Modelling binaural speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired subjects

Von der Fakultät für Mathematik und Naturwissenschaften

der Carl-von-Ossietzky-Universität Oldenburg

zur Erlangung des Grades und Titels eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

angenommene Dissertation

von Herrn

Dipl.-Phys. Rainer Beutelmann

geboren am 23. Juni 1977

in Mainz

Gutachter: Prof. Dr. Dr. Birger Kollmeier

Zweitgutachter: Prof. Dr. Volker Mellert

Tag der Disputation: 3. November 2008

Abstract

Speech intelligibility in complex situations, the so-called "cocktail party problem" (Cherry, 1953), is strongly affected by the ability of the listener to use both ears, that is to use binaural hearing. Differences in sound source location between target and interferer may cause differences in the speech reception threshold (the signal-to-noise ratio at which an intelligibility of 50% is achieved) of up to 12 dB in anechoic conditions (Bronkhorst, 2000). The number and position of sound sources, reflections and reverberation, and several other factors influence the amount of binaural unmasking. Especially for hearing-impaired listeners, this benefit due to binaural hearing or its absence can be essential.

Being able to predict the binaural speech intelligibility from given knowledge of the situation, for example a binaural recording at the place of the listener, is valuable for the acoustical design of rooms, for audiology and hearing-aid fitting, and of course generally for the understanding of the underlying mechanisms.

This dissertation presents the development of a model of binaural speech intelligibility and its evaluation for selected binaural conditions. The model uses a multi-band equalization-cancellation process based on the principle by Durlach (1963) as a binaural front end for the speech intelligibility index (ANSI, 1997). The model was extended for the prediction of binaural speech intelligibility in fluctuating noise and the validity of the multi-band approach with independent binaural processing in different frequency bands was examined. Generally, the model is capable of predicting the binaural speech reception threshold for normal-hearing and hearing-impaired subjects in situations with one steady-state, speech-shaped noise source at different azimuths in the horizontal plane and under different room acoustical conditions. The prediction of binaural speech

intelligibility in fluctuating noise is less accurate, but reasonable as a proof of concept. About 70% of the variance due to individual hearing-impairment can be predicted using the hearing threshold input parameter to the model, the remaining variance may be attributed to other, presumably supra-threshold aspects of the impairment. A critical experiment was able to show that the hypothesis of independent binaural processing in adjacent frequency bands cannot be rejected.

Kurzfassung

Sprachverständlichkeit in komplexen Situationen, das sogenannte “Cocktail-Party-Problem” (Cherry, 1953), wird stark von der Fähigkeit des Hörers beeinflusst, beide Ohren benutzen zu können, das heißt, binaurales Hören auszunutzen. Unterschiede der Schallquellenposition zwischen Sprache und Störgeräusch können Unterschiede in der Sprachverständlichkeitsschwelle (das Signal-Rausch-Verhältnis, bei dem eine Verständlichkeit von 50% erreicht wird) von bis zu 12 dB bewirken (Bronkhorst, 2000). Die Zahl und Position der Schallquellen, Reflexionen und Nachhall, sowie etliche weitere Faktoren beeinflussen die Größe des binauralen Gewinns. Insbesondere für Schwerhörende ist der binaurale Gewinn oder seine Abwesenheit entscheidend.

Die Fähigkeit, binaurale Sprachverständlichkeit aus vorgegebenem Wissen über die Situation vorherzusagen, zum Beispiel einer binauralen Aufnahme am Ort des Hörers, ist nützlich für das akustische Design von Räumen, für die Audiologie und Hörgeräteanpassung, und natürlich grundsätzlich für das Verständnis der zugrundeliegenden Mechanismen.

Diese Dissertation stellt die Entwicklung eines Modells der binauralen Sprachverständlichkeit und seine Überprüfung für ausgewählte binaurale Situationen vor. Das Modell benutzt einen Equalization-Cancellation-Prozess in mehreren Bändern basierend auf dem Prinzip von Durlach (1963) als binaurale Vorstufe für den Speech Intelligibility Index (ANSI, 1997). Das Modell wurde für die Vorhersage von binauraler Sprachverständlichkeit in fluktuierendem Rauschen erweitert, und die Gültigkeit der Annahme, dass die binaurale Verarbeitung in verschiedenen Bändern unabhängig voneinander ist, wurde überprüft. Insgesamt ist das Modell in der Lage, binaurale Sprachverständlichkeitsschwellen für Normalhörende und Schwerhörende in Situa-

tionen mit einer stationären, sprachsimulierenden Störquelle an verschiedenen Azimutwinkeln in der Horizontalebene und unter verschiedenen raumakustischen Bedingungen vorherzusagen. Die Vorhersage binauraler Sprachverständlichkeit in fluktuierendem Rauschen ist weniger genau, aber als Machbarkeitsstudie angemessen. Ungefähr 70% der Varianz infolge des individuellen Hörverlusts lässt sich auf Basis des Audiogramms vorhersagen, die verbleibende Varianz ist wahrscheinlich von anderen, überschwelligen Faktoren des Hörverlusts abhängig. Mit einem Kontrollexperiment konnte gezeigt werden, dass die Hypothese der unabhängigen binauralen Verarbeitung in benachbarten Frequenzbändern nicht abgelehnt werden kann.

Contents

1. General Introduction

Speech is an important means of communication and a key to many aspects of social life. It is a multi-faceted topic, whose details are studied in various scientific fields. Speech carries emotions and factual information and is used to express relationships. Speech sounds are some of the first sensory perceptions of an embryo and speech is the first sophisticated communication medium that a child learns, long before for example reading and writing. Speech separates human conversation from the acoustical interaction of animals, because it transmits complex symbolic content on multiple levels. Nevertheless, or maybe actually because speech is such a fundamental part of life, people usually rather think about the content of the message than bother about how to produce or to receive speech, except when the situation is exceedingly adverse or for example if the listener is affected by a hearing impairment. This is remarkable, given that already on the acoustical level it is unlikely or even impossible to achieve an undisturbed transmission from the speaker to the listener. Although a lot of acoustically caused errors are compensated for by redundancies on various levels of speech itself (e.g., acoustic, syntactic, semantic), a considerable amount of effort is needed by the listener in order to gather useful speech information from the noisy signals which are received by the ears. In this context, a very essential requirement is an unimpaired ear, because a hearing loss substantially decreases the ease of conversation, especially in loud environments.

This dissertation approaches the role of the receiver in speech communication from a psychoacoustical, especially binaural, point of view, considering the function of the auditory system in the receiver's task. A very descriptive term was coined by Cherry (1953), who called it the "cocktail party problem": How does the auditory system extract the desired speech from a mixture of the target signal, other speakers, and ambient noise, often additionally distorted by room reflections and reverberation? It is generally assumed, that the auditory system of a listener performs some kind of "auditory scene analysis" (Bregman, 1990). This means segregating the received signals into components and grouping them into "auditory objects" according to attributes like spatial location, common onset or comodulation (similar signal envelope in different frequency regions), harmonicity, etc. and trying to anticipate and follow the progress of these "auditory objects", which is called streaming.

The core of this dissertation is a model of the contribution of binaural hearing to the solution of the "cocktail party problem", a *model of binaural speech intelligibility*. The model is based on previous work by vom Hövel (1984) and Zurek (1990) and was implemented as a numerical model (in contrast to the largely analytical approaches in literature) in MATLAB®. It is intended for the prediction of speech reception thresholds (SRTs, the signal-to-noise ratio at which an intelligibility of 50 % is achieved) for binaural signals, taking potential effects of sound source locations, reverberation, hearing impairment, and noise modulation into account. The long-term object is to develop and validate the model of binaural speech intelligibility to include as many aspects of the "cocktail party problem" as possible. While being empirical and descriptive and rather based on data from speech intelligibility experiments and signal processing theory than on detailed physiological knowledge, it is a top-down approach which is supposed to complement more exact but also more complex

bottom-up approaches based on physiology (e.g., Colburn, 1977a). But regardless of the phenomenological nature of the model and observed data, specifically chosen experimental paradigms are not only used to substantiate model parameters or to indicate the need for extensions, but may also give a hint at underlying details of the auditory system.

A model of binaural speech intelligibility has several benefits. It serves as a tool for fundamental research on binaural hearing and speech perception. Furthermore, the experimental data collected for the validation of the model are useful independently of the model. Knowing how the unimpaired auditory system works makes it easier to provide help for the hearing-impaired. In audiology, the model may be used as a reference for the expected loss of speech intelligibility based on other measures of the hearing-impairment. It can reduce measurement time in the daily work of an audiologist or, in comparison with an actual measurement, indicate exceptional types of hearing loss. The model can furthermore be applied to asses the performance, with regard to speech intelligibility, of binaural hearing aid algorithms or similar binaural algorithms and devices in general audio technology. It may also be helpful in room acoustics, by estimating specific binaural effects in real or simulated rooms, in addition to already existing monaural measures like the Speech Intelligibility Index (SII, ANSI, 1997) or the Speech Transmission Index (STI IEC, 1998), thus saving the expense of subjective measurements.

The prediction of binaural speech intelligibility spans a bridge between a number of research fields, which separately provide a very good basis of comprehensive studies. Monaural speech intelligibility prediction has a long tradition with the Articulation Index (AI, ANSI, 1969; French and Steinberg, 1947; Kryter, 1962) and its successor, the Speech Intelligibility Index (SII, ANSI, 1997), both of which have become standardized.

They are based on the concept that the fraction of maximally possible information that is delivered to the listener is a function of the signal-to-noise ratio (SNR) in narrow frequency bands. The contribution of each frequency band is weighted by its importance, which is empirically determined and dependent on specific speech material and test conditions. Hearing loss is included in form of an additional, "internal" noise, which sets an upper limit for the SNR, if the external noise level is below the individual hearing threshold. It has been shown, that the prediction of monaural speech intelligibility for hearing-impaired subjects is basically possible (Smoorenburg, 1992; Pavlovic et al., 1986), but the residual variance is still quite large and more factors than only the hearing threshold have to be considered (Plomp and Mimpen, 1979). The SII has also successfully been extended in order to predict the effect of noise fluctuations (Rhebergen et al., 2005; Wagener, 2003). A more elaborate way to determine the amount of information in each frequency band is realized in the Speech Transmission Index (STI, IEC, 1998; Houtgast and Steeneken, 1973), which is widely used in room acoustics. The STI is based on the modulation transfer function between the original (speech) signal and the (monaural) signal at the position of the listener, including interfering noise and reverberation. The advantage compared to the SII is that it is not necessary to record speech and interference separately and that detrimental effects of reverberation on speech itself are correctly considered as decreasing the effective SNR. Both, SII and STI, integrate the calculated band-wise fractions of information without consideration of interaction or correlation between the frequency bands. The Speech Recognition Sensitivity (SRS, Müsch and Buus, 2001) includes these effects.

An extension of speech intelligibility prediction towards binaural signals is, of course, based on the thorough research that has been performed on binaural phenomena, mostly on binaural tone detection. Zurek (1990) was able to predict a large set of speech

intelligibility data taken from literature quite well by assuming that the monaural SNR in a given frequency band could be increased for a certain binaural configuration by the binaural masking level difference (BMLD, the difference between the masked threshold of the binaural configuration and a monaural or diotic reference condition) of a pure tone at that frequency and in the same binaural configuration. He calculated the BMLD using an empirical formula (from Colburn, 1977a) and a simple head model in order to determine the interaural phase difference. Culling et al. (2004) reported similar results with previously measured subjective BMLDs. Several models of binaural interaction exist in literature and they are differently well suited for a practicable model of binaural speech intelligibility. Assuming that a model of the frequency-dependent BMLD is a key to binaural speech intelligibility prediction, binaural models which only deal with lateralization or which are not capable of quantitative predictions of BMLDs are of limited use. That excludes the use of models like the ones by Jeffress (1948) or Lindemann (1986), or simple "count-comparison" models (in the terminology of Colburn and Durlach, 1978) like for example the ones by von Békésy (1930) and van Bergeijk (1962). Most promising is the equalization-cancellation (EC) theory by Durlach (1972), which offers the possibility to be used as a signal processor. Other models, which are based on the interaural cross-correlation, for example as presented by Osman (1971), are mathematically very close to the EC theory (cf. Zerbs, 2000).

The most straightforward approach which was chosen in this dissertation (based on the work by vom Hövel, 1984), was to combine a multi-frequency-band EC process as a binaural processor with the SII. The SII, although simpler than the other mentioned methods of speech intelligibility prediction, matches the SNR-maximising principle of the EC process and offers the possibility of including hearing impairment in an easy way. Future work may still imply to combine the model presented in this dissertation

with advantages of, for example, the STI or the SRS model.

The content of this dissertation splits up into three parts. The first part (chapter 2, published in the Journal of the Acoustical Society of America; Beutelmann and Brand, 2006) introduces the basic concept of the model, the combination of the binaural EC principle with the monaural SII and details of a first straightforward numerical implementation. Modifications of the original idea by vom Hövel (1984) are explained, above all the introduction of an internal masking noise based on individual audiogram data in order to incorporate hearing impairment in form of the hearing threshold. The model predictions are verified with experimental data from normal-hearing and hearing-impaired subjects. The measurement conditions comprise several spatial arrangements of a speech source in front of the listener and a stationary speech-shaped noise source at various azimuths in three differently reverberant rooms.

The second part (chapter 3) deals with two main issues. The first one is a review of the analytical basis of the EC principle and how to reformulate it in order to reduce the amount of time needed to compute the predictions. The second issue is an extension of the model from chapter 2 that includes a way to predict binaural speech intelligibility not only in stationary, but also in fluctuating noise. The model implementation in chapter 2 was simple and good for general validation, but not efficient enough for practical use. Therefore, a possibility was sought to eliminate unnecessary computation by analysis of the EC formula and subsequent transformation into a numerically favorable form. A useful side effect of the reformulation was that the role of binaural signal parameters like the interaural level difference and interaural correlation in the calculation of the signal-to-noise ratio after EC processing was pointed out without detailed assumptions about the input signals. Furthermore, a first step towards the prediction of combined effects of noise modulation and binaural

hearing on speech intelligibility was examined, since interfering noises in "cocktail party" situations are often non-stationary, for example when only a small number of other talkers are nearby. It has to be expected that both effects interact, because room reverberation influence modulations and interaural correlation. The approach was based on previous work by Brand and Kollmeier (2002b) and Rhebergen et al. (2005) and basically consists of calculation of the model for several short-time frames and averaging over the results. The model predictions were compared to observed data of normal-hearing and hearing-impaired subjects in anechoic and reverberant conditions.

The third part (chapter 4) is focused on an important detail of the model: the auditory filter bank. Speech and noise interferers are mostly broad-band signals and their model evaluation requires division into rather narrow auditory bands according to the concept of auditory filters (Fletcher, 1940; Patterson, 1976). Binaural signal detection models for pure tones in noise usually use only a single auditory filter (although this appears to be inaccurate for certain conditions (Hall et al., 1983)), but for speech signals, a spectral region which is wider than a single auditory filter has to be considered. This poses the question, whether the binaural processing in adjacent auditory filters can be assumed to be independent of each other, and whether this hypothesis is valid for the model. Furthermore, there is no complete agreement in literature about the effective bandwidth of auditory filters in the binaural case (Kohlrausch, 1988; Kollmeier and Holube, 1992; Holube et al., 1998), which leaves another uncertainty. Therefore an experimental setup with strongly frequency-dependent interaural phase differences was used, which was suitable for answering the above-mentioned questions for the binaural speech intelligibility model and additionally permitted to examine the interaction of binaural information between remote auditory filters.

2. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners[1]

Abstract

Binaural speech intelligibility of individual listeners under realistic conditions was predicted using a model consisting of a gammatone filter bank, an independent equalization-cancellation (EC) process in each frequency band, a gammatone resynthesis, and the speech intelligibility index (SII). Hearing loss was simulated by adding uncorrelated masking noises (according to the pure-tone audiogram) to the ear channels. Speech intelligibility measurements were carried out with eight normal-hearing and 15 hearing-impaired listeners, collecting speech receptions threshold (SRT) data for three different room acoustic conditions (anechoic, office room, cafeteria hall) and eight directions of a single noise source (speech in front). Artificial EC processing errors derived from binaural masking level difference data using pure tones were incorporated into the model. Except for an adjustment of the SII-to-intelligibility mapping function, no model parameter was fitted to the SRT data of this study. The overall correlation coefficient between predicted and observed SRTs was 0.95. The dependence of the SRT of an individual listener on the noise direction and on room acoustics was predicted with

[1]This chapter has been published in the present form in the Journal of the Acoustical Society of America (Beutelmann and Brand, 2006).

a median correlation coefficient of 0.91. The effect of individual hearing impairment was predicted with a median correlation coefficient of 0.95. However, for mild hearing losses the release from masking was overestimated.

2.1. Introduction

A binaural model, capable of predicting speech intelligibility under the influence of noise, reverberation, and hearing loss, may help in understanding the underlying mechanisms of binaural hearing and may assist in the development and fitting of hearing aids. In this study, a binaural model of speech intelligibility based on an approach by vom Hövel (1984) is presented and the model predictions are compared to measurement data. It combines two established models, the binaural equalization-cancellation (EC) processing (Durlach, 1963) with the monaural speech intelligibility index (SII, ANSI, 1997).

A number of studies are concerned with measuring the effects of spatial unmasking of speech. A detailed overview can be found in a review by Bronkhorst (2000). Research has focused on the influence of synthetic and natural spatial cues on speech intelligibility (Platte and vom Hövel, 1980; Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1988; Peissig and Kollmeier, 1997), on the influence of reverberation (Moncur and Dirks, 1967; Haas, 1972; Nábělek and Pickett, 1974) and hearing loss (Duquesnoy, 1983; Festen and Plomp, 1986; Irwin and McAuley, 1987; Bronkhorst and Plomp, 1989).

Spatial unmasking of speech is based on spatial differences between target talker and interfering sources and can cause a benefit of speech reception threshold (SRT) of up to 12 dB (Bronkhorst, 2000). The basic cues for binaural processing are interaural time differences (ITD) due to the distance between the ears and interaural level differences

(ILD) mainly due to the head shadowing effect. There are also spectral cues, mainly caused by the geometry of the pinna, but they play a less important role in spatial unmasking of speech (Mesgarani et al., 2003).

A number of standardized methods of monaural speech intelligibility prediction exist in the literature, for instance the articulation index (AI; ANSI, 1969; Fletcher and Galt, 1950) and the speech intelligibility index (SII, ANSI, 1997), which was derived from the AI. A recent development by Müsch and Buus (2001a,b); Müsch and Buus (2004), the speech recognition sensitivity (SRS) model, incorporates interactions between frequency bands which were neglected by the AI and SII. In this study, the standardized SII (ANSI, 1997) was used. However, the binaural part of the model is independent of the method for speech intelligibility prediction. Consequently, other methods can be used as well.

Models of binaural interaction in psychoacoustics, such as the models by Jeffress (1948), Osman (1971), Colburn (1977a) and Lindemann (1986), provide a basis for some binaural speech intelligibility models. Zerbs (2000) and Breebaart et al. (2001a) each described a binaural signal detection model that uses peripheral preprocessing (modelled outer/middle ear, basilar membrane and haircells) which converts the signals arriving at the ears into an internal representation. The binaural processing is done by an equalization-cancellation (EC) type of operation according to the theory by Durlach (1972). Both models differ in details, mainly of the way the internal inaccuracies are handled. The model presented here also makes use of the EC theory, but is kept simpler by omitting the peripheral preprocessing and working directly on the signals.

The model of Culling and Summerfield (1995) in some way spans the gap between rather psychoacoustic binaural models and models related to binaural speech perception. It has been used to predict the release of masking for vowel intelligibility, but only

qualitatively in the form of processed vowel spectra, where certain features could be identified or not. It incorporates most of the elements which were also used in this study, namely waveforms as input signals, a peripheral filter bank and an equalization-cancellation type mechanism. Particularly, it features independent delays in each frequency band. There was no need for level equalization, because the stimuli contained only binaural time or phase differences.

Existing models of binaural speech intelligibility (Levitt and Rabiner, 1967; Zurek, 1990; vom Hövel, 1984) have certain common elements. They act as a preprocessing unit for monaural speech intelligibility models like the AI (Levitt and Rabiner, 1967; Zurek, 1990) or a modified version of the AI (vom Hövel, 1984). The benefit due to binaural interaction is expressed as a reduction of masking noise level after binaural processing. The models differ in the way they calculate the release of masking. Levitt and Rabiner (1967) used frequency dependent binaural masking level differences (BMLD) for interaurally phase reversed tones in diotic noise, taken from Durlach (1963), and subtracted these from the masking noise level. Zurek (1990) calculated the release of masking with the help of an equation from Colburn (1977b), using measured interaural level differences and an analytical expression for interaural phase differences. Vom Hövel (1984) derived an expression for the increase in signal-to-noise ratio based on EC theory. He used interaural parameters from actual transfer functions and incorporated a coarse estimate of the influence of reverberation.

The model presented in this study processes signal waveforms. Two uncorrelated internal masking noises accounting for the individual hearing thresholds of the two ears are added prior to dividing the binaural input signals into frequency bands and further processing. Independent EC stages in each band with artifical errors, which simulate human inaccuracy, calculate residual monaural signals consisting of speech

and noise with the best possible signal-to-noise ratio. These signals are resynthezised into one broadband signal and with the aid of the SII a speech reception threshold is computed. Speech and noise have to be available as separate signals.

The goal of the present work was to determine the ability of such a straightforward functional model to predict binaural speech intelligibility under realistic conditions such as spatial sound source configuration, reverberation and hearing loss. Model predictions were compared to observed SRTs for various combinations of noise source azimuths, room acoustic conditions and hearing losses. To begin with, the idea by vom Hövel (1984) was maintained as far as possible, i.e. the combination of EC and SII and especially the original EC parameters. Only the SII-to-intelligibility mapping function was adjusted to measurement data from normal-hearing subjects without binaural and room acoustic cues, all other parameters were taken from literature and not fitted to speech intelligibility measurement data. Particular attention was paid to which of the listeners' individual characteristics (such as the pure tone audiogram) were necessary as parameters to produce accurate predictions. As a compromise between realistic situations and easy handling, measured manikin head related transfer functions including room impulse responses have been used.

2.2. Methods

2.2.1. Model of binaural hearing

The model used in this study applies the Equalization-Cancellation principle (similar to the one proposed by Durlach, 1963), combined with the Speech Intelligibility Index (ANSI, 1997) in order to predict binaural speech reception thresholds (SRT) in noise.

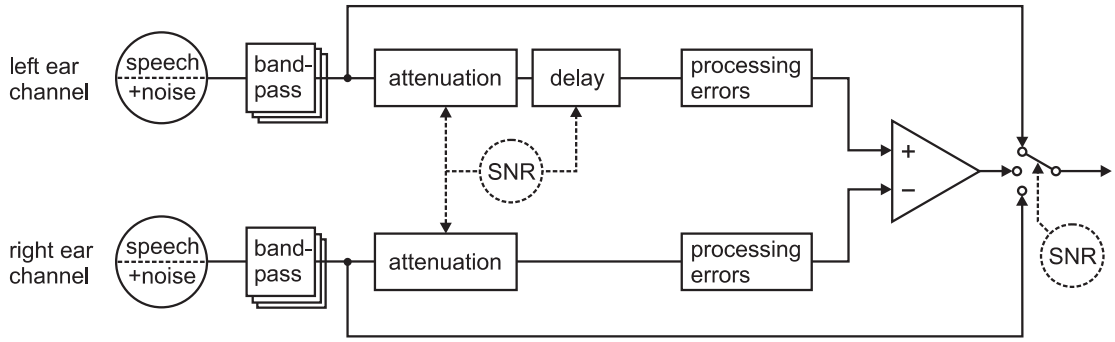


FIG. 2.1. *Binaural processing using the modified, multi frequency channel EC-model according to vom Hövel (1984). The speech and noise signals are processed identically, but separately for exact SNR calculation. The noise signal part includes the internal masking noise. Attenuation is only applied to one of the channels, depending on which of them contains more noise energy compared to the other.*

Additional masking noises were used to simulate the effects of hearing impairment. The binaural part is shown schematically in Fig. 2.1.

In the following, the inputs from the left and right ears will be referred to as "left ear channel" and "right ear channel", respectively. Each ear channel includes both speech and noise. Different parts of the interfering noise signal (cf. 2.2.2) from the Oldenburg Sentence Test (Wagener et al., 1999a,b,c) filtered with the respective HRTFs were used as speech input signals and as noise input signals. These signals had the same long term spectrum as the speech material used in the speech intelligibility measurements (important for the SII), speech and noise were uncorrelated (important for the EC model) and the variations of the actual sentences in level and spectrum were avoided. The speech and noise signals were supplied separately to the model to allow for exact SNR calculation. There was no difference between processing the sum of speech and noise or both signals separately and summing afterwards, since all processing steps were linear. The entire model was implemented using MATLAB® (MathWorks, 2002).

The SII part was based on a MATLAB® implementation of the one-third octave band SII procedure by Müsch (2003).

Gammatone filter bank analysis

The input signals were split into 30 frequency bands. Each band was one ERB (equivalent rectangular bandwidth, Glasberg and Moore, 1990) wide with center frequencies from 146 Hz to 8346 Hz using a gammatone filter bank (Hohmann, 2002). Frequency components beyond this range were considered irrelevant for speech intelligibility. The gammatone filter transfer functions are based on the shape and bandwidth of the auditory filters of the basilar membrane (Patterson, 1976). The gammatone filter bank provides complex analytical output signals, which can be resynthesized after the binaural model processing with negligible artefacts.

Internal masking noise

Individual hearing thresholds were modelled by adding uncorrelated (between the left and right ear channel) Gaussian noise signals as internal masking noise to the external masking signals. The spectral shape of the internal masking noise was determined by the individual pure tone audiogram for the left and right ear, respectively. In each frequency band of the gammatone filter bank, the total noise energy equaled the energy of a sine tone 4 dB above the individual hearing threshold level at the corresponding band center frequency (Breebaart et al., 2001a; Zwicker and Fastl, 1999).

EC stage

The equalization-cancellation processing takes advantage of the fact that signals from different directions result in different interaural time and level differences. It aims at maximizing the signal-to-noise ratio (SNR) in each frequency band. A simple way to

maximize the SNR is to choose the ear channel with the largest SNR, but in many cases it is possible to utilize the time and level differences to exceed the SNR obtainable with a monaural signal.

The binaural processing (shown schematically in Fig. 2.1) is carried out in the model as follows: In each frequency band, the ear channels are attenuated and delayed[2] with respect to each other (*equalization* step), and then the right channel is subtracted from the left (*cancellation* step). The gain and delay parameters for the equalization step are chosen such that after cancellation step the SNR is maximal[3]. Thus there is no need for explicit decision between the two possible strategies of either minimizing the noise level or maximizing the speech level. The actual amplitude equalization is always realized by means of attenuating the correct ear channel rather than amplifying the other, because in this way a seamless transition to the monaural case is achieved with increasing attenuation.

The noise level is minimized by subtracting one ear channel from the other, because all correlated noise components which are aligned after the equalization step can be eliminated due to destructive interference. Assuming that only the time and level differences of the noise signals are completely compensated for, but not the differences

[2]The time delay of one channel relative to the other one was realized by means of fast fourier transformation and multiplication with a phase factor in the frequency domain. This allowed delay times smaller than the sample period. The signals were padded with sufficient zero samples (about 3.5 ms) at both ends to avoid circular aliasing.

[3]A numerical optimization procedure (simplex-based MATLAB® function `fminsearch`) was used to find the optimum gain and delay values, which yielded maximum SNR. The SNR was calculated via the RMS difference between the resulting speech and noise signal after subtraction of the amplified and delayed left ear channel from the right one. Suitable initial gain and delay values for the optimization procedure were estimated by evaluating a short section of the noise signal: the RMS difference between the ear channels was used as gain initial value, the delay was initialized with the lag of the cross correlation maximum. The SNR as a function of gain and delay exhibits local maxima due to the periodic structure of the bandpass filtered signals. To find the global maximum (assumed that a first search may have only found a local maximum) the optimization procedure was started again with initial parameters close to neighboring local maxima. These could be found at delay intervals calculated from the center frequency of the current bandpass ($1/f_c$).

of the speech signals (when noise and speech come from different directions), more speech than noise remains in the resulting signal, which effectively increases the SNR. If the best possible SNR after binaural processing was still lower than the largest SNR of the monaural signal pairs, the best monaural signal pair was used in the SII calculation.

