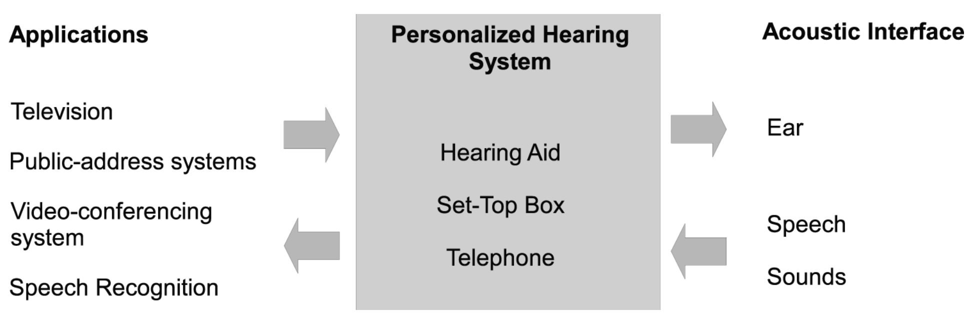
Since consumer electronics and communication systems merge more and more into multifunctional multimedia platforms, nowadays a strict differentiation between phone, PC or TV is not given any more. At the same time, these multifunctional systems can be personalized by individual graphical themes, ring tones or additional applications - called *apps*, *gadgets* or *plugins*. If human-machine interfaces for persons with hearing problems are considered, a further personalization can be achieved by using a signal playback that is adapted to the specific preferences or needs of the user. Preferences like the sound of a system may be easy to adapt, e.g. by equalizers known from common hi-fi systems. However, if the compensation of hearing deficiencies is concerned, a special challenge lies in the adaptation of the sound to the hearing ability and hearing sensation of the individual user without involvement of professionals like audiologist, who normally help to adjust such systems as hearing aids (Appell et al., 2007; Baumgartner et al., 2009; Rohdenburg et al., 2009).

The research and development in the combination of conventional audio devices with medical hearing support leads to new solutions and applications to support people of all age groups and hearing deficiencies in various hearing situations. Systems like TV or telephones may, thus, be individualized and enabled to playback sound signals pre-processed in a way that increases the intelligibility of the content by incorporating the individual hearing loss as illustrated in Figures 1 and 2. Hence, the authors worked on integration of supporting technologies for hearing impaired on a TV platform (Appell et al., 2007; Hearing At Home, 2010). Within this project an integration of audio-visual hearing-support technologies within a common digital TV/set-top-box was developed. This way, the acceptance barrier that has to be faced by conventional hearing aids was lowered to a minimum. Persons suffering from hearing losses sometimes refuse to wear hearing aids or are not even equipped with a hearing aid partly due to psycho-social factors that are associated with hearing loss like stigmatization, lack of comfort, cost or effort. This general acceptance problem can be tackled by applying supportive hearing technology into (highly accepted) home-entertainment devices, allowing the elderly people to participate in communication without using hearing aids (Appell et al., 2007; Rohdenburg et al., 2009). The fitting of such a supporting electronic device to the specific needs of the hearing-impaired person was investigated e.g. by Baumgartner et al. (2009). One suggestion for the interface of the fitting process was a simple sequence of test signals the users listened to and subsequently adjusted the volume to achieve a certain perception of these sounds (e.g. a comfortable listening level). A proper design of the test signals could extract the most relevant information about the individual hearing loss of the user without professional support. Such a system can even support persons suffering from mild hearing losses. In principle, personalized hearing aid technology can be introduced into any electronic device interacting with the user via acoustic signals as illustrated in Figure 9. The personalized hearing system, implemented e.g. in hearing aids, set-top boxes or telephones, controls the acoustic interaction between interface and application (e.g. television, public-address systems, video-conferencing, or speech recognition). Exemplarily, the described technologies for supporting people suffering from hearing deficiencies have been integrated by the authors in different electronic devices, such as a television, a tele-conference system and an I-Phone™/I-Pod Touch™ for telephony and for listening to music or MP3s. Various other areas of application are possible wherever a hearing-impaired person is not equipped with a hearing aid or is unwilling to wear it. The developed algorithms for hearing support are computationally efficient and easily adaptable to all processor based electronic devices.

