Two-channel noise reduction algorithms motivated by models of binaural interaction

Vom Fachbereich Physik der Universität Oldenburg
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation.

Thomas Wittkop
geb. am 9. September 1968
in Hamburg

Erstreferent: Prof. Dr. Dr. Birger Kollmeier
Korreferent: Prof. Dr. Volker Mellert
Tag der Disputation: 12. März 2001

Two-channel noise reduction algorithms motivated by models of binaural interaction

Vom Fachbereich Physik der Universität Oldenburg
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation.

Thomas Wittkop
geb. am 9. September 1968
in Hamburg

Erstreferent: Prof. Dr. Dr. Birger Kollmeier
Korreferent: Prof. Dr. Volker Mellert
Tag der Disputation: 12. März 2001

Abstract

In this thesis, signal processing strategies for the reduction of undesired interfering signals in binaurally recorded signals are derived und described. The properties of the different processing strategies are discussed and the processing performance is investigated using artificial signals. Two hearing aid algorithms are described that combine different noise reduction strategies and provide a complete processing of a simulated, digital hearing aid. Furthermore, a method is described and applied that allows for the optimization of "critical" processing parameters with respect to the subjectively perceived signal quality. Finally, particular audiological properties of the algorithms are investigated and compared, i.e., the influence of the processing on signal quality and on speech intelligibility in noise is measured with hearing impaired subjects.

In chapter 2, several strategies and algorithms for binaural noise reduction that have been described in the past in the literature are reviewed. Additionally, methods and hardware setups for the evaluation of such algorithms either using off line or real time processing are described.

In chapter 3, an algorithm is described that employs two different noise reduction strategies, i.e. a dereverberation technique and a suppression of lateral sound sources in a fixed combination of both strategies. Parameters of the processing are investigated and optimized with respect to signal quality for different acoustical conditions.

A measure of the "complexity" of the actual acoustical situation, i.e. the diffusiveness of the sound field is introduced and described in chapter 4. This measure allows for a continuous rating of the acoustical situation within a binaural hearing aid algorithm in order to automatically adapt the processing to the respective situation. This may be realized by an automatic selection of appropriate processing strategies or an optimization of processing parameters depending on the actual situation.

The algorithm described in chapter 3 and the measure of the diffusiveness of the actual acoustical situation introduced in chapter 4 are the basis for the development of a new, strategy-selective algorithm in chapter 5. This algorithm combines three different noise reduction strategies which are either based on existing processing techniques or have been theoretically derived for particular acoustical situations. The application of two of the processing strategies is depending on the actual acoustical situation which is rated using the measure of the diffusiveness. All strategies are described and evaluated using artificial signals. Particular processing parameters are optimized with respect to the subjectively perceived signal quality.

In chapter 6, the strategy-selective algorithm introduced in chapter 5 is evaluated and compared to the algorithm with the fixed processing described in chapter 3. The evaluation includes subjective preference judgements and speech intelligibility measurements with hearing impaired subjects. The strategy-selective algorithm is shown to be superior or at least comparable to the other algorithm in all investigated situations. The strategy-selective algorithm is found to improve the signal quality in the situation with diffuse background noise. It is also found that the algorithm is able to improve speech intelligibility under certain conditions, although no significant improvement of the speech reception threshold was found in the free-field listening conditions. The effect of the processing, however, appears to depend on the type of the hearing loss. Taken together, the strategy-selective noise reduction algorithm and the parameter optimization procedures employed in this work are promising candidates for the development of futural "intelligent" hearing aids.

Kurzfassung

In dieser Dissertation werden Signalverarbeitungsstrategien zur Reduktion von Störgeräuschen in binaural aufgenommenen Signalen hergeleitet und beschrieben. Die Eigenschaften der verschiedenen Verarbeitungsstrategien werden diskutiert und die Verarbeitungsleistung wird mit künstlichen Signalen untersucht. Zwei Algorithmen für Hörgeräte werden beschrieben, die jeweils verschiedene Strategien miteinander zu einer kompletten Verarbeitung in einem simulierten, digitalen Hörgerät kombinieren. Weiterhin wird ein Verfahren zur Optimierung "kritischer" Parameter der Verarbeitung bezüglich der subjektiv wahrgenommenen Signalqualität beschrieben und angewandt. Abschließend werden bestimmte audiologischen Eigenschaften der Algorithmen untersucht und verglichen, d.h. der Einfluss der Verarbeitung auf Signalqualität und Sprachverständlichkeit im Störgeräusch wird mit schwerhörenden Probanden gemessen.

Kapitel 2 gibt eine Übersicht über einige Strategien und Algorithmen zur binauralen Störgeräuschreduktion, die in der Vergangenheit in der Literatur beschrieben wurden. Außerdem werden Methoden und Versuchsaufbauten zur Evaluation solcher Algorithmen mittels Signalvorverarbeitung oder Echtzeitsignalverarbeitung beschrieben.

In Kapitel 3 wird ein Algorithmus beschrieben, der zwei Strategien zur Störgeräuschreduktion, d.h. Enthallung und Seitenschallsuppression in einer festen Kombination verwendet. Verarbeitungsparameter werden untersucht und bezüglich der subjektiv wahrgenommenen Signalqualität in verschiedenen akustischen Situationen optimiert.

Ein Maß für die "Komplexität" der aktuellen akustischen Situation, d.h. die Diffusität des Schallfeldes wird in Kapitel 4 entwickelt und beschrieben. Dieses Maß erlaubt eine kontinuierliche Beurteilung der akustischen Situation in einem binauralen Hörgerätealgorithmus, um die Verarbeitung automatisch an die jeweilige Situation anpassen zu können. Dies kann durch eine Auswahl geeigneter Verarbeitungsstrategien erfolgen oder durch die Anpassung von Parametern an die jeweilige Situation.

Der Algorithmus aus Kapitel 3 und das Maß der Diffusität aus Kapitel 4 bilden die Grundlage für die Entwicklung eines neuen, Strategie-selektiven Algorithmus in Kapitel 5. Dieser Algorithmus kombiniert drei verschiedene Verarbeitungsstrategien, die entweder auf bereits existierenden Strategien beruhen oder theoretisch für bestimmte akustische Situationen hergeleitet werden. Zwei dieser Verarbeitungsstrategien werden abhängig von der aktuellen akustischen Situation ein- bzw. ausgeschaltet. Alle verwendeten Verarbeitungsstrategien werden beschrieben und mit künstlichen Signalen evaluiert. Einzelne Parameter der Verarbeitung werden bezüglich der subjektiv wahrgenommenen Signalqualität optimiert.

In Kapitel 6 wird der Strategie-selektive Algorithmus aus Kapitel 5 weiter evaluiert und mit dem Algorithmus mit der festen Verarbeitung aus Kapitel 3 verglichen. Die Evaluation umfaßt subjektive Präferenzurteile und Sprachverständlichkeitsmessungen mit Schwerhörenden. Der Strategie-selektive Algorithmus erzielt dabei in allen untersuchten Situationen bessere oder zumindest gleichwertige Ergebnisse als der andere Algorithmus. Die Signalqualität wird z.B. in der Cafteria-Situation mit diffusem Störschall durch die Verarbeitung verbessert. Außerdem kann der Algorithmus unter bestimmten Bedingungen die Sprachverständlichkeit verbessern, allerdings kann keine signifikante Verbesserung des SRT unter Freifeld-Bedingungen nachgewiesen werden. Die Ergebnisse scheinen jedoch durch die Art des Hörverlustes beinflusst zu werden. Insgesamt erscheinen der Strategie-selektive Algorithmus und die verwendeten Methoden zur Parameteroptimierung vielversprechend für die Entwicklung zukünftiger, "intelligenter" Hörgeräte.

Contents

Chapter 1

General Introduction

Hearing impairment is clinically often classified by the pure tone hearing threshold and the speech-reception threshold (SRT) in quiet. These measures are used to determine the medical indication of a hearing aid provision (cf. Kießling, 1997). However, it is well known that the ability to understand speech in noise, i.e., in the presence of background noise or interfering speakers, is considerably affected in the hearing impaired (cf. Plomp, 1978). It is also well known that binaural hearing plays an important role for understanding speech in noise, especially when speech and noise are spatially separated (cf. Bronkhorst and Plomp, 1989; Peissig and Kollmeier, 1997).

Conventional hearing aids include an amplification stage to compensate for the shift of hearing threshold and optionally dynamic compression to compensate for a reduced dynamic range in one or more frequency channels. They provide almost complete restoration of speech intelligibility in quiet to the level of normal hearing, but they are not able to restore speech intelligibility in noise (cf. Plomp, 1978; Marzinzik and Kollmeier, 1999). This can be explained by the fact that these hearing aids amplify speech as well as noise and thus do not compensate for any kind of distortion process due to the hearing loss. Furthermore, monaural and binaural hearing is affected by the level and phase transfer characteristics of the hearing instruments, by distortions due to technical restrictions of the devices and by occlusion effects due to the ear molds. These effects rather decrease speech intelligibility than improve it.

To overcome this problem, various noise reduction strategies for monaural, i.e. single hearing aids have been developed. They comprise simple strategies like generally attenuating low frequencies (Kates, 1986) as well as sophisticated techniques like voice separation (Parsons, 1976; Stubbs and Summerfield, 1991), spectral subtraction (Boll, 1979; Ephraim and Malah, 1985; Cappé, 1994) or the so-called ZETA Noise Blocker (Graupe *et al.*, 1984), cf. overviews given by Lim (1983) and Boll (1991). However, although some techniques are reported to improve speech intelligibility in noise under certain conditions (e.g., Graupe *et al.*, 1987), monaural noise reduction strategies are mainly reported to rather improve subjective speech quality than speech intelligibility under realistic acoustical noise conditions (cf. Humes *et al.*, 1997; Stubbs and Summerfield, 1988; Elberling *et al.*, 1993; Marzinzik and Kollmeier, 1999).

As an alternative, directional microphones are often used which statically attenuate signals emitted by lateral or backward sound sources and which are able to improve speech intelligibility while maintaining a good signal quality (Nielsen and Ludvigsen, 1978). Since directional microphones usually employ two closely positioned omnidirectional microphones or a single diaphragm with some acoustic delay, their useable effect is limited by the physical size of the directional microphone, and the obtained directionality is frequency dependent. Because of the good practical results obtained with directional microphones, they should be used in hearing aids whenever possible and in addition to any kind of noise reduction processing.