Artificial processing errors

Durlach (1963) proposed an artificial variance of the gain and delay parameters used in the EC process in order to model human inaccuracy. The model presented here used a modified way of calculation according to vom Hövel (1984). The underlying assumption is that the EC processing in a given channel is carried out simultaneously in a number of parallel, equivalent processing units, which only differ in their (time invariant) processing errors. The final result is averaged over the outputs of all processing units (see below).

The gain errors (ε_L, ε_R) and delay errors (δ_L, δ_R) of the left and right ear channel were Gaussian distributed, ε_L and ε_R on a logarithmic scale (level), δ_L and δ_R on a linear scale (time). Their standard deviations, σ_ε and σ_δ depended on the actual gain (α) and delay (Δ) settings in each frequency band of the EC process defined by the following equations:

$$\sigma_\varepsilon = \sigma_{\varepsilon 0}\left[1 + \left(\frac{|\alpha|}{\alpha_0}\right)^p\right] \qquad \sigma_\delta = \sigma_{\delta 0}\left(1 + \frac{|\Delta|}{\Delta_0}\right) \tag{2.1}$$

with $\sigma_{\varepsilon 0} = 1.5\,\mathrm{dB}$, $\alpha_0 = 13\,\mathrm{dB}$, $p = 1.6$ and $\sigma_{\delta 0} = 65\,\mu\mathrm{s}$, $\Delta_0 = 1.6\,\mathrm{ms}$. Vom Hövel (1984) calculated these parameters by fitting BMLD predictions to results from measurements with pure tones in noise using a single frequency band ($f_0 = 500\,\mathrm{Hz}$) of his model with

the above-mentioned processing errors. In this way, vom Hövel (1984) was able to predict BMLD data in S_0N_τ and $S_\pi N_\tau$ situations (Langford and Jeffress, 1964) with less deviation from the data than with the original model of Durlach (1963), which only limited the delay values to $|\Delta| < (2f_0)^{-1}$ in order to introduce artifical inaccuracy. Particularly in the S_0N_τ situation, the original model prediction had discontinuities which did not occur in the data of Langford and Jeffress (1964) and in the predictions of vom Hövel (1984). For the gain errors, BMLD data in S_mN_a situations (Blodgett et al., 1962; Egan, 1965) were used, with monaural presentation (m) of the signal and various noise ILDs (a). These, too, could be predicted with the model of vom Hövel (1984) with deviations in the range of about 1 dB, while the original model of Durlach (1963) predicted BMLD values which were way to small and did not even fit qualitatively to the measured data.

In this study, the artificial errors were taken into account using a Monte Carlo method by generating 25 sets of Gaussian distributed random numbers for each of the 30 frequency bands with standard deviations according to Eq. (2.1) and adding them to the previously found optimal gain and delay values. All subsequent processing steps were carried out repeatedly for each of the 25 sets of errors resulting in a set of SRTs from which a mean SRT was calculated. Each SRT prediction is derived from 750 random values (i.e. 30 frequency channels times 25 Monte Carlo drawings), which supplies a sufficient statistical basis.

Gammatone filter bank synthesis

The resulting speech and noise signals from each frequency band were resynthesized as described in Hohmann (2002) into a broadband speech and noise signal after the EC stage. The resynthesis step consisted of a phase and group delay adjustment in order

to equalize the analysis filters according to Hohmann (2002), followed by a simple addition of the frequency bands. The broadband monaural signals were then used in the calculation of the speech intelligibility index. The signals could also be listened to or could be used to examine the benefit of the model binaural processing for human speech intelligibility using SRT measurements.

Speech intelligibility index

The SII was calculated from the resulting speech and noise spectra according to ANSI S3.5-1997 using the one-third octave band computational procedure (ANSI S3.5-1997, Table 3) with the band importance function "SPIN" (ANSI S3.5-1997, Table B.2). The hearing threshold level was set to -100 dB HL in the SII procedure, because the effect of hearing threshold was already taken into account by the internal masking noise (cf. 2.2.1).

Intelligibility scores for a number of overall speech levels (at constant noise level) were calculated from the corresponding SII values using a mapping function derived from the mapping function for "sentence intelligibility (I)" from Fletcher and Galt (1950, Table III, p. 96, and Fig. 7, p. 99). An adjustment of the SII-to-intelligibility mapping function is necessary to account for differences between the articulation of different speech materials. In this study, the adjustment was based on the anechoic S_0N_0 situation (cf. 2.2.2), since in this situation no binaural (same HRTF for speech and noise) or room acoustical effects are involved. First, a suitable analytical function ($P(SII)$, the intelligibility score in percent as a function of the SII, Eq. (2.2)) was chosen, which described the original mapping function as close as possible.

$$P(SII) = \frac{m}{a + e^{-b \cdot SII}} + c,\ P(0) = 0,\ P(1) = 1 \tag{2.2}$$

For the SRT calculation, only the SII at 50 % intelligibility is important, therefore only the parameter $a = 0.01996$ was fitted to the anechoic S_0N_0 measurement data of the normal-hearing subjects. b was set to 20, which yields a slope (at the SRT) of the resulting psychometric function (intelligibility against SNR) close to the one measured by Wagener et al. (1999c) for the Oldenburg Sentence Test in noise (17.1 %/dB). $m = 0.8904$ and $c = -0.01996$ are defined by the boundary conditions. The parameters for the original mapping function from Fletcher and Galt (1950) were $a = 0.1996$, $b = 15.59$, $m = 0.2394$ and $c = -0.1996$. The SRT was obtained by a simple search algorithm, which iteratively calculated an estimate of the psychometric function from the previously determined intelligibility scores and stopped, if the difference between the actual intelligibility at the estimated SRT and 50 % was below a certain threshold (0.1 %).

2.2.2. Measurements

Subjects

A total number of 10 normal-hearing and 15 hearing-impaired subjects participated in the measurements. Their ages ranged from 21 to 43 years (normal-hearing) and from 55 to 78 years (hearing-impaired).

The hearing levels of the normal-hearing subjects exceeded 5 dB HL at four or less out of 11 audiometric frequencies and 10 dB HL at only one frequency. None of the thresholds exceeded 20 dB HL.

The hearing-impaired subjects had various forms of hearing loss, including symmetric and asymmetric, flat, sloping and steep high frequency losses. They are listed in Table 2.1. Their (monaural) pure tone average (PTA, at 1 kHz, 2 kHz and 4 kHz) ranged

TABLE 2.1. *Hearing threshold at 500 Hz, pure tone average (mean of the hearing threshold in dB HL over 1 kHz, 2 kHz and 4 kHz), hearing loss type and noise level in dB SPL used for the sentence tests of all hearing-impaired subjects participating in this study.*

subject	left ear			right ear			noise
number	500 Hz	PTA	type	500 Hz	PTA	type	level
1	10.0	15.0	high freq	50.0	31.7	flat	65
2	5.0	33.3	steep	5.0	26.7	steep	50
3	35.0	40.0	flat	35.0	35.0	flat	60
4	45.0	58.3	flat	5.0	18.3	high freq	65
5	15.0	41.7	high freq	20.0	43.3	high freq	60
6	35.0	50.0	sloping	25.0	41.7	sloping	60
7	15.0	46.7	sloping	50.0	58.3	sloping	65
8	15.0	43.3	high freq	50.0	63.3	flat	65
9	30.0	63.3	sloping	30.0	55.0	sloping	70
10	45.0	56.7	sloping	45.0	65.0	sloping	75
11	25.0	31.7	flat	55.0	91.7	steep	65
12	35.0	58.3	steep	60.0	68.3	flat	65
13	60.0	68.3	flat	55.0	66.7	flat	75
14	30.0	48.3	high freq	75.0	88.3	flat	70
15	55.0	76.7	sloping	55.0	60.0	flat	65

from 15 dB HL to 92 dB HL. 12 hearing losses were only sensorineural, three had an additional conductive component. The subjects were paid for their participation.

Sentence test procedure

Speech intelligibility measurements were carried out using the HörTech Oldenburg Measurement Applications (OMA), version 1.2. As speech material, the Oldenburg Sentence Test in noise (Wagener et al., 1999a,b,c) was used. Except for the convolution with binaural room impulse responses, the signals complied with the commercial version. A test list of 20 sentences was selected randomly from 45 such lists to obtain each

observed SRT value. Each sentence consisted of five words with the syntactic structure *name verb numeral adjective object*. The subjects' task was to repeat every word they recognized after each sentence as closely as possible. The subjects responses were analyzed using word scoring. An instructor marked the correctly repeated words on a touch screen display connected to a computer, which adaptively adjusted the speech level after each sentence to measure the SRT level of 50 % intelligibility. The step size of each level change depended on the number of correctly repeated words of the previous sentence and on a "convergence factor" that decreased exponentially after each reversal of presentation level. The intelligibility function was represented by the logistic function, which was fitted to the data using a maximum-likelihood method. The whole procedure has been published by Brand and Kollmeier (2002a, A1 procedure). At least two sentence lists with 20 sentences each were presented to the subjects prior to each measurement session for training purposes.

The noise used in the speech tests was generated by randomly superimposing the speech material of the Oldenburg Sentence Test. Therefore, the long-term spectrum of this noise is similar to the mean long-term spectrum of the speech material. The noise was presented simultaneously with the sentences. It started 500 ms before and stopped 500 ms after each sentence. The noise level was kept fixed at 65 dB SPL (for the normal-hearing subjects). The noise levels for the hearing-impaired subjects were adjusted to their individual most comfortable level. They are listed in Table 2.1. All measurements were performed in random order. The measurements with the hearing-impaired listeners were performed in the laboratory of Jürgen Kießling at the University of Gießen, Germany.

TABLE 2.2. *Azimuth angles used for the presentation of noise signal. Negative values: left side, positive values: right side, from the subject's viewpoint*

Location	Angles							
Anechoic room & office room	-140°	-100°	-45°	0°	45°	80°	125°	180°
Empty cafeteria	-135°	-90°	-45°	0°	45°	90°	135°	180°

Acoustical conditions and calibration

Speech and noise signals were presented via headphones (Sennheiser HDA200) using HRTFs (head related transfer functions) in order to simulate different spatial conditions. The speech signals were always presented from the front (0°). The noise signals were presented from the directions shown in Table 2.2. The terminology used here is S_0N_x for a situation where the speech signal was presented from front (0°) and the noise signal from an azimuth angle of x degrees. For example S_0N_{-45} is: speech from front (0°), noise from 45° to the left.

The speech and noise signals had been filtered with a set of HRTFs to reproduce both direction and room acoustics. Three different acoustical environments were used in the measurements: an anechoic room, an office room (reverberation time 0.6 s) and an empty cafeteria (reverberation time 1.3 s).

The headphones were free–field equalized according to international standard (ISO/ DIS 389-8), using a FIR filter with 801 coefficients. The measurement setup was calibrated to dB SPL using a Brüel & Kjær (B&K) 4153 artificial ear, a B&K 4134 ½" microphone, a B&K 2669 preamplifier, and a B&K 2610 measuring amplifier.

The anechoic HRTFs were taken from a publicly available database (Algazi et al., 2001) and had been recorded with a KEMAR manikin. The office room and cafeteria HRTFs were own recordings with a B&K manikin using maximum length sequences.

The sequences were played back by Tannoy System 800a loudspeakers and recorded with a B&K 4128C manikin and a B&K 2669 preamplifier. HRTF calculations were done using MATLAB® on a standard PC equipped with an RME ADI-8 PRO analog/digital converter.

In the office room, the loudspeakers were placed in a circle with a radius of 1.45 m around the head center of the manikin which was seated in the middle of the room. The centers of the concentric loudspeaker diaphragms were adjusted to a height of 1.20 m, the height of a sitting, medium-height person's ears. In the cafeteria, a single loudspeaker was placed at different locations around the manikin seated in front of a table. A large window front, tilted from floor to ceiling, was situated at about 3 m distance from the manikin's head, making this situation rather asymmetric.

2.3. Results and Discussion

2.3.1. Normal-hearing subjects

"Anechoic room" condition

Figure 2.2, left panel, shows predicted SRTs (open circles and crosses) and observed SRT data (filled circles) from eight normal hearing subjects (means and interindividual standard deviations) in anechoic conditions. The observed SRT for 0° noise azimuth (-8.0 dB) differed slightly from the reference value for monaural speech and noise presentation (-7.1 dB, Wagener et al., 1999c). The SRT was approximately 1 dB lower than for noise from the front, if the noise was presented from 180° (from behind), but the difference was not significant. Lateral noise azimuths led to substantially lower SRTs. Maximum release from masking (difference to reference situation S_0N_0) was reached at a noise azimuth of -100° and could be as large as 12 dB.

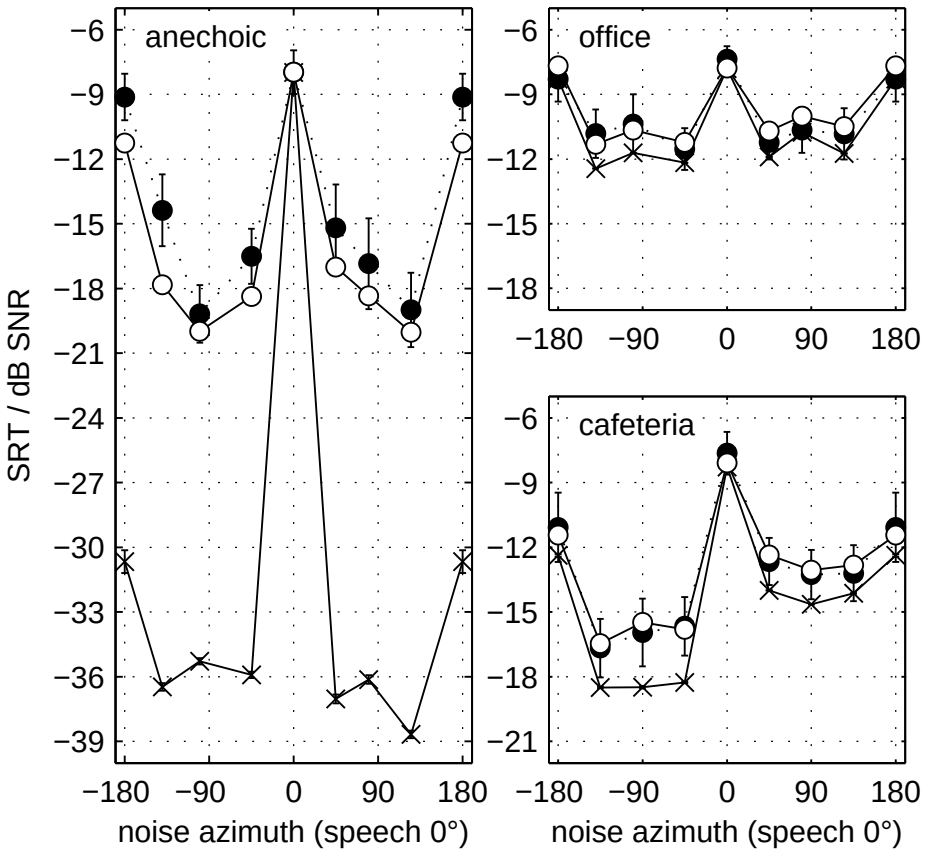


FIG. 2.2. *SRTs for the Oldenburg sentence test with noise from different directions and speech from front (0°) in three room acoustic conditions. Data from eight normal hearing subjects. Filled circles: measurement data, mean and interindividual standard deviation. Open circles: prediction with internal processing errors. Crosses: prediction without internal processing errors. The SRTs for 180° have been copied to -180° in the figure in order to point out the graph's symmetry. Left panel: anechoic room, upper right panel: office room, lower right panel: cafeteria.*

The predicted SRT *including* internal processing errors (open circles) are lower than the observed values for all noise azimuths except 0°, which was the reference value for the adjustment of the SII-to-intelligibility mapping function. The prediction error (i.e. the absolute difference between predicted SRT and the corresponding observed SRT) has a mean of 1.9 dB for the individual data and 1.6 dB if both predictions and

observed data are averaged across subjects. Although there are differences (≤ 20 dB) between the normal-hearing subjects in the individual audiograms (which have been taken into account by the model), these are not reflected in the predictions.

The model predictions *without* internal processing errors σ_ϵ and σ_δ (see Eq. (2.1)) of the EC model (crosses) resulted in SRTs that were much too low.

"Office room" conditions

Figure 2.2, upper right panel, shows predicted SRTs (open circles and crosses) and observed SRT data (filled circles) from eight normal hearing subjects in office conditions. The observed SRTs for noise from front (0°) as well as from behind (180°) did not significantly differ from the corresponding values in anechoic conditions (Fig. 2.2, left panel), but the release from masking in this situation was reduced to about 3 dB for all other noise azimuths (lateral angles).

The difference between model predictions with (open circles) and without internal processing errors (crosses) decreased compared to anechoic conditions to about 1 dB and less. In this room condition the prediction errors have a mean of 0.9 dB (individual data) and 0.5 dB (data averaged across subjects).

"Cafeteria" conditions

Figure 2.2, lower right panel, shows the predicted (open circles) and observed SRTs (filled circles) in reverberant empty cafeteria conditions. The difference of the observed SRT data compared to the office room and anechoic conditions at 0° noise azimuth was not significant. But there was a clear difference between this room and the others at 180° noise azimuth. The graph also shows a remarkable asymmetry between negative (left) and positive (right from the subject's viewpoint) azimuths. The release from masking at negative azimuths reached about 9 dB, but for positive azimuths the

maximal release from masking was only 6 dB. The SRTs for the left side even fall into the range of the corresponding values for anechoic conditions. This asymmetry is probably caused by the asymmetric listening situation with the window front on the left side and the open cafeteria on the other side and will be discussed later.

Like in the office conditions, the difference between model predictions without internal processing errors (crosses) and predictions with internal processing errors (open circles) is much smaller for the cafeteria conditions than for anechoic conditions. The mean prediction error in the cafeteria is 1.1 dB (individual data) and 0.3 dB (data averaged across subjects).

Statistical Analysis

An analysis of variance (ANOVA) of the observed data for the normal-hearing subjects showed a significant effect (at the 1 % level) of both parameters (noise azimuth, room condition) and for interactions of noise azimuth and room condition. In the predicted data for normal-hearing subjects, significant effects (at the 1 % level) were found for noise azimuth, room condition and their interaction.

2.3.2. Hearing-impaired subjects

In Fig. 2.3, three examples of individual predictions for hearing-impaired subjects are shown. All examples show a difference between observed (filled circles) and predicted (open circles) SRTs. Possible reasons for this difference will be discussed later. Subjects 7 and 4 have asymmetric hearing losses, with the better ear on the left side for subject 7 and on the right side for subject 4. The influence of these asymmetries can be seen, for instance in the anechoic condition. It leads to a substantial binaural benefit, if the noise source is close to the worse ear, because then the external SNR is larger at the

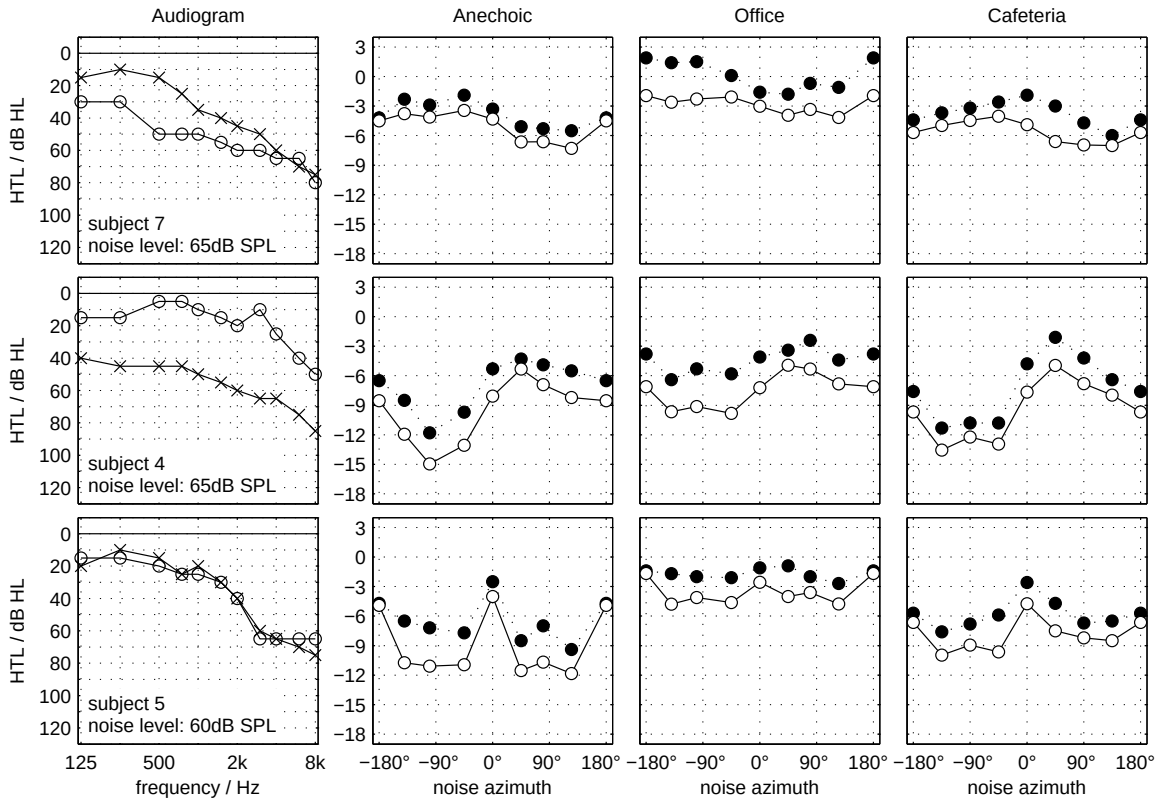


FIG. 2.3. *Three examples of individual predictions of hearing-impaired subject data. Each row contains the results of one subject. The leftmost column shows the individual hearing loss of three listeners and the reference noise level used (crosses: left ear, circles: right ear). The other columns show individual observed SRTs (filled circles) and model predictions (open circles) for each of the three rooms (indicated by the titles). The speech signal was always at 0°. The SRTs for 180° have been copied to -180° in the figure in order to point out the graph's symmetry.*

better ear due to the head shadow. Therefore, subject 7 shows a large binaural benefit for noise at the right side and subject 4 for noise at the left side, which can be predicted very well by the model. Due to the large difference of hearing loss between the left and right ear of subject 4, the external SNR at the right, better ear determines most of the speech intelligibility. This is a simple task for the model, which only had to choose the ear with the better internal SNR (in each frequency band), which occurs at the

right ear in most situations. The predictions for the symmetric hearing loss of subject 5 overestimate the binaural benefit in anechoic conditions. In the office situation, the binaural benefit is very small. For subject 7, the binaural benefit can even be negative at negative azimuths in anechoic and office conditions, which is also found qualitatively in the model predictions, although the prediction error is quite large for some angles. A stronger binaural effect than in the office condition could be found in the cafeteria condition, which is consistent with the results of the normal-hearing listeners.

Figure 2.4 shows predicted and observed SRTs for all hearing-impaired subjects plotted against each other, with each condition on a separate panel. There are three blocks of panels, each for one of the room acoustic conditions. In each panel, the observed SRTs of all subjects for one of the noise azimuths (indicated in the lower right corner) are plotted against the respective predicted SRTs. The dotted line in each panel represents identity.

The individual observed SRTs in each panel vary due to the different hearing losses and extend from values close to the ones measured in normal-hearing subjects in the corresponding situation to thresholds of almost +6 dB SNR, even in situations where a binaural release of masking should be expected. The maximal increase of SRT due to hearing loss (related to the corresponding mean SRT of all normal-hearing subjects) was 22 dB.

Clear correlations (coefficients greater than 0.9 except for Office/S_0N_{180} and Cafeteria/S_0N_0, > 0.8) between predicted and observed SRTs were found. The lower correlations are mainly due to the small variance of observed and predicted data. In anechoic conditions and situations with noise from lateral positions, the binaural benefit was often overestimated by the model, indicated by the wider spread of dots towards lower predicted SRTs at low observed SRTs in the two leftmost columns of Fig. 2.4. This

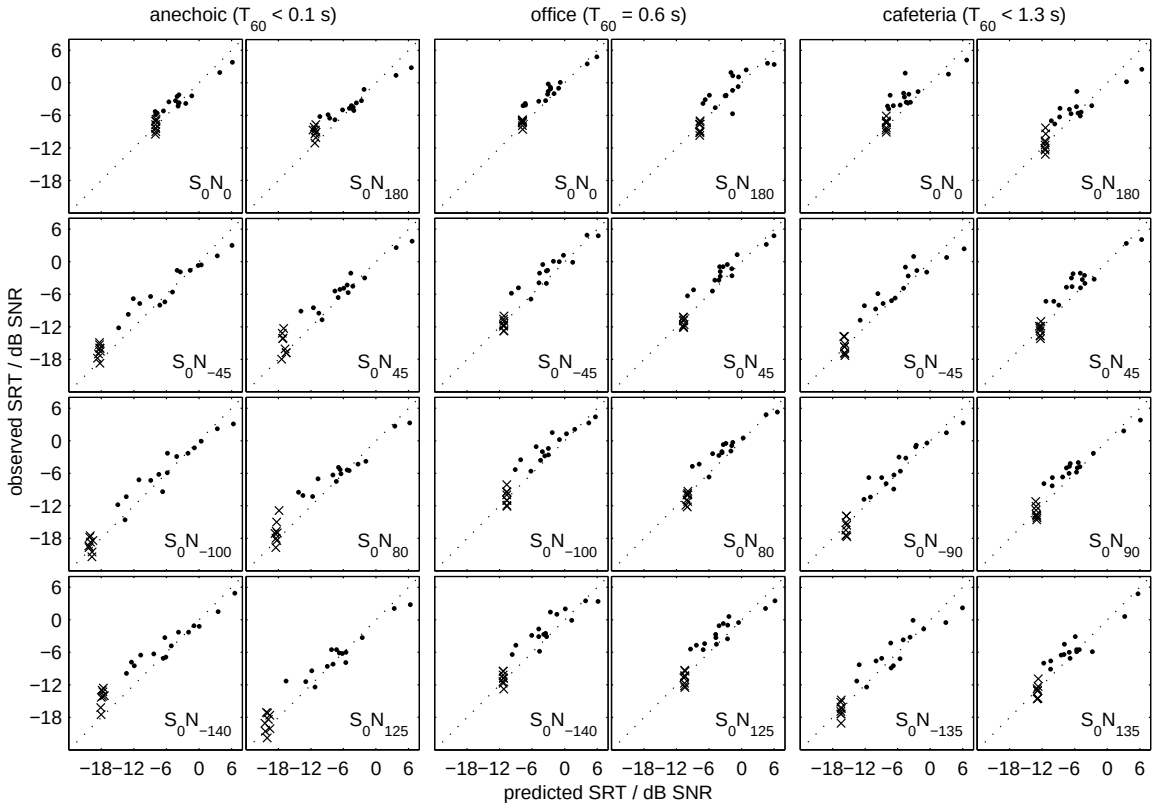


FIG. 2.4. *Predicted and observed SRTs for all hearing-impaired subjects (dots) in this study. The observed SRTs are plotted against the predicted SRT values. Each panel contains the SRTs of 15 hearing-impaired subjects measured at one of the noise source azimuths which are indicated in the lower right corner. There are two columns of panels for each room condition, marked by the respective room names. The SRTs of the normal-hearing subjects (crosses) have been added for comparison*

could not be related to hearing loss and/or noise level. The mean prediction errors for the rooms are 1.7 dB, 1.9 dB, and 1.9 dB (individual data, anechoic, office and cafeteria, respectively).

An ANOVA for the hearing-impaired subjects showed significant main effects (at the 1 % level) for all parameters (noise azimuth, room condition, subject) as well as for all interactions of two parameters in both observed and predicted data.

2.3.3. Correlations

The overall correlation coefficient between all predicted and observed data shown in this study is 0.95. Regarding individual subjects, the correlation coefficients range from 0.69 to 0.99 with a median of 0.91. There is one subject with non-significant correlation (at the 5 % level). This is due to the negligible release from masking (≤ 2 dB) caused by the subject's large hearing losses at both ears (subject 15 in Table 2.1) in combination with a noise level close to the subject's threshold rather than to an insufficient prediction.