SUMMARY AND CONCLUSION

In this chapter, personalized acoustic interfaces for human-computer interaction based on models of the human auditory system were discussed and concrete realizations of such interfaces were introduced. The importance of personalized and adaptable user-interfaces is commonly accepted in the field of ambient assisted living and personal tele-health. The demographic change puts even more pressure on the development of accessible user interfaces also for people with impairments. However, concrete realizations of such concepts are difficult to realize and the specific implementation often remains unclear partly due to the many degrees of freedom of personalized and adaptable human-computer interfaces. In the field of acoustic human-machine interaction, many technologies exist which can improve communication. Several individual approaches were discussed which could be used to enhance the signal quality, use personalized signal processing

Figure 9. Schematic usage of personalized hearing systems in acoustic human-machine interaction. Applications transmit and receive acoustic information, which is processed in the personalized hearing system for interaction with the user



and can detect and quantify acoustically difficult situations. In combination, these approaches can significantly improve communication systems and human-computer interfaces, e.g. for speech input or individualized signal presentation. The key towards such a combined system including the different approaches is a hearing perception model. This model can transform information about the acoustic environment, the current acoustic context and the individual user into a prediction of the user's ability to communicate with the given system. This model-based assessment of individual communication quality can provide relevant information for control and adjustment of human-machine interaction. Thus, it represents a holistic approach towards adaptable and personalized acoustic interfaces for human-computer interaction. This concept can be considered as an example also for other modalities. This may be important since accessibility can be reduced due to a number of factors, many of which are more prevalent in older persons. In cases of visual impairment or a reduced mobility, the same approach could help to increase the accessibility of user interfaces. The crucial element of this approach is a perception model, which can reliably predict the *individual* abilities of the user based on a limited set of information.

REFERENCES

Alexandersson, J., Zimmermann, G., & Bund, J. (2009). How Can I Satisfy All Users? In *Proc. Ambient Assisted Living - AAL. 2. Deutscher Kongress mit Ausstellung/ Technologien - Anwendungen - Management* (pp. 5–9). Berlin, Germany: User Interfaces for AAL.

Allen, J. B., & Berkley, D. A. (1979). Image Method for Efficiently Simulating Small-Room Acoustics. *The Journal of the Acoustical Society of America*, *65*, 943–950. doi:10.1121/1.382599

American Speech-Language-Hearing Association. (2009). *Type, degree, and configuration of hearing loss*. Retrieved September, 3, 2009 from http://www.asha.org/public/hearing/disorders/types.htm.

ANSI. (1997). *ANSI S3.5-1997. Methods for calculation of the speech intelligibility index*. American National Standards Institute.

ANSI. (2007). *ANSI S3.4-2007. American national standard procedure for the computation of loudness of steady sounds*. American National Standards Institute.

Appell, J.-E. (2002). *Loudness Models for rehabilitative Audiology*. Doctoral dissertation, University of Oldenburg, Germany.

Appell, J.-E., Hohmann, V., Gabriel, B., & Kollmeier, B. (2002). Evaluation of different 3-channel dynamic compression schemes in a field-test with a wearable dsp prototype hearing aid. *Audiological Acoustics.*, *41*, 22–46.

Appell, J.-E., Hohmann, V., Schulz, A., & Hein, A. (2007). Hearing at home. In *Fortschritte der Akustik* (pp. 629–630). Berlin: DAGA.

Baumgartner, H., Schulz, A., Hein, A., Holube, I., & Herzke, T. (2009). A fitting method for headphones to compensate individual hearing impairments. In *Proc. 3rd International ICST Conf. on Pervasive Computing Technologies for Healthcare,* London, UK. Benesty, J., Sondi, M.M., & Huang, Y.A. (Eds.). *Springer Handbook of Speech Processing*. Berlin: Springer.