More information about the signal is available by employing at least two microphones with a certain distance between them, such as, e.g., a binaural hearing aid supply with a central processor for both microphone inputs. In general, there are two different approaches for binaural noise reduction. The first approach employs a two-microphone input and delivers a single output. Various studies consider the adaptive filtering of the input signals, the so-called "adaptive beamformer" (e.g., Strube, 1981; Griffiths and Jim, 1982; Peterson *et al.*, 1987; Greenberg and Zurek, 1992; Kompis and Dillier, 1994; Berghe and Wouters, 1998). A considerable improvement in signal-to-noise ratio (SNR) can be obtained with this adaptive filtering in a variety of laboratory conditions. However, the improvement usually drastically decreases with an increasing number of interfering sound sources and reverberation (cf. Lu and Clarkson, 1993). Additionally, if an adaptive filter is designed to perform well at low signal-to-noise ratios (below 0 dB), its performance often is unsatisfactory at high signal-to-noise ratios (above 0 dB). The influence of the head related transfer functions (HRTFs) on the signals recorded with hearing instruments also rather deteriorates the performance of an adaptive beamformer in comparison to the free-space condition. In more recent studies, adaptive beamformer techniques providing binaural output have been introduced (cf. Asano *et al.*, 1996; Desloge *et al.*, 1997; Welker *et al.*, 1997; Suzuki *et al.*, 1999). These techniques already belong to the second binaural approach to noise reduction that delivers a binaural output in order to provide binaural hearing aid supply. A different concept belonging to this second approach is the directional filter based on the evaluation of interaural differences in level and phase as introduced by Gaik and Lindemann (1986). Peissig (1993) described algorithms for binaural hearing aids based on this concept. These algorithms require reference values for interaural level and phase differences for particular sound incidence directions, which, however, can be easily obtained. Similar algorithms are also used in signal processors for cochlear implants (Goldsworthy, 1998). Bodden (1993) described a so-called "cocktail-party processor" for binaural noise reduction, based on the interaural cross-correlation and contralateral inhibition model for human sound source localisation introduced by Lindemann (1986a, 1986b). Furthermore, Kollmeier and Koch (1994) introduced an algorithm which evaluates binaural differences in the modulation frequency domain instead of employing the time or frequency domain. All the latter algorithms including some kind of directional filtering based on binaural parameters have been reported to improve speech intelligibility under certain conditions, although the effect also decreases with increasing number of competing sound sources and reverberation.

Another approach for noise reduction is the use of microphone arrays with more than two microphones combined with a beamforming processing to produce a directivity with respect to a desired incidence direction (Soede *et al.*, 1993; Hoffman *et al.*, 1994; cf. an overview by Zurek *et al.*, 1996). The microphones are usually mounted on glasses or other devices fixed to the head. Multi-microphone arrays are reported to produce considerable directivity, but the size and the required additional devices like glasses or other head-worn devices make an application as every-day-life hearing aids in general very difficult.

Among the various types of noise reduction strategies, the binaural approach with binaural output signals seems to be not only recommended from the audiologic point of view (in case of a binaural hearing loss), but also very promising with respect to expectable benefits in speech intelligibility and signal quality in noise. Binaural input signals are at least "potentially" available for any kind of binaural hearing aid supply, independent of the type of hearing aid (Behind-The-Ear = BTE, In-The-Ear = ITE or Complete-In-the Canal = CIC) and the particular geometry (although ITE or CIC devices seem to be recommended for strategies employing the individual HRTFs). No additional devices are necessary to carry microphone arrays. However, the connection between the two devices located at both ears is still an unsolved problem. Wire connections are usually used in laboratory systems or prototype wearable devices (cf. e.g. Rass and Steeger, 1999; Wittkop *et al.*, 1997; Gingsjö, 1996; Sone *et al.*, 1995; Gelnett *et al.*, 1995; Terry *et al.*, 1994), but they are not suitable for commercial products. Only the development of an appropriate wireless connection will allow for the application of binaural signal processing strategies in commercial hearing aids. In consideration of the progress of mobile communication devices in the last few years, however, it seems quite possible that this connection will be available in the future. And since understanding speech in noise is still a severe problem with present hearing instruments, the investigation of binaural noise reduction strategies seems also worth the effort, even though the connection to the central processor is not yet applicable for every-day-life products.

This study describes the development and evaluation of several signal processing techniques for the use in noise reduction algorithms for binaural hearing aids. These techniques aim at reducing the noisy part of the mixture of ambient noise and target signal in the binaural microphone signals in order to restore the undegraded target signal and thus increase speech intelligibility under noisy conditions. However, each of the noise reduction technique makes one or more assumptions about the statistical properties or the spatial configuration of the interfering noise and the target signal. For instance, one technique assumes that there is only a single interfering sound source present. One problem is that the actual situation the noise reduction technique is applied to has to meet the assumptions, otherwise the signal processing might not yield a benefit for the hearing aid user or even degrades the signal quality by introducing audible processing artefacts. A new measure of the overall diffusiveness of the acoustical situation is therefore introduced which is employed to control and switch off signal processing techniques if the underlying acoustical situation is assumed not to be suitable for the respective technique, i.e., if the processing is assumed to rather decrease signal quality than to yield any benefit.

Chapter 2 of this thesis gives a brief overview of the work done in the medical physics group over the past years concering binaural noise reduction algorithms for hearing aids. The stationary and wearable signal processing devices employed for the development and evaluation of these algorithms are also described. Furthermore, two different algorithms are compared with respect to their effect on the signal-to-noise ratio for different spatial noise conditions.

Chapter 3 deals with the optimisation of the most recent implementation of the dereverberation and direction filtering algorithm which already has been briefly described in Chapter 2. A method is described to systematically vary processing parameters with the possibility of a comprehensive statistical evaluation of the results and their consistency. Selected parameters of the algorithm are then systematically varied and their influence on the subjectively perceived signal quality is investigated in detail in different spatial noise conditions with normal hearing listeners in order to find appropriate values for different acoustical conditions.

In Chapter 4, a new measure for the diffusiveness of the acoustical situation is described and evaluated. This measure is a monotonous function of parameters of the acoustical situation like the number of present interfering sound sources and the amount of reverberation. This new measure allows decision units in hearing aid algorithms to continuously estimate or rank the complexity of the acoustical situation and to control the influence of different noise reduction techniques on the processing depending on this estimate.

In Chapter 5, a new, strategy-selective algorithm for binaural noise reduction is described. This algorithm combines three different signal processing techniques for noise reduction. Furthermore, the measure introduced in Chapter 4 is employed to control the processing by selecting, i.e. switching on or off particular processing techniques, depending on the current acoustical situation. The effect of the processing techniques is evaluated technically and with respect to the sound quality perceived by normal hearing listeners under appropriate noise conditions.

In Chapter 6, the strategy-selective algorithm described in Chapter 5 is investigated and evaluated with respect to sound quality and speech intelligibility for hearing impaired listeners. The two different algorithms described in this work will be compared in three different acoustical situations with different interfering signals. The results demonstrate that the new, strategy-selective algorithm is superior to the other algorithm in all situations and at least equal to or even better than the unprocessed condition in the investigated situations with hearing impaired listeners. The new algorithm can thus be shown to be not only a promising further development of former algorithms, but also successfully employing an automatic processing strategy selection for different acoustical situations.

Chapter 2

Noise reduction motivated by models of binaural interaction[1]

Abstract

Several signal processing techniques are reviewed that aim at reducing ambient noise and enhancing the "desired" speech signal in complex acoustical environments ("cocktail-party processing"). These algorithms are motivated by models of binaural interaction in the normal human auditory system and try to simulate several different aspects of normal auditory function that are typically impaired in hearing-impaired listeners. All algorithms assume input signals from microphones located near the ears of a subject and one or two output signals to be presented. The first class of algorithms performs a directional filtering with respect to the forward direction and a reduction of the perceived reverberation. The second class of algorithms performs an analysis in the modulation frequency domain and combines binaural cues with cues from modulation frequency analysis to perform a noise-robust directional filtering. The third class of algorithms simulates a localization process in a way comparable to neurophysiological findings in the barn owl, while the fourth class of algorithms combines cues from binaural interaction and fundamental frequency analysis. The respective psychoacoustical and physiological motivation of these algorithms as well as their advantages and shortcomings are outlined. In addition, the hardware and software required for implementing and testing these algorithms in real-time are introduced and discussed. Since most of these algorithms are shown to provide significant benefit by increasing the "effective" signal-to-noise ratio in different acoustical situations, a combination of these algorithms appears promising for future "intelligent" digital hearing aids.

2.1 Introduction

Restoring the "desired" speech signal from a mixture of speech and background noise is one of the oldest, still elusive goals in speech processing and communication systems research.

[1] This Chapter was published as paper named "Speech processing for hearing aids: Noise reduction motivated by models of binaural interaction", written together with Stephan Albani, Volker Hohmann, Jürgen Peissig, William S. Woods and Birger Kollmeier, see Wittkop *et al.* (1997).

The possible applications of such techniques range from enhancing the communication conditions in all kinds of communication systems (such as, e.g., telephones and video conferences), to automatic speech recognition and to "intelligent" hearing aids. One of the main problems for separating "target" speech from a background signal is the variability of the target speech as well as the wide range of possible background noise sources and acoustical conditions. Since the normal listener's auditory system is fairly well capable of performing this task even under very unfavourable acoustic conditions characterized by speech masked by speech in a reverberant environment (for example, in a cocktail party), this type of processing has often been referred to as "cocktail-party processing".