The correlation coefficients for the data pooled across room conditions are 0.97, 0.94, and 0.94 (anechoic, office, cafeteria). If the average individual prediction error is subtracted from the prediced SRTs, all correlations increase to 0.98.

Pooled across noise azimuth, the correlation coefficients range from 0.90 to 0.97 with a median of 0.95. With the average individual prediction error subtracted, the median increases to 0.98 (0.94–0.99).

2.4. General Discussion

Although the correlations between model predictions and observed data are high, there are discrepancies between predicted and observed SRTs. A number of reasons for these discrepancies have to be considered and lead to several possibilities to improve the model predictions. Because the goal was to base the whole model on literature data, namely BMLD data of sinusoidals in noise and the standardized SII (ANSI, 1997), only the SII-to-intelligibility mapping function has been adjusted and all other discrepancies have not yet been corrected for in this study. Further work on the model has to include

adjustment of internal parameters and possibly the use of further individual external parameters.

The predictions of data in the present study showed an individual average prediction error of -4.1 dB to +2.5 dB. Although the difference between the mean prediction errors of normal-hearing and hearing-impaired subjects is small (0.5 dB), it is significant (at the 1 % level) and the predicted SRTs for hearing-impaired subjects are too low in most cases. It is known from literature (Pavlovic, 1984; Plomp, 1978), that not all of the decrease of monaural speech intelligibility due to hearing loss can be explained only by the individual hearing threshold. The question is, whether the binaural part of the model needs to be fed with additional individual data or only the monaural back-end. The latter would mean, that binaural processing itself is not affected by the hearing loss, but simply has to deal with the incomplete information coming from the impaired ear. It is still surprising, how much of the binaural speech intelligibility measured in this study seems to be determined by audibility. This may be due to the fact, that the noise level was adjusted to the individual most comfortable level and was clearly audible, but often close to the hearing threshold, which emphasizes the influence of the threshold.

The predictions for all S_0N_0 situations with and without processing errors are almost equal, which means that an adjustment of the processing error parameters would not change the prediction at S_0N_0. In anechoic conditions, the prediction error for S_0N_0 is smaller than at other noise azimuths, above all S_0N_{180}. 180° and 0° azimuth both result in ITDs and ILDs around zero, and the differences between the HRTFs at 0° and 180° may have been small, but still of use for the binaural model. Since the HTRFs used for speech and noise in the S_0N_0 situation were exactly the same, not much of an effect of binaural processing could be expected.

The artificial processing errors assumed by the model turn out to be crucial for correct predictions. In reverberant situations there is only a small difference between predictions with and without processing errors. In the anechoic situation, however, the processing errors have a large influence. The differences between the mean prediction errors of the different room conditions (anechoic: 2 dB office/cafeteria: about 1 dB) for normal hearing subjects appear to be related to the different influence of the processing errors. Moreover, the predictions overestimate the binaural benefit for all subject groups particularly in situations with a strong effect of binaural processing, i.e. when large binaural benefit occurs and for hearing-impaired subjects with symmetric hearing loss, where the better SNR is not necessarily determined by the better ear. Changing the processing error parameters should change the prediction error mainly in the above mentioned situations where the prediction error is large and thus may improve predictions of absolute SRTs as well as equalize the difference between room conditions. A preliminary study has shown that variation of $\sigma_{\varepsilon 0}$ and $\sigma_{\delta 0}$ by a common factor between 0.5 and 2 leads to continuous changes in the predictions of situations with a large influence of the processing error. Nevertheless, there is no quick solution, all error parameters have to be considered.

For normal hearing subjects, no strong dependence of the SRTs on the hearing threshold in both prediction and measurement data would be expected. Although there is only a small difference between individual predicted SRTs, the observed SRTs vary across subjects. The typical standard deviation of the Oldenburg sentence test of about 1 dB (Wagener et al., 1999a,b,c) cannot explain all of this variance. Other factors which cannot be modelled and which are difficult to control experimentally, such as individual attention and motivation, are probably responsible. In this light it is remarkable that the prediction error standard deviations in the different rooms are

almost the same for normal-hearing and hearing-impaired subjects.

It is surprising that in the room with the largest reverberation time (cafeteria hall, $T_{60} = 1.3\,\mathrm{s}$) the release from masking is larger than in the office room, which has only half the reverberation time (0.6 s). Using another room acoustical measure related to the energy in the early parts of the room impulse response, definition or D_{50}, gives a hint why the SRTs are generally lower in the cafeteria than in the office room. The D_{50} is calculated in octave bands and is the ratio between the energy arriving in the first 50 ms and the energy of the whole impulse response. The D_{50} is a common measure used for characterizing rooms in terms of speech perception (ISO 3382; CEN, 2000). Bradley and Bistafa (2002) have shown, that early/late ratios can be a quite good predictor of speech intelligibility in rooms. The D_{50} values averaged over all eight azimuths do not differ significantly between office room and cafeteria at 1-8 kHz (all > 0.9), but they are generally higher for the cafeteria in the low frequency bands (office/cafeteria 125 Hz: 0.70/0.76, 250 Hz: 0.75/0.89, 500 Hz: 0.84/0.88), which would correctly predict better intelligibility in the cafeteria. The reduced release from masking at positive noise azimuths (to the right of the listener) in relation to negative noise azimuths can be attributed to the reflection of a large window front to the left of the listener. It creates a second, virtual noise source, if the actual noise source is located on the opposite side, which hampers the binaural processing. As it can be seen from the predictions, the model is capable of taking these effects into account.

2.4.1. Comparison with literature data

In Fig. 2.5, the observed SRT difference compared to the S_0N_0 situation for various noise azimuths and normal-hearing subjects that were obtained in this study are compared to data from a number of similar experiments in literature (Platte and vom

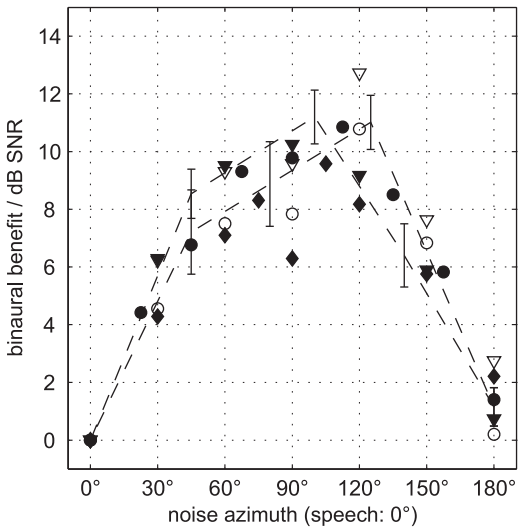


FIG. 2.5 *Release from masking for various noise azimuths with a single noise source and speech presented from the front (0°) relative to the SRT in the S_0N_0 situation. Observed release from masking for eight normal-hearing listeners measured in this study shown with dashed lines (left and right side of the listener) and interindividual standard deviation. The other data points are taken from Platte and vom Hövel (1980, open circles and triangles), Plomp and Mimpen (1981, filled circles), Bronkhorst and Plomp (1988, filled triangles), Peissig and Kollmeier (1997, diamonds) according to Bronkhorst (2000).*

Hövel, 1980; Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1988; Peissig and Kollmeier, 1997; Bronkhorst, 2000). All studies used a single, speech-shaped noise source as an interferer. Regardless of the differences in measurement procedures (speech material, noise level, realization of the binaural configuration), the data from literature show a clear trend of release from masking being dependent on noise azimuth. The maximum benefit is found at azimuths of about 105°–120° rather than at 90°where it might be expected. The data from Peissig and Kollmeier (1997) even shows a dip at 90°, due to interference effects. The data from this study fits very well into the range of values found in the literature.

2.4.2. Comparison to other models

The model presented here extends the model proposed by vom Hövel (1984). The basic principle, multi-frequency band equalization and cancellation, followed by a monaural speech intelligibility model, is the same. Extending the model in order to predict

data of hearing-impaired subjects was possible by adding a masking noise. It yielded encouraging results without changes in the basic principle, but still needs improvement. The handling of early reflections was left to the EC process instead of explicit division of the room impulse response into useful and detrimental sections like in the model by vom Hövel (1984). Although the effect of room acoustics on the noise signal seems to dominate the binaural perception in the approach of the current study with a rather close speech source and a limited amount of reverberation, care must be taken, if the disturbance of the speech signal itself due to reverberation becomes as strong as the effect of the external noise. Solutions to this shortcoming are discussed below. The present model's advantage over models like the ones according to vom Hövel (1984) or Zurek (1990) is that it is, in principle, not limited to known HRTFs or spatial configurations, but is still relatively simple.

In the binaural part, the present model is very similar to psychoacoustic models like the ones from Zerbs (2000) or Breebaart et al. (2001a), because they are all based on the EC principle. This similarity, and the independence of front-end (EC process) and back-end (SII) in the current model, facilitates the transfer of developments and knowledge between psychoacoustical models and the speech model presented in this study. For example, the present model does not incorporate any peripheral preprocessing like a hair cell model or compression. These could replace the somewhat arbitrary binaural processing errors, because half-wave rectification and low-pass filtering smear the high frequency signal components in a manner similar to the delay processing errors have in high-frequency bands. Compression also introduces decorrelation between the ears especially if large ILDs are involved (Breebaart et al., 2001b) and thus acts in a similar manner to the amplitude processing errors.

The present model goes beyond the model by Culling and Summerfield (1995) by

actually using the output from the binaural processing to predict speech intelligibility quantitatively. Culling and Summerfield (1995) were able to decide from their recovered spectra (activity in each frequency band after applying the best delay for each band independently), if certain vowel features were present or not. These recovered spectra were an expression of the effects of binaural hearing, but to predict actual speech intelligibility, the frequency dependent weighting of the SII (or similar models) is necessary. For the predictions in the present study, other parameters of binaural coincidence detectors like the shape of temporal integration windows, as Culling and Colburn (2000) mentioned, were obviously not crucial or implicitly included in the internal noise parameters by vom Hövel (1984).

In the same way as Culling and Summerfield (1995), the EC processing in the present model implies little or no interaction between the frequency bands. This is in accordance with the findings by Akeroyd (2004), who has found that binaural detection experiments with complex tones in noise in different binaural configurations yield thresholds, which are more consistent with free ITD equalization across different frequency bands than with ITD equalization using the same delay for all frequency bands. Edmonds and Culling (2005) also found that speech intelligibility measurements with opposed ITD of speech and noise ($\pm 500\,\mu$s) yield the same thresholds when the ITDs are fixed over the whole frequency range and when the ITDs of speech and noise are swapped at frequencies exceeding a certain splitting frequency between 750 and 3000 Hz. While this study focused on the binaural processing of different simultaneous spatial cues, another matter is the time needed to switch between binaural processing strategies or to select one of several possibilities (cf. Kohlrausch, 1990). However, it should still be investigated whether the EC parameters are completely independent across frequency bands or if there is a remaining interaction, even when it is weak.

Measurements with artificial interfering noise that require gain and/or delay parameters in the EC processing which differ widely between neighboring frequency bands would help in determining the importance of band interaction at the EC stage of the model.

2.4.3. Possible extensions

Overall, the results show that the model is capable of predicting the influence of room acoustics on speech intelligibility. Strictly speaking, this only holds for the influence of room acoustics on the noise (for instance, the emergence of additional "mirror" noise sources caused by early reflections). Since the model assumes the whole speech energy as being useful, it only holds for near field speech, because the disturbance of the speech itself caused by reverberation is not taken into account. It might be possible to solve this shortcoming using the speech transmission index (STI, IEC, 1998), which could be used either instead of the SII or as a kind of correction factor. Since the STI considers the modulation transfer function, it is very successful in predicting the influence of room acoustics on speech intelligibility.

In the light of a possible application of the model as a signal processing device, it would be desirable to remove the constraint of separated speech and noise signals. The need for separate speech and noise signals originates only from the way the SNR is calculated in the EC step. Any other way of calculating a sufficiently accurate SNR from the combined speech and noise signals can be principally incorporated into the model and would remove the constraint.

A further step towards a more comprehensive model that takes attention mediated processes into account is probably much more difficult. The fact, that the model needs speech and noise in separate recordings, implies that the listener is able to distinguish perfectly between speech and noise. Therefore, the experimental setup

of this study, using non-modulated speech-shaped noise, certainly supported the accordance between predictions and observations. Maskers that involve informational masking, like competing voices, are clearly much more challenging for models of speech intelligibility.

Nevertheless, even in its present form, the model shows a strong relationship between tone audiogram and binaural speech intelligibility, which might help audiologists to classify clinical results. A recent study (Brand and Beutelmann, 2005) applied the model to a clinical database of 238 hearing-impaired subjects. This large number of different hearing impairments will certainly help in the further development of the model.

2.5. Conclusions

1. A relatively straightforward functional model of binaural speech intelligibility consisting of a gammatone filter bank (Hohmann, 2002), an independent equalization-cancellation process (Durlach, 1963) in each frequency band, a gammatone resynthesis and the speech intelligibility index (SII, ANSI S3.5-1997) yielded high correlations between predictions and measurements of binaural SRT data for spatial arrangement of noise and speech sources (within the horizontal plane) in anechoic as well as reverberant room environments. In order to simulate the limited human accuracy, pure tone in noise BMLD data has been used to determine the maximum precision of the EC-process (vom Hövel, 1984). Only the SII-to-intelligibility mapping function has been adjusted and no other parameters have been fitted to speech intelligibility data, but because it was not possible to predict all absolute SRTs accurately, an adjustment of model parameters to match predictions and measurement data should be considered.

2. Without changes, the model yields similar correlations between predicted and observed SRTs for both normal-hearing and hearing-impaired subjects and the same order of magnitude in prediction accuracy of relative binaural effects. Regarding absolute SRTs, there is a difference between normal-hearing and hearing-impaired subjects, which probably originates from suprathreshold effects of the hearing impairment, which are not treated by the model.

3. Early reflections that lead to "mirror" noise sources disrupt binaural unmasking more strongly than long reverberation tails of the room impulse response. This was consistent with the model predictions.

4. The human processing errors assumed in the EC stage were highly relevant in the anechoic condition. In the conditions with reverberation the predictions were hardly influenced by the processing errors.

Acknowledgements

We are very grateful to Birger Kollmeier for his substantial support and contribution to this work. We would like to thank Birgitta Gabriel, Daniel Berg, Jürgen Kießling, and Matthias Latzel for organizing and performing the measurements. This work was motivated by helpful discussions with many colleagues, including Kirsten Wagener, Volker Hohmann, and Jesko Verhey. We would also like to thank the editor, Armin Kohlrausch, and two anonymous reviewers for their thorough and helpful reviews. This work was supported by BMBF (Kompetenzzentrum HörTech) and the European 6th Framework Programme "HEARCOM".

3. Revision, extension, and evaluation of a binaural speech intelligibility model (BSIM)[4]

Abstract

This study presents revision, extension, and evaluation of a binaural speech intelligibility model (Beutelmann and Brand, 2006, J. Acoust. Soc Am. 120(1), 331–342) that yields accurate predictions of speech reception thresholds (SRTs) in presence of a stationary noise source at arbitrary azimuths and in different rooms. The modified model is based on an analytical expression of binaural unmasking for arbitrary input signals and is computationally more efficient, while maintaining the prediction quality of the original model. An extension for non-stationary interferers was realized by applying the model to short time frames of the input signals and averaging over the predicted SRT results. The extended model predictions were compared to binaural speech intelligibility data from eight normal-hearing and twelve hearing-impaired listeners, incorporating all combinations of four rooms, three source setups and three noise types. Depending on the noise type, the correlation coefficients between observed and predicted SRTs were 0.80-0.93 for normal-hearing subjects and 0.59-0.80 for hearing-impaired subjects. The mean absolute prediction error was 3 dB for the mean normal-hearing data, and

[4]This chapter has been submitted in the present form for publication to the Journal of the Acoustical Society of America (Beutelmann et al., 2008a).

4 dB for the individual hearing-impaired data. 70% of the variance of the SRTs of hearing-impaired listeners could be explained by the model, which is based only on the audiogram.

3.1. Introduction

The task of understanding speech in complex environments, which has been termed "cocktail party problem" by Cherry (1953), is affected by many factors. These factors include, among others, the location of the speech and interferer sources, room acoustics, the type of interferer, and a potential hearing loss of the listener. It has been shown that the ability to use binaural information in order to segregate interferer and target signal is very important for solving the "cocktail party problem" (Bronkhorst, 2000). A comprehensive model of speech intelligibility in complex situations, which might give more insight into the underlying mechanisms (and which may be used for example in audiology or room acoustics) should incorporate as many of the involved factors as possible, especially binaural hearing.

This study extends the binaural speech intelligibility model presented by Beutelmann and Brand (2006), which combined the equalization-cancellation (EC) model by Durlach (1963) with the standard speech intelligibility index (SII, ANSI, 1997), based on the work by vom Hövel (1984). The original model was able to predict speech reception thresholds (SRTs) of sentences in steady state noise for different noise source locations, different room acoustics, and different degrees of hearing loss. The extension of the model in this study is a first appraoch at predicting binaural SRTs also for modulated interferers. Furthermore, the model was mathematically reformulated, in order to make it simpler and more efficient.

The original model implementation by Beutelmann and Brand (2006) was very straightforward and employed the EC model as a signal-processing front-end in order to process binaural input signals. Speech and noise input signals were split into 30 frequency bands using a gammatone filter band (Hohmann, 2002). In each frequency band, the EC process was performed with independent values for gain and delay. A Monte-Carlo simulation was used in order to calculate the effect of the binaural processing errors (Durlach, 1963; vom Hövel, 1984). The processing errors controlled the maximal performance of the model in situations in which the model would otherwise be able to eliminate the noise perfectly. The output of the binaural front-end was a monaural signal with improved signal-to-noise ratio (SNR), from which an SRT was calculated using the Speech Intelligibility Index (SII, ANSI, 1997) as a monaural speech intelligibility prediction back-end. The two model stages operate independently. Therefore, the EC front end might in theory be replaced by other binaural models (e.g. Breebaart et al., 2001a; Osman, 1971; Zerbs, 2000; Nitschmann and Verhey, 2007) and the speech intelligibility prediction back-end might be replaced by another speech intelligibility predictor, for example the Speech Transmission Index (STI, IEC, 1998), the speech recognition sensitivity (SRS, Müsch and Buus, 2001a), or speech intelligibility prediction based on automatic speech recognition (Holube and Kollmeier, 1996). The original model's components, which were well established in literature, made it easy to implement and to experiment, but the whole model was very slow and difficult to interpret in terms of psychoacoustics. In the terminology of Colburn (1996), the model would be called a "black box" model, meaning without explicit relation to physiology. A combination of the binaural model by Zerbs (2000) and the (monaural) speech intelligibility model by Holube and Kollmeier (1996) could thus be a future step towards a more physiologically oriented model. Both are based on the

same auditory preprocessing model, and the binaural part of the model by Zerbs (2000) is based on EC theory, but it would require some fundamental modifications of the speech intelligibility prediction part, if open-set sentence intelligibility test results, as measured in this study, need to be predicted.

One striking difference between a binaural speech intelligibility model and binaural psychoacoustical models is, that the former requires parallel processing in multiple frequency bands, because both target signal (speech) *and* interferer are broad-band, whereas the latter typically use signals (at least for the target) that are constrained to a single critical band (e.g., Durlach, 1963; Zerbs, 2000; Breebaart et al., 2001a). In addition, for speech intelligibility prediction a different back-end than for detection or discrimination tasks has to be used. The EC stage of the binaural speech intelligibility model presented by Beutelmann and Brand (2006) can - in principle - deal with arbitrary signals, including non-speech, although the validity of the predictions has so far only been tested for speech in a restricted set of conditions. The binaural configuration (i.e., directions or interaural relations of target and interferer, as well as room acoustics) needs not to be known explicitly, because the optimal equalization parameters are estimated by the model by optimizing the signal-to-noise ratio.

A number of studies have investigated different combinations of aspects of the "cocktail party problem" with a special focus on modulated or speech-like interferers. An early study by Miller and Licklider (1950) investigated the masking effect of interrupted broadband noise and noise bursts on speech reception compared to stationary noise. They found an increase of intelligibility for interrupted noise compared to stationary noise, which was dependent on the frequency of interruption and the signal-to-noise ratio (SNR) during the noise bursts. The largest increase was found for interruption frequencies between 4-10 Hz. Other studies have also shown that there is a decrease

in speech reception threshold (SRT) for modulated noises or for speech maskers compared to stationary noise (Dubno et al., 2002; Gustafsson and Arlinger, 1994; Festen and Plomp, 1990; Wagener, 2003). In theses studies, the SRT decrease was up to 10 dB, depending on the modulation frequency, the modulation depth and the type of modulation (broadband or frequency-dependent, regularly or random). The release of masking due to fluctuations in the masker is significantly lower or absent for hearing-impaired listeners (Festen and Plomp, 1990; Gustafsson and Arlinger, 1994; Peters et al., 1998; Wagener and Brand, 2006) and there is an additional effect of age which is not related to the hearing threshold (Dubno et al., 2002; Peters et al., 1998; Festen and Plomp, 1990). There is also evidence that linear amplification does not restore the release of masking due to fluctuations in the masker (Peters et al., 1998; Gustafsson and Arlinger, 1994). Among possible reasons mentioned by Festen and Plomp (1990) for the detriment of hearing-impaired listeners are reduced temporal resolution and reduced comodulation masking release (Hall et al., 1984), although the amount of comodulation masking release on speech recognition as opposed to speech detection appears to be small (Festen, 1993; Grose and Hall, 1992). For diagnostic purposes, on the other hand, fluctuating maskers can even have an advantage, because hearing-impaired subjects show larger inter-individual differences in speech-modulated noise than in stationary noise (Wagener and Brand, 2006; Versfeld and Dreschler, 2002; Smits and Houtgast, 2007).

While the studies mentioned so far considered only monaural or diotic signals, others have additionally taken binaural aspects into account. Especially the interaction between the binaural release of masking and the beneficial effect of modulated maskers is of interest, both for normal-hearing and hearing-impaired listeners. Generally, it has been found that there is a combined benefit of location and modulation of the masker for

normal hearing subjects, but the single effects do not simply add up. It depends on the spatial distribution and number of interferers as well as their degree of comodulation, if the combined effect is larger or smaller than the sum of the single effects (Hawley et al., 2004; Peissig and Kollmeier, 1997; Duquesnoy, 1983). Hearing-impaired subjects have only little or no benefit from masker fluctuations, even if they can use a binaural advantage (Bronkhorst and Plomp, 1992; Duquesnoy, 1983; Peissig and Kollmeier, 1997).

In some of the above mentioned studies (Festen and Plomp, 1990; Peters et al., 1998; Dubno et al., 2002), the Articulation Index (AI, ANSI, 1969) has been used to assess approximative first order predictions of speech intelligibility results. The focus lay mainly on the influence of audibility for hearing-impaired subjects and less on the effect of modulated or speech-like maskers. Predictions that were especially aimed at the prediction of speech intelligibility in modulated interferers were presented by Wagener (2003) and Rhebergen and Versfeld (2005). The former included the noise level dependence of the SRT (Plomp, 1978) and a context model for phonemes and words, while the latter extended the Speech Intelligibility Index (SII, ANSI, 1997) for modulated noises by frame-wise calculation and subsequent averaging of the results per frame. The frame-wise calculation principle was also used in this study. Culling et al. (2004) measured the amount of binaural unmasking for pure tones in noise in different spatial configurations of target and interferer sources and at different target frequencies. They then used the results to successfully predict the increase of speech intelligibility in speech shaped noise in the same spatial configuration by calculating the expected SNR increase from the binaural masking level differences.

Some factors not mentioned above, as for example “informational” masking, fundamental frequency differences between target and masker speaker (cf. Hawley et al.,

2004) or inter-individual cognitive differences not related to the auditory periphery, are not considered in this study. Although they are definitely important in certain situations, they are still very difficult to model and too complex to be included at the current state of the model presented here.

The purpose of the current study was (1) to analytically simplify the binaural speech intelligibility model presented in Beutelmann and Brand (2006) and (2) a first approach toward the extension of the model in order to predict binaural SRTs not only in stationary noise, but also in modulated interferers. The simplification has the advantage of making the mathematical description of the model more concise, and it points out the role of binaural signal parameters like the interaural level difference and interaural correlation in the calculation of the signal-to-noise ratio after EC processing without detailed assumptions about the input signals. The formulas are closely related to the expressions derived by Durlach (1963) for tone detection in special binaural conditions and by vom Hövel (1984) for a basic binaural speech intelligibility model, but they remain more universally valid. Simplifying the model has also accelerated its practical use: with the help of analytical and numerical optimizations, the computing time of the model can be considerably reduced. In order to verify that the reformulated model provides at least the prediction quality as the original model, the new model was evaluated with the data from Beutelmann and Brand (2006). The prediction quality, in terms of correlation with the observed SRTs and absolute prediction error, remains the same as for the predictions of the model in Beutelmann and Brand (2006). The results of this evaluation are summarized in section 3.3.2.

It is expected, that the effects of non-stationary interferers on the SRT interact with the factors which are already incorporated in the model, namely the effects of spatial separation of target and interferer sources, reverberation and a possible hearing

loss of the listener. Therefore, a set of reference data for the model extension was measured from 8 normal-hearing and 12 hearing-impaired subjects, which incorporates all combinations between the above mentioned parameters. It includes four room types spanning a range of reverberation times between 0 s and 8.8 s, three spatial setups of target and interferer sources and three noise types with different degrees of modulation. The measurement parameters are described in detail in section 3.4.1. The observed data was used to evaluate an extension of the binaural speech intelligibility model, which is described in section 3.2.3. In order to distinguish between the original model, the revised model, and the extension for modulated noises, the abbreviations "EC/SII", "BSIM" (for binaural speech intelligibility model), and "stBSIM" (for short-time BSIM), respectively, are used.

3.2. Model development

3.2.1. Analytic revision

The input signals $x_k(t)$ of the binaural speech intelligibility model (with $k \in \{L, R\}$ representing the left or right ear, respectively), are assumed to be a linear superposition

$$x_k(t) = s_k(t) + n_k(t) \tag{3.1}$$

of the target speech signals $s_k(t)$ and the noise signals $n_k(t)$. This assumption is valid as long as nonlinearities in the transmission paths from the sound sources to the ears can be neglected, which is especially true for natural sound sources in reverberant rooms or their simulation via HRTFs. The noise signals are assumed to be a superposition

$$n_k(t) = \nu_k(t) + \mu_k(t) \tag{3.2}$$

of the external noise signals $\nu_k(t)$, and internal masking noises $\mu_k(t)$. The latter are simulating the hearing threshold for the left and right ear, respectively. The internal masking noises $\mu_k(t)$ are regarded throughout the derivation such that the cross-correlation function is always exactly zero between $\mu_L(t)$ and $\mu_R(t)$, as well as between one of them and each other input signal. This was done in order to ensure that the masking noises cannot be eliminated by the binaural processing.

The basic idea of the EC mechanism is to attenuate the external noise signal, if possible, by destructive interference between the left and right channel. For this purpose, a residual signal

$$x_{EC}(t) = \alpha x_L(t+\tau) - x_R(t), \tag{3.3}$$

is calculated from the input signals by applying an attenuation factor α and a relative time shift τ to one of the signals and subtracting the other signal, thus eliminating signal components with amplitude ratio α and time difference τ.

Eq. (3.3) is symmetric in the sense that $x_L(t)$ and $x_R(t)$ may be swapped, if α is replaced by α^{-1} and τ by $-\tau$, resulting only in a sign change of $x_{EC}(t)$. This can be expressed more clearly by symmetrizing Eq. (3.3), which gives

$$x_{EC}(t) = e^{\gamma/2} x_L(t+\tau/2) - e^{-\gamma/2} x_R(t-\tau/2) \qquad \text{with} \quad \alpha = e^{\gamma} \tag{3.4}$$

The level equalization factor $e^{\gamma/2}$ is restricted to positive values. This represents the assumption that a simple addition of the channels is impossible, an assumption made originally by Durlach (1963) in order to explain the differences in binaural masking level difference (BMLD) between a π-phase-shifted pure tone in diotic noise and a

diotic pure tone in π-phase-shifted noise.