Beutelmann, R., & Brand, T. (2006). Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *120*, 331–342. doi:10.1121/1.2202888

Bitzer, J., & Simmer, K. U. (2001). Superdirective microphone arrays. In Brandstein, M. S., & Ward, D. (Eds.), *Microphone Arrays: Signal Processing Techniques and Applications* (pp. 19–38). Berlin: Springer.

Bormann, V., Sust, A. S., Heinecke-Schmidt, R., Fuder, G., & Lazarus, H. (2005). Schwerhörigkeit und Sprachkommunikation am Arbeitsplatz (In German). *Technical Report Fb 1041, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin,* Dortmund/Berlin/Dresden, Germany.

Brand, T., & Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. *The Journal of the Acoustical Society of America*, *112*, 1597–1604. doi:10.1121/1.1502902

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica*, *86*, 117–128.

Cappe, O. (1994). Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor. *IEEE Transactions on Speech and Audio Processing*, *2*(2), 345–349. doi:10.1109/89.279283

Chalupper, J. & Fastl, H. (2002). Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acustica united with Acustica*, 88, 378-386.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*, 975–979. doi:10.1121/1.1907229

Davis, A. (2003). Population study of the ability to benefit from amplification and the provision of a hearing aid in 55-74 year old first time hearing aid users. *International Journal of Audiology*, 42, 2S39-2S52.

DIN. (1991). *DIN 45631. Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum, Verfahren nach E. Zwicker (Procedure for calculating loudness level and loudness)*. Deutsches Institut für Normung.

Doblinger, G. (2006). Localization and Tracking of Acoustical Sources. In Hänsler, E., & Schmidt, G. (Eds.), *Topics in Acoustic Echo and Noise Control* (pp. 91–122). Berlin, Heidelberg: Springer.

Duquensoy, A. J., & Plomp, R. (1980). Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis. *The Journal of the Acoustical Society of America*, *62*, 537–544.

Ephraim, Y., & Malah, D. (1985). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *33*(2), 443–445. doi:10.1109/TASSP.1985.1164550

European Ambient Assisted Living Innovation Alliance. (2009). *Ambient Assisted Living Roadmap*. VDI/VDE-IT AALIANCE Office.

Felchlin, I., & Hohmann, B. W. (1997). Gehörgefährdung durch "Walkman"-Geräte (In German). In *Fortschritte der Akustik - DAGA '97. Oldenburg: Deutsche Gesellschaft für Akustik e.V.* (pp. 493-494).

Goetze, S., Kallinger, M., & Kammeyer, K.-D. (2005). Residual echo power spectral density estimation based on an optimal smoothed misalignment for acoustic echo cancelation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC-2005),* (pp. 209-121), Eindhoven, The Netherlands.

Goetze, S., Kallinger, M., Mertins, A., & Kammeyer, K.-D. (2008). Multi-channel listening-room compensation using a decoupled filtered-X LMS algorithm. In *Proc. Asilomar Conf. on Signals, Systems, and Computers,* Pacific Grove, USA.

Goetze, S., Mildner, V., & Kammeyer, K.-D. (2006). A psychoacoustic noise reduction approach for stereo hands-free systems. In *Proc. Audio Engineering Society (AES), 120th Convention,* Paris, France.

Gustafsson, S. (1999). *Enhancement of Audio Signals by Combined Acoustic Echo Cancellation and Noise Reduction*. Doctoral dissertation, Aachen University of Technology, Aachen/Mainz: Wissenschaftsverlag. Aachener Beiträge zu digitalen Nachrichtensystemen, Band 11.

Habets, E. A. P. (2007). *Single and Multi-Microphone Speech Dereverberation using Spectral Enhancement*. Unpublished doctoral dissertation, University of Eindhoven, The Netherlands.

Hänsler, E., & Schmidt, G. (2004). *Acoustic echo and noise control: a practical approach*. Hoboken, NJ: Wiley. doi:10.1002/0471678406

Hänsler, E., & Schmidt, G. (Eds.). (2008). *Speech and Audio Processing in Adverse Environments*. Berlin, Heidelberg: Springer.