Sensorineural hearing-impaired patients suffer severely from their loss in understanding speech especially in noisy environments. This also holds when they use conventional hearing aids that perform an amplification and dynamic compression of the signals received at one or both ears. Since an improvement of their every-day communication situation only appears to be possible by introducing efficient "cocktail-party processing" strategies into future "intelligent" hearing aids, this contribution focusses on this application without limiting the processing to other applications. For a review of the specific problems encountered in signal processing for hearing-impaired listeners, the reader is referred to, e.g., Allen (1996), Hohmann and Kollmeier (1996) and Verschuure and Dreschler (1996). Noise reduction techniques developed for hearing aids in the past can be divided into procedures using a single microphone as input or multiple microphones. Examples of the single-input approach use a directional microphone (the small dimensions of which, relative to sound wavelengths in the audio frequency range, hardly provide any directivity at low frequencies and limit the useful directivity to higher frequencies, cf. Soede, 1990), an attenuation of certain frequencies (e.g., low frequencies, cf. overview given by Steeger, 1996), or a spectral subtraction technique (Boll, 1979; Graupe *et al.*, 1987). Although such systems have been employed in commercial hearing aids, their effect is rather limited because the inherent assumption about the stationarity of the background noise is not always met. Other single-microphone techniques therefore use assumptions about the target signal, such as the periodicity of voiced parts of the speech signal. Then cepstral filtering or harmonic selection enhances the appropriate components of the incoming signal (cf., Summerfield and Stubbs, 1990). Such a technique has been shown to be successful in certain laboratory situations, but has not yet been implemented in hearing aids because it requires control information (the value of the target's fundamental frequency) that cannot be unambiguously obtained from the input signal.

Taken together, the one-microphone approaches appear to be rather restricted in their applicability to real-world situations. Therefore, multiple-microphone techniques appear to be more promising, since more information about the target speaker and background noise can be obtained by sampling the sound field at different points in space simultaneously. Using arrays of multiple microphones, the directivity and frequency range over which the directivity is maintained can be largely improved in comparison with single directional microphones (cf., Soede *et al.*, 1993). However, the physical dimensions of the microphone array again impose a frequency dependend limitation on the directivity and the effective shape of the directional characteristic. More sophisticated approaches use adaptive filters to combine the signals from different microphones to form an "adaptive beam former" (Strube, 1981; Griffiths and Jim, 1982; Brey *et al.*, 1987; van Campernolle, 1990; Kompis and Dillier, 1994; Zurek *et al.*, 1996). A considerable gain in signal-to-noise ratio can be obtained with these approaches in certain laboratory situations. However, the gain decreases with increasing number of interfering noise sources and reverberation. In addition, the computational complexity is rather high, which limits the practical application of these approaches in digital hearing aids.

An alternative way to overcome these problems in a manner similar to the "effective" processing performed by a normal listener's auditory system is through the use of approaches suggested in the literature that employ certain types of binaural signal processing. That is, they exploit the input signals to both ears of a subject or a dummy head and perform a processing similar to the "effective" noise reduction processing performed by the binaural system. This general concept of applying knowledge of the normal human auditory system to technical speech communication systems has been introduced by Schroeder *et al.* (1979). One type of "binaural" signal processing can be described as directional filtering, i.e., suppressing sounds emanating from "undesired" directions and restoring sound from a "desired" direction (Gaik and Lindemann, 1986; Peissig, 1993; Kollmeier *et al.*, 1993; Lindemann, 1995). Another type of binaural signal processing performs a reduction of the subjective impression of reverberation (Allen *et al.*, 1977; Peissig, 1993) or more elaborate versions of cocktail-party processing that include a more or less detailed model of human binaural interaction (Bodden, 1993; Sullivan and Stern, 1993; Kollmeier and Koch, 1994). These approaches were shown to operate successfully in a variety of acoustical environments. Although they appear to yield less artifacts and are more stable for low signal-to-noise ratios than the beamforming algorithms mentioned above, they encounter similar problems, i.e., their performance decreases with increasing number of sound sources and reverberation. In addition, they require a high amount of computational power which restricts the usage in wearable hearing aids.

To overcome these problems, the current contribution describes further developments that are based on these algorithms and that aim at increasing the robustness of the algorithms in diffuse noise and reverberation. In addition, the development of real-time signal processing systems is described. Such systems are used to implement these algorithms for testing their application with hearing-impaired listeners under realistic communication situations in real-time. First, a review of the binaural model-motivated algorithms is given. The subsequent Section compares the performance of different types of processing, and the last Section describes the systems developed for real-time implementation.

2.2 Binaural signal processing techniques

2.2.1 Directional filter and dereverberation algorithm

Figure 2.1 sketches the algorithm developed by Peissig (1993) for suppressing lateral noise sources and dereverberation in binaural microphone signals. In the original implementation, an overlap-add technique was used for digital frequency analysis and resynthesis (Allen and Rabiner, 1977) using an FFT of 512 samples with Hanning-windowed segments of 408 samples and an overlap rate of 0.5 at a sample rate of 25 kHz. In the modification evaluated in Section 2.3, however, a filter-bank approach was used instead. In each frequency channel the interaural differences in phase and level, and the interaural coherence function at zero lag, are determined from the short-term autocorrelation in both channels and the cross-correlation between them (see Peissig, 1993, for more details). In a successive stage, a weighting factor g is derived for each frequency band that is used to either suppress the respective frequency band or to leave it unchanged. The idea behind this weighting function is that sound sources that emanate directly from the front should be passed through unchanged (i.e., the target is assumed to be in front of the listener). Thus, values of g near unity are applied to frequency bands exhibiting interaural phase and level differences equal to or near those expected for sound sources directly in front (or a certain angular region in front of the subject). Conversely, a low value of g should be obtained for interaural differences that deviate from this "desired" range. A similar concept was described by Gaik and Lindemann (1986). Since interaural time and intensity differences are not unambiguous indicators of a signal's direction of incidence in diffuse sound fields, the interaural coherence function is used to decide if the respective frequency channel is due to components of a direct incident sound (high interaural correlation) or diffuse sound. A low interaural correlation coefficient is obtained for the reverberant (diffuse) part of a signal which should result in an attenuation of the respective frequency channel and a low value of g. A temporal and spectral average of this weighting factor g is performed to reduce the artifacts (see Peissig, 1993, for details).

The general outline of the algorithm is motivated by the Jeffress (1948) model of binaural interaction, which assumes first a peripheral filtering process (similar to the FFT or filter-bank analysis performed here), followed by a delay line arrangement of neural units for the signal from both sides with coincidence detectors and subsequent integrators. This yields a kind of cross-correlation function between the respective bandpass filtered signals from both ears. The shape of the main lobe of this cross-correlation function corresponds inversely to the diffuseness of the incoming sound field (measured here using the value of the coherence function at zero lag), whereas the displacement of the maximum value along the interaural delay axis (corresponding to the interaural time or phase difference evaluated here) is related to the lateral displacement of the sound source from midline. One difference between the Jeffress model and the kind of processing introduced here (besides the different implementation methods) is the evaluation of interaural intensity differences. They are coded into corresponding interaural time delays in the Jeffress model, whereas they are independently evaluated here.

The directivity pattern of the algorithm given in Figure 2.1 can be selected from a wide range by manipulation of the weighting function g. Figure 2.2 shows the directivity pattern for a speech-spectrum-shaped noise as input signal as produced by both microphones of the dummy head employed (upper left panel) and the algorithm described above (upper right panel) with a particular weighting function (see Peissig, 1993, for details). The upper right panel shows that the directivity pattern can be restricted to a very narrow range of target directions. However, since the processing is highly nonlinear, the "effective" directivity is changed as soon as more than one sound source is present: The lower left and right panels of Figure 2.2 show effective directivity patterns using the same noise as before, but presenting one additional fixed noise source at 105 degrees azimuth, and two fixed sound

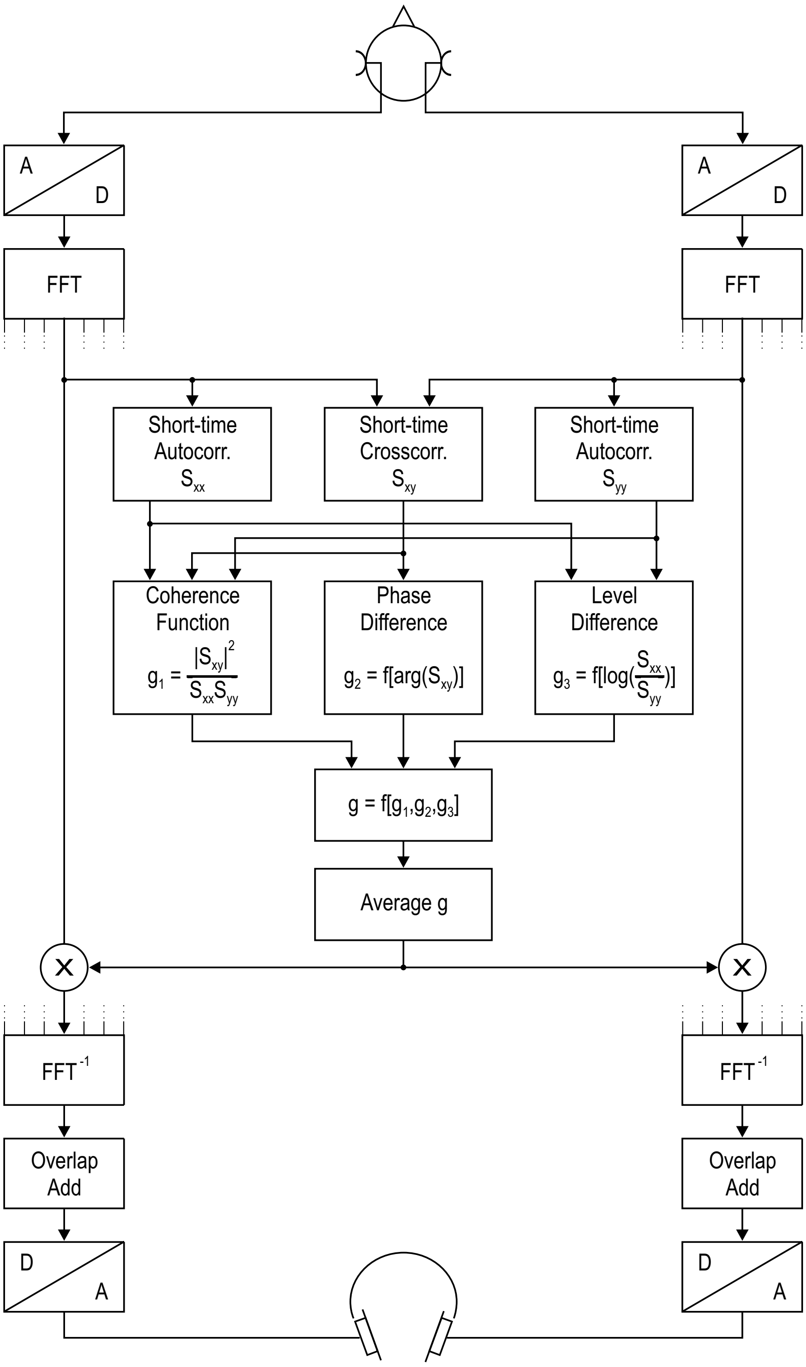


Figure 2.1: Block diagram of the algorithm for dereverberation and suppressing lateral noise sources after Peissig (1993).

sources (at 105 and 255 degrees azimuth), respectively. The directivity pattern is clearly broadened relative to the condition with only one sound source. However, an attenuation of sounds from the side is maintained.