For pure tone signal detection, modeling BMLDs usually only requires to examine a single auditory filter band centered on the target signal - contrary to speech reception, where the bandwidth of the target signal is almost always larger than a single auditory frequency band. It has been shown that the binaural system is able to evaluate frequency-dependent interaural time and level differences (Akeroyd, 2004; Edmonds and Culling, 2005), suggesting independent binaural processing in different frequency bands. Within a single auditory filter, however, it is typically assumed (e.g., Durlach, 1972) that the interaural parameters of a binaural model may be considered to be constant. The conclusion for this model is that the input signals $x_L(t)$ and $x_R(t)$ are filtered into B narrow auditory frequency bands with center frequencies Ω_b, where $b \in [1, B]$. The transfer function magnitudes of the auditory filters are assumed to be negligible beyond a certain bandwidth β_b around Ω_b. In each frequency band, the SNR is maximized using an independent EC process with a separate set of equalization parameters $\alpha_b = e^{\gamma_b}$ and τ_b. The following derivations are performed in the frequency domain and represent the output of *one* of the B auditory filters, without loss of generality. In order to avoid overly complex expressions, the index b was omitted. Upper case letters represent the filtered spectrum of time domain signals with respective lower case letters, for example $X_L(\omega) = H(\omega)\mathcal{F}\{x_L(t)\}$ etc., where $H(\omega)$ is the transfer function of the respective auditory filter, and ω is the angular frequency[5]. The EC process in Eq. (3.4) expressed in the frequency domain is

$$X_{EC}(\omega) = e^{\gamma/2+i\omega\tau/2}X_L(\omega) - e^{-\gamma/2-i\omega\tau/2}X_R(\omega). \tag{3.5}$$

[5]The normalization factors $(2\pi)^{-1/2}$ for the Fourier transform when using ω as the frequency variable are applied to both the transform and the inverse transform.

In EC theory, the signals are assumed to be subject to uncertainties in level and time, expressed by normally distributed processing errors ϵ_k and δ_k. These processing errors have been adapted by vom Hövel (1984) from the concept by Durlach (1963). Every quantity derived from the residual signal

$$X_{EC}(\omega) = e^{\gamma/2+\epsilon_L+i\omega(\tau/2+\delta_L)} X_L(\omega) - e^{-\gamma/2+\epsilon_R-i\omega(\tau/2-\delta_R)} X_R(\omega), \tag{3.6}$$

especially the signal intensity $I(X_{EC})$ (as defined in Eq. (3.8), see below), is assumed to be the expectation value of this quantity with respect to distributions of the processing errors. The distributions of ϵ_k and δ_k have a mean of zero and standard deviations dependent on the actual equalization parameters: $\sigma_\epsilon(\alpha)$ and $\sigma_\delta(\tau)$ [6].

Speech intelligibility prediction using the Speech Intelligibility Index (SII) is based on the band-wise signal-to-noise ratio (SNR)

$$\mathrm{SNR} = \frac{I(S_{EC})}{I(N_{EC})}, \tag{3.7}$$

with the intensity I of a band pass signal with center frequency Ω and bandwidth β defined in the frequency domain as

$$I(X) = \int_{\Omega-\beta/2}^{\Omega+\beta/2} |X(\omega)|^2 \, d\omega. \tag{3.8}$$

[6] The standard deviations of the processing errors are defined as: $\sigma_\epsilon(\alpha) = \sigma_{\epsilon 0}\left[1 + (|\alpha|/\alpha_0)^p\right]$ and $\sigma_\delta(\tau) = \sigma_{\delta 0}\left[1 + |\tau|/\tau_0\right]$ with $\sigma_{\epsilon 0} = 1.5$, $\alpha_0 = 13\,\mathrm{dB}$, $p = 1.6$, $\sigma_{\delta 0} = 65\,\mu\mathrm{s}$, and $\tau_0 = 1.6\,\mathrm{ms}$. These values have been fitted to pure tone BMLD measurement data (Blodgett et al., 1962; Langford and Jeffress, 1964; Egan, 1965; vom Hövel, 1984; Beutelmann and Brand, 2006).

A comprehensive derivation, which is carried out in detail in Appendix A, leads to a closed-form expression for the SNR,

$$\mathrm{SNR} = (M_L M_R)^{1/2} \; \frac{e^{\sigma_\epsilon^2} \cosh(\gamma + \Delta_S) - \lambda(\tau) * \mathrm{Re}(\rho_S(\tau))}{e^{\sigma_\epsilon^2} \cosh(\gamma + \Delta_N) - \lambda(\tau) * \mathrm{Re}(\rho_N(\tau))}, \tag{3.9}$$

where $\mathrm{Re}(\rho)$ denotes the real part of ρ, and $*$ denotes the convolution. All new variables will be defined and explained in the following: The first two factors in Eq. (3.9),

$$M_L = \frac{I(S_L)}{I(N_L)} \quad \text{and} \quad M_R = \frac{I(S_R)}{I(N_R)}, \tag{3.10}$$

represent the monaural SNRs at each ear. The second summands in the argument of the cosh-functions,

$$\Delta_S = \frac{1}{2} \ln\left(\frac{I(S_L)}{I(S_R)}\right) \quad \text{and} \quad \Delta_N = \frac{1}{2} \ln\left(\frac{I(N_L)}{I(N_R)}\right), \tag{3.11}$$

represent the interaural level difference (ILD) of the speech and noise signals, respectively (except for a scaling factor, they are equivalent to the ILD in dB). $\rho_S(\tau)$ is defined as the normalized cross-correlation function between the left and right ear for the speech signal

$$\rho_S(\tau) = \frac{2\pi}{\sqrt{I(S_L) I(S_R)}} \int_{\Omega-\beta/2}^{\Omega+\beta/2} S_L(\omega) S_R^*(\omega) e^{i\omega\tau} d\omega \tag{3.12}$$

and $\rho_N(\tau)$ is defined analogously for the noise[7]. Both are smoothed by convolution with a Gaussian window

$$\lambda(\tau) = \frac{1}{\sigma_\lambda \sqrt{2\pi}} e^{-\frac{1}{2}\tau^2 \sigma_\lambda^{-2}}, \tag{3.13}$$

whose width is defined by the standard deviation of the time processing errors $\sigma_\lambda = \sigma_\delta \sqrt{2}$. Note, that this is equivalent to a low pass filter in the frequency domain (with a likewise Gaussian transfer function).

The aim of the EC process is to maximize the SNR given in Eq. (3.9). It can be easily shown by expanding the cosh-functions, that the SNR converges to the left monaural SNR M_L as γ goes to positive infinity and that the SNR converges to the right monaural SNR M_R as γ goes to negative infinity. This means that the trivial case of "better ear listening", that is using only the signal at the ear with the favorable SNR, is implicitly included in Eq. (3.9). However, depending on the properties of the input signals, the parameters γ and τ can be used to achieve an additional benefit exceeding the "better ear" SNR, that is a true binaural release from masking.

Since the cosh function is symmetric with a minimum value of one at zero in the argument, and because the absolute value of the cross-correlation terms (even after convolution with the normalized smoothing window) is always equal or less than one, the fraction in Eq. (3.9) is always equal to or greater than zero. Equality is only achieved, if σ_ϵ is zero and $\rho_S(\tau)$ is one for a certain value of τ. Otherwise, both enumerator and denominator are always finite, thus only a finite benefit compared to the "better ear" SNR can be achieved. This corresponds to the purpose of the processing errors, that is to restrict the performance of the EC process by preventing

[7] S^* denotes the complex conjugate of S throughout this paper and Re() the real part of the argument.

perfect cancellation of the noise signal. The internal masking noise $\mu_k(t)$ is another reason why the noise signal cannot be perfectly canceled out. Although it is present in the combined noise signal $n_k(t)$, it does not contribute to the correlation between the ears in $\rho_N(\tau)$. Therefore, $\rho_N(\tau)$ can never reach an absolute value of one. Details about the internal noise are specified in the next section and further discussion of the parameters and their meaning can be found in Sec. 3.5.

3.2.2. Implementation

The practical implementation of the new model, which is called "BSIM" (Binaural Speech Intelligibility Model) in the following, involved some aspects which are important to mention, because they concern essential parts of the model or contributed considerably to the acceleration. These modifications of the original "EC/SII" model (Beutelmann and Brand, 2006) include a new matched frequency band scheme for the SII, the way how the internal threshold noise is included, and the search method for the optimal SNR in each band.

The number and bandwidth of the SII calculation bands was adapted to the gammatone filter bank (Hohmann, 2002) which was used to divide the input signals into auditory frequency bands. The basic calculation procedure of the SII was not changed, only the band importance functions had to be adapted to the new center frequencies. Although this implies a deviation from the standard SII, it was considered to be more accurate than using a different filter bank for the binaural part of the model, or interpolating the output SNR of the binaural part to one of the standard frequency schemes. Because the transfer function relating SII to percent intelligibility is dependent on the speech material and type of presentation, the SII corresponding to 50% intelligibility at the SRT needed to be adjusted to a reference condition. The

revised model's modified SII procedure was adjusted to the monaural presentation of the original Oldenburg Sentence Test in noise (cf. Sec. 3.4.1) at 65 dB SPL, which yields an SII of 0.2 at the reference SRT of -7.1 dB SNR (Wagener et al., 1999c) . This differed from the procedure of Beutelmann and Brand (2006), that was adjusted to a quasi-diotic anechoic condition, in which speech and noise both came from the front.

The hearing threshold was simulated by adding a pair of constant intensity values corresponding to 1 dB above the hearing level to the noise intensities used for the calculation of Δ_N and the normalization of ρ_N in each frequency band. This replaced the actual internal noise signals in the original model of Beutelmann and Brand (2006), which were spectrally shaped in such a way that the noise energy in an auditory filter band was 4 dB above the energy of a pure tone at the band center frequency with the respective hearing level at that frequency (cf. Breebaart et al., 2001a), and added to the external noise signals. The threshold criterion of 1 dB instead of 4 dB was chosen, because it provided a better correlation between the predicted and the observed SRTs in the reference data (cf. Sec. 3.3.2)[8].

The optimal γ in Eq. (3.9), that is the γ leading to the best SNR for a given τ, can be calculated analytically if the error variances are both set to zero. The optimal τ is searched for each band independently by calculating the interaural cross-correlation functions in Eq. (3.9) with the help of a fast Fourier transform as a first coarse estimate. Inter-sample interpolation was achieved by quadratic approximation at the maximum of Eq. (3.9) with respect to τ, which is possible because the input signals are band-limited and therefore quasi-periodic.

[8]This may partly be due to the fact that the assumption of perfectly uncorrelated internal noise channels, and internal and external signals, respectively, has to be relaxed (cf. Diercks and Jeffress, 1962; Osman, 1971).

3.2.3. Extension for modulated noises

The "EC/SII" model (Beutelmann and Brand, 2006) used long signals (between 1 and 3 s, i.e., about the length of a test sentence, see Sec. 3.4.1) to calculate a single SRT with a single set of EC parameters. This has the advantage, that the result is not dependent on the (residual) signal statistics of the stationary interferer and that the EC parameters can be estimated very reliably, if the binaural parameters of the input signals are constant. For modulated interferers, however, the SNR and hence potentially also the choice of optimal EC parameters varies over time. Thus, the signal level statistics need to be considered explicitly. In a first approach, we therefore calculated the model for short time frames of the input signals and averaged across the frame-wise SRTs in order to obtain the final SRT prediction. Rhebergen and Versfeld (2005) showed that this approach is sufficient for good predictions of monaural SRT data in modulated noise, even with a fixed frame length across all frequency bands. A frame length of 1024 samples at 44100 Hz sampling rate was used with a hann window and a frame shift of half the frame length. Considering that the equivalent rectangular duration of a hann window is only half of its full length, the effective frame length of this model is about 12 ms, which is close to the best fitting frequency-independent frame length found by Rhebergen and Versfeld (2005). The extended model is called "stBSIM" (short-time BSIM) in the following. It is rather a proof of concept than an elaborate model for the combination of binaural speech intelligibility and fluctuating noise and may be refined with knowledge from monaural models (Rhebergen et al., 2006; Plomp, 1978) in future studies.

3.3. Evaluation with reference data

3.3.1. Methods

In order to ensure that the prediction quality of the original "EC/SII" model from Beutelmann and Brand (2006) is maintained, the predicted SRTs from both the revised "BSIM" and the original "EC/SII" model for the measurement data from Beutelmann and Brand (2006) were compared. The SRTs had been measured with the Oldenburg Sentence Test in noise (cf. Sec. 3.4.1), with the speech source always in front of the listener and a single, stationary speech-shaped noise source at one of eight azimuths. The measurements were performed in three different simulated room acoustical conditions. The rooms had reverberation times (T_{60}) of 0 s ("anechoic"), 0.6 s ("office") and 1.3 s ("cafeteria"). The subjects taking part in the original study included eight normal-hearing and 15 hearing-impaired listeners with different degrees and types of hearing loss. (see Beutelmann and Brand, 2006, for further details). The predictions with the BSIM were calculated with a frame length of about 2.9 s to test the consistence with the original model. Additionally, the predictions for the same data were calculated with the "stBSIM" using a frame length of about 12 ms in order to assess if the model for modulated noises yields the same results as for the original model for stationary noise data.

3.3.2. Results

Table 3.1 gives an overview of the correlation coefficients and root mean squared prediction errors of the original "EC/SII" model from Beutelmann and Brand (2006), of the long-frame "BSIM" and the short-time "stBSIM". Despite the already high correlation coefficients and low mean absolute predictions errors in the original EC/SII

TABLE 3.1. *Correlation coefficients R between predicted and observed SRTs and root mean squared prediction errors ε in dB for the EC/SII model from Beutelmann and Brand (2006), for the revised model (BSIM) and the modified, short-time model (stBSIM). The models reference SII was set according to section 3.2.2. The subject group "NH" are individual normal-hearing subject data, "NH mean" averaged normal-hearing subject data and "HI" individual hearing-impaired subject data and their respective predictions.*

subject group	EC/SII		BSIM		stBSIM	
	R	ε / dB	R	ε / dB	R	ε / dB
all	0.95	2.1	0.96	1.7	0.96	1.8
NH	0.91	1.7	0.93	1.3	0.93	1.4
NH mean	0.97	1.2	0.99	0.5	0.99	0.6
HI	0.92	2.3	0.94	1.9	0.93	2.0

model, the predictions of both revised models show higher correlations and lower prediction errors than the original model. The small deviations of the "stBSIM" compared to the long-frame "BSIM" are probably due to a larger variance of level and time parameters across the short-time frames.

3.4. Evaluation with modulated interferer

3.4.1. Methods

Sentence Test Procedure

The speech intelligibility measurements were carried out using the HörTech Oldenburg Measurement Applications (OMA), version 1.2. As speech material, the Oldenburg Sentence Test in noise (Wagener et al., 1999a,b,c) convolved with room impulse responses was used. Except for the convolution with binaural room impulse responses, the signals complied with the commercial version. Each sentence of the Oldenburg Sentence Test consists of five words with the syntactic structure 'name verb numeral

adjective object'. For each part of the sentence, ten alternatives are available, each of which occurs exactly twice in a list of 20 sentences, but in random combination. This results in syntactically correct, but semantically unpredictable sentences. The subjects' task was to repeat each word they recognized after each sentence as closely as possible. The subjects responses were analyzed using word scoring. An instructor marked the correctly repeated words on a touch screen display connected to a computer, which adaptively adjusted the speech level after each sentence to measure the SRT level of 50% intelligibility. The step size of each level change depended on the number of correctly repeated words of the previous sentence and on a "convergence factor" that decreased exponentially after each reversal of presentation level. The intelligibility function was represented by the logistic function, which was fitted to the data using a maximum-likelihood method. The details of this procedure have been published by Brand and Kollmeier (2002a, A1 procedure). A test list of 20 sentences was selected from 45 such lists to obtain each observed SRT value. Two sentence lists with 20 sentences each were presented to the subjects prior to each measurement session for training purposes. At the beginning of the first session of each subject, three training lists were presented. The test lists were balanced across subjects and conditions, and all measurements except for the training lists were performed in random order.

The noise signals used in the speech tests will be described in detail in section 3.4.1. The noise token, with its starting point randomly selected within the whole noise signal, was presented simultaneously with the sentences. It started 500 ms before and stopped 500 ms after each sentence. The noise level was kept fixed at 65 dB SPL for the normal-hearing subjects. For the hearing-impaired subjects, the noise levels were adjusted to their individual hearing loss. The noise level was first set to 55 dB SPL plus half the individual hearing loss averaged across 500 Hz and 4 kHz (in steps of 5

dB). No level was set lower than 65 dB SPL or higher than 85 dB SPL. The subjects were asked whether the level was uncomfortably loud during the first training sentence and the noise level was decreased in steps of 5 dB if necessary.

The headphones (Sennheiser HDA 200) were free-field equalized according to international standard (ISO/DIS 389-8), using an FIR filter with 801 coefficients. The measurement setup was calibrated to dB SPL using a Brüel & Kjær (B&K) 4153 artificial ear, a B&K 4134 1/2" microphone, a B&K 2669 preamplifier, and a B&K 2610 measuring amplifier.

Interferer Noises

Three different noise types were used in the measurements: stationary speech-shaped noise ("stationary"), 20-talker babble noise ("babble"), and a single-talker modulated noise ("single-talker"). As stationary speech-shaped noise, the original noise from the Oldenburg Sentence Test was used. It has been generated by randomly superimposing the speech material of the sentence test. Therefore, the long-term spectrum of this noise is very close to the mean long-term spectrum of the speech material. The multi-talker babble noise was taken from the Auditec CD "CD101RW2" (Auditec, 2006) and is a mixture of 20 speakers simultaneously reading different passages. The single-talker modulated noise is based on the "ICRA5" noise (Dreschler et al., 2001). The "ICRA5" noise has been created to eliminate intelligibility of the speaker as far as possible while preserving the modulation features of a single speaker in multiple frequency bands. The speech pause durations in this noise have been limited to 250 ms (Wagener and Brand, 2006). The long-term spectra of stationary noise and the single-talker noise are similar, but the babble noise was attenuated by about 16 dB at frequencies higher than 5 kHz with a slope of about 5 dB/oct between 500 Hz and 5 kHz. Although this

TABLE 3.2. *Basic room acoustic parameters of the three realistic (non-anechoic) rooms used in the measurements for two distances (3 m and 6 m) between the speech source and the (omnidirectional) receiver at the listener's position. The values given are average values across octave bands from 63 Hz to 8 kHz calculated by the ODEON software. The STI values only include the room acoustics, but not the noise interferers used in this study. For a detailed description see section 3.4.1 (Rooms and Setups)*

Room	distance / m	T_{30} / s	EDT / s	C80 / dB	D50	STI
listening	3	0.40	0.35	13.2	0.88	0.81
room	6	0.40	0.41	11.4	0.82	0.77
classroom	3	0.94	0.48	10.1	0.83	0.77
	6	0.92	0.62	8.1	0.77	0.72
church	3	8.78	7.38	2.8	0.57	0.60
	6	8.69	7.91	0.9	0.48	0.52

was originally due to a missing headphone equalization, it was kept, because this was a way to test the model with substantially differing speech and noise spectra.

Rooms and Setups

Room acoustics and sound source locations were realized by using virtual acoustics over headphones. The stimuli were prepared by convolving the original sentence material as well as the noise signals with binaural room impulse responses, which had been calculated using the ODEON software, Version 8.0 (Christensen, 2005). Four simulated rooms were used for the measurements: an anechoic room, a "listening room" (7.8 m×5 m×3 m, appr. 115 m^3), a typical classroom (9.7 m×6.9 m×3.2 m, appr. 210 m^3) and a church (outer dimensions: 63 m×32 m×22 m, appr. 22.000 m^3). The listening room was designed according to IEC 268-13 (IEC, 1985) and the church was a model of Grundtvig's Church in Copenhagen. Table 3.2 lists basic room acoustic

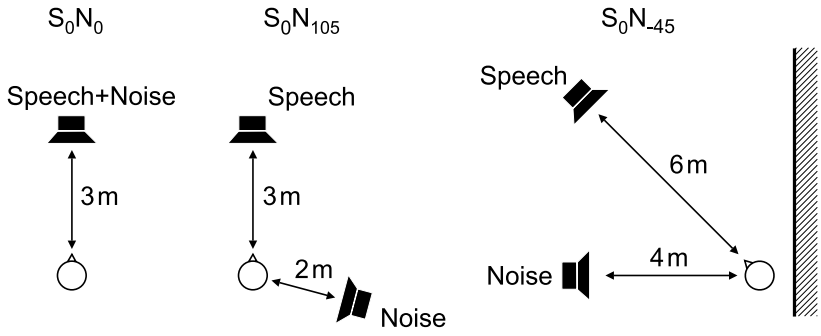


FIG. 3.1 *Speech and noise source locations in the three setups used in these measurements. The third setup was always placed very close to a wall at the right side of the listener.*

parameters[9] of the three realistic (i.e., non-anechoic) rooms. The parameters were calculated for two different speech source distances relative to the listener (which was replaced by an omnidirectional receiver), that are used in the sound source setups described below. In each room, three different spatial setups were used: S_0N_0 (i.e. the speech source at 0° and the noise source at 0°), S_0N_{105} and S_0N_{-45}. The configurations are shown in Figure 3.1. In the S_0N_{-45} situation in each room (except for the anechoic case), the listener was positioned very close to a wall opposite to the noise source, as illustrated in Figure 3.1. This was done to include the potentially disturbing effect of the direct reflections from the wall in this situation.

Subjects

A total number of 8 normal-hearing and 12 hearing-impaired subjects participated in the measurements. The ages of the normal-hearing subjects ranged from 25 to 31 years (median: 26.5 years) and the ages of the hearing-impaired subjects from 36 to 80

[9]The reverberation time T_{30} is based on the decay time of the room impulse response from -5 dB to -35 dB below the level of the direct sound, but expressed as the time after which the level has decreased by -60 dB. The early decay time EDT is calculated in a similar way, but for the first 10 dB of the decay curve. C80 ("Clarity") and D50 ("Definition") are measures which are related to the balance between early and late arriving sound energy in the room impulse response. C80 is the ratio between the energy arriving within the first 80 ms and the energy arriving later than 80 ms expressed in dB, while D50 is the (linear) ratio between the energy arriving in the first 50 ms and the total energy of the room impulse response (cf. CEN, 2000). STI denotes the Speech Transmission Index (IEC, 1998).

TABLE 3.3. *Summarized hearing losses of the hearing-impaired subjects and individual noise levels that were used in the SRT measurements. The pure tone averages (PTA) are the mean hearing thresholds in dB HL across the audiometric frequencies from 125 Hz to 750 Hz (PTA low), from 1 kHz to 3 kHz (PTA mid) and from 4 kHz to 8 kHz (PTA high). The subjects are grouped by similarity of their hearing losses: group I is a mild hearing loss, group II steep high-frequency, group III reverse sloping, group IV moderate sloping, and group V severe.*

		left ear PTA			right ear PTA			noise level
Group	Subject	low	mid	high	low	mid	high	dB SPL
I	1	8	13	20	10	12	27	70
II	2	6	29	63	10	35	63	70
	3	15	49	85	17	49	73	80
III	4	64	50	37	53	53	38	80
IV	5	34	49	67	24	49	68	75
	6	26	46	62	28	48	67	75
	7	33	51	62	34	55	62	75
	8	18	52	57	22	45	55	70
	9	33	53	57	30	48	45	70
	10	43	60	68	29	53	65	75
V	11	53	59	77	55	63	73	80
	12	58	61	70	66	66	60	85

years (median: 67 years). None of the hearing levels of the normal-hearing subjects exceeded 10 dB HL. Seven of the hearing-impaired subjects had similar, moderately sloping hearing losses. The remaining five subjects had various shapes and degrees of hearing loss. All subjects were paid for their participation. The hearing losses of the 12 hearing-impaired subjects are summarized in Table 3.3. The subjects are grouped by similarity of their hearing losses, in ascending order of severity. The frequencies for the calculation of the pure tone averages (PTAs) have been chosen according to the principal component analysis of audiograms by Smoorenburg (1992). They were 125 Hz, 250 Hz, 500 Hz, and 750 Hz for the low frequency component, 1 kHz, 1.5 kHz,

2 kHz, and 3 kHz for the mid frequency component, and 4 kHz, 6 kHz, and 8 kHz for the high frequency component.

Statistical Analysis

The statistical significance of the measured effects was analysed by means of an ANOVA of the observed SRTs, which was performed separately for normal-hearing and hearing-impaired subjects. The significance level was always 5 %. The parameters for the ANOVA of the normal-hearing subjects' data were the room condition, the spatial setup, and the noise type. Post-hoc comparisons of single parameter values were performed with Bonferroni corrections for multiple comparison. For the hearing-impaired subjects, the groups given in Table 3.3 were included as an additional parameter.