Hearing At Home. (2010). *Project website*. Retrieved from http://www.hearing-at-home.eu.

Huang, Y. A., Benesty, J., & Chen, J. (2006). *Acoustic MIMO Signal Processing*. Berlin, Heidelberg: Springer.

IEC. (1998). *IEC 60268-16. Sound System Equipment - Part 16: Objective rating of speech intelligibility by speech transmission index*. International Electrotechnical Commission.

ISO. (1990). [Acoustics - determination of occupational noise exposure and estimation of noise induced hearing impairment. *International Organization for Standardization*.]. *ISO*, *1999*, 1990.

Johansson, M. S. K., & Arlinger, S. D. (2003). Prevalence of hearing impairment in a population in Sweden. *International Journal of Audiology*, *42*, 18–28. doi:10.3109/14992020309056081

Jürgens, T., Brand, T., & Kollmeier, B. (2009). Consonant recognition of listeners with hearing impairment and comparison to predictions using and auditory model. In *Proc. Intern. Conf. on Acoustics DAGA/NAG 2009* (pp.1663-1666).

Katz, J., Medwetsky, L., Burkard, R. F., & Hood, L. J. (Eds.). (2009). *Handbook of clinical audiology*. Philadelphia, PA: Lippincott Williams & Wilkins.

Knapp, C. H., & Carter, G. C. (1976). The Generalized Correlation Method for Estimation of Time Delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *24*(4), 320–327. doi:10.1109/TASSP.1976.1162830

Kollmeier, B. (1992). *Moderne Verfahren der Sprachaudiometrie*. Heidelberg: Median. (In German)

Kühler, R., Rennies, J., Appell, J.-E., & Kollmeier, B. (2010, *in press*). Lautheitsabhängige Sprachverständlichkeit (In German). In *Fortschritte der Akustik DAGA 2010. Berlin: Deutsche Gesellschaft für Akustik e.V.*

Löllmann, H. W., & Vary, P. (2008). Estimation of the reverberation time in noisy environments. In *International Workshop on Acoustic Echo and Noise Control (IWAENC), Seattle, USA*.

MacWilliams, F. J., & Sloane, N. J. A. (1976). Pseudo-random sequences and arrays. *Proceedings of the IEEE*, *64*(12), 1715–1729. doi:10.1109/PROC.1976.10411

Martini, A. (Ed.). (1996). European working group on genetics of hearing impairment, *European Commission Directorate, Biomedical and Health Research Programme (HEAR). Infoletter 2, November 1996.*

Marzinzik, M., & Kollmeier, B. (2002). Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. *IEEE Transactions on Speech and Audio Processing*, *10*(2), 109–118. doi:10.1109/89.985548

Mertins, A., Mei, T., & Kallinger, M. (2010). Room Impulse Response Shortening/Reshaping with Infinity and p-Norm Optimization. *IEEE Transactions Audio, Speech, and Language Processing*, *18*(2), 249–259. doi:10.1109/TASL.2009.2025789

Meyer, R., Brand, T., & Kollmeier, B. (2007). Predicting speech intelligibility in fluctuating noise. In *Proc. 8th EFAS Congress / 10th Congress of the German Society of Audiology, Deutsche Gesellschaft für Audiologie e.V., Heidelberg*.

Mildner, V., Goetze, S., & Kammeyer, K.-D. (2006). Multi-Channel Noise-Reduction-Systems for Speaker Identification in an Automotive Acoustic Environment. In *Proc. Audio Engineering Society (AES), 120th Convention*, Paris, France.

Monzingo, R. A., & Miller, T. W. (1980). *Introduction to Adaptive Arrays*. New York: John Wiley & Sons.