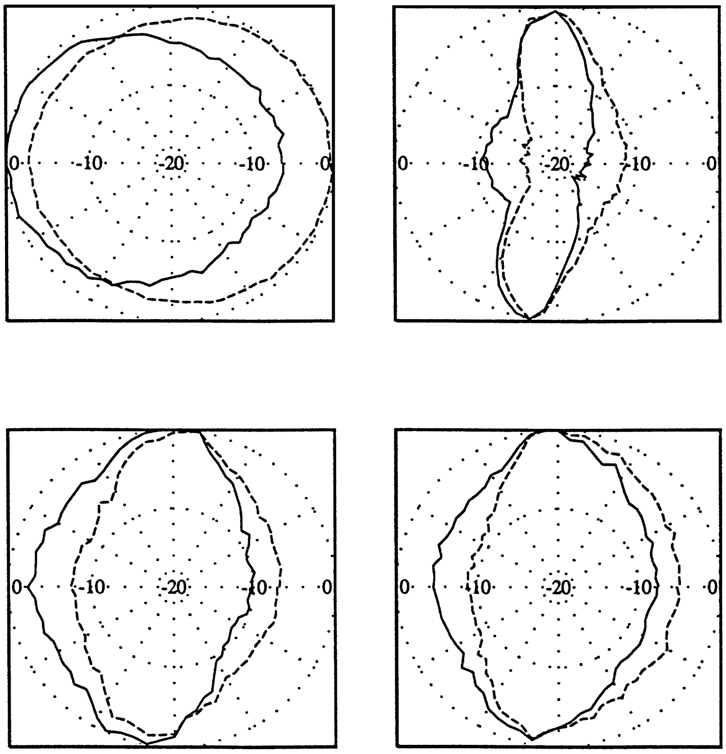


Figure 2.2: Directionality of the algorithm from Figure 2.1, obtained with speech-spectrum-shaped noise. The resulting signal levels of the left ear (solid line) and the right ear (dashed line) are given as a function of the sound incidence direction (see text for details). The numbers on the abscissa denote the attenuation (in dB) corresponding to the respective circles relative to the direction exhibiting the least attenuation. Upper left panel: no processing. Upper right panel: processing without interfering noise. Lower left: processing with one fixed interfering noise source at 105 degrees azimuth. Lower right: Two interfering noise sources fixed at 105 and 255 degrees azimuth.

Peissig (1993) and Kollmeier *et al.* (1993) demonstrated that the directional filtering algorithm provides a significant gain in intelligibility both in a non-reverberant environment (using up to three interfering speakers from different directions) and in a highly reverberant environment (using one target speaker and one interfering speaker), both for normal-hearing listeners and hearing-impaired listeners. Thus, the algorithm appears to be very promising for an application in a digital hearing aid. However, the algorithm tends to fail if the number of interfering sound sources increases and if the reverberation time becomes larger. In these situations, the interaural time and level differences alone are not a good and valid estimator for determining the spectral shape of the target speaker. This introduces processing artifacts in the current algorithm and limits the performance. Therefore, the usage of additional information appears to be necessary (see below). A further evaluation of the potentials and shortcomings of this algorithm is given in Section 2.3.

2.2.2 Binaural processing in the modulation frequency domain

Figure 2.3 gives the block diagram of an algorithm proposed by Kollmeier and Koch (1994), which is similar in general structure to the algorithm presented in Figure 2.1. However, it differs in the aspect that interaural differences are extracted in a modulation frequency domain. A modulation frequency analysis is performed by first extracting the temporal envelope within each band pass channel and then determining the modulation spectrum for each band pass channel and each side. The comparison of interaural time (or phase) and level difference with a "target" difference is performed for each combination of center frequency and modulation frequency (right most portion of Figure 2.3). The underlying assumption is that the spectral analysis in the modulation frequency domain provides a better separation between different acoustical objects emitting energy in the same frequency range but with a differing temporal envelope and hence different modulation spectrum. Hence, the preservation of only those combinations of modulation frequencies and center frequencies that exhibit the "desired" range of interaural phase and level difference should result in a more robust noise reduction. The reconstruction of the enhanced signal is then performed by first averaging across the weighting factors in the time, frequency and modulation frequency domains, and using these weights on the modulation spectra to determine a "desired" envelope. The "desired" envelope is reconstructed by an inverse Fourier transform of the modified modulation spectra, and the output signal is determined by modulating the original bandpass-filtered signal with the desired envelope divided by the original envelope. The time signal is finally reconstructed by an overlap-add synthesis of these filtered bandpass signals. Alternatively, a filterbank summation technique can be used for the spectral analysis and reconstruction (see Section 2.3).

The model motivating this kind of processing is based on neurophysiological findings of a modulation frequency analysis which is found to be represented orthogonally to the center frequency analysis in different stations of the auditory system in certain animals (cf., Langner, 1992). A different set of physiological evidence exists for a representation of interaural disparities orthogonal to the tonotopic organization (cf., Casseday and Covey, 1987). The functional model of binaural signal processing proposed by Kollmeier and Koch (1994) is a specific realization of an interaction that might exist between modulation frequency analysis and binaural analysis in each center frequency band. Such interactions are used to group the energy falling in each critical band into different internal "objects". Each sound source is assumed to be characterized by specific patterns of interaural disparity and modulation across center frequencies. Decomposition of received energy into objects is facilitated by this assumed linkage between binaural information and modulation spectrum information. These motivating assumptions are also compatible with the feature linkage model of v. d. Malsburg and Buhmann (1992). They argued that different feature detectors in the brain link the respective features of each object by synchronous oscillations. Under the motivating assumptions here, synchronous oscillations (i.e., similar patterns in modulation spectra across center frequency) leads to linkage of different features (for further discussion see Kollmeier and Koch, 1994). The algorithm depicted

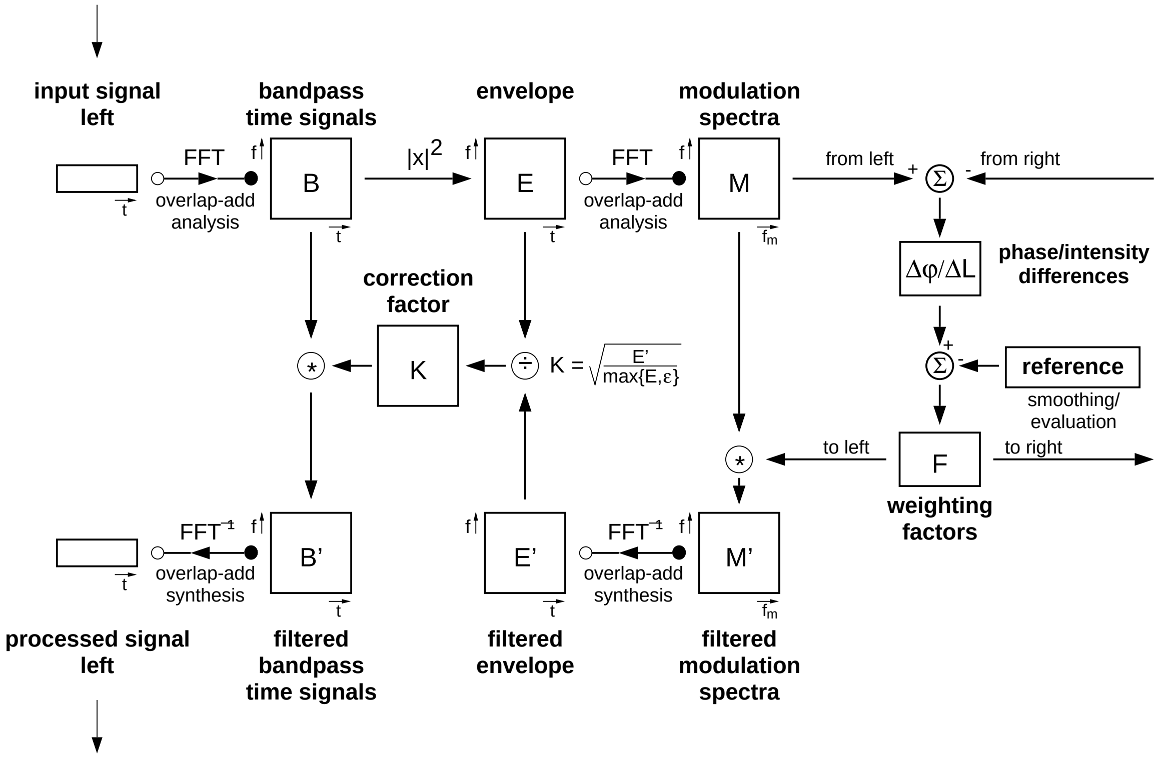


Figure 2.3: Block diagram of the algorithm employed for suppressing noise emanating from "undesired" directions after Kollmeier and Koch (1994). Only the left stereo channel is shown.

in Figure 2.3 was shown to provide a significant improvement in speech intelligibility and listening comfort for normal-hearing listeners in a variety of different acoustical situations, in both anechoic and echoic environments and with one to four interfering sound sources. It should be noted that a positive effect was even observed for very unfavourable signal-to-noise ratios. For example, a 15% increase in sentence intelligibility was observed in a reverberant environment with two interfering talkers at a signal-to-noise ratio of -8 dB (Kollmeier and Koch, 1994). A further evaluation of this algorithm and a comparison with the algorithm described above is given in Section 2.3.