3.4.2. Results

Normal-Hearing Subjects

Figure 3.2 shows the observed SRTs for the NH subjects (filled symbols, dashed lines). The observed data for stationary noise is replotted in the panels for the other noise types (dotted lines) for comparison. The mean observed SRT for stationary noise in anechoic/S_0N_0 conditions (no binaural difference between speech and noise) of -7.3 dB SNR is very close to the reference value for the Oldenburg Sentence Test for monaural presentation of speech and noise (-7.1 dB SNR, Wagener et al., 1999c). In anechoic conditions, the NH subjects show a considerable SRT difference between corresponding S_0N_0 and S_0N_{105} or S_0N_{-45} conditions, respectively, of up to 18 dB (in babble noise and at S_0N_{105}). The SRT difference depending on the noise type in the simplest situation (anechoic/S_0N_0), that is quasi-diotic and without room acoustics, is on average 11 dB (up to 15.5 dB for individual subjects) between single-talker noise and stationary noise,

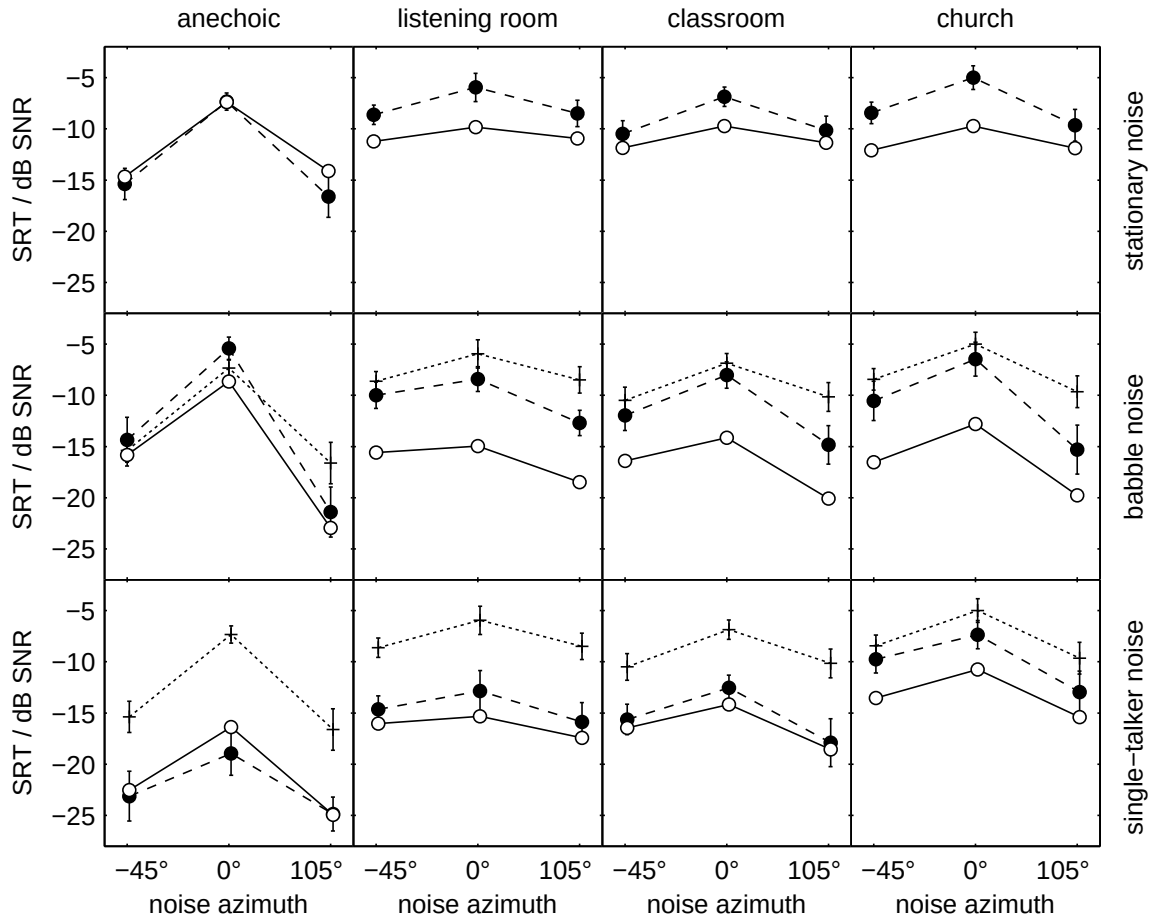


FIG. 3.2. *SRTs of the normal hearing (NH) subjects, observed (filled symbols, dashed lines) and predicted (open symbols, solid lines) data. The observed SRTs are shown as mean with interindividual standard deviations. The panels are arranged in columns per room and rows per noise type. The data for stationary noise is replotted (dotted lines) in the respective panels for babble and single talker noise for comparison.*

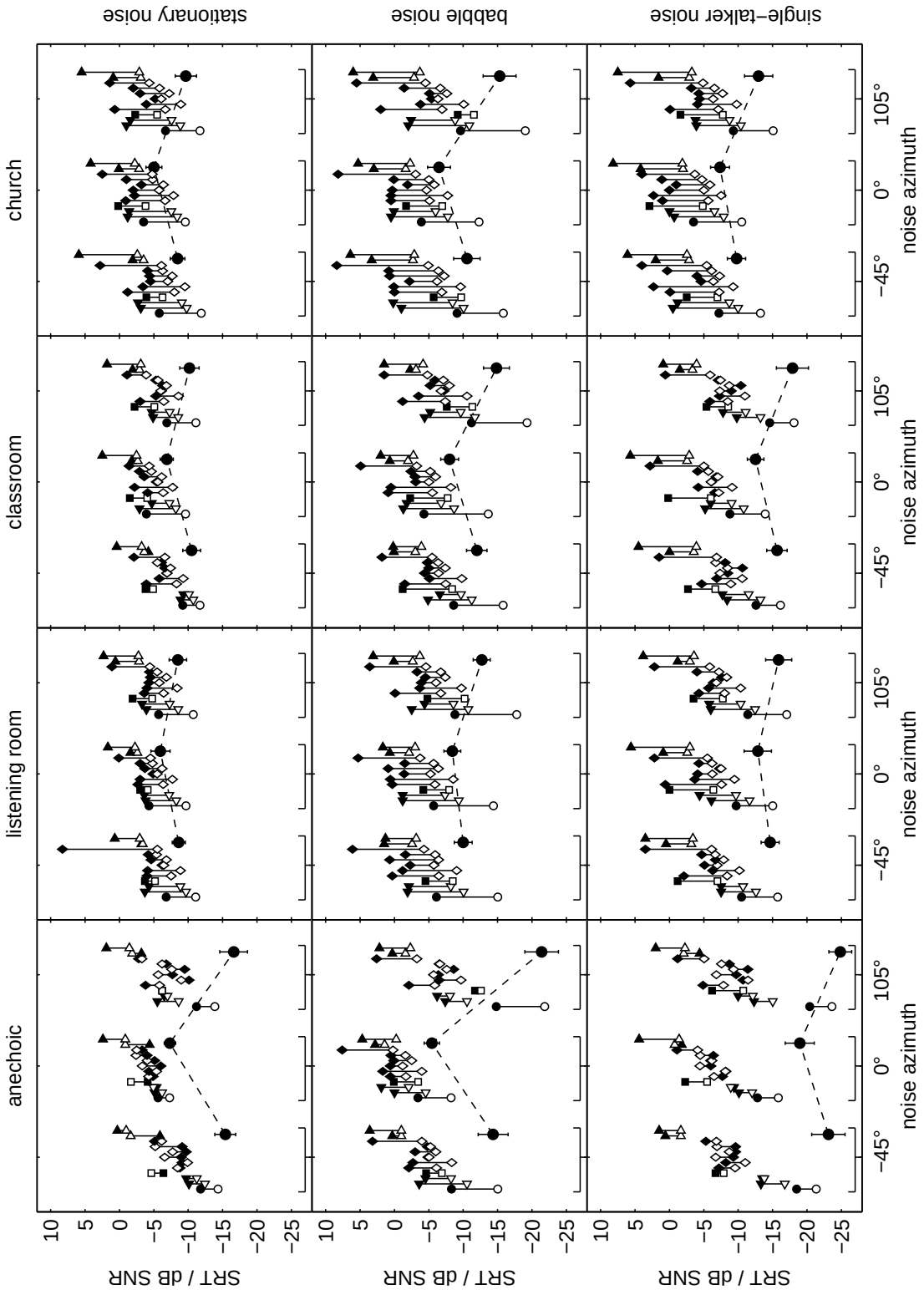
but it is non-significant between babble noise and stationary noise. With increasing reverberation time, the difference between the SRTs for S_0N_0 in single-talker noise and stationary noise decreases and becomes non-significant in the church condition, i.e. the effect of noise modulation is reduced by the reverberation. The difference between the SRTs for S_0N_0 in babble noise and stationary noise are never significant in any room.

FIG. 3.3. *(right page) Observed SRTs of individual hearing-impaired subjects (small filled symbols), corresponding individual predicted SRTs (open symbols), and mean and standard deviation of the normal hearing subjects' observed SRTs (large filled circles and dotted lines). The panels are arranged in columns per room and rows per noise type. The symbols of the hearing-impaired subjects correspond with their group in Tab. 3.3 (I: circle, II: left-pointing triangles, III: square, IV: diamonds, V: right-pointing triangle).*

The effect of noise source location, that is the difference between SRTs in the S_0N_0 condition and the other conditions, differs between noise types and is largest for the babble noise and the S_0N_{105} situations. It decreases generally between the anechoic room condition and all other three rooms, but no clear dependence on reverberation time is seen in the non-anechoic rooms on the effect of noise source location.

Hearing-Impaired Subjects

Figure 3.3 shows the individual observed SRTs of the hearing-impaired (HI) subjects (small filled symbols) and the mean observed SRTs of the normal-hearing subjects (large filled circles and dashed lines with interindividual standard deviation). A general trend of higher SRTs with increasing severity of the hearing loss (i.e., increasing group number) can be found, but the intra-group variance is in the same order of magnitude as the inter-group variance so that a larger number of subjects per group would be necessary in order to find significant correlations between subject group and results. In the anechoic/S_0N_0 condition, the SRTs are not more than 3 dB higher than the SRTs of the normal-hearing subjects for most HI subjects. But for the other interferer locations, the difference in SRT relative to the S_0N_0 condition is considerably smaller than for normal-hearing subjects for some of the HI subjects, especially in anechoic conditions and stationary noise. In contrast to the stationary noise, both modulated


anechoic
listening room
classroom
church
single−talker noise
babble noise
stationary noise
SRT / dB SNR
noise azimuth
−45°
0°
105°
10
5
0
−5
−10
−15
−20
−25

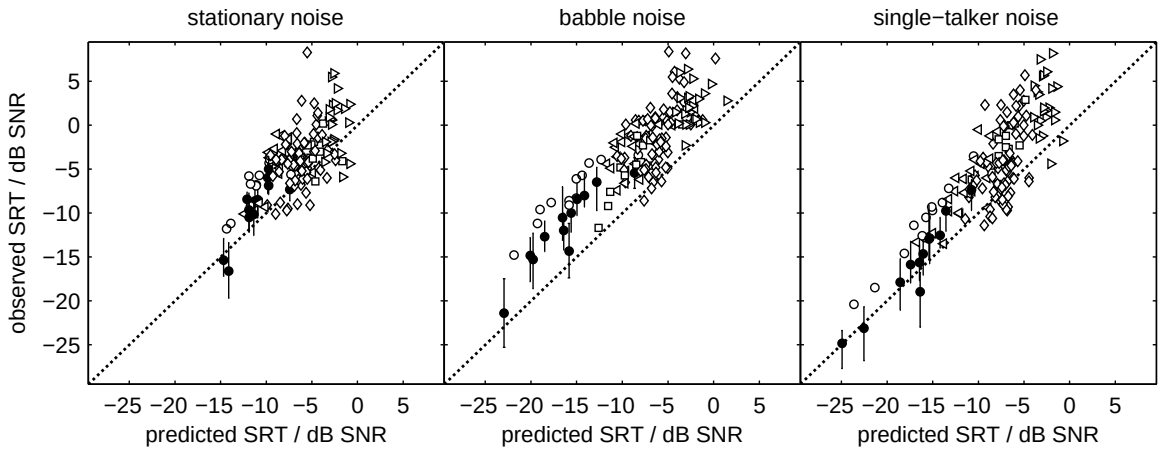


FIG. 3.4. *Scatter plots of the observed SRTs versus the predicted SRTs. Mean normal-hearing data are denoted with filled circles and lines for minimum and maximum individual SRTs, individual hearing-impaired data are denoted with open symbols. All parameter combinations are included in the plots. Each panel contains the data for one noise type. The symbols of the hearing-impaired subjects correspond with their group in Tab. 3.3 (I: circle, II: left-pointing triangles, III: square, IV: diamonds, V: right-pointing triangle).*

noises, babble and single-talker, differentiate more between the individual HI subjects, which can be seen especially in anechoic conditions and single-talker noise. This is in line with the findings of Wagener and Brand (2006). Strong reverberation, like in the church conditions, reduces the noise modulation depth and thus this differentiating effect.

Model Predictions

Figure 3.2 shows the predicted SRTs for the NH subjects (open circles, solid lines). Error bars are not shown, because there is no difference in the model predictions between individual NH subjects despite small differences in the audiograms of the NH subjects. The prediction error (i.e., the absolute difference between predicted and observed SRTs) is very small for the anechoic room condition and stationary

and single-talker noise types. The predictions for the babble noise exhibit an overall prediction error and the predicted SRTs are always too low. Additionally, there is a room-dependent prediction error in all situations and for all noise types. The effect of spatial unmasking is nevertheless predicted quite well by the model, if the room- and noise-type-dependent prediction error is removed, that is if the predicted SRTs are shifted to match the S_0N_0 condition in each panel separately. Possible reasons for these prediction errors are discussed below.

Figure 3.3 shows the individual predicted SRTs for the HI subjects (open symbols). The general trend of higher SRTs with increasing severity of the hearing loss is reflected in the predictions. In anechoic conditions, the absolute predicted SRTs are very close to the observed SRTs, but especially in the church conditions, there is a large difference between predicted and observed SRTs, an indication for particular detriment in strong reverberation not included in the model predictions, which are only based on the audiogram.

Figure 3.4 shows scatter plots of the observed SRTs versus the predicted SRTs. Filled symbols denote mean NH data, minimum and maximum of individual NH data are denoted by the error bars, and open symbols denote individual HI data. The symbols of the hearing-impaired subjects correspond with their group in Tab. 3.3 (I: circle, II: left-pointing triangles, III: square, IV: diamonds, V: right-pointing triangle). The noise-type-dependent prediction error can be observed as a parallel shift of the data points away from the unity line. There is a remaining variance in the HI data which can not be explained by the model so far, shown by the spread of data points around the unity line. This variance is slightly larger than the residual variance of the NH data that is not related with the pure tone audiogram.

In Table 3.4, the correlation coefficients for different subsets of the data are summa-

TABLE 3.4. *Correlation coefficients between predicted and observed SRTs for different subsets of the data. For "Mean NH", the normal-hearing data was averaged across subjects before calculation of the correlation.*

	all noise types	stationary	babble	single-talker
All subjects	0.88	0.80	0.92	0.93
Mean NH	0.88	0.86	0.92	0.96
Individual NH	0.84	0.80	0.86	0.91
Individual HI	0.72	0.59	0.77	0.80

rized. They correspond with the scatter plots in Figure 3.4. The leftmost column in Table 3.4 combines all plots.

3.5. Discussion

3.5.1. Model Revision

The binaural speech intelligibility model by Beutelmann and Brand (2006) was revised with the aim of simplifying the model calculation and thus reducing the processing time, and to point out the role of binaural signal parameters like the interaural level difference and interaural correlation in the calculation of the signal-to-noise ratio after EC processing without detailed assumptions about the input signals. The original model implemented the EC process as a signal processing device, using the basic EC equations with explicit delays and gain factors, and calculated an actual residual, single channel signal, which was then used as an input for the standard SII. While this was very straightforward and relatively easy to realize by combining standard elements, it involved a lot of redundant calculations. With the refinement of the model, most of these redundancies could be removed, but the model still remains independent from explicit knowledge of binaural parameters. One example of such a redundancy is the

Monte-Carlo simulation of the binaural processing errors in the original model, which calculated the residual signal several times with different processing errors randomly drawn from their distributions, and averaged over the SRTs derived from these signals with the SII. The Monte-Carlo simulation was replaced by the analytic expectation values of the speech and noise intensities after the EC process (cf. App. A) without the need for an actual residual signal. Another example is the simplification of the search process for the optimal EC parameters, gain and delay. In most psychoacoustical models, the binaural parameters are known in advance from the experimental design and do not need to be searched explicitly. The binaural speech intelligibility model is supposed to be useful even in complex situations, which can for example arise in conditions with more than one noise source or due to early reflections in room acoustics. Therefore, it is not always possible to specify the optimal binaural parameters in advance. Nevertheless, since it is possible to find the gain γ which maximizes the SNR in Eq. (3.9) for a given delay τ analytically, the number of dimensions of the search process was reduced to a single one, the delay parameter τ. All these optimizations lead to a significant reduction of computing time by a factor of about 60, that is from 10-20 min to 10-20 s on a standard PC, depending on the signal length and computing speed. Apart from the general value of short computation times, this was beneficial for the development of the extension for modulated noises, which needs to calculate the model a large number of times for small time frames of the input signals.

Expressing the result of the EC process in the form of Eq. (3.9) makes it easier to evaluate its properties than in its original form. The cosh summands in numerator and denominator are only dependent on the signal intensity and the γ parameter, while the other summands (i.e., $\lambda(\tau) * \mathrm{Re}\{\rho(\tau)\}$) are only dependent on the signal correlation and the τ parameter. This indicates that the influences of the interaural level differences

and the interaural time/phase differences on unmasking can be partly separated. If the values of the ITD/IPD-related terms are very small, that is if the interaural correlations of speech and noise are close to zero, only the intensity differences represented by the cosh terms are responsible for the effect of the EC process. If, on the other hand, one or both of the cosh terms results in a value close to one, because γ compensates the ILD (Δ_S or Δ_N), the value of the corresponding cross-correlation functions becomes very important. Hence, a noise signal cross-correlation coefficient close to one leads to a small denominator and thus to a large SNR. A speech signal cross-correlation coefficient close to -1 (equivalent to a phase inversion between the left and the right ear) would also lead to a gain in SNR, provided that the noise cross-correlation coefficient is larger than -1 at the same time.

The influence of the processing errors can also be interpreted easily in Eq. (3.9). They both control the maximal SNR benefit achievable by the binaural processing. Although the gain error factor $e^{-\sigma_\epsilon^2}$ is assigned to the cosh functions, because it is dependent on the gain parameter γ, it controls the effect of the cross-correlation function, together with the delay error factor (in the frequency domain) $e^{-\omega^2\sigma_\delta^2}$. The error factors determine the frequency dependence of the maximally achievable SNR by reducing the effective phase coherence between the ears, which can also be interpreted as a very simple model of the reduced phase locking on the auditory nerve at high frequencies. Although the filter shape and cutoff frequency differ from other peripheral models (e.g., Breebaart et al., 2001a), the error parameters used here have been taken from vom Hövel (1984), who derived them from predictions of pure tone BMLD data. Nevertheless, an adjustment of the error parameters could be used in the future to improve the prediction quality of the binaural speech intelligibility model.

In literature (Durlach, 1963, 1972; Sieben, 1979), expressions similar to Eq. (3.9) have

already been derived, but mostly for certain binaural configurations and limited to tonal target signals, while the mathematical prerequisites for Eq. (3.9) are less restricted. Eq. (3.9) can be transformed into the expression that Durlach (1963) derived for the "EC factor" f_j (Eq. (6) on p. 1210 in Durlach, 1963), because they are based on the same principle. For this purpose, the target signal is assumed to be a pure tone and the noise signal to be white Gaussian noise passed through an auditory filter centered at the target signal frequency. Both target and noise signals have constant, but not necessarily equal ILDs and ITDs and γ and τ are set to equalize the ILD and ITD of the noise signal. Although the amplitude errors are expressed in a different form in this paper ($e^{\sigma_\epsilon^2} = 1.03$) and by Durlach (1963) ($1 + \sigma_\epsilon^2 = 1.06$), their values are very similar and $1 + \sigma_\epsilon^2$ can be regarded as the first order series approximation of $e^{\sigma_\epsilon^2}$.

The revised model ("BSIM") predicts the reference SRTs from Beutelmann and Brand (2006) at least as good as the original ("EC/SII") model. This justifies to modify the standard SII with a new frequency band scheme, which matches the gammatone filter bank used in the binaural part of the model. Another possibility would have been to interpolate the results of the binaural part to one of the standard frequency band schemes, for example the critical band scheme, or to calculate the binaural part with one of the frequency band schemes from the standard. Both alternatives are possible, but the former would have been a loss of information and the latter would be less close to physiology than with the gammatone filter bank.

3.5.2. Binaural speech intelligibility in modulated noise

Regarding the observed SRTs for the normal-hearing subjects, the expectation was met, that there is an interaction between the parameters, namely sound source location, room acoustics and type of interferer. The interaction between spatial unmasking and

room acoustics, which has already been investigated by Beutelmann and Brand (2006), can also be seen in the data presented in this study. The effect of spatial unmasking is much smaller in rooms with reverberation (especially with strong early reflections) than in anechoic conditions. The difference between the respective S_0N_0 condition and the other two conditions (S_0N_{105} and S_0N_{-45}), that is the amount of spatial unmasking, differs only significantly between the three non-anechoic rooms for babble noise and in the S_0N_{105} condition. This difference is, against first expectation, larger in rooms with higher reverberation time. The reason might be a distraction by strong early reflections of the interferer in small rooms, where the distance to the walls is low. The fact that the reciprocal dependence on reverberation is strongest for the babble noise might also come from the spectral differences between the babble noise and the other two noise types. These spectral differences were up to 15 dB above 5 kHz (cf. Sec. 3.4.1), and they put a higher weight on the influence of the low frequencies in the interferer.

The influence of room acoustics on the S_0N_0 condition in stationary noise is not significant, but there is a tendency towards higher SRTs with larger reverberation time. This correlates very roughly with the decreasing room related speech intelligibility measures C80, D50 and STI in Table 3.2, but in the situations measured in this study, the masking caused by the interferer noise is obviously considerably larger than the masking caused by the reverberated target speech and thus dominates the results (cf. Lavandier and Culling, 2007).

Three different noise type have been used in this study. A stationary, speech-shaped noise ("stationary"), a 20-talker babble noise ("babble"), and a noise with one-speaker-like modulations in three frequency bands and reduced speech pauses ("single-talker"). The effect of noise modulation occurs in the considerable difference between the SRTs in stationary noise and in single-talker noise in anechoic conditions. For the babble noise,

there is only a significant difference between SRTs in the S_0N_{105} condition, which could also be an effect of the spectral difference, because there is no significant difference in all other conditions between babble noise and stationary noise. This is not surprising taking into account that the babble noise is a mixture of 20 talking persons and that, for example, Bronkhorst and Plomp (1992) have shown that SRT benefits due to fluctuations in mixtures of six or more speech-like modulated noise maskers are very small with respect to the SRTs in steady-state noise. No explanation has been found for the relatively high SRT in babble noise in the anechoic/S_0N_0 condition. However, this value is not significantly different from the same condition in stationary noise.

The spatial unmasking in anechoic conditions for single-talker noise is lower than for stationary noise. This could be a threshold effect, because the lowest instantaneous noise levels at gaps in the interferer are low enough that the hearing threshold determines the intelligibility, even if the spatial unmasking would have a larger effect, if the overall noise level was higher.

The reverberation has a deteriorating effect on the benefit due to the modulated interferer. In the church condition, the SRTs for S_0N_0 and S_0N_{-45} do not significantly differ between single-talker noise and stationary noise, which means that the reverberation has filled the gaps in the modulated interferer and reduced the effective modulation depth.

The SRTs of the hearing-impaired subjects are generally between 1 dB and 27 dB higher than the average SRTs of the normal-hearing subjects, depending on the condition. There is a general trend in the hearing-impaired data that the SRT difference to the normal-hearing subjects increases with increasing hearing loss, but this is not always the case. Hearing-impaired group I (cf. Table 3.3), with the least severe hearing loss, has in most cases the lowest SRTs compared to the other hearing-impaired, while

group V, the most severe hearing-losses, typically has the highest SRTs. The other groups, with intermediate hearing losses, do not differ much in their results, except for the effect of noise type. In stationary noise, there is no significant difference between groups II–IV, but the differences are larger and significant between all groups except III and IV in single-talker noise. This is in line with the results from Wagener and Brand (2006), that modulated noises differentiate more between hearing losses. Comparing the maximum benefit from the modulated interferer for normal-hearing subjects (15.5 dB) with the results of the hearing-impaired subjects shows, that some hearing-impaired subjects still have a benefit from the modulated interferer, even though it is reduced, but some even have a disadvantage due to the interferer modulations.

In some conditions, especially anechoic and stationary noise, the effect of spatial unmasking is considerably reduced or even absent for hearing-impaired subjects (e.g., in the top-left panel in Fig. 3.3, group V, right-pointing triangles). If there is only little spatial unmasking in the normal-hearing data, it is just as well not expected that hearing-impaired subjects can benefit from the interferer location. One subject (number 10 in Table 3.3) shows striking results: In babble noise, subject 10 has always the highest SRT and in some conditions (listening room/stationary noise, anechoic/single-talker noise and church/single-talker noise), the reduction of spatial unmasking is very large for this subject. This can probably be explained by the special hearing loss with a rising slope toward high frequencies, which is especially detrimental in the babble noise with its relatively high SNR at high frequencies, because the hearing loss nullifies the favorable external SNR at this frequency range.

3.5.3. Prediction of SRTs in modulated noise

The aim of predicting binaural speech intelligibility in modulated noise was generally successful, although there are discrepancies between the predicted and observed SRTs which need to be explained. The predictions in anechoic conditions agree well with the observed data, regarding the effect of spatial unmasking and of noise modulation, including their interaction. There is a small remaining prediction error in anechoic conditions for babble noise, where the predicted SRTs in the S_0N_0 condition are slightly too low, and for the single-talker noise, where the predicted SRT in the S_0N_0 is slightly to high. Both prediction errors are small compared to the discrepancies in the other rooms.

The effect of the frame length can be shown by comparing the model predictions presented here with predictions calculated for the same data, but with the long frame "BSIM". The predictions with the latter model for stationary noise and babble noise differ from the predictions with the short frame length between -2 dB and +1 dB. For the single-talker noise however, the predictions with a long frame length are equal to the predictions for stationary noise, showing that the prediction of the effect of modulated noise is indeed affected by the frame length.

In the non-anechoic rooms, the prediction errors are larger than for anechoic conditions. The prediction error depends on the room and on the noise type, but much less on the spatial setup. The prediction error for stationary noise and single-talker noise is most probably due to the effect of room acoustics on the target speech itself, because the distance between speech and listener is quite large, even in the rooms with low reverberation times. The D50 and STI values in Table 3.2 support this. For these conditions, a correction factor for the detrimental components of the speech signals

is necessary, but it can not simply be derived from the rooms acoustical measures, because it interacts with the influence of the noise masker.

The babble noise predictions exhibit an additional prediction error, which is difficult to explain. It seems to be caused by the spectral differences between babble noise and stationary noise or single-talker noise, which lead to high SNRs in the high frequency region. These high SNRs could be overinterpreted by the SII, but the changes needed in the band importance function of the SII in order to achieve correct results for the babble noise are too extreme and would increase the prediction error in other conditions.

Overall, the simple extension of the binaural speech intelligibility model for modulated noise has shown that the frame-wise calculation procedure is a reasonable approach. On this basis, future studies are needed to eliminate the remaining prediction errors and to improve details of the model. Next steps could be, for example, to introduce a frequency-band-dependent frame length and the effects of forward masking, as it has been shown to be effective in the (monaural) extended SII for fluctuating interferers by Rhebergen et al. (2006).

3.6. Conclusions

1. The binaural speech intelligibility model (Beutelmann and Brand, 2006) was analtically simplified and expressed more concisely. Thus, along with numerical optimizations, the practical use of the model was considerably accelerated (by a factor of about 60), while maintaining equivalent predictions compared to the original model. The correlation coefficients between predictions of the revised model and the observed SRTs are larger than the correlation coefficients of the original “EC/SII” model. The root mean squared prediction error of the “BSIM”

(normal-hearing subjects: 1.3 dB, hearing-impaired subjects: 1.9 dB) was less than for the original "EC/SII" model (NH: 1.7 dB, HI: 2.3 dB). An additional simplification and acceleration is possible for the prediction of results for several subjects measured in the same binaural condition, because the time-consuming search for the maximal SNR is independent of the audiogram and needs to be done only once for each binaural condition.

2. The binaural EC errors are mathematically equivalent to a low pass filter reducing the interaural fine structure correlation for high frequencies. This is analogous to the low pass filter in more physiological (hair cell) models.

3. Binaural speech intelligibility in modulated noise interferers can, in principle, be predicted by calculating the model in short time frames and averaging the resulting SRTs. The correlation coefficients between predicted and observed SRTs range from 0.80-0.91 (depending on noise type) for individual normal-hearing subjects, 0.85-0.96 for mean normal-hearing data and 0.57-0.75 for hearing-impaired subjects.

4. In situations with modulated noise interferer, a large SRT benefit of up to 15.5 dB relative to unmodulated noise was measured for normal-hearing subjects. The benefit decreases to zero with increasing reverberation time, but the interaction with sound source location was rather small compared to the overall size of the effect. Hearing-impaired subjects generally have less benefit, in certain cases they are even disturbed by modulated interferers and their SRTs are higher than with stationary interferers.

5. About 70 % of the variance in SRTs of hearing-impaired listeners relative to the mean SRT of the normal-hearing listeners in the same condition can be predicted with the binaural model presented here based the audiogram alone. The remaining variance presumably requires a more detailed model of hearing loss that includes more factors than just the hearing threshold.

Acknowledgements

This research was supported by grants from the European Union FP6, Project No. 004171 HEARCOM. The authors wish to thank Claus Lynge from Ørsted DTU, Acoustic Technology, Technical University of Denmark for providing the binaural room impulse responses.

4. Prediction of binaural speech intelligibility with frequency-dependent interaural phase differences[10]

Abstract

The aim of this study was to test the hypothesis of independent processing strategies in adjacent binaural frequency bands underlying current models for binaural speech intelligibility in complex configurations and to investigate the effective binaural auditory bandwidth in broadband signals. Speech reception thresholds were measured for binaural conditions with frequency-dependent interaural phase differences (IPDs) of speech and noise. Threshold predictions with the binaural speech intelligibility model by Beutelmann and Brand (2006, J. Acoust. Soc Am. 120(1), 331–342) were compared with the observed data. The IPDs of speech and noise had a sinusoidal shape on a logarithmic frequency scale. The bandwidth between zero crossings of the IPD function was varied from 4 to 1/8 octaves. Speech and noise had either the same IPD function (reference condition) or opposite signs of the IPD function (binaural condition). Each condition had two subconditions with alternating and non-alternating signs, respectively, of the IPD function. The SRT benefit with respect to the reference condition decreased from 6 dB to zero with decreasing IPD bandwidth for the alter-

[10]This chapter has been submitted in the present form for publication to the Journal of the Acoustical Society of America (Beutelmann et al., 2008b).

nating condition while it stayed significantly larger than zero for the non-alternating condition. The observed results were well predicted by the model with an analysis filter bandwidth of 2.3 ERB.