Moore, B. C. J. (1985). Frequency selectivity and temporal resolution in normal and hearing-impaired listeners. *British Journal of Audiology*, *19*, 189–201. doi:10.3109/03005368509078973

Moore, B. C. J. (1995). *Perceptual Consequences of Cochlear Damage*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780198523307.001.0001

Moore, B. C. J. (1996). Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear and Hearing*, *17*, 133–161. doi:10.1097/00003446-199604000-00007

Moore, B. C. J., & Glasberg, B. R. (1997). A model of loudness perception applied to cochlear hearing loss. *Auditory Neuroscience*, *3*, 289–311.

Moore, B. C. J., & Glasberg, B. R. (2007). Modeling binaural loudness. *The Journal of the Acoustical Society of America*, *121*, 1604–1612. doi:10.1121/1.2431331

Moore, B. C. J., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness and partial loudness. *Journal of the Audio Engineering Society. Audio Engineering Society*, *45*, 224–239.

Müller, S., & Massarani, P. (2001). Transfer-Function Measurement with Sweeps. *Journal of the Audio Engineering Society. Audio Engineering Society*, *49*, 443–471.

Naylor, P. A., & Gaubitch, N. D. (Eds.). (2010). *Speech Dereverberation*. London: Springer.

Neely, S. T., & Allen, J. B. (1979). Invertibility of a Room Impulse Response. *The Journal of the Acoustical Society of America*, *66*, 165–169. doi:10.1121/1.383069

Pintelon, R., & Schoukens, J. (2001). *Identifikation of Linear Systems - A Practical Guideline to Accurate Modeling*. New York: IEEE Press.

Rennies, J., Brand, T., & Kollmeier, B. (2010a, *in press*). Modellierung binauraler Sprachverständlichkeit in verhallter Umgebung. In *Fortschritte der Akustik DAGA 2010. Berlin: Deutsche Gesellschaft für Akustik e.V.*

Rennies, J., Goetze, S., & Appell, J.-E. (2009a). How can audio technology improve working conditions? In *Fachtagung zur Bewältigung des demografischen Wandels (CHANGE 2009)*, Emden, Germany.

Rennies, J., Goetze, S., & Appell, J.-E. (2010b, *in press*). Automatic live monitoring of communication quality for normal-hearing and hearing-impaired listeners. In *Proc. 12th International Conference on Computers Helping People with Special Needs,* Vienna, Austria.

Rennies, J., & Verhey, J. L. (2009). Temporal weighting in loudness of broadband and narrowband signals (L). *The Journal of the Acoustical Society of America*, *126*, 951–954. doi:10.1121/1.3192348

Rennies, J., Verhey, J.L., Chalupper, J., & Fastl, H. (2009b). Modeling temporal effects of spectral loudness summation. *Acta Acustica united with Acustica*, 95, 1112-1122.

Rennies, J., Verhey, J.L., & Fastl, H. (2010c, in press). Comparison of loudness models for time-varying sounds. *Acta Acustica united with Acustica*.

Rhebergen, K. S., & Versfeld, N. J. (2005). A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, *117*, 2181–2192. doi:10.1121/1.1861713

Rife, D., & Vanderkooy, J. (1989). Transfer-function measurement with maximum-length sequences. *Journal of the Audio Engineering Society. Audio Engineering Society*, *37*, 419–444.

Rohdenburg, T. (2008). *Development and Objective Perceptual Quality Assessment of Monaural and Binaural Noise Reduction Schemes for Hearing Aids*. Doctoral dissertation, University of Oldenburg, Germany.

Rohdenburg, T., Goetze, S., Hohmann, V., Kammeyer, K.-D., & Kollmeier, B. (2008). Objective Perceptual Quality Assessment for Self-Steering Binaural Hearing AidMicrophone Arrays. In *Proc. IEEE Int.Conf. on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 2449-2452), Las Vegas, USA.

Rohdenburg, T., Hohmann, V., & Kollmeier, B. (2005). Objective Measures for the Evaluation of Noise Reduction Schemes. In *International Workshop on Acoustic Echo and Noise Control,* Eindhoven, the Netherlands.

Rohdenburg, T., Huber, R., van Hengel, P. W. J., Bitzer, J., & Appell, J.-E. (2009). Hearing aid technology and multi-media devices. In *ITG Fachtagung für Elektronische Medien, 13*. Dortmund, Germany: Dortmunder Fernsehseminar.