2.2.3 Binaural signal processing with a localization model

Figure 2.4 gives the block diagram of a noise-reduction algorithm (Albani *et al.*, 1996) which is based on a localization algorithm motivated by physiological data from the barn owl. The general structure of the algorithm is comparable to those depicted in Figures 2.1 and 2.3. This refers to the binaural input, the decomposition into frequency bands, the evaluation of interaural level and interaural phase differences, the generation of weighting factors, and the reconstruction of the time signal with an overlap-add technique. However, the algorithm differs from the others by an across-frequency processing of interaural phase and level differences, motivated by neurophysiological findings of frequency-specific "localization maps" in the barn owl (Konishi *et al.*, 1988; Brainard *et al.*, 1992). Within each frequency band, the activity of neurons in this map reflects the probability that the sound source emanates from a direction that the respective neuron is most sensitive to. However, since the interaural phase and level differences within a given frequency band are ambiguous with respect to the direction to the active sound source (an ambiguity which is described by the "cone of confusion"), an unambiguous localization decision can only be made if an interaction occurs across the spatial maps at each frequency. This interaction ensures that only that direction is preserved which shows the maximum excitation consistently across all different frequency bands.

In the algorithm shown in Figure 2.4 the frequency-specific maps are generated by comparing the actual frequency-specific interaural phase and level difference with a set of reference interaural differences that are previously recorded from a fixed set of azimuthal and elevational angles of sound incidence. Those directions with the best match between actual differences and reference differences receive the highest "activity" values. The activity from each map is then weighted with the average level within the respective frequency band and summed across bands to yield a global localization map. In this map, the position with the highest activity is considered to be the direction of the sound source. In order to stabilize the localization judgement and to implement properties similar to the "law of the first wave front" (precedence effect), a feed-back mechanism is introduced which uses the resulting global map for "presetting" (i.e., altering the activity of) the frequency-specific maps in the subsequent time frame. This resembles a neuronal facilitation process and helps to build more stable localization decisions across time. The acoustical "target" object can be selected from the "global map" (for example, the sound source coming from the front), and only those frequency bands receive a high weighting factor where a contribution to the respective activity in the global map occurred. The resulting weighting is applied to the original input signal to yield the noise-reduced output signal.

It has been demonstrated that the algorithm yields a stable and reliable localization performance for up to four different sound sources both in anechoic and in reverberant environments (Albani *et al.*, 1996), and can track the spatial position of a target speaker as a function of time. Tests of noise-reduction schemes based on this representation have yet to be performed. Figure 2.5 demonstrates the ability of the model to separate the signals from two concurrently speaking talkers who are located at 270 degrees azimuth and 330 degrees azimuth. The instantaneous energy (envelope) of the signal emitted by each respective speaker in isolation is drawn as solid line in the upper panel for a speaker at 270 degrees azimuth and in the middle panel for a speaker at 330 degrees azimuth. Also shown are the temporal activity patterns of the resulting "global map" in the situation where both speakers are concurrently active. The upper and middle panel display the activity patterns (dashed lines) that correspond to 270 and 330 degrees azimuth, respectively, whereas the lower panel illustrates the activity patterns for three other arbitrary azimuths, i.e. 90, 230 and 160 degrees.

Obviously, the extracted activity pattern at the respective position corresponds very well with the temporal envelope of each speaker. Conversely, the activity corresponding to other directions is much lower than the activity of the respective "correct" directions and

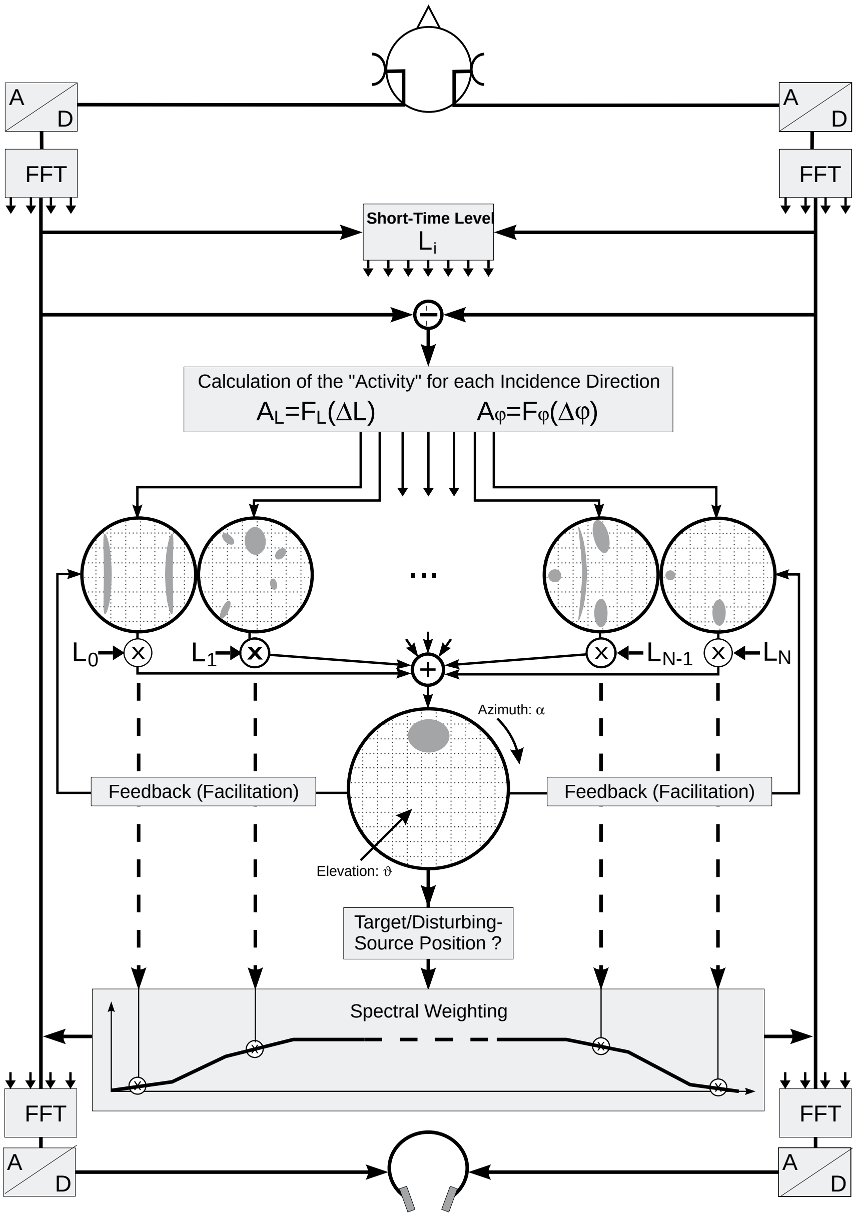


Figure 2.4: Schematic diagram of the localization model serving as a preprocessing stage for noise reduction after Albani *et al.* (1996). The different azimuthal angles are schematically represented by the circular angle α, whereas the different elevations are schematized by the respective diameter of the circle.

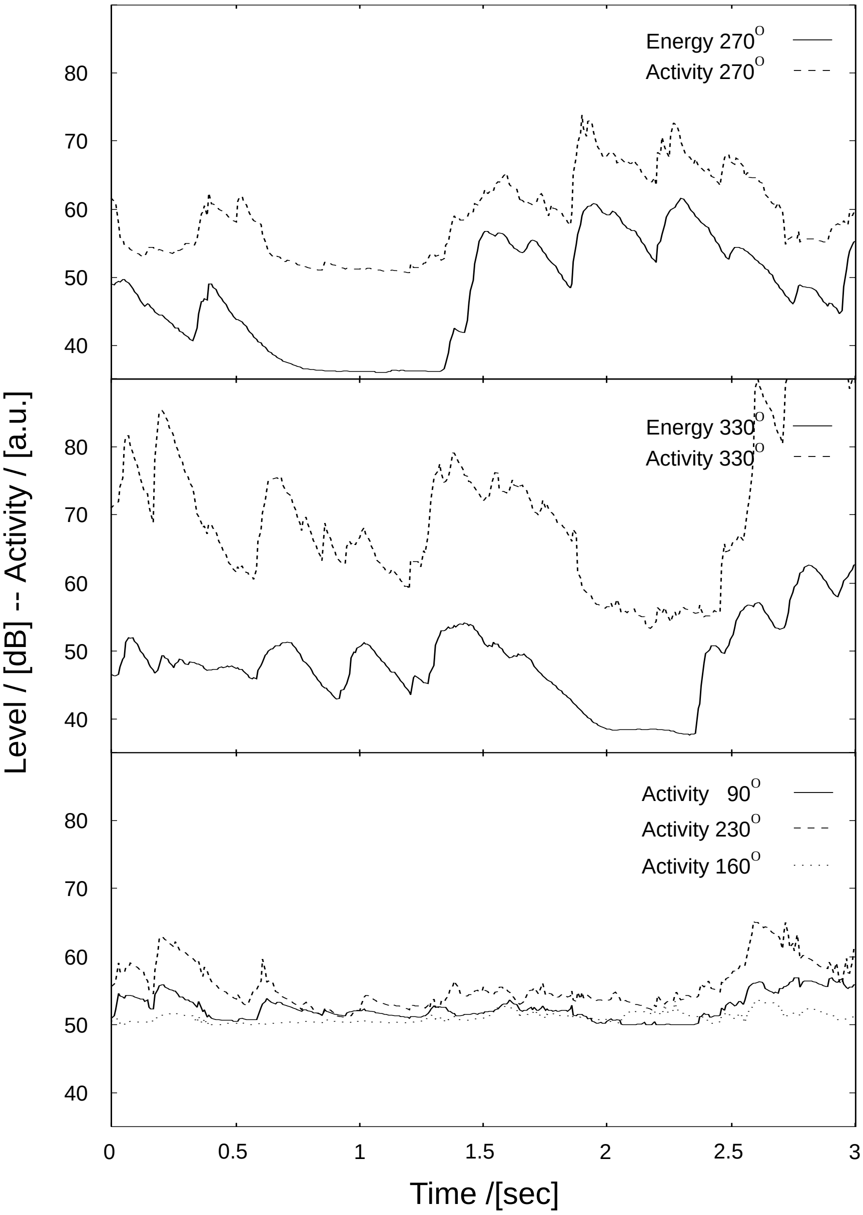


Figure 2.5: Temporal activity patterns for the algorithm depicted in Figure 2.4 operating on two concurrent speakers located at 270 and 330 degrees azimuth, respectively. The curves (dashed lines) denote the activity in the "global map" corresponding to the azimuths 270 degrees (upper panel), 330 degrees (middle panel) and a pattern with 90, 230 and 160 degrees (lower panel). For comparison, the temporal energy pattern (envelope) of the speakers in isolation are plotted as solid lines in the upper and middle panels.

much less correlated with any of the envelopes of the speakers employed. Thus, it should be possible to increase the signal-to-noise ratio in the combined acoustical signal from both speakers using the time-varying information in the "global map". The evaluation of such a noise-reduction system has yet to be performed.