4.1. Introduction

Binaural hearing plays an important role in solving the "cocktail party problem" (Cherry, 1953), a term used for the task of understanding speech in complex environments. This is one of the reasons, why the effects of binaural hearing have received considerable attention in literature. The classical way of quantifying this effect is the binaural masking level difference (BMLD) which is used to describe the threshold level difference for the detection of a pure tone target in noise between a binaural condition and a (typically diotic) reference condition. When trying to predict binaural speech intelligibility on the basis of binaural tone detection experiments, however, it has to be observed that the latter typically employ a narrow band target signal, while the target signal in most speech intelligibility experiments is broad band, as well as the interferer signals.

It has been shown, nevertheless, that binaural masking level differences (BMLDs) at different frequencies, whether they are subjectively measured or predicted by a psychoacoustical model, are a good predictor for the frequency-dependent effective signal-to-noise-ratio (SNR) enhancement for speech in noise (vom Hövel, 1984; Zurek, 1990; Culling et al., 2004; Beutelmann and Brand, 2006). This study is concerned with the questions that arise with the division of the broad-band speech and noise signals into narrow frequency bands. Although the concept of the auditory filter and integration of auditory processes within a certain, finite frequency range is generally accepted,

it is not clear whether (1) the effective auditory filters and their bandwidths for binaural processes are different to their monaural counterparts and (2) the hypothesis of independent binaural processing in adjacent or remote auditory frequency bands is true. This study has the intention of assessing the relevance of these questions for binaural speech intelligibility prediction models. Both items are related, because a large effective bandwidth also affects the independence of adjacent frequency regions.

The effective bandwidth in binaural tone detection has been measured using band-widening and broadband approaches that differ in their effective binaural bandwidths. The most common experiment is to measure tone detection thresholds in noises with different bandwidths centered around the target tone (Wightman, 1971; Sever and Small, 1979; Hall et al., 1983; Cokely and Hall, 1991). If the overall noise level is kept constant, the threshold remains constant for noise bandwidths below the effective critical bandwidth and decreases if the noise bandwidth exceeds the effective critical bandwidth. If the power spectral density of the noise is kept constant, the threshold increases for increasing noise bandwidth up to the effective critical bandwidth and stays constant for higher bandwidths. The effective critical bandwidths measured for an antiphasic tone in homophasic noise (dichotic condition) appear to be 1.5 to 4 times larger than the effective critical bandwidths measured with homophasic tone and noise (diotic condition). In the dichotic conditions, the effective critical bandwidth is furthermore dependent on the noise power spectral density and increases with increasing level (Hall et al., 1983). In broadband conditions, however, the differences between the effective monaural and binaural bandwidths are much smaller. Hall et al. (1983) also measured the critical bandwidth with a notch of variable width in broadband noise, centered on the target tone. In the diotic case, the notched-noise and bandlimiting critical bandwidths are about the same, but in the dichotic case, the effective critical

bandwidth measured with the notched-noise paradigm is considerably lower than with the bandlimiting paradigm. Nitschmann and Verhey (2007) presented a successful approach which was able to model these results using weighted sums of neighboring auditory filters and thus increasing the effective binaural bandwidth.

Another paradigm was employed by Sondhi and Guttman (1966) and Holube et al. (1998). In theses studies, the noise spectra were broadband and flat, but the interaural phase was inverted in a rectangular region of variable width centered around the target tone. The interaural phase was either the same as the interaural phase of the target tone (0 or π), or it was the opposite. The estimated effective binaural bandwidth depended on the assumed filter shape and on the fitting method, but it was in all cases significantly larger for conditions with a phase difference between target and on-frequency noise band than in the other conditions with the same phase of target and on-frequency noise band.

In other studies, a single, sharp transition between 0 and π interaural phase difference (IPD) in an otherwise flat-spectrum, broadband noise was used. Kohlrausch (1988) varied the target tone frequency and thus the influence of the interaural phase edge on the detection of the target tone, while Kollmeier and Holube (1992) varied the edge frequency and the target tone frequency was fixed. Kohlrausch (1988) concluded, that the effective peripheral critical bandwidth for binaural processes might not be larger than the monaural critical bandwidth, but that the effects found in binaural bandwidth experiments are a consequence of different detection mechanisms for monaural and binaural hearing. A similar conclusion was drawn by Kollmeier and Holube (1992), although in this study there was a significant difference in binaural and monaural bandwidth by a factor of 1.2. They furthermore pointed out, that the estimate of the bandwidth is critically dependent on the filter shape and in which way the bandwidth

of the respective filter shape is calculated. Holube et al. (1998) used another paradigm similar to the one used by Houtgast (1977) for monaural auditory filters. The interaural correlation was changed sinusoidally with frequency and the detection thresholds were measured as a function of the periodicity in the (linear) frequency domain. The estimated effective binaural critical bandwidths were larger than the ones estimated from the rectangular and stepwise interaural correlation changes in the same study.

While the so far mentioned studies concern the effective binaural bandwidth, the hypothesis of independent binaural processing channels is examined in studies with multiple target tones or speech with frequency-dependent interaural phase or time differences: Akeroyd (2004) showed, that detection thresholds of multi-component tone complexes of up to 17 components stretching from 200 Hz to 1 kHz in broadband, white noise were the same for S_0N_{180}, $S_{180}N_0$ and $S_{270}N_{90}$, where the index of S denotes the target interaural phase difference (IPD) in degrees and the index of N denotes the noise IPD. If the binaural system was constrained to eliminate only noise with a single interaural time difference (ITD) across all frequencies, the thresholds would have been different.

There are also studies which use speech or speech-like sounds as targets in binaural experiments. While the processing of strongly frequency dependent interaural level differences (ILDs) seems to be dominated by the ear with the better SNR (Edmonds and Culling, 2006), there is evidence that (in speech intelligibility experiments) it is possible to process different ITDs and IPDs in high- and low-frequency regions separately (Culling and Summerfield, 1995; Edmonds and Culling, 2005). This is obviously true, as long as the binaural cues are not needed for localization and subsequent streaming of different auditory objects (Best et al., 2007). For speech in stationary noise without further speech-like distractors, this should be the case, because especially

the harmonicity of speech sounds is a stronger cue than spatial location (Buell and Hafter, 1991).

The experimental results raise the question, whether the existing binaural models can predict the findings. Metz et al. (1968) needed to include a bandwidth dependence in the binaural processing errors of the EC model (Durlach, 1963) in order to accommodate for the noise bandwidth dependence of binaural detection thresholds. Sondhi and Guttman (1966) found that basic changes in the concept of the EC model would be needed in order to predict the data from experiments in their study with rectangularly inverted spectral phase. The binaural model of Breebaart et al. (2001b), however, was able to explain the wider binaural bandwidth of bandlimiting experimental paradigms without explicit adjustment of the model parameters quite well.

The binaural speech intelligibility model of Beutelmann and Brand (2006) applied in this study uses a gammatone filter bank (Hohmann, 2002) to split the input signals into auditory ERB-wide frequency bands (Glasberg and Moore, 1990). In each frequency band, the maximally possible SNR enhancement due to interaural differences is calculated using the equalization-cancellation (EC) principle proposed by Durlach (1963) with error parameters adapted from vom Hövel (1984). The equalization parameters are independent in each frequency band, but the gammatone filters overlap to a large extent, thus the processing is not completely independent between frequency bands. After that, a speech reception threshold (SRT) is calculated from the band-wise speech and noise levels with the help of the Speech Intelligibility Index (SII, ANSI, 1997). This model has yielded good SRT predictions for a single, stationary noise source at various azimuths and in different room acoustics and a simple extension for modulated interferers yielded also promising results (Beutelmann and Brand, 2006).

The purpose of the current study was to test the performance of the binaural speech

intelligibility model presented by Beutelmann and Brand (2006) in conditions with strongly frequency dependent interaural phase differences of both speech and noise. The intention was to examine the so far reasonable and successful choice of the gammatone filter bank by Hohmann (2002), especially the filter bandwidth. The paradigm chosen was a sinusoidally varying IPD based on Houtgast (1977) and Holube et al. (1998) to create frequency-dependent binaural cues that require independent binaural processing in different frequency regions in order to achieve a binaural benefit. Assuming that the hypothesis of independent binaural channels is true, a variation in the spectral spacing of the conflicting binaural cues was introduced as a parameter. If the spectral distance between conflicting binaural cues is smaller than the effective binaural integration bandwidth, a decrease in binaural benefit is expected. This allows for an estimate of the correct filter bandwidth of the binaural speech intelligibility model.

In this study, the sinusoidal variation of the IPD was defined on a logarithmic frequency axis in order to be consistent with the roughly constant ratio of auditory bandwidth and center frequency. Speech reception thresholds (SRTs) were measured in the described binaural condition with opposite IPD signs for speech and noise, varying the IPD periodicity. As reference, conditions with the same IPD function but equal IPD signs for speech and noise were measured. In these conditions, no effect of binaural unmasking was expected. The same conditions, but with one channel switched off, were included in order to assess any effects of the monaural phase distortion. The observed SRTs were compared with the predictions of the binaural speech intelligibility model by Beutelmann and Brand (2006) for the same conditions, which was calculated with various filter bandwidths.

4.2. Methods

4.2.1. Sentence Test Procedure

The speech intelligibility measurements were carried out using the HörTech Oldenburg Measurement Applications (OMA), version 1.2. The Oldenburg Sentence Test in noise (Wagener et al., 1999a,b,c) was used as speech material. Except for the convolution with the filters that produced the binaural conditions as described in section 4.2.2, the signals complied with the commercially available version. Each sentence of the Oldenburg Sentence Test consists of five words with the syntactic structure 'name verb numeral adjective object'. For each part of the sentence, ten alternatives are available, each of which occurs exactly twice in a list of 20 sentences, but in random combination. This results in syntactically correct, but semantically unpredictable sentences. The subjects' task was to repeat each word they recognized after each sentence as closely as possible. An instructor marked the correctly repeated words on a touch screen display connected to a computer, which adaptively adjusted the speech level after each sentence to measure the SRT level of 50% intelligibility. The step size of each level change depended on the number of correctly repeated words of the previous sentence and on a "convergence factor" that decreased exponentially after each reversal of presentation level. The intelligibility function was represented by the logistic function, which was fitted to the data using a maximum-likelihood method. The whole procedure has been published by Brand and Kollmeier (2002a, A1 procedure). A test list of 20 sentences was selected from 45 such lists to obtain each observed SRT value. Two sentence lists with 20 sentences each were presented to the subjects prior to each measurement session for training purposes. The test lists were balanced across subjects and conditions and all measurements except for the training lists were performed in random order.

The noise used in the speech tests was generated by randomly superimposing the speech material of the Oldenburg Sentence Test (Wagener et al., 1999a,b,c). Therefore, the long-term spectrum of this noise is very similar to the mean long-term spectrum of the speech material. The noise token was presented simultaneously with the sentences. It started 500 ms before and stopped 500 ms after each sentence. The starting point of the noise token was randomly selected within the whole noise signal of about 3.7 s which was looped to its beginning if necessary. The noise level was kept fixed at 65 dB SPL.

The headphones (Sennheiser HDA 200) were free-field equalized according to international standard (ISO 389-8), using an FIR filter with 801 coefficients. This free-field equalization is already inherent in the standard signals of the Oldenburg Sentence Test for the HDA 200. The measurement setup was calibrated to dB SPL using a Brüel & Kjær (B&K) 4153 artificial ear, a B&K 4134 1/2" microphone, a B&K 2669 preamplifier, and a B&K 2610 measuring amplifier.

4.2.2. Stimuli

Both speech and noise signals were presented to the listeners with frequency-dependent interaural phase differences (IPDs). The IPD $\phi(f)$ as a function of frequency was given by

$$\phi(f) = \phi_0 \sin\left[4\pi \left(B \log \frac{f_h}{f_l}\right)^{-1} \log \frac{f}{f_l}\right]. \tag{4.1}$$

$|\phi_0|$ was always $\pi/2$, whereas the sign of ϕ_0 was varied according to the condition and the respective signal. The speech and noise signals were bandpass filtered between $f_l = 250\,\mathrm{Hz}$ and $f_h = 4000\,\mathrm{Hz}$ and $\phi(f)$ was set to zero below f_l and above f_h in order

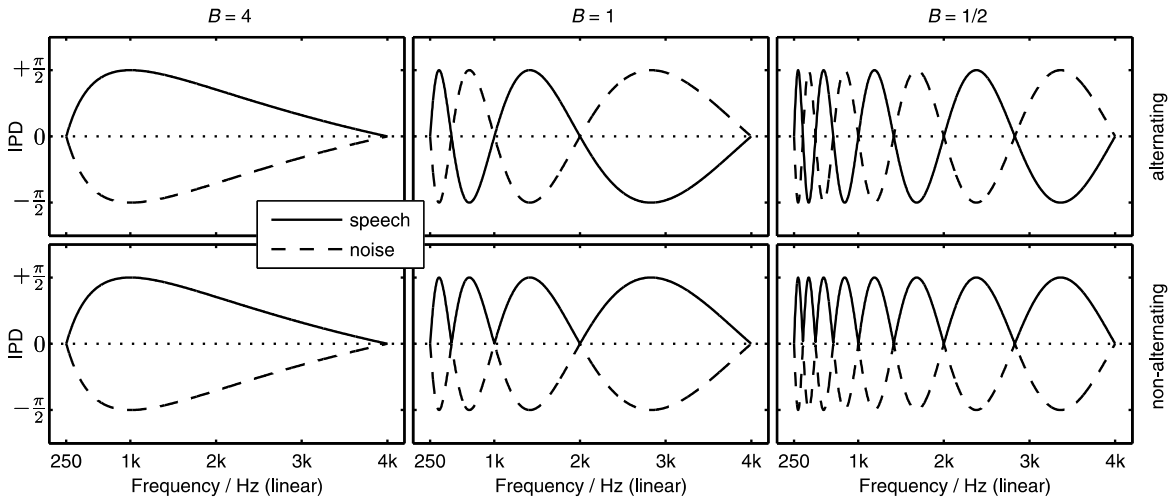


FIG. 4.1. *Schematic display of the interaural phase difference (IPD) function $\phi(f)$ for three different examples of the parameter B which corresponds to the IPD bandwidth in octaves. Solid lines show the IPD of the speech signal, dashed lines the IPD of the noise signal. The upper panels show the IPD functions used in the conditions, in which periods of positive and negative IPD signs alternate. The lower panels show IPD functions of the non-alternating conditions, in which the speech signal has always a positive IPD and the noise signal always a negative IPD. The signals were bandpass filtered between 250 Hz and 4 kHz, IPDs outside this range (considering finite filter slopes) were always zero.*

to avoid edge effects because of the finite filter slopes. The parameter B was used to control the bandwidth of the half periods of $\phi(f)$ or, in other words, the distance between zeros crossings of the IPD. B corresponds to the frequency ratio between zero crossings measured in octaves. The values used for B were: 0.125, 0.25, 0.5, 1, 2, and 4. At these values of B, $\phi(f)$ is equal to zero at f_l and f_h. Examples of $\phi(f)$ for different values of B are displayed in Fig. 4.1. An important distinction is made in the following between the "IPD bandwidth", which is controlled by the parameter B, and the "filter bandwidth", which denotes the filter bandwidth of the model. The IPDs were realized by fast convolution of the speech and noise signals with finite impulse response filters. The filters were digitally generated in the frequency domain and had

a length of 65536 samples ($\approx 1.49\,\mathrm{s}$ at a sampling rate of 44100 Hz). The phase shift creating the IPD was divided symmetrically among both ears in order to reduce the monaural phase distortions. Thus, the frequencies at a maximum or minimum of $\phi(f)$ were shifted by $\pm\pi/4$ in the left ear and $\mp\pi/4$ in the right ear with respect to the frequencies at which $\phi(f)$ was zero. The amplitude function of the filter was flat between 250 Hz and 4000 Hz and decreased linearly to zero within a third octave below and above this region. The correct actual interaural phase difference (with respect to phase distortions in the headphones and due to the headphones' placement) was controlled by recording the output of the headphones with an artificial head (B&K HATS 4128C) several times and removing and repositioning the headphones after each recording. A frequency-dependent IPD deviation from the desired value was measured, which is mostly due to an asymmetry of the artificial head in combination with the limited reproducibility of headphones placement. The maximal absolute IPD deviation in the frequency range used in the stimuli (250 Hz to 4 kHz) was $\pi/6$. This deviation is not expected to affect the results substantially, because the exact IPD is not critical in the design of this experiment, as long as the IPD difference between speech and noise as well as the alternating IPD signs are reproduced correctly. Therefore, the difference between adjacent IPD maxima and minima was measured in the recordings. The deviation from the desired value of π was below $\pi/50$.

SRTs were measured in six conditions for each value of B. The conditions were a combination of the amount of binaural cues for segregation of speech and noise present in the stimuli (monaural, reference, binaural) and the characteristics of the IPD function (alternating, non-alternating). The IPD functions used for each condition are listed in Tab. 4.1. In the binaural conditions, the IPDs of speech and noise exhibit differences of up to $\pm\pi$ which may be used as a cue for binaural unmasking. In the

TABLE 4.1. *Conditions and their respective IPD functions used for the speech and noise signals, where $\varphi(f)$ is given in Eq. 4.1. The monaural conditions are equivalent to the binaural conditions except that the right headphone was switched off and only the monaural phase distortion due to the IPD filter was present in the left ear. In the reference and the binaural conditions, stimuli were presented to both ears.*

		monaural	reference	binaural
alternating	speech IPD	$+\varphi(f)$	$+\varphi(f)$	$+\varphi(f)$
	noise IPD	$-\varphi(f)$	$+\varphi(f)$	$-\varphi(f)$
non-alternating	speech IPD	$+\lvert\varphi(f)\rvert$	$+\lvert\varphi(f)\rvert$	$+\lvert\varphi(f)\rvert$
	noise IPD	$-\lvert\varphi(f)\rvert$	$+\lvert\varphi(f)\rvert$	$-\lvert\varphi(f)\rvert$

non-alternating binaural condition, the difference was always positive, while in the alternating binaural condition, the sign of the IPD difference between speech and noise changed at each zero crossing of the IPD function. The alternating binaural condition requires independent binaural processing in different frequency bands for maximal binaural unmasking. In the reference conditions, speech and noise had the same IPD at all frequencies and thus no binaural unmasking could be expected, neither for the alternating reference condition nor for the non-alternating reference condition. In the monaural conditions, the stimuli were only presented to the left ear, the right ear channel was switched off. The presented left channel contained the same phase shifts that were necessary to generate the IPDs in the binaural conditions in order to asses the effect of monaural phase distortions on the SRT.

4.2.3. Subjects

A total number of 6 subjects with normal hearing participated in the measurements. Their ages ranged from 24 to 32 years. Their hearing levels did not exceed 15 dB HL. All subjects had little or no prior experience in sentence tests. Three of them

were members of the research group, the other three subjects were paid for their participation.

4.2.4. Model

A detailed description of the binaural speech intelligibility model used to predict the measurement data of this study can be found in Beutelmann and Brand (2006). Here, only a short overview of the important features is given. The binaural speech intelligibility model processes separately binaural speech and noise input signals. The signals are split into 30 ERB-wide frequency bands (Glasberg and Moore, 1990) between 140 Hz and 9 kHz with a gammatone filter bank (Hohmann, 2002). In each frequency band, an equalization-cancellation (EC, Durlach, 1963) model process is used to estimate the best signal-to-noise ratio (SNR) achievable by binaural interaction. The performance of the process is limited by both an additional internal noise that represents the hearing threshold and artificial inaccuracies of the EC process (cf. Durlach, 1963; vom Hövel, 1984) that constrain the maximum SNR benefit due to binaural interaction. The band-wise SNRs are then used as input into the Speech Intelligibility Index (SII, ANSI, 1997), from which a speech intelligibility and finally an SRT is computed. A special SII frequency band scheme in deviation from the standard was employed in order to match the center frequencies of the SII with the gammatone filter bank. The SII calculation procedure was left unchanged except for the computation of the spread of masking between the frequency bands, which was skipped. This was done because the gammatone filters used in the binaural model are overlapping and already incorporate the spread of masking as it is computed explicitly in the standard SII for non-overlapping bands. The importance function was adapted from the standard importance function for speech in noise, “SPIN”, by interpolating

the bandwidth-weight product, which is practically constant across all standardized SII frequency band schemes.

In order to examine the influence of filter bandwidth, the model calculations were repeated with different bandwidths of the filters in the gammatone filter bank. The filter bandwidths were varied in steps of 0.1 ERB between the original value of 1 ERB and 4 ERB, while the center frequencies remained unchanged. Furthermore, the calculations were repeated with the model that was forced to use a constant time delay ($\tau = \text{const.}$) or a constant phase delay ($\varphi = \omega_k \tau_k = \text{const.}$, where ω_k is the center frequency and τ_k the equalization delay of the k-th band) across all bands, while calculating the best possible SRTs for each type of delay.

4.3. Results

4.3.1. Measurement Data

Fig. 4.2 shows the speech reception thresholds (SRTs) of all conditions. The observed SRTs are displayed as medians of six subjects with error bars showing the respective upper and lower quartile of the data. Filled symbols represent the alternating conditions and open symbols represent the non-alternating conditions. The data at a bandwidth of 4 octaves are the same in alternating and non-alternating conditions, because one single IPD half period spans the complete frequency range used in this experiment and there is no difference between the alternating and non-alternating conditions at this IPD bandwidth. As expected, no binaural unmasking was found in all monaural and reference conditions (leftmost and middle panel in Fig. 4.2). The SRTs in the alternating binaural condition are strongly dependent on the IPD bandwidth as opposed to the SRTs in the non-alternating binaural condition.

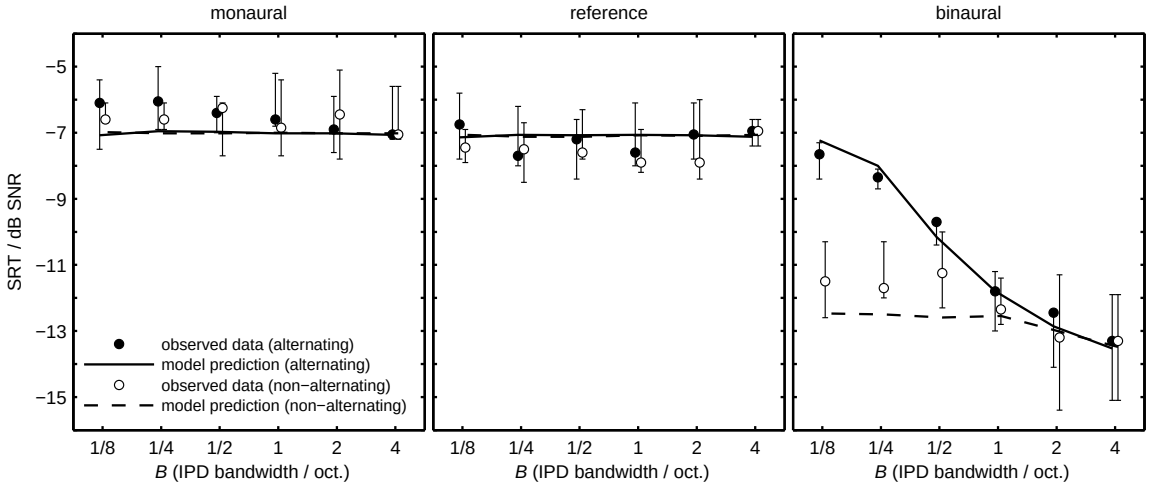


FIG. 4.2. *Observed speech reception thresholds (SRTs) of six subjects (circles, median with upper and lower quartiles) and model predictions (lines). Filled symbols and solid lines represent alternating conditions, open symbols and dashed lines represent non-alternating conditions. The leftmost panel shows SRTs for the monaural conditions, the middle panel for the reference conditions with equal IPD for speech and noise, and the rightmost panel for the binaural conditions with opposite IPD signs for speech and noise.*

An ANOVA of the observed SRTs with the three factors IPD bandwidth, condition, and subject showed a large, significant (at the 5% level) main effect of the factor "subject" and no significant effect of the other factors and of any two-way interaction for the conditions in the two left panels in Fig. 4.2. Post-hoc comparisons with Bonfferoni adjustments for multiple comparisons show the following results (all significances at the 5% level): In the binaural conditions (right panel), a significant difference was found between alternating and non-alternating at 0.125, 0.25, and 0.5 octave IPD bandwidth. At 0.125 and 0.25 octaves IPD bandwidth, the SRTs in alternating binaural conditions are not significantly different from the respective SRTs in the alternating reference conditions, which means that no significant binaural unmasking was found in these conditions for IPD bandwidths lower than 0.5 octaves. The amount of binaural

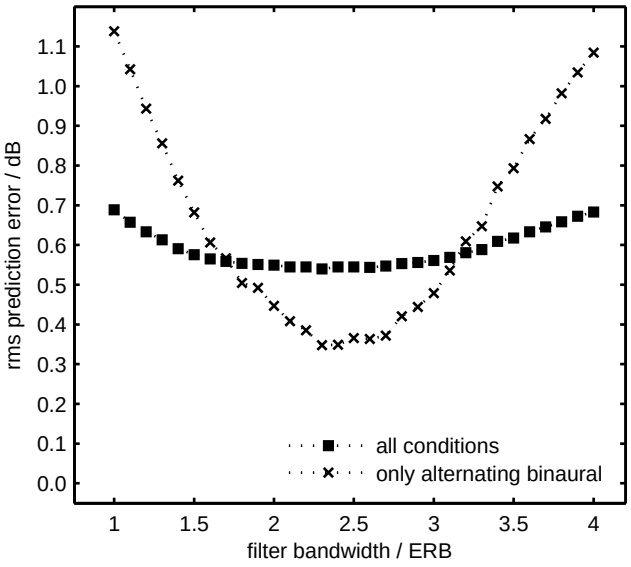


FIG. 4.3 *Root mean squared errors between model predictions and observed data (median across subjects) as a function of filter bandwidth for all conditions (square symbols) and for the alternating binaural conditions (cross symbols).*

unmasking in the non-alternating binaural conditions, however, remains significantly larger than zero for all IPD bandwidths.

In the monaural conditions (left panel), no significant difference was found between alternating and non-alternating conditions at all IPD bandwidths. The same is true for the reference conditions (middle panel). Between monaural and reference conditions, only one significant difference was found for the combination of 0.25 octaves IPD bandwidth and alternating sign.

4.3.2. Model Predictions

In Fig. 4.2, the predicted SRTs are shown with solid (alternating) and dashed (non-alternating) lines for a filter bandwidth of 2.3 ERB. This filter bandwidth resulted in the lowest overall root mean squared error between predicted and observed data (median across subjects) of 0.5 dB. This is the minimum of the overall root mean squared prediction error as a function of filter bandwith shown with square symbols in Fig. 4.3. The overall correlation coefficient between predicted and median observed data is 0.98

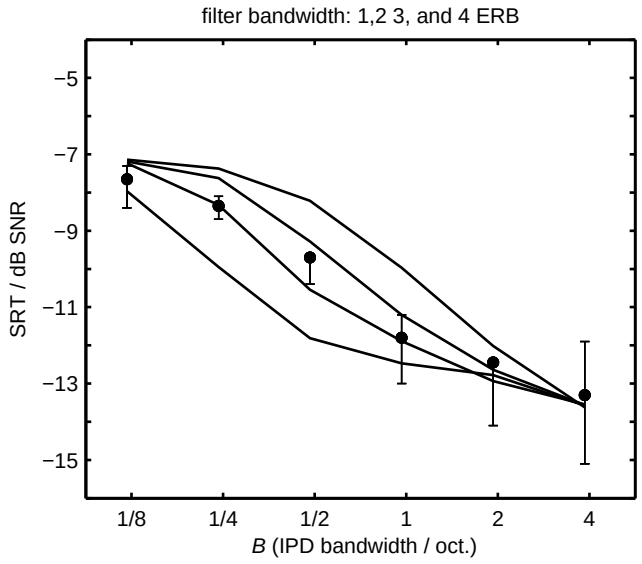


FIG. 4.4 *Model predictions with a filter bandwidth of 1, 2, 3, and 4 ERB (lines, from left to right) and observed SRTs of six subjects (circles, median with upper and lower quartiles) for the alternating binaural conditions. The symbols and line styles and the observed data correspond to the rightmost panel in Fig. 4.2.*

in Fig. 4.2. Model predictions at filter bandwidths of 1, 2, 3, and 4 ERB, are shown in Fig. 4.4. The rms errors at filter bandwidths of 1 ERB and 4 ERB, respectively, were both 0.2 dB higher than at 2.3 ERB and the error values increase monotonically from the minimum as a function of filter bandwidth (square symbols in Fig. 4.3). The range of the error values is small, because it is dominated by the errors in the monaural and reference conditions and by the non-alternating binaural conditions (dashed line, rightmost panel in Fig. 4.2). The difference between predictions and observed data in the non-alternating binaural conditions is larger than for most other conditions, especially at low IPD bandwidths.