Shield, B. (2005). *Evaluation of the social and economic costs of hearing impairment*. Report for Hear-It. Retrieved October 12, 2009, from www.hear-it.org/multimedia/Hear_It_Report_October_2006.pdf.

Simmer, K. U., Bitzer, J., & Marro, C. (2001). Post-filtering techniques. In Brandstein, M., & Ward, D. (Eds.), *Microphone Arrays* (pp. 39–60). Berlin: Springer.

Sohn, W. (2001). Schwerhörigkeit in Deutschland, Repräsentative Hörscreening-Untersuchung bei 2000 Probanden in 11 Allgemeinpraxen [In German]. *Zeitschrift fur Allgemeinmedizin*, *77*, 143–147.

Sukowski, H., Brand, T., Wagener, K., & Kollmeier, B. (2009). Untersuchung zur Vergleichbarkeit des Freiburger Sprachtests mit dem Göttinger Satztest und dem Einsilber-Reimtest nach von Wallenberg und Kollmeier [In German]. *HNO*, *57*, 239–250. doi:10.1007/s00106-008-1727-9

Uimonen, S., Huttunen, K., Jounio-Ervasti, K., & Sorri, M. (1999). Do we know the real need for hearing rehabilitation at the population level? Hearing impairments in the 5- to 75-year old cross-sectional Finnish population. *British Journal of Audiology*, *33*, 53–59.

van Hengel, P. W. J., & Anemüller, J. (2009). Audio event detection for in-home care. In *Proc. Int. Conf. on Acoustics (NAG/DAGA 2009)*.

van Hengel, P. W. J., Huisman, M., & Appell, J.-E. (2009). Sounds like trouble. In deWaard, D., Godthelp, J., Kooi, F. L., & Brookhuis, K. A. (Eds.), *Human Factors - Security and Safety* (pp. 369–375). Maastricht: Shaker Publishing.

Vanderkooy, J. (1994). Aspects of MLS Measuring Systems. *Journal of the Audio Engineering Society. Audio Engineering Society*, *42*, 219–231.

Wagener, K., Brand, T., & Kollmeier, B. (1999a). Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests (In German). *Zeitschrift für Audiologie/ Audiological Acoustics*, 38, 44-56.

Wagener, K., Brand, T., & Kollmeier, B. (1999b). Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests (In German). *Zeitschrift für Audiologie/ Audiological Acoustics*, 38, 86-95.

Wagener, K., Kühnel, V., & Kollmeier, B. (1999c). Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (In German). *Zeitschrift für Audiologie/ Audiological Acoustics*, 38, 4-15.

Yost, W. A. (2000). *Fundamentals of Hearing: An Introduction, 4th edition*. London: Academic Press.

Zwicker & Fastl. H. (1999) *Psychoacoustics*. Berlin: Springer.

ADDITIONAL READING

Cohen, I., Benesty, J., & Gannot, S. (Eds.). (2010). *Speech Processing in Modern Communication*. Berlin: Springer.

Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). *Voice User Interface Design*. Amsterdam: Addison-Wesley Longman.

Gay, S. L., & Benesty, J. (Eds.). (2000). *Acoustic Signal Processing for Telecommunication*. Berlin: Springer.

Gelfand, S. A. (2001). *Essentials of Audiology, 2nd edition*. NewYork: Thieme.

Holube, I., Wesselkamp, M., Dreschler, W. A., & Kollmeier, B. (1997). Speech intelligibility prediction in hearing-impaired listeners for steady and fluctuating noise. In Jesteadt, W. A. (Ed.), *Modeling sensorineural Hearing Loss* (pp. 447–459). Mahwah, NJ: Lawrence Erlbaum Assoc.

Kahrs, M., & Brandenburg, K. (Eds.). (1998). *Applications of Digital Signal Processing to Audio and Acoustics*. Boston: Kluwer Academic Press.