2.2.4 Combination of different noise suppression algorithms

Figure 2.6 describes schematically an architecture proposed by Woods *et al.* (1996b) that combines the processing described in Section 2.2.1 (directional filtering) with a cepstral filtering algorithm similar to the one proposed by Stubbs and Summerfield (1991). The general outline of the algorithm again is very similar to the schemes discussed above, i.e., binaural input signals are transformed with a short-term fast Fourier transform and the modified spectrum is converted back to the time domain with an overlap-add technique. However, the estimate of the "target spectrum" is based on a combination of the respective estimates from both the directional filter algorithm and the cepstral filtering algorithm operating on each stereo channel. This combination is based on a "confidence" value computed by each separate algorithm for the frame-by-frame estimate of the target spectrum. The final estimate of the target is a combination of the spectral estimates of the preliminary algorithms and the unprocessed input spectrum. The final estimate follows that of a preliminary estimator when the estimator signals high "confidence" in its estimate, but passes the input spectrum unprocessed if no estimator has high confidence values. If past estimates are also used in the final combination, the general operation of the system can be made similar to that of a Kalman filter. That is, the target estimate will be mainly driven by the input data if the input signal-to-noise ratio (S/N) is high (i.e., when the preliminary estimators yield high "confidence"), and will be the "best guess" the system can make (either derived from past estimates or simply the unprocessed input) when the input S/N is low. This architecture also has the advantage that an arbitrary number of different noise suppression algorithms can be combined, provided that they deliver both a spectral estimate of the target and a confidence value. This confidence value can be interpreted as the degree to which the assumptions of the algorithm, concerning the target and noise signals, are fulfilled by the received signal conditions.

Similar to the algorithm specified in Figure 2.4, a facilitation process is introduced that helps to stabilize the current estimate of the fundamental frequency of the target: The combined estimate and its confidence value are used to preset the estimate of the fundamental frequency in the subsequent frame (denoted as "attributes" in Figure 2.6). This is motivated by the continuity of fundamental frequencies to be expected in natural speech and helps to unambiguously determine a target fundamental frequency from several possible fundamentals.

The underlying hypothesis of this algorithm is that the auditory system is thought to use different types of cues to decompose the summed acoustical signals into different auditory "objects". Each object is characterized by a certain set of attributes (such as its perceived position or its pitch). Physical constraints on natural sound sources require that these attributes follow certain rules (e.g., "spatial position must change smoothly"), and these constraints are used in determining the confidence values, and allow the feedback of information from central to more peripheral processing in the model. Figure 2.6 shows

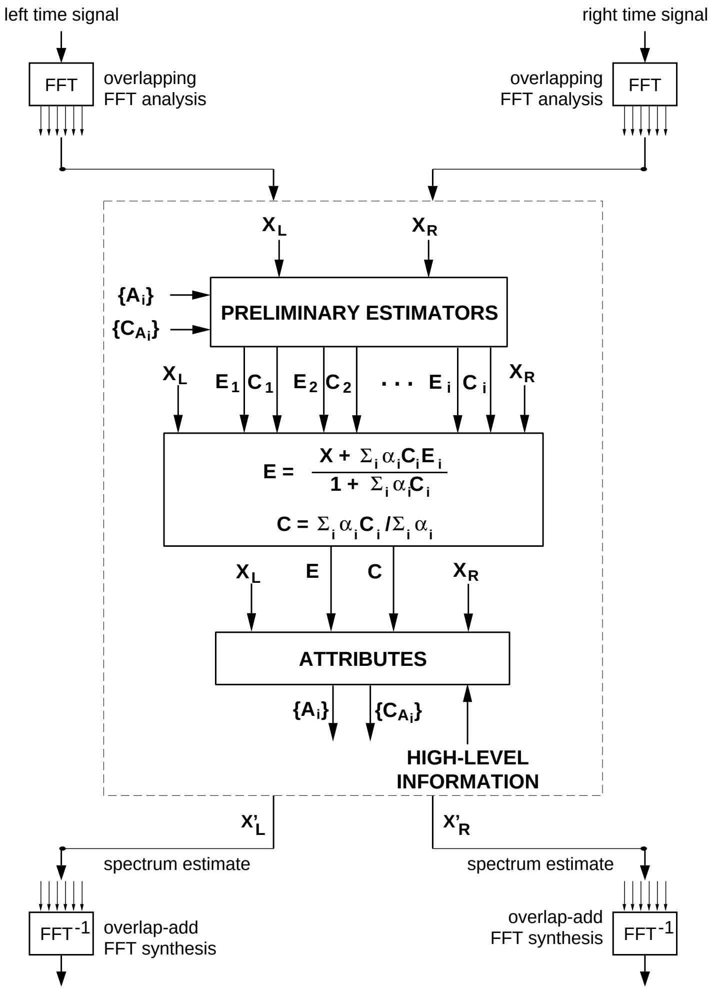


Figure 2.6: Block diagram of architecture for combining algorithm outputs. Two received short-term signal spectra X_L and X_R are processed by a bank of target estimation/noise suppression routines. Each routine produces a target magnitude-estimate E_i, and a scalar "confidence" value C_i ranging between 0 and 1. In the current implementation, a binaural directional filter algorithm in combination with two monaural cepstral filtering algorithms is used. The estimates and confidences are combined in weighted averages to form a final spectral estimate E and final confidence C. If any estimate has high confidence, the input in that channel is suppressed and the output comprises mostly the estimate. If no estimate has high confidence, the input signal is passed through unprocessed. The final estimate and confidence are used to determine parameters ("attributes" A_i) and their respective confidence values (C_{A_i}) to be used as "a priori" information in the subsequent time frame. At this stage also "high-level" information (such as expected pitch range or position of the target) is incorporated.

only the outline of a specific implementation of these ideas. Algorithms based on different cues could easily be integrated into the general structure.

The algorithm depicted in Figure 2.6 reliably tracked the fundamental frequency of two target sounds with a different spatial position even if the fundamental frequency contour of the sound sources intersected (Woods *et al.*, 1996b). This indicates that the system can make use of both algorithms concurrently. In addition, Woods *et al.* (1996a) demonstrated that the signal-to-noise ratio (S/N) at the output is increased by the combined algorithms to the same degree or further than the least effective of the two algorithms. This holds even for unfavourable signal-to-noise ratios of the input signal (e.g., an estimated 4.5 dB improvement in S/N for input S/N of -4 dB). Formal tests of the possible improvement in speech intelligibility are planned.

2.3 Comparison between two signal processing techniques

To obtain more insights into the advantages and shortcomings of the algorithms described so far, the binaural noise reduction algorithms described in Sections 2.2.1 and 2.2.2 were optimized and implemented in a uniform off-line signal processing environment. The first algorithm (denoted as algorithm F) has been implemented following Peissig (see Section 2.2.1) and the second (based on the modulation spectrum and denoted as algorithm M) is an implementation following Koch (see Section 2.2.2). The current implementation differs from those described in Section 2.2 with respect to the frequency analysis and synthesis technique employed: Instead of using an overlap-add FFT technique, a filter bank summation technique was employed using a FIR frequency sampling filter bank (Stearns, 1991). The only free parameter that has to be adjusted for this technique is the frequency-sampling period Δf (see below). The center frequencies of this filterbank are equally spaced across the frequency range. They were grouped and combined into critical-band-wide filter outputs that are no longer equally spaced in linear frequency but are equally spaced on a Bark scale.

Speech intelligibility measurements were performed with both algorithms in three target speech/interfering noise configurations (see Figure 2.7) with normal hearing and hearing-impaired listeners. Sentences from the Göttingen sentence test (Wesselkamp *et al.*, 1992) were used as speech signals. Speech simulating noise (Wesselkamp *et al.*, 1992) was employed as noise in configuration hr1, while in configurations hs2 and hs4 several interfering speakers, each reading aloud a different text, were employed as noise signals. These configurations are the same as employed by Kollmeier and Koch (1994).

For each configuration, the unprocessed and six differently processed versions were tested (three versions for each algorithm). These three versions differed in the frequency resolution Δf of the employed frequency-sampling filter bank. When the sampling period Δf in the frequency domain decreases, a better separation of the various filter outputs of the filter bank is achieved. Thus, for the processing scheme used here, a better suppression of one "unwanted" critical band is possible even if the adjacent bands are not attenuated. However, when increasing Δf, the impulse responses of the filters are elongated and the

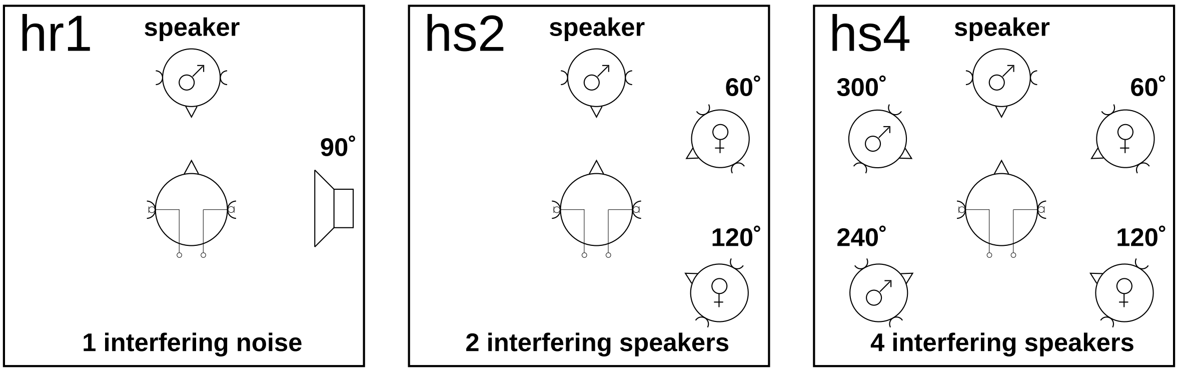


Figure 2.7: Sketch of the acoustical configurations employed. The dichotic signals were recorded using the Göttingen dummy head (Damaske and Wagener, 1969; Peissig, 1993) in reverberant environment (broadband reverberation time 1.33 s).

computation time is increased. Thus, to minimize the computation time for the filters, Δf should be as large as possible without exceeding the auditory critical bandwidth and without causing any decrease in algorithm performance. Values of 10, 33 and 100 Hz for Δf were investigated.