Varying the filter bandwidth has practically only an effect on the predictions in the alternating binaural conditions (solid line, rightmost panel in Fig. 4.2), because the maximal difference between corresponding model predictions with different filter bandwidths is below 0.3 dB in all other conditions. In the alternating binaural conditions, however, the maximal difference is up to 3.6 dB (cf. Fig. 4.4). The effect of filter bandwidth becomes also more apparent, if only the rms error across the alternating

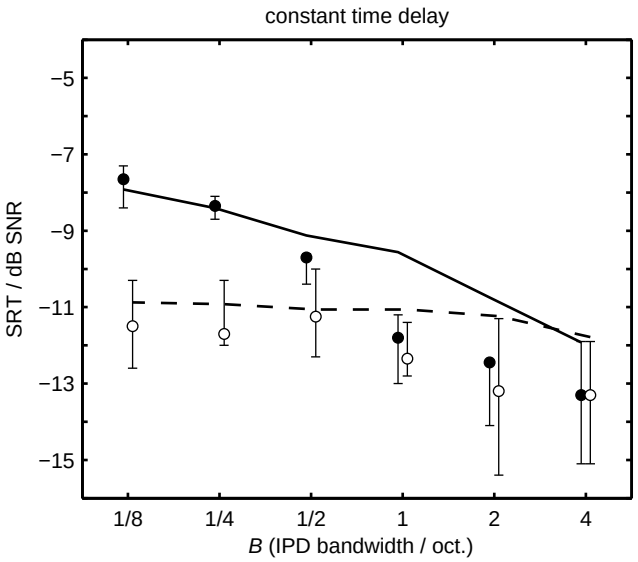


FIG. 4.5 *Model predictions (lines) with constant time delay across all frequency bands and observed SRTs of six subjects (circles, median with upper and lower quartiles) for the binaural conditions. The symbols and line styles and the observed data correspond to the rightmost panel in Fig. 4.2.*

binaural conditions is shown (cross symbols in Fig. 4.3). The predictions with forced constant time or phase delay are shown in Figures 4.5 and 4.6, respectively. Both predictions underestimate the binaural benefit at high IPD bandwidths in the alternating conditions (solid lines), the predictions with constant phase delay slightly more than the predictions with constant time delay. While the predictions of the non-alternating conditions (dashed lines) with constant phase delay differ only negligibly from the predictions with independent frequency bands, the predicted SRTs with constant time delay are about 1.7 dB higher than all other predicted SRTs in these conditions. This is especially striking at an IPD bandwidth of 4 octaves. All other predictions are not dependent of filter bandwidth and the prediction error with respect to the median observed data is very low.

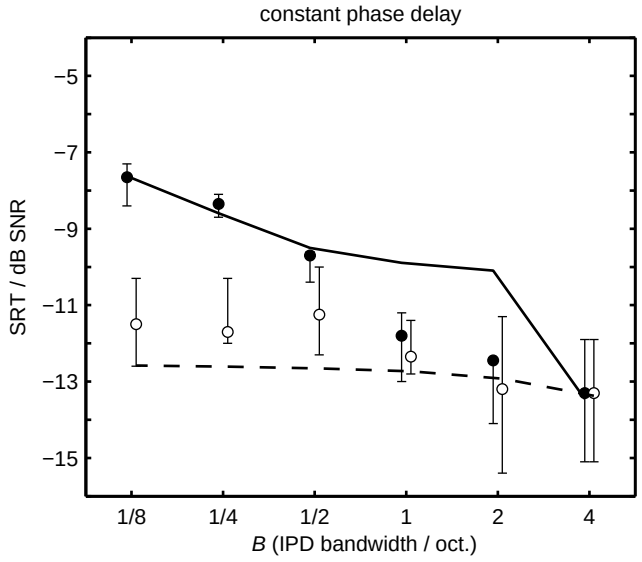


FIG. 4.6 *Model predictions (lines) with constant phase delay ($\varphi = \omega_k \tau_k = const.$, where ω_k is the center frequency and τ_k the equalization delay of the k-th band) across all frequency bands and observed SRTs of six subjects (circles, median with upper and lower quartiles) for the binaural conditions. The symbols and line styles and the observed data correspond to the rightmost panel in Fig. 4.2.*

4.4. Discussion

4.4.1. Measurement Results

The most striking result of this study is the dependence of the binaural unmasking due to differences in IPD between the speech and the noise signals on the IPD bandwidth. If the distance between adjacent sign changes in the alternating IPD function was smaller than about 0.5 octaves, no significant binaural unmasking was found, while at IPD bandwidths above this threshold, a binaural unmasking effect occurred of up to 6 dB at very large IPD bandwidths of 4 octaves. This allows for the conclusion that interaural phase differences are integrated over a certain bandwidth, but are processed independently in regions whose distance exceeds this bandwidth. The results in the monaural control conditions, which are not significantly dependent on the IPD bandwidth, show that this is a true binaural effect and not mainly caused by the monaural phase distortions.

If the IPD was consistent across frequency bands, as in the binaural non-alternating

conditions, the binaural unmasking was not strongly dependent on the IPD bandwidth. The results in the alternating and non-alternating binaural conditions were not significantly different for IPD bandwidths above 0.5 octaves, which suggests that the binaural cues can be utilized equally well in both situations, that is if their spectral distance is sufficiently large.

All subjects were able to benefit from the binaural cues in the respective situations, although they had not been specifically instructed or trained to listen to certain binaural features of the signals. Some of them even participated in binaural experiments for the first time. Nevertheless, a few subjects reported a diffuse lateralization of different spectral components of speech and noise when they were asked after completing the speech tests.

Comparing the monaural and reference conditions, there were only few conditions showing significantly different SRTs. This may be due to the relatively low number of subjects. Some trends that can already be found may become significant with a larger number of subjects. One of those trends is the overall slightly higher SRT in the monaural conditions, which is not reflected in the model predictions. It is known from literature, that even in situations without binaural differences between target speech and interfering noise there may be an SRT gain of about 1 dB from monaural to binaural presentation of the signals (Bronkhorst and Plomp, 1988). The observed difference between monaural and reference conditions may be explained by the above mentioned monaural-binaural SRT gain, even though the signals in the monaural conditions and the respective ear of the reference conditions are not the same (opposite IPD of speech and noise in the monaural conditions).

Slight differences between conditions with different IPD bandwidth (and otherwise fixed parameters) may arise from the unequal distribution of speech cues (e.g. formants)

across the spectrum and their relative position with respect to to the zero crossings of the IPD function. Strong phase distortions in the proximity of a distinct speech cue could be more disturbing than if they appear in other, less important regions. The importance weighting of the frequency regions is far less detailed in the model, so that these differences cannot be found in the model predictions.

4.4.2. Model Predictions

The general trend of the observed data, the break-down of the binaural benefit at small IPD bandwidths for the alternating binaural conditions and the roughly constant binaural benefit for the corresponding non-alternating conditions, is qualitatively predicted well at all tested filter bandwidths by the binaural speech intelligibility model. However, in order to achieve the best prediction of the exact relation between IPD bandwidth and binaural benefit, a filter bandwidth of 2.3 ERB had to be used instead of the filter bandwidth of 1 ERB in the original implementation of the model. Given the relatively large spread of individual observed SRTs, the value of 2.3 ERB may need to by adjusted, if data from more subjects is added, but the order of magnitude appears to be correct. The prediction error (Fig. 4.3) differs only slightly from the minimum within a range of about 0.5 ERB.

The predictions at the lowest and highest IPD bandwidths remain nearly the same with increasing filter bandwidths, while the slope of the SRTs as a function of IPD bandwidth shifts from lower to higher IPD bandwidths (Fig. 4.4), indicating a continuous relation between the filter bandwidth and the IPD bandwidth resolution of the model. The prediction error and the predictions themselves behave smoothly across different filter bandwidths. This allows for the interpretation that there are no artifacts because of the relative position of IPD zero-crossings and filter bands.

By far the largest prediction error occurs in the non-alternating binaural conditions. This may be attributed to the fact, that the steep zero crossings of the ideal non-alternating IPD function cannot be reproduced exactly by the measurement equipment and thus provide less useful binaural information to the listener than to the model. In the monaural and reference conditions, this is not relevant and therefore the model predictions are more accurate.

Forcing the model to use only one time delay which is constant across all frequency bands can be regarded as a case with extremely wide filters. Thus it is not surprising that the predictions with constant time delay (Fig. 4.5) are similar to the predictions with 4 ERB filter bandwidth (Fig. 4.4, rightmost line) and underestimate the binaural benefit even at higher IPD bandwidths. The predictions for an IPD bandwidth of 4 octaves are not as accurate as with the independent-band model indicating that a constant time delay across all frequency bands is not sufficient for the correct prediction of even the condition with the least variation in IPD. The predictions with constant phase delay, that is with a constant $\varphi = \omega_k \tau_k$ in each frequency band with the center frequencies ω_k, are as good as the independent-band model for the IPD bandwidth of 4 octaves and for all non-alternating binaural conditions, but they also underestimate the binaural benefit in the binaural alternating conditions for high IPD bandwidths. This is especially remarkable at an IPD bandwidth of 2 octaves, because in this case, the first zero crossing of the IPD function is at 1 kHz, and it is usually expected that the contribution of IPDs to binaural unmasking is by far more important in the frequency range below 1 kHz than above. Thus, the optimal strategy would be to choose the phase delay for equalization that yields good binaural unmasking in the low frequency range. The error made by this strategy in the high frequency range should be negligible, if the contribution of binaural unmasking due to IPD differences between speech and

noise in the frequency range above 1 kHz was small compared to the contribution at frequencies below 1 kHz. The fact that the predictions with constant phase delay and the observed data at an IPD bandwidth of 2 octaves differ significantly shows that the contribution of high frequencies has to be taken into account.

Cross-checking the model with the normal-hearing subjects' data from Beutelmann and Brand (2006) between the models with 1 ERB and 2.3 ERB results only in minor changes of correlation coefficients and rms prediction errors. The rms prediction error of the hearing-impaired subjects' data rises from 1.9 dB to 2.6 dB, mainly because of an overall prediction offset (mean difference between predicted and observed SRTs) of -2 dB. The reason for this is not clear and should be examined in further studies.

In this study, no interaural level differences (ILDs) were present in the stimuli. The relation between model filter bandwidths and IPD bandwidths is therefore purely based on the processing of IPDs and may be different for similar experiments, which employ frequency-dependent ILDs or combined IPDs and ILDs. While there is evidence for less independent processing of ILDs in adjacent frequency bands (Edmonds and Culling, 2006), the combination of ILDs and IPDs should be examined in further studies and may be crucial for the development of broad band binaural models like the binaural speech intelligibility model presented here. Related to this is the question, how the larger binaural filter bandwidths should be combined with the usual monaural bandwidths in those models. The filter bandwidth has obviously only an influence on the prediction of conditions with very extreme spectral changes in the IPD, but not on the predictions of the monaural and reference conditions. Nevertheless, it is worthwhile examining more closely if there is a need for multiple bandwidths in binaural speech intelligibility models for monaural and binaural conditions, particularly with regard to potentially different auditory bandwidths of hearing-impaired subjects. Nitschmann

and Verhey (2007) approached this issue, for example, by using the monaural filter bandwidth for signal analysis, but combining the information of the target-centered band with neighboring bands for binaural processing.

Another question concerns the difference between interaural time difference (ITD) and interaural phase difference (IPD). Would the results of this study be similar, if the frequency-dependent IPDs were replaced by frequency-dependent ITDs? This question applies to the currently ongoing discussion about the way how binaural timing disparities are represented in the brain. The assumption of the very popular and successful model by Jeffress (1948) was that ITD is coded by the activation of neurons, which are tuned to a certain best ITD due to the difference of axonal propagation time between the left and the right ear. ITDs are displayed by coincident arrival of spikes at certain neurons, each of which represents a certain ITD. Although there is anatomical evidence for this kind of structure in birds (Carr and Konishi, 1990), recent studies (David McAlpine and Palmer, 2001; McAlpine and Grothe, 2003) have cast doubt on this "delay line" hypothesis in mammals. In tone detection experiments, IPD and ITD of the target tone are virtually indistinguishable, but for the interferer (apart from sine tones used as interferers), a constant IPD leads to a frequency-dependent ITD and vice versa. If the frequency band that needs to be considered is sufficiently small, Breebaart et al. (1998) has shown that the difference between the effect of constant ITD and IPD, respectively, on binaural unmasking is rather small. For broad-band target signals as in binaural speech intelligibility experiments, however, it is certainly necessary to distinguish between IPD and ITD. Whereas the IPD is unambiguously defined as a function of frequency, the ITD as a function of frequency can be either defined as a *phase* delay, $\varphi(\omega)/\omega$, where $\varphi(\omega)$ is the IPD as a function of angular frequency ω, or as a *group* delay, $d\varphi(\omega)/d\omega$. The values of the ITD according to these

two definitions are only equal, if the ITD is constant across all considered frequencies. Phase delay generally acts on the fine structure of a signal, while group delay effects its envelope. Using a windowed sinusoid as the signal, for example, a constant phase delay shifts the zero crossings of the sinusoid without changing the window position, while a constant group delay shifts the maximum of the window. The phase delay ITDs calculated from the IPD functions used in this study do not change the functional form and the sign of the IPD functions, they are only multiplied by a factor of $1/\omega$. The group delay ITDs in the alternating conditions have different zero-crossing frequencies than the IPD functions (due to the derivative of the sin-function in Eq. (4.1)), but the general periodic form of the function is similar between IPD and group delay ITD function. Most interesting is the group delay ITD in the non-alternating conditions, because from the mathematical point of view, the group delay ITD in these conditions is still alternating between positive and negative signs across frequencies. Binaural processing exclusively based on group delay would not be expected to result in the different dependence of binaural SRTs on IPD bandwidth in the alternating and the non-alternating conditions observed in this study, because the distinction between alternating and non-alternating signs is not given in the group delay ITDs calculated from the IPD function in Eq. (4.1).

The auditory bandwidth factor of 2.3 of binaural processing relative to monaural processing estimated in this study is generally in line with other results from the literature. It matches very well the factor of about 2.5 found by Hall et al. (1983) for binaural tone detection in band-limited noise and with a spectral level of 30 dB/Hz, which is close to the average noise spectral level used in this study. Sondhi and Guttman (1966) found a factor of about 2 for the frequency band centered on 500 Hz, with a paradigm of noise bands with binaural cues closely embedded in noise bands

without binaural cues, which is similar to the paradigm used in this study. In the study of Holube et al. (1998), the similar periodic variation of binaural cues on a linear frequency scale resulted in binaural bandwidth factors of about 1.6, which is smaller than the value from this study, but would still lead to tolerable predictions with the binaural speech intelligibility model.

Exact comparisons of the binaural filter bandwidth would need to consider not only the bandwidth, but also the filter shape, as described in Kollmeier and Holube (1992). As a compromise, the -10 dB-bandwidth or even better, the bandwidth, which encompasses 90% of the integrated filter function, were suggested instead of the -3 dB-bandwidth. This is reflected in the comparison of this study and the model of Nitschmann and Verhey (2007). The latter is aimed at predicting differences between effective binaural bandwidths calculated from bandlimiting and notched-noise experiments as performed by Hall et al. (1983). The -3 dB-bandwidth is nearly the same in this study and in Nitschmann and Verhey (2007), while the -10 dB-bandwidth of the 2.3 ERB wide fourth order gammatone filters used in this study is about 30% larger than the weighted combination of three adjacent 1 ERB wide third order gammatone filters used by Nitschmann and Verhey (2007).

The consequences for binaural modelling that can be drawn from this study are (1) the hypothesis of independent processing in different auditory frequency bands cannot be rejected and (2) the binaural processing of broad-band target and interferer signals with frequency-dependent IPDs is subject to a larger auditory integration bandwidth than typically used in monaural detection models.

4.5. Conclusions

1. A periodically frequency-dependent IPD *difference* between speech and noise resulted in an SRT benefit of up to 6 dB relative to the SRT for equal IPDs of speech and noise. If the IPD difference had alternating signs in adjacent frequency bands, the SRT benefit strongly depended on the bandwidth of the IPD periods and was only significantly larger than zero for IPD bandwidths larger than a third octave. If the sign of the IPD difference was non-alternating, that is consistent across all frequencies, the SRT benefit showed only little variation and was significantly larger than zero even for IPD bandwidths below a third octave.

2. The binaural speech intelligibility model by Beutelmann and Brand (2006) predicted the binaural SRT benefit due to the IPD differences between speech and noise very well. The predictions correctly exhibit the decrease in benefit with decreasing IPD bandwidth for alternating sign of the IPD difference and the stable benefit for non-alternating sign of the IPD difference. Although the original choice of the gammatone filterbank (Hohmann, 2002) including the filter bandwidth and spacing of one ERB, respectively, already yielded a good prediction of the general trend in the data, it was possible to improve the prediction quality, in terms of root mean squared prediction error, by increasing the model filter bandwidth to 2.3 ERB.

3. The assumption of constant equalization parameters across all frequency bands is not sufficient for good predictions. Provided that the filter bandwidth is within the limits mentioned above, it appears reasonable that binaural processing in each frequency band is virtually independent of the adjacent bands (i.e., the

equalization parameters can be chosen independently). Thus, the "independent binaural processing channel" hypothesis cannot be rejected.

Acknowledgements

This study was supported by the Deutsche Forschungsgemeinschaft within the SFB TRR 31 "The active auditory system".

5. Summary and general conclusions

The primary aim of this dissertation was to develop a model of binaural speech intelligibility in complex situations, so-called "cocktail party situations" (Cherry, 1953). The complexity of these situations arises from the spatial arrangement of target speech and interferer sources, from early reflections and reverberation in rooms, and from properties of the interferers, like spectrum and modulation. An individual hearing-impairment of the listener can make these complex situations even more difficult to cope with. There are more parameters which affect speech intelligibility, for example informational masking (i.e., not exclusively attributable to physical signal parameters) or cognitive factors like linguistic complexity of the target speech or non-native language, but they have not been considered in detail in this dissertation.

The principle of the binaural speech intelligibility model, the core of the work presented here, was based on the thesis of vom Hövel (1984). The idea was to use the equalization-cancellation principle proposed by Durlach (1963) for binaural tone-in-noise unmasking to calculate the amount of binaural unmasking which is possible in the given signal configuration (in terms of signal-to-noise ratio), and to use its results from multiple frequency bands as input for the (monaural) speech intelligibility index (ANSI, 1997). The equalization-cancellation principle uses an amplitude and time delay adjustment between the left and the right ear channel with subsequent subtraction of the channels. Depending on the interaural correlation of speech and interferer

and their relative spatial location, an optimal set of equalization parameters can be found, that eliminates the maximal possible amount of the interferer by destructive interference and thus increases the signal-to-noise ratio. An essential element of the original equalization-cancellation model, the study by vom Hövel (1984), and the model developed in this dissertation is an internal binaural noise, that controls the maximal unmasking in signal configurations that would in theory allow for the complete elimination of the interferer. The parameters of this internal noise, which is realized in form of artifical inaccuracies of the equalization parameters, can be used to adjust the model to comply with human performance.

This dissertation consists of three parts, which have been published (chapter 2, Beutelmann and Brand, 2006) in or submitted for publication (chapters 3 and 4, Beutelmann et al., 2008a,b, respectively) to the Journal of the Acoustical Society of America in their present form, apart from some minor layout changes. Each part provides a different point of view on the central topic. The first part (chapter 2) deals with the basic implementation of the model and its extension to the prediction of the influence of the hearing threshold on binaural speech intelligibility. The second part (chapter 3) presents on one hand an analytically optimized version of the model, and on the other hand another extension of the model aimed at the prediction of binaural speech intelligibility in fluctuating noise. The third part (chapter 4) is concerned with the hypothesis of independent binaural processing of broadband input signals in adjacent auditory filters and the choice of parameters of the filter bank which is used to split the input signals of the model into narrow frequency bands.

In chapter 2, it was shown that a straightforward combination of a gammatone filter bank (Hohmann, 2002), an independent equalization-cancellation process (Durlach, 1963) in each frequency band, resynthesis of the frequency bands into a waveform

signal, and the speech intelligibility index (ANSI, 1997) results in good predictions of binaural SRT data with a high correlation coefficient of 0.95 between predictions and measured data. The measurement conditions included a steady-state, speech-shaped noise source at different azimuths in the horizontal plane and three room conditions (anechoic, office room with $T_{60} = 0.6\,s$, and cafeteria with $T_{60} = 1.3\,s$) and the speech source was always in front of the listener. The mean absolute prediction error for the average normal-hearing data was between 0.3 dB and 1.6 dB, depending on the room condition. It was shown, that the internal binaural errors are indeed essential for the correct prediction of binaural speech reception thresholds, which were much too low, if the internal binaural errors were omitted. Incorporating the individual hearing threshold in form of a masking noise added to the external noise signal led to almost equally good predictions of the individual observed data from hearing-impaired subjects with correlation coefficients above 0.9 and mean prediction errors of 1.7–1.9 dB, depending on the room condition.

The first part of chapter 3 presented an analytical optimization and revision of the model from chapter 2. The first model approach was a simple combination of signal processing components and included redundant calculations. While this was an easy way to start, the practical application of the model was limited because of its inefficiency. The analytical optimization removed most of the redundant calculations, provided a more efficient search procedure for the best equalization parameters, and resulted in a formal expression of the signal-to-noise ratio after the equalization-cancellation process, which emphasizes the role of the interaural level and time differences of the speech and noise signals in the process. With an additionally improved implementation, the computing time was reduced by a factor of about 60 (from 10–20 min to 10–20 s on a standard PC) while maintaining the same prediction quality as with the original model.

In the second part of this chapter, an extension of the revised model was presented, which was a first approach toward the prediction of binaural speech intelligibility in fluctuating noise. Based mainly on Rhebergen et al. (2006), the model was calculated in short-time frames and the predicted short-time SRTs were averaged to obtain the final result. Although this was rather a proof of concept than an elaborate model, it was shown that it is in principle possible to predict the effect of fluctuating noise on binaural speech intelligibility with a short-time frame model. Further possibilities of improvement are discussed below. As a side result of this chapter, it was found that strong spectral differences between the speech and noise signals may result in reduced prediction quality, but this is mainly a monaural effect attributed to the concept of the SII. This lead to a mean absolute prediction error of 3 dB for the mean normal-hearing data and 4 dB for the hearing-impaired data. Overall, the predictions of SRTs in fluctuating noise had a correlation coefficient with the observed data of 0.88 for the mean normal-hearing data and 0.72 for the individual hearing-impaired data.

While the studies in chapter 2 and 3 were concerned with implementations and extensions of the model and their evaluation with experimental data, chapter 4 was aimed at testing the so far implicit hypothesis of independent binaural processing in adjacent auditory filters as well as the question of the effective binaural auditory bandwidth. At the same time, this was a verification of the auditory filter bank parameters that are used in the model. A critical binaural speech intelligibility experiment was designed that incorporated strongly frequency-dependent interaural phase differences and the spectral distance of conflicting binaural cues was varied as a parameter. Achieving a large binaural benefit would require significantly different equalization parameters in adjacent filter bands of the model. This binaural benefit was found in the observed data from normal-hearing subjects and only a model with

independent binaural processing in adjacent filter bands was able to predict it properly. Nevertheless, increasing the filter bandwidth of the model by a factor of about 2.3 compared to the common monaural filter bandwidth of 1 ERB (Glasberg and Moore, 1990) led to the lowest prediction error.

In addition to the binaural speech intelligibility model, which forms the center of this dissertation, the individual chapters have some more in common. All measurements share a basic principle, because they were all performed using the Oldenburg Sentence Test in noise. The speech and noise signals were filtered with the appropriate binaural room impulse responses (BRIRs) or head related transfer functions (HRTFs), respectively, depending on the required condition. The anechoic HRTFs were taken from a publicly available database (Algazi et al., 2001), and the BRIRs were own measurements with a manikin (in chapter 2), or simulated in a room-acoustical software (Christensen, 2005, in chapter 3).

The benefit of this work, beyond the gain of scientific knowledge about binaural speech intelligibility in complex situations, is that the model can be used as a tool for the prediction of binaural speech intelligibility, in order to reduce the need for time-consuming and expensive subjective tests. It might be used in room acoustics, for example for the planning of auditoria or class rooms, in audiology as an estimate for the loss of speech intelligibility based on other measures and for the assessment of the expected benefit of bilateral hearing aids, and it might be used for predicting the benefit of binaural algorithms in hearing aids or audio devices.

Altogether, this dissertation project has produced a model of binaural speech intelligibility, which is on its way to practical application and is well evaluated, albeit in a limited range of conditions. Nevertheless, the chances are that conditions, which are basically the same as the conditions tested in this dissertation and only differ in

their parameters (e.g., sound source azimuth or reverberation time), are predicted equally well. Although a number of questions could be answered in this dissertation, a lot of open ones remain to be solved. They start at rather technical issues, for example the so far inevitable separation of the input signals into (useful) speech and (detrimental) noise parts, which is unfavorable for predictions after non-linear signal processing. It could be solved by a reliable SNR estimate from the combined signal. Another issue linked to this is the insufficient inclusion of the detrimental effect of strong reverberation on speech itself. Although it was not a substantial problem in the work presented here, it has to be considered in future studies. There are solutions based on the speech transmission index (van Wijngaarden and Drullman, 2008), but a combination of both approaches would need more effort. A refinement of the short-time binaural model for fluctuating noises, including frequency-dependent frame lengths and forward masking, is obvious and should be considered. In combination with this, the prediction of time-varying binaural cues could be interesting, because the fixed binaural configuration used in all experiments of this dissertation is simple, but hardly realistic. A first approach for this could be to transfer the experiment of chapter 4 from the frequency domain to the time domain, that is to generate periodically changing binaural cues over time and to vary the period length as a parameter. A future application might lie in the prediction of the speech intelligibility benefit of adaptive beam-forming algorithms. Maybe the most important question could only partly dealt with in this dissertation: the influence of hearing impairment on speech intelligibility. Although it is possible to predict the binaural speech intelligibility of hearing-impaired subjects if the noise level is low and close to the hearing threshold (as in chapter 2), it was not possible to predict the supra-threshold deficits sufficiently accurate (as can be seen in chapter 3, because the noise levels were considerably higher). These problems

concern monaural as well as binaural speech intelligibility and are an incentive for comprehensive future work.

Appendix A.

Detailed derivation of the analytical expression for the SNR after the EC process

The EC process described in Eq. (3.5) is a linear operation on the input signals. Together with Eq. (3.1) and the assumption, that the speech and external noise signals are available separately, the residual signal after the EC process

$$X_{EC}(\omega) = S_{EC}(\omega) + N_{EC}(\omega) \tag{A.1}$$

can be split up into the residual speech signal and the residual noise signal.