Launer, S., Hohmann, V., & Kollmeier, B. (1997). Modeling loudness growth and loudness summation in hearing-impaired listeners. In Jesteadt, W. (Ed.), *Modeling sensorineural hearing loss* (pp. 175–185). Mahwah, NJ: Lawrence Erlbaum Assoc.

Lehnhardt, E., & Laszig, R. (Eds.). (2001). *Praxis der Audiometrie* (8th ed.). Stuttgart: Thieme. (in German)

Loizou, P. C. (Ed.). (2007). *Speech Enhancement: Theory and Practice*. Boca Raton, FL: CRC Press.

Lotter, T. (2004). *Single and Multimicrophone Speech Enhancement for Hearing Aids*. Doctoral dissertation, Aachen University of Technology, Aachen/Mainz: Wissenschaftsverlag. Aachener Beiträge zu digitalen Nachrichtensystemen, Band 18.

Maj, J.-B. (2004). *Adaptive Noise Reduction Algorithms for Speech Intelligibility Improvement in Dual Microphone Hearing Aids*. Doctoral dissertation, Katholieke Universiteit Leuven, Belgium.

Martin, R., Heute, U., & Antweiler, C. (Eds.). (2008). *Advances in Digital Speech Transmission*. West Sussex, UK: John Wiley and Sons. doi:10.1002/9780470727188

Moore, B. C. J. (1995). *Hearing*. San Diego: Academic Press.

Moore, B. C. J. (1999). Application of a loudness model to the fitting of hearing instruments. In Rasmussen, A. N., Osterhammel, P. A., Anderson, T., & Poulsen, T. (Eds.), *Auditory Models and Non-linear Hearing Instruments*. Denmark: Holmens Trykkeri.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). San Diego: Academic Press.

Moore, B. C. J. (2005). The use of a loudness model to derive initial fittings for hearing aids: validation studies. In Anderson, T., Larsen, C. B., Poulsen, T., Rasmussen, A. N., & Simonsen, J. B. (Eds.), *Hearing Aid Fitting*. Denmark: Holmens Trykkeri.

Moore, B. C. J., Glasberg, B. R., & Vickers, D. A. (1996). Factors influencing loudness perception in people with cochlear hearing loss. In Kollmeier, B. (Ed.), *Psychoacoustics, Speech and Hearing Aids*. Singapore: World Scientific.

Quakenbusch, S. R., Barnwell, T. P. III, & Clemens, M. A. (1988). *Objective Measures of Speech Quality*. Englewood Cliffs, NJ: Prentice Hall.

Shneidemann, B., & Plaisant, C. (2009). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (5th ed.). Amsterdam: Addison-Wesley Longman.

Srinivasan, S. (2005). *Knowledge-based Speech Enhancement*. Doctoral dissertation, KTH Royal Institute of Technology, Stockholm, Sweden.

Vary, P., & Martin, R. (2006). *Digital Speech Transmission*. West Sussex, UK: John Wiley and Sons. doi:10.1002/0470031743

KEY TERMS AND DEFINITIONS

Adaptability: The property of an acoustic user interface to allow for different means of interaction depending on the current context.

Audiogram: More specifically called pure-tone audiogram. Indicates the level of pure tones needed to be just audible relative to a standardized normal-hearing population.

Cocktail-Party Effect: Ability of the hearing system to separate different sound sources in a complex acoustic environment comprising different sound sources.

Dereverberation: The process of removing those parts of an acoustic signal that result from reflections of the sound at boundaries of an enclosed space.

Dynamic Range: The range of levels between the threshold of hearing and the uncomfortable level. In most cases reduced in hearing-impaired persons.

Loudness Recruitment: Abnormal perception of loudness in consequence of a hearing loss. Typically, loudness of soft sounds increases faster with level than normal.

Noise Reduction: Also called de-noising. The process of removing unwanted signal components from an acoustic signal (e.g. machinery noise, driving noise, concurrent speech).

Personalization: Inclusion of information on the individual user (e.g. the individual hearing loss) in the processing of acoustic signals.