2.3.1 Comparison at a fixed signal-to-noise ratio

In this experiment, six normal-hearing subjects aged between 25 and 31 (1 female, 5 male) participated voluntarily. All signals were presented to the subjects via headphones (Sennheiser HD 25) without free field equalization in a sound-insulated booth. The maximum peak level of the presented material was approximately 80 dB SPL. Since the evaluation of all the different test conditions with traditional speech intelligibility scoring techniques would have exceeded the available number of test items in a sentence test, the computational resources available and the measurement time available for each subject, a more resource-saving subjective speech intelligibility assessment method was employed. As shown in Figure 2.8, the subject's task was to compare each test stimulus ② with a reference stimulus ① and to adjust the signal-to-noise ratio (S/N) of the reference stimulus until test and reference were judged to be equal in intelligibility. To do so, the subjects first listened to a sequence of a fixed test stimulus and a variable reference stimulus. They then had to depress one out of five responses ("LOUDER", "louder", "play", "softer", "SOFTER") to change the S/N of the reference stimulus, or "OK" to end the trial. The S/N of the reference version was altered by +3, +1, 0, -1 or -3 dB in response to the above categories, respectively. The response categories were displayed on a handheld touchscreen response box (Epson EHT-10S). The validity of such a subjective speech intelligibility assessment method and its relation to "objective" speech intelligibility scoring methods has been investigated by Wesselkamp (1994) and Kollmeier and Wesselkamp (1996).

The reference stimulus consisted of the same test sentence as the test stimulus plus a diotic anechoic speech simulating noise (Wesselkamp *et al.*, 1992) as reference noise signal. The test stimuli were generated each with a fixed S/N that yielded approximately 50%

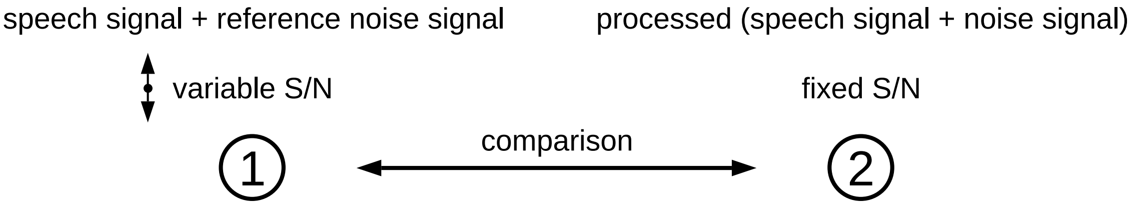

speech signal + reference noise signal
processed (speech signal + noise signal)
variable S/N
fixed S/N
comparison
①
②

Figure 2.8: Sketch of the measurement procedure at a fixed S/N (signal-to-noise ratio). The subjects adjust the S/N of the reference version ① until their subjectively assessed intelligibility of both reference version ① and test version ② is equal.

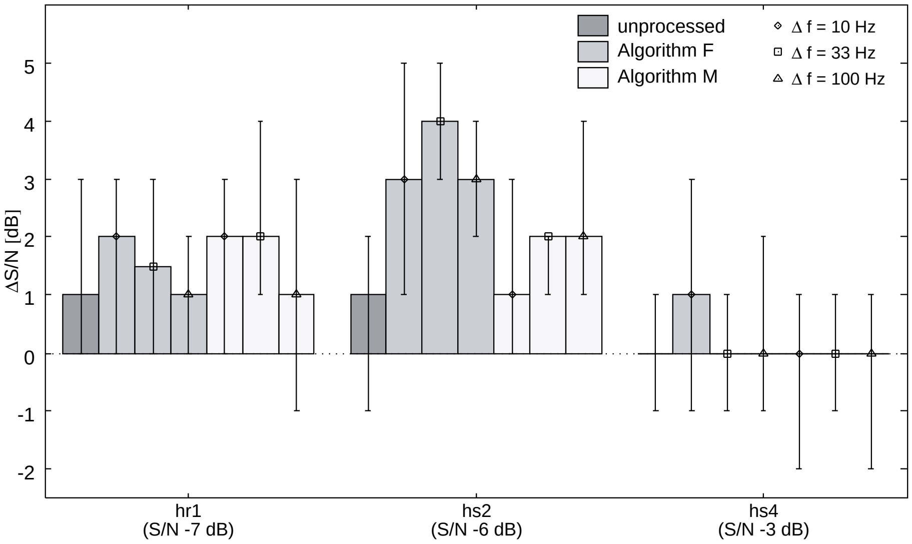


Figure 2.9: Results for six normal-hearing subjects for comparison at fixed S/N (see Section 2.3.1 for details). The abscissa shows the three configurations hr1, hs2 and hs4 with one unprocessed and six processed versions each. For each configuration, the S/N (RMS value) of the unprocessed version is given below the abscissa. On the ordinate the median values and interquartile ranges of the deviations ΔS/N between the adjusted S/N and the S/N of the unprocessed version (RMS value) are given. Positive values (bars pointing upwards) denote an improvement.

intelligibility before processing, and were computed off-line in advance. The signals were presented diotically to the subjects (i.e., the right and left stereo channels were added and presented to both ears) in order to avoid any further binaural noise reduction performed by the auditory system of the subjects. With each subject a total of four measurements were performed, i.e., each tested version was tested twice with each of two test sentences. The same set of two sentences was employed for all measurements.

The resulting values of the deviation ΔS/N (averaged across subjects and test sentences) between the S/N adjusted by the subjects and the real S/N (RMS value) of the unprocessed signals are given in Figure 2.9. Positive ΔS/N values (bars pointing upwards) denote an improvement in intelligibility with respect to the unprocessed version. The real S/N values employed for each reference version are given in parentheses below the abscissa. Note that the subjects adjusted the S/N of the unprocessed version very consistently close to its actual value (within less than 1 dB deviation on the average). Also, the measurements were performed at low or very low S/N, but considerably above the masked threshold.

Analysis of variance (ANOVA) tests with respect to the factors algorithms and Δf revealed that the results are highly significant ($p < 0.001$) for configuration hs2 with algorithm F and marginally significant ($p < 0.05$) with algorithm M. Factor Δf has no significant influence on the results (except for a marginally significant effect for configuration hr1). A distinct improvement of the intelligibility could be found for configuration hs2. Here algorithm F is more effective than algorithm M.

2.3.2 Comparison at a variable signal-to-noise ratio

In this experiment, six normal-hearing subjects aged between 24 and 31 (2 female, 4 male) and three sensorineural hearing-impaired subjects aged between 41 and 73 (1 female, 2 male) participated voluntarily. The hearing-impaired subjects exhibited symmetric, moderate to severe hearing losses of up to 60 to 80 dB at higher frequencies. The test procedure for the hearing-impaired subjects was the same as for the normal-hearing, except for the absolute presentation level and the application of a digital master hearing aid. This was used to compensate for the hearing loss by linear amplification within three channels. The amplification within each channel was adjusted in a way that the most comfortable loudness levels for normal-hearing listeners were amplified to the most comfortable loudness level of the individual hearing-impaired subject. As above, the signals were presented diotically to the subjects via headphones. For the hearing-impaired subjects, the presentation level was adjusted individually in advance to a level the subject judged to be comfortable.

As before, a speech intelligibility assessment method was used that allowed to test a series of different conditions without exceeding the available test materials, computational resources and measurement time for each subject. As shown in Figure 2.10, the subject's task was to adjust the before-processing S/N of the test version ① until the speech intelligibility was judged to be 50%. To do so, the respective test stimulus was presented diotically to the subject via headphones. The subjects task was to depress one out of five response alternatives (”LOUDER”, ”louder”, ”play”, ”softer”, ”SOFTER”) to change the before-processing S/N of the test version ①, or ”OK” to end the trial. The before-processing S/N was altered by +3, +1, 0, -1 or -3 dB in response to the respective categories listed above. The response categories were displayed on a handheld touchscreen response box (Epson EHT-10S).

This method of adjusting speech intelligibility has been reported to yield stable and reliable estimates of speech intelligibility thresholds. Since an individual subjective threshold criterion influences the absolute value of the assessed threshold for each subject, only differences in threshold with respect to a fixed reference condition should be reported (see Discussion). Both the relative size, the intrasubject and intersubject variability of these differences are comparable to results obtained with standard scoring methods, but require less measurement time (Wesselkamp, 1994; Peissig and Kollmeier, 1997; Kollmeier and Wesselkamp, 1996).

The processing of the tested versions had to be performed off-line in advance for all

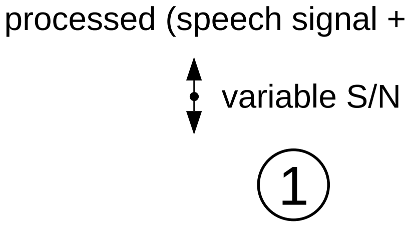


Figure 2.10: Sketch of the measurement procedure using variable S/N. The subjects adjust the before-processing S/N of the test version ① until 50% speech intelligibility is attained according to their subjective criterion.