In order to compute the SNR that is needed for the SII (Eq. (3.7)), the overall intensity of the residual speech and noise signals has to be calculated. In the following, the derivation is only shown for the speech signal, because it is performed analogously for the noise signal. By using $|x - y|^2 = |x|^2 + |y|^2 - 2\mathrm{Re}(xy^*)$ on Eq. (3.6) inserted into the definition of the intensity (Eq. (3.8)), the absolute square in the integral can

be expanded

$$I(S_{EC}) = \int_{\Omega-\beta/2}^{\Omega+\beta/2} |S_{EC}(\omega)|^2 \, d\omega \tag{A.2}$$

$$= \int_{\Omega-\beta/2}^{\Omega+\beta/2} \left| e^{\gamma/2+\epsilon_L} e^{+i\omega(\tau/2+\delta_L)} S_L(\omega) - e^{-\gamma/2+\epsilon_R} e^{-i\omega(\tau/2+\delta_R)} S_R(\omega) \right|^2 d\omega \tag{A.3}$$

$$\begin{aligned} &= e^{\gamma+2\epsilon_L} \int_{\Omega-\beta/2}^{\Omega+\beta/2} |S_L(\omega)|^2 \, d\omega + e^{-\gamma+2\epsilon_R} \int_{\Omega-\beta/2}^{\Omega+\beta/2} |S_R(\omega)|^2 \, d\omega \\ &\quad - 2e^{\epsilon_L+\epsilon_R} \operatorname{Re}\left(\int_{\Omega-\beta/2}^{\Omega+\beta/2} S_L(\omega) S_R^*(\omega) e^{i\omega(\delta_L+\delta_R)} e^{i\omega\tau} d\omega \right) \end{aligned} \tag{A.4}$$

into three summands. The first two summands are only dependent on the overall intensity of the left and right channel, respectively, while the third summand is a cross-correlation term, which is strongly dependent on the phase information available in the signals. As described in section 3.2.1, the EC processing errors are incorporated by calculating the expectation value of the intensity with respect to processing error variables. With $\langle e^{2\epsilon} \rangle_\epsilon = e^{2\sigma_\epsilon^2}$ and $\langle e^{\epsilon} \rangle_\epsilon = e^{\sigma_\epsilon^2/2}$ for normally distributed ϵ, follows that

$$\left\langle I(S_{EC}) \right\rangle_{\epsilon_L,\epsilon_R,\delta_L,\delta_R} \tag{A.5}$$

$$= e^{2\sigma_\epsilon^2} e^{\gamma} I(S_L) + e^{2\sigma_\epsilon^2} e^{-\gamma} I(S_R) - 2e^{\sigma_\epsilon^2} \operatorname{Re}\left(\int_{\Omega-\beta/2}^{\Omega+\beta/2} S_L(\omega) S_R^*(\omega) e^{-\omega^2\sigma_\delta^2} e^{i\omega\tau} d\omega \right), \tag{A.6}$$

leading to a Gaussian low pass filter $e^{-\omega^2\sigma_\delta^2}$ on the cross-correlation term, i.e. on the phase information available as a function of frequency. The cross-correlation term can be normalized by extracting the square root of the product of both channel intensities

$$\begin{aligned}\Big\langle I(S_{EC})\Big\rangle_{\epsilon_L,\epsilon_R,\delta_L,\delta_R} \\ = 2e^{\sigma_\epsilon^2}\sqrt{I(S_L)I(S_R)}\left[e^{\sigma_\epsilon^2}\frac{1}{2}\left(e^{\gamma}\sqrt{\frac{I(S_L)}{I(S_R)}}+e^{-\gamma}\sqrt{\frac{I(S_R)}{I(S_L)}}\right)\right. \\ \left.-\mathrm{Re}\left(\frac{1}{\sqrt{I(S_L)I(S_R)}}\int_{\Omega-\beta/2}^{\Omega+\beta/2}S_L(\omega)S_R^*(\omega)e^{-\omega^2\sigma_\delta^2}e^{i\omega\tau}d\omega\right)\right],\end{aligned} \tag{A.7}$$

leaving a symmetric expression for the first two summands, that can be transformed into a cosh function

$$\begin{aligned}\Big\langle I(S_{EC})\Big\rangle_{\epsilon_L,\epsilon_R,\delta_L,\delta_R} \\ = 2e^{\sigma_\epsilon^2}\sqrt{I(S_L)I(S_R)}\left[e^{\sigma_\epsilon^2}\cosh\left(\gamma+\ln\sqrt{\frac{I(S_L)}{I(S_R)}}\right)\right. \\ \left.-\mathrm{Re}\left(\frac{1}{\sqrt{I(S_L)I(S_R)}}\int_{\Omega-\beta/2}^{\Omega+\beta/2}S_L(\omega)S_R^*(\omega)e^{-\omega^2\sigma_\delta^2}e^{i\omega\tau}d\omega\right)\right].\end{aligned} \tag{A.8}$$

The argument of the cosh is simplified with the definition of Δ_S (cf. Eq. (3.11)) for the interaural level difference of the signal. The low pass function $e^{-\omega^2\sigma_\delta^2}$ can be extracted from the cross-correlation term by using the convolution theorem of the Fourier transform,

$$\begin{aligned}\Big\langle I(S_{EC})\Big\rangle_{\epsilon_L,\epsilon_R,\delta_L,\delta_R} \\ = 2e^{\sigma_\epsilon^2}\sqrt{I(S_L)I(S_R)}\Big[e^{\sigma_\epsilon^2}\cosh\left(\gamma+\Delta_S\right) \\ -\frac{\sqrt{\pi}}{\sigma_\delta}e^{-\frac{\tau^2}{4\sigma_\delta^2}} * \mathrm{Re}\left(\frac{1}{\sqrt{I(S_L)I(S_R)}}\int_{\Omega-\beta/2}^{\Omega+\beta/2} S_L(\omega)S_R^*(\omega)e^{i\omega\tau}d\omega\right)\Big].\end{aligned} \tag{A.9}$$

The inverse Fourier transform of the cross-correlation term is then also carried out, resulting in the normalized cross-correlation function in the time domain. Because of the convention used for the normalization of the Fourier transform pair, a factor of $(2\pi)^{-1}$ arises, which is included in the definition of the low pass filter or Gaussian smoothing window $\lambda(\tau)$ (cf. Eq. (3.13))

$$\begin{aligned}\Big\langle I(S_{EC})\Big\rangle_{\epsilon_L,\epsilon_R,\delta_L,\delta_R} \\ = 2e^{\sigma_\epsilon^2}\sqrt{I(S_L)I(S_R)}\Big[e^{\sigma_\epsilon^2}\cosh\left(\gamma+\Delta_S\right) - \lambda(\tau) * \mathrm{Re}(\rho_S(\tau))\Big]\end{aligned} \tag{A.10}$$

Together with the same derivation for the noise intensity, this results in Eq. (3.9).

Bibliography

Akeroyd, M. A. (**2004**). "The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking," J. Acoust. Soc. Am. **116**, 1135–1148.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," in *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*.

ANSI (**1969**). "Methods for the calculation of the articulation index," American National Standard S3.5–1969, Standards Secretariat, Acoustical Society of America.

ANSI (**1997**). "Methods for the calculation of the speech intelligibility index," American National Standard S3.5–1997, Standards Secretariat, Acoustical Society of America.

Auditec (**2006**). "CD101RW2," Audio CD, Auditec of St. Louis, 2515 South Big Bend Blvd, St. Louis MO 63143, www.auditec.com (date last viewed 07/31/08).

Best, V., Gallun, F. J., Carlile, S., and Shinn-Cunningham, B. G. (**2007**). "Binaural interference and auditory grouping," J. Acoust. Soc. Am. **121**, 1070–1076.

Beutelmann, R. and Brand, T. (**2006**). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 331–342.

Beutelmann, R., Brand, T., and Kollmeier, B. (**2008**a). "Revision, extension, and evaluation of a binaural speech intelligibility model (BSIM)," J. Acoust. Soc. Am. submitted.

Beutelmann, R., Brand, T., and Kollmeier, B. (**2008**b). "Prediction of binaural speech intelligibility with frequency-dependent interaural phase differences," J. Acoust. Soc. Am. submitted.

Blodgett, H. C., Jeffress, L. A., and Whitworth, R. H. (**1962**). "Effect of noise at one ear on the masked threshold for tone at the other," J. Acoust. Soc. Am. **34**, 979–981.

Bradley, J. S. and Bistafa, S. R. (**2002**). "Relating speech intelligibility to useful-to-detrimental sound ratios," J. Acoust. Soc. Am. **112**, 27–29.

Brand, T. and Beutelmann, R. (**2005**). "Examination of an EC/SII based model predicting speech reception thresholds of hearing-impaired listeners in spatial noise situations," in *Proc. of the 21st Danavox Symposium "Hearing Aid Fitting"*, edited by A. N. Rasmussen, T. Poulsen, T. Andersen, J. B. Simonsen, and C. B. Larsen.

Brand, T. and Kollmeier, B. (**2002**a). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**, 2801–2810.

Brand, T. and Kollmeier, B. (**2002**b). "Vorhersage der Sprachverständlichkeit in Ruhe und Störgeräusch aufgrund des Reintonaudiogramms (prediction of speech intelligibility in quiet and in noise based on the pure tone audiogram)," Z. Audiol., Suppl. 5 .

Breebaart, J., van de Par, S., and Kohlrausch, A. (**1998**). "Binaural signal detection with phase-shifted and time-delayed noise maskers," J. Acoust. Soc. Am. **103**, 2079–2083.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**a). "Binaural processing model based on contralateral inhibition. II. dependence on spectral parameters," J. Acoust. Soc. Am. **110**, 1089–1104.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**b). "Binaural processing model based on contralateral inhibition. III. dependence on temporal parameters," J. Acoust. Soc. Am. **110**, 1105–1117.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT Press, Cambridge/ Massachusetts).

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

Bronkhorst, A. W. and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A. W. and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1374–1383.

Bronkhorst, A. W. and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Buell, T. N. and Hafter, E. R. (**1991**). "Combination of binaural information across frequency bands," J. Acoust. Soc. Am. **90**, 1894–1900.

Carr, C. E. and Konishi, M. A. (**1990**). "A circuit for detection of interaural time differences in the brainstem of the barn owl," J. Neurosci. **10**, 3227–3246.

CEN (**2000**). "Messung der Nachhallzeit von Räumen mit Hinweis auf andere akustische Parameter (Measurement of the reverberation time of rooms with reference to other acoustical parameters)," European Standard EN ISO 3382, Europäisches Komitee für Normung.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.

Christensen, C. L. (**2005**). "ODEON," Room Acoustics Modelling Software v8.0, ODEON A/S, www.odeon.dk (date last viewed 07/31/08).

Cokely, J. A. and Hall, J. W. (**1991**). "Frequency resolution for diotic and dichotic listening conditions compared using the bandlimiting measure and a modified bandlimiting measure," J. Acoust. Soc. Am. **89**, 1331–1339.

Colburn, H. S. (**1977**a). "Theory of binaural interaction based on auditory-nerve data. II. detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.

Colburn, H. S. (**1977**b). "Theory of binaural interaction based on auditory-nerve data. II. detection of tones in noise. supplementary material," AIP document no. PAPS JASMA-91-525-98 .

Colburn, H. S. (**1996**). *Computational Models of Binaural Processing* (Springer, New York), *Springer Handbook of Auditory Research*, vol. 6, chap. 8, 332–400.

Colburn, H. S. and Durlach, N. I. (**1978**). *Models of Binaural Interaction* (Academic Press), *Handbook of Perception*, vol. IV, "Hearing", chap. 11, 467–518.

Culling, J. F. and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise." J. Acoust. Soc. Am. **107**, 517–527.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.

Culling, J. F. and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds - absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.

David McAlpine, D. J. and Palmer, A. R. (**2001**). "A neural code for low-frequency sound localization in mammals," Nat. Neurosci. **4**, 396–401.

Diercks, K. J. and Jeffress, L. A. (**1962**). "Interaural phase and the absolute threshold for tone," J. Acoust. Soc. Am. **34**, 981–984.

Dreschler, W., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). "Icra noises: artificial noise signals wioth speech-like spectral and temporal properties for hearing instrument assessment," Audiology **40**, 148–157.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," J. Acoust. Soc. Am. **111**, 2897–2907.

Duquesnoy, A. J. (**1983**). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739–743.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**.

Durlach, N. I. (**1972**). *Binaural signal detection: Equalization and Cancellation Theory* (Academic Press, New York, London), vol. II, chap. 10, 371–462.

Edmonds, B. A. and Culling, J. F. (**2005**). "The spatial unmasking of speech: evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am. **117**, 3069–3078.

Edmonds, B. A. and Culling, J. F. (**2006**). "The spatial unmasking of speech: evidence for better-ear listening." J. Acoust. Soc. Am. **120**, 1539–1545.

Egan, J. P. (**1965**). "Masking-level differences as a function of interaural disparities in intensity of signal and of noise," J. Acoust. Soc. Am. **38**, 1043–1049.

Festen, J. M. (**1993**). "Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice," J. Acoust. Soc. Am. **94**, 1295–1300.

Festen, J. M. and Plomp, R. (**1986**). "Speech-reception threshold in noise with one and two hearing aids," J. Acoust. Soc. Am. **79**, 465–471.

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

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Fletcher, H. and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.

French, N. I. and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Glasberg, B. R. and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched noise data," Hear. Res. **47**, 103–138.

Grose, J. H. and Hall, J. W. (**1992**). "Comodulation masking release for speech stimuli," J. Acoust. Soc. Am. **91**, 1042–1050.

Gustafsson, H. A. . and Arlinger, S. D. (**1994**). "Masking of speech by amplitude-modulated noise," J. Acoust. Soc. Am. **95**, 518–529.

Haas, H. (**1972**). "The influence of a single echo on the audibility of speech," J. Audio Eng. Soc. **20**, 146–159.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hall, J. W., Tyler, R. S., and Fernandes, M. A. (**1983**). "Monaural and binaural auditory frequency resolution measured using bandlimited noise and notched-noise masking," J. Acoust. Soc. Am. **73**, 894–898.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a gammatone filterbank," Acust. Acta Acust. **88**, 433–442.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," J. Acoust. Soc. Am. **104**, 2412–2425.

Holube, I. and Kollmeier, B. (**1996**). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," J. Acoust. Soc. Am. **100**, 1703–1716.

Houtgast, T. (**1977**). "Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker." J. Acoust. Soc. Am. **62**, 409–415.

Houtgast, T. and Steeneken, H. J. M. (**1973**). "The modulation transfer function in room acoustics as a predictor of speech intelligibility," Acustica **28**, 66–73.

IEC (**1985**). "Sound systems equipment, listening tests on loudspeakers," International Standard 268-13, International Electrotechnical Commission.

IEC (**1998**). "Sound system equipment - part 16: Objective rating of speech intelligibility by speech transmission index," International Standard IEC 60268-16 (1998), International Electrotechnical Commission.

Irwin, R. J. and McAuley, S. F. (**1987**). "Relations among temporal acuity, hearing loss, and the perception of speech distorted by noise and reverberation," J. Acoust. Soc. Am. **81**, 1557–1565.

Jeffress, L. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol **41**, 35–39.

Kohlrausch, A. (**1988**). "Auditory filter shape derived from binaural masking experiments," J. Acoust. Soc. Am. **84**, 573–583.

Kohlrausch, A. (**1990**). "Binaural masking experiments using noise maskers with frequency-dependent interaural phase differences. II: Influence of frequency and interaural-phase uncertainty," J. Acoust. Soc. Am. **88**, 1749–1756.

Kollmeier, B. and Holube, I. (**1992**). "Auditory filter bandwidths in binaural and monaural listening conditions," J. Acoust. Soc. Am. **92**, 1889–1901.

Kryter, K. D. (**1962**). "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34**, 1689–1697.

Langford, T. L. and Jeffress, L. A. (**1964**). "Effect of noise crosscorrelation on binaural signal detection," J. Acoust. Soc. Am. **36**, 1455–1458.

Lavandier, M. and Culling, J. F. (**2007**). "Speech segregation in rooms: Effects of reverberation on both target and interferer," J. Acoust. Soc. Am. **122**, 1713–1723.

Levitt, H. and Rabiner, L. R. (**1967**). "Predicting binaural gain in intelligibility and release from masking for speech," J. Acoust. Soc. Am. **42**, 820–828.

Lindemann, W. (**1986**). "Extension of a binaural cross-correlation model by contrallateral inhibition. I. simulation of lateralization for stationary signals," J. Acoust. Soc. Am. **80**, 1608–1622.

MathWorks (**2002**). "MATLAB® 6.5," .

McAlpine, D. and Grothe, B. (**2003**). "Sound localization and delay lines - do mammals fit the model?" Trends Neurosci. **26**, 347–350.

Mesgarani, N., Grant, K. W., Shamma, S., and Duraiswami, R. (**2003**). "Augmented intelligibility in simultaneous multi-talker environments," in *Proceedings of the 2003 International Conference on Auditory Display* (International Community for Auditory Display, Boston, MA, USA), 71–74.

Metz, P. J., von Bismarck, G., and Durlach, N. I. (**1968**). "Further results on binaural unmasking and the EC model. II. noise bandwidth and interaural phase," J. Acoust. Soc. Am. **43**, 1085–1091.

Miller, G. A. and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.

Moncur, J. P. and Dirks, D. (**1967**). "Binaural and monaural speech intelligibility in reverberation," J. Speech Hear. Res. **10**, 186–195.

Müsch, H. and Buus, S. (**2001**). "Using statistical decision theory to predict speech intelligibility. II. measurement and prediction of consonant-discrimination performance," J. Acoust. Soc. Am. **109**, 2910–2920.

Müsch, H. and Buus, S. (**2004**). "Using statistical decision theory to predict speech intelligibility. III. effect of audibility on speech recognition sensitivity," J. Acoust. Soc. Am. **116**, 2223–2233.

Müsch, H. (**2003**). "MATLAB® implementation of core aspects of the american national standard "Methods for calculation of the speech intelligibility index" ANSI S3.5-1997," Downloadable MATLAB® Script: http://www.sii.to/html/programs.html (date last viewed 07/31/08).

Müsch, H. and Buus, S. (**2001**a). "Using statistical decision theory to predict speech intelligibility. I. model structure," J. Acoust. Soc. Am. **109**, 2896–2909.

Müsch, H. and Buus, S. (**2001**b). "Using statistical decision theory to predict speech intelligibility. II. measurement and prediction of consonant-discrimination performance," J. Acoust. Soc. Am. **109**, 2910–2920.

Nábělek, A. K. and Pickett, J. M. (**1974**). "Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids," J. Acoust. Soc. Am. **56**, 628–638.

Nitschmann, M. and Verhey, J. L. (**2007**). "Experimente und Modellrechnungen zur binauralen spektralen Selektivität (Experiments and model calculations on binaural spectral selectivity)," in *Fortschritte der Akustik, DAGA 2007* (Deutsche Gesellschaft für Akustik e.V. (DEGA), Berlin), 371–372.

Osman, E. (**1971**). "A correlation model of binaural masking level differences," J. Acoust. Soc. Am. **6**, 1494–1511.

Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.

Pavlovic, C. V. (**1984**). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**, 1253–1258.

Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). "An articulation index based procedure for predicting the speech recognition performance for hearing-impaired individuals," J. Acoust. Soc. Am. **80**, 50–57.

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

Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.

Platte, H.-J. and vom Hövel, H. (**1980**). "Zur Deutung der Ergebnisse von Sprachverständlichkeitsmessungen mit Störschall im Freifeld (On the interpretation of speech intelligibility measurement results in noise in free-field conditions)," Acustica **45**, 139–150.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Plomp, R. and Mimpen, A. M. (**1979**). "Speech-reception threshold for sentences as a function of age and noise level," J. Acoust. Soc. Am. **66**, 1333–1342.

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

Rhebergen, K. S. and 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," J. Acoust. Soc. Am. **117**, 2181–2192.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2005**). "Release from informational masking by time reversal of native and non-native interfering speech," J. Acoust. Soc. Am. **118**, 1274–1277.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," J. Acoust. Soc. Am. **120**, 3988–3997.

Sever, J. C. and Small, A. M. (**1979**). "Binaural critical masking bands," J. Acoust. Soc. Am. **66**, 1343–1350.

Sieben, U. (**1979**). "Binaraurale Mithörschwellen bei aus Tönen und Rauschen zusammengesetzten Maskierern (Binaural masking thresholds with maskers composed of tones and noise)," Diploma thesis, Drittes Physikalischen Institut, Georg-August-Universität zu Göttingen.

Smits, C. and Houtgast, T. (**2007**). "Recognition of digits in different types of noise by normal-hearing and hearing-impaired listeners," Int. J. Audiol. **46**, 134–144.

Smoorenburg, G. F. (**1992**). "Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram," J. Acoust. Soc. Am. **91**, 421–437.

Sondhi, M. M. and Guttman, N. (**1966**). "Width of the spectrum effective in the binaural release of masking," J. Acoust. Soc. Am. **40**, 600–606.

van Bergeijk, W. A. (**1962**). "Variation on a theme of Békésy: A model of binaural interaction," J. Acoust. Soc. Am. **34**, 1431–1437.

van Wijngaarden, S. J. and Drullman, R. (**2008**). "Binaural intelligibility prediction based on the speech transmission index," The Journal of the Acoustical Society of America **123**, 4514–4523.

Versfeld, N. J. and Dreschler, W. A. (**2002**). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," J. Acoust. Soc. Am. **111**, 401–408.

vom Hövel, H. (**1984**). "Zur Bedeutung der Übertragungseigenschaften des Außenohrs sowie des binauralen Hörsystems bei gestörter Sprachübertragung (On the importance of the transmission properties of the outer ear and the binaural auditory system in disturbed speech transmisson)," Ph.D. thesis, Fakultät für Elektrotechnik, RTWH Aachen.

von Békésy, G. (**1930**). "Zur Theorie des Hörens," Phys. Z. **31**, 857–868.

Wagener, K. (**2003**). "Factors influencing sentence intelligibility in noise," Ph.D. thesis, Carl-von-Ossietzky-Universität Oldenburg.

Wagener, K., Brand, T., Kühnel, V., and Kollmeier, B. (**1999**a). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a sentence test for the german language I: Design of the oldenburg sentence test)," Z. Audiol. **38**, 4–14.

Wagener, K., Brand, T., Kühnel, V., and Kollmeier, B. (**1999**b). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests (Development and evaluation of a sentence test for the german language II: Optimization of the oldenburg sentence test)," Z. Audiol. **38**, 44–56.

Wagener, K., Brand, T., Kühnel, V., and Kollmeier, B. (**1999**c). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a sentence test for the german language III: Evaluation of the oldenburg sentence test)," Z. Audiol. **38**, 86–95.

Wagener, K. C. and Brand, T. (**2006**). "The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners," Int. J. Audiol. **45**, 26–33.

Wightman, F. L. (**1971**). "Detection of binaural tones as a function of masker bandwidth," J. Acoust. Soc. Am. **50**, 623–636.

Zerbs, C. (**2000**). "Modelling the effective binaural signal processing in the auditory system," Ph.D. thesis, Carl-von-Ossietzky-Universität Oldenburg.

Zurek, P. M. (**1990**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hockberg (Allyn and Bacon, Boston), chap. 15, 255–276, 2. ed.

Zwicker, E. and Fastl, H. (**1999**). *Psychoacoustics - Facts and Models* (Springer, Berlin), 2nd ed.

Danksagung

Ich danke meinem Erstreferenten Prof. Dr. Dr. Birger Kollmeier für die Möglichkeit, diese Dissertation in seiner Arbeitsgruppe anzufertigen und für die vielfältige Unterstützung auf dem Weg dahin. Das schließt insbesondere die außerordentlich guten Arbeitsbedingungen ein, und die häufigen Gelegenheiten, meine Arbeiten vor Kollegen und Fachpublikum zu präsentieren. Außerdem war er immer bereit, Ideen und Ergebnisse durch konstruktive Kritik auf feste Füße und in einen größeren Rahmen zu stellen.

Prof. Dr. Volker Mellert möchte ich danken, dass er freundlicherweise das Korreferat dieser Dissertation übernommen hat, sowie Prof. Dr. Georg M. Klump für die Übernahme des Vorsitzes der Prüfungskommision.

Dr. Thomas Brand gebührt ganz besonderer Dank, denn er hat mich schon seit der Diplomarbeit mit Rat und Tat begeleitet und stand praktisch jederzeit als Ansprech- und Diskussionspartner zur Verfügung. Sein scharfer Blick auf die Modellvorhersagen hat manchmal ohne Blick in den Quellcode Fehler im Programm entlarvt. Insbesondere sein unermüdlicher Einsatz immer wenn es darum ging, Publikationen (also auch diese Dissertation) zu schreiben, war mir eine große Hilfe und hat wesentlich zu ihrer Qualität beigetragen. Herzlichen Dank, Tom!

Auch den Mitgliedern des “SprAud”-Bläschens möchte ich danken für die Gelegenheiten, in kleiner und entspannter Runde alltägliche und machmal auch nicht so alltägliche wissenschaftlich-technische Probleme zu wälzen und Daten zu diskutieren. Dass es dabei nicht immer bitter ernst zuging hat die Arbeit durchaus erleichtert.

Die gute Zusammenarbeit mit Iris Arweiler, Prof. Dr. Torben Poulsen und Dr. Rainer Huber im HearCom-Projekt, in dessen Rahmen Teile dieser Dissertation entstanden sind, weiß ich sehr zu schätzen und bin dafür sehr dankbar.

In die Geheimnisse der Oldenburger Messsoftware und ihre bislang unentdeckten Features hat mich Dr. Daniel Berg eingeführt. Ohne seine Beratung und Hilfe wären die meisten “Spezial”-Messungen und damit viele Ergebnisse dieser Arbeit nicht möglich gewesen. Vielen Dank!

Ebenso wäre diese Dissertation sehr arm an Daten, wenn nicht eine Unzahl von geduldigen Versuchspersonen stundenlang spannende OlSa-Sätze im Rauschen aus allen möglichen Raumrichtungen gehört und zu verstehen versucht hätten. Dankeschön! Dr. Kirsten Wagener, Dr. Birgitta Gabriel und Prof. Dr. Jürgen Kießling danke ich für die Organisation eines Teils der Messungen und den fleißigen und freundlichen wissenschaftlichen Hilfskräften für die Durchführung.

Ein ganz dickes Dankeschön möchte ich auch Susanne Garre, Ingrid Wusowski, Anita Gorges und Frank Grunau aussprechen, ohne die vieles nicht rund laufen würde und die immer ein offenes Ohr und praktische Problemlösungen zur Hand hatten.

Und auch allen namentlich nicht genannten Mitgliedern der Arbeitsgruppe Medizinische Physik und ihrem gesamten Umfeld sowie den Mitgliedern des Graduiertenkollegs

"Neurosensorik" möchte ich danken, nicht nur für ein wissenschaftlich fruchtbares Umfeld, sondern auch für das angenehme Miteinander, konstruktive Diskussionen, ideenreiche Kaffepausen und nicht zuletzt für viele schöne nicht-wissenschaftliche gemeinsame Aktivitäten. Es hat mir viel Spaß gemacht!

Meine Eltern und meine Schwester möchte ich an dieser Stelle kräftig drücken und ihnen dafür danken, dass sie immer an mich geglaubt haben, immer für mich da waren (auch aus räumlicher Entfernung!) und mich mit allen idellen und materiellen Mitteln unterstützt haben. Ohne euch wäre das alles nicht möglich gewesen!

Lebenslauf

Rainer Beutelmann

geboren am 23. Juni 1977
in Mainz

Staatsangehörigkeit: deutsch

seit Oktober 2003	Universität Oldenburg, Promotion in der Arbeitsgruppe “Medizinische Physik”. Stipendiat des Graduiertenkollegs “Neurosensorik” bis Juli 2005, danach assoziiertes Mitglied
September 2003	Diplom, Titel der Diplomarbeit: “Sprachverständlichkeit in räumlichen Störgeräuschsituationen”
Oktober 1999 bis September 2003	Universität Oldenburg, Hauptstudium Physik, Nebenfach Informatik
Oktober 1997 bis September 1999	Universität Marburg, Grundstudium Physik, Nebenfach Informatik, Vordiplom
September 1996 bis September 1997	Zivildienst
Juni 1996	Abitur am Staatlichen Gymnasium Nieder-Olm

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig verfasst und nur die angegebenen Hilfsmittel verwendet habe. Die Dissertation hat weder in Teilen noch in ihrer Gesamtheit einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorgelegen. Teile der Dissertation wurden bereits veröffentlicht bzw. sind zur Veröffentlichung eingereicht, wie an den entsprechenden Stellen angegeben. Der Anteil der Koautoren an den Veröffentlichungen bestand in der Betreuung der Arbeit und Korrektur der Manuskripte. Planung, Vorbereitung und Durchführung der Experimente sowie die Entwicklung der Modellskripte lagen in meiner Hand. Die Bedienung der Messapparaturen und Betreuung der Versuchspersonen wurde teilweise von wissenschaftlichen Hilfskräften unter meiner Überwachung übernommen.

Oldenburg, den 6. Dezember 2009

.....
Rainer Beutelmann