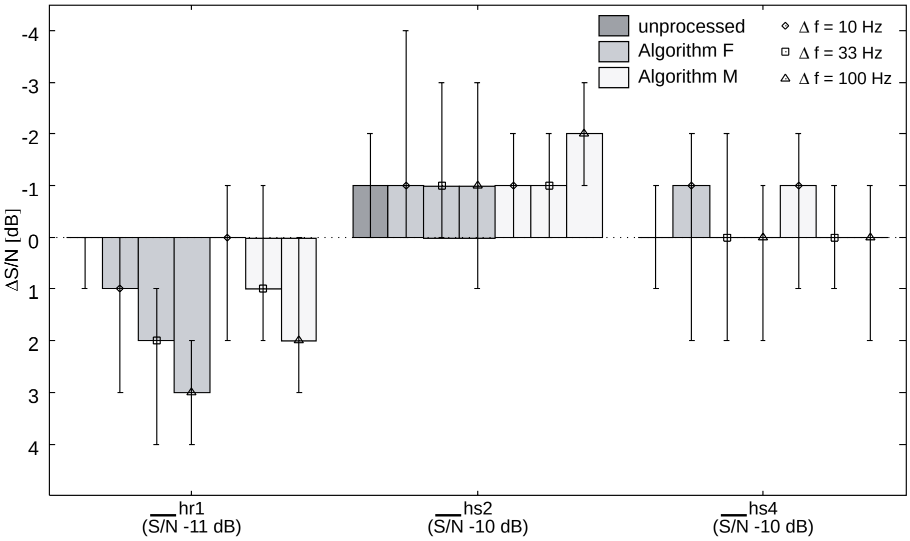


Figure 2.11: Results for six normal-hearing subjects for comparisons at variable S/N (see Section 2.3.2 for details). The abscissa shows the three configurations hr1, hs2 and hs4 with one unprocessed and six processed versions each. For each configuration, the mean adjusted S/N of the unprocessed version is given below the abscissa. The ordinate gives the median values and interquartile ranges of the deviations ΔS/N between the adjusted S/N of the different versions and the adjusted S/N of the unprocessed version. Negative values (bars pointing upwards) denote an improvement.

possible S/N values required during the experiment. Thus, the computation time was considerably higher than in the first experiment. Again, a total of four measurements were performed with each subject.

Figure 2.11 displays the deviations ΔS/N (averaged across subjects and test sentences) between the adjusted S/N of each different version and the adjusted S/N of the unprocessed version for the normal-hearing subjects. To obtain the deviation for the unprocessed version itself, the S/N was adjusted twice for this particular version. Note that the subjects adjusted the S/N of the unprocessed version very consistently in both cases (within less than 1 dB deviation on the average). The mean adjusted S/N values of the unprocessed version of each configuration are given in parentheses below the abscissa. They show that the measurements of this experiment resulted in very low S/N values that are close to the masked threshold for speech detection. An improvement in S/N at threshold with respect to the unprocessed version is denoted by negative values (bars pointing upwards).

The results for the normal-hearing subjects are highly significant ($p < 0.001$) for configuration hr1 with algorithm F ($\Delta f \geq 33$ Hz), and significant ($p < 0.01$) with algorithm M ($\Delta f = 100$ Hz). The factor Δf is highly significant for configuration hr1. No significant improvement of speech intelligibility was found, and no significant deterioration occurred, at least not with $\Delta f = 10$ Hz. Apparently, the algorithms are not able to take advantage of the little cues which may be available at these low S/Ns. However, algorithm M appears to perform slightly better than algorithm F (in configuration hr1).

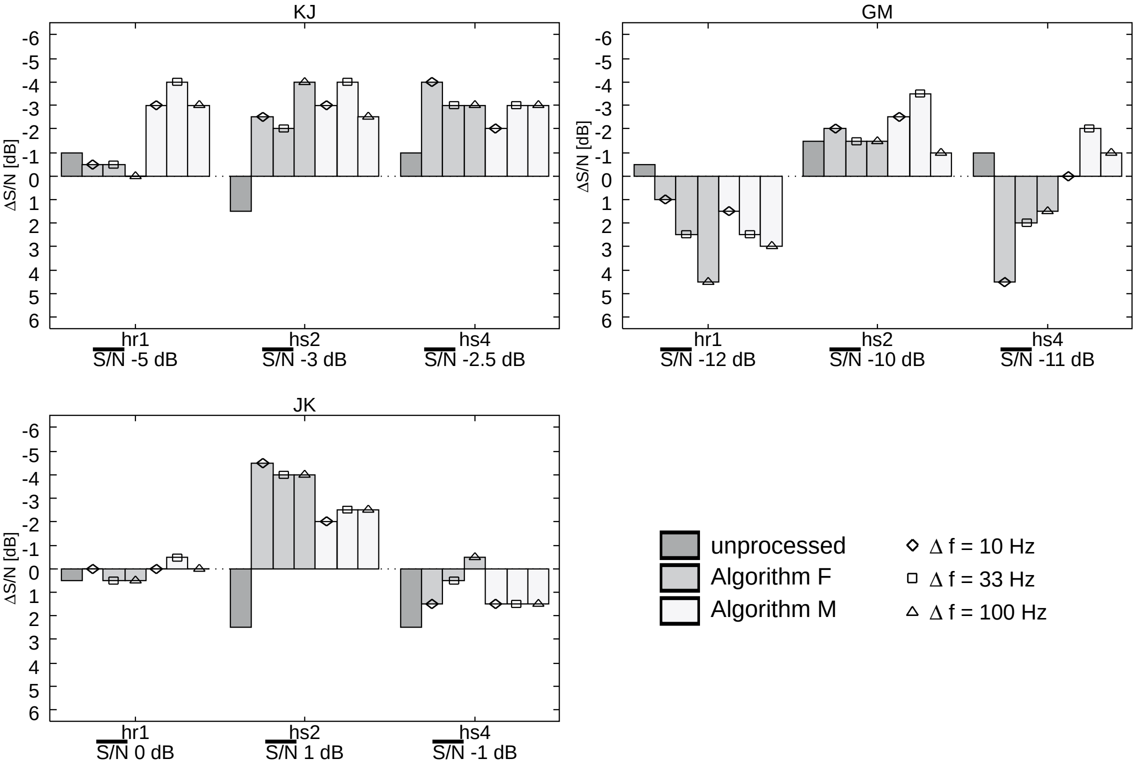


Figure 2.12: Results for three sensorineural hearing-impaired subjects KJ, GM and JK (see Section 2.3.2 for details). Each panel shows the results of the subject given above. For further explanation, see Figure 2.11.

The results of the hearing-impaired subjects, given in Figure 2.12, show considerable interindividual variability. For subject KJ, who exhibited the largest hearing loss, the results reveal an improvement in almost every processed version. For subject GM, the results (including the absolute values of the adjusted S/N) are similar to the results of the normal-hearing subjects. Algorithm M tends to yield slightly better results. For subject

JK, both algorithms resulted in a considerable improvement for configuration hs2, where algorithm F was better than algorithm M.

2.3.3 Discussion

The main results from the comparison between the two different algorithms with normal-hearing and hearing-impaired listeners can be summarized as follows:

a) The subjective adjustment methods employed here provide a reliable estimate of the benefit (or deterioration) resulting from the processing schemes. This is advantageous because relatively little speech material is required for these methods.

b) Both algorithms provide a significant benefit in certain acoustical configurations (one target speaker and one or two interfering noise sources at moderate signal-to-noise ratios), but fail to provide a significant benefit at low signal-to-noise ratios and for more complex acoustical configurations, i.e., four interfering noise sources in a reverberant environment. This general finding holds both for normal-hearing and hearing-impaired listeners if the different signal-to-noise ratios employed are taken into consideration.

c) The frequency resolution Δf of the filterbank analysis employed in the algorithms only plays a marginal role for the values tested here.

d) Only small differences were observed between the two algorithms. In general, algorithm F appears to provide the larger effect at moderate signal-to-noise ratios, whereas algorithm M (involving processing in the modulation frequency domain) appears to be a bit more robust at low signal-to-noise ratios.

Finding a) can be deduced from the consistent and reproducible judgement of the subjects in the reference conditions as well as the comparatively low interindividual and intraindividual deviations for the conditions tested here. These deviations are of the same order as values expected for traditional "objective" scoring methods. However, the methods employed here require only two sentences to be processed in every condition tested and every S/N employed. Traditional scoring methods would require a complete test list to be processed for all conditions. In addition, a considerably larger group of subjects would have to be employed with traditional sentence tests, since each sentence may only be presented once to each subject. Therefore, the large variety of test conditions employed in this study would constitute an impractically large measurement effort if performed with traditional speech scoring methods.

Results of the subjective intelligibility assessment methods employed here are highly correlated to those obtained with traditional, "objective" scoring methods (cf., Cox *et al.*, 1991; Kollmeier and Wesselkamp, 1996). However, they can not replace these traditional methods, since they do not yield absolute scores or threshold values. One problem with the subjective method employed here, is the influence of the individual subjective criterion for being "equally intelligible" or for a sentence to be "50% intelligible". To eliminate this subjective criterion, only the difference between any two conditions (i.e., the test condition and the reference condition) has been employed here. The intraindividual standard deviations would be much larger without such a difference measure. Another problem with this method is its tendency to yield a smaller effect in terms of signal-to-noise ratios than observable with traditional scoring methods (Peissig and Kollmeier, 1997; Kollmeier and Wesselkamp, 1996). This may be related to the subjects tendency to subjectively assess the decrease in "noisiness", which can be - due to processing artefacts - less pronounced than the increase of the "objective" speech intelligibility score produced by the algorithm. Therefore, the benefit in terms of signal-to-noise ratio reported here is not directly the same as the benefit in speech reception thresholds obtainable with traditional scoring methods. However, the rank order of the thresholds and the relative magnitude of the effects are presumed to be preserved. Thus, the conclusions derived from the present data are given in these terms rather than in absolute threshold improvement data.

The general finding of an improvement in certain acoustical configurations (finding b) is in agreement with previous studies with the respective algorithms (Peissig, 1993; Kollmeier and Koch, 1994), although no direct comparison was performed in these previous studies. In addition, the more favourable signal-to-noise ratios and different acoustical configurations employed in the previous studies resulted in somewhat larger effects than those obtained here. The smaller effects found here may be due both to the measurement method employed here (see above) and the fact that the benefit obtainable from the algorithms depends on the signal-to-noise ratio of the respective acoustical configuration. It is expected that the benefit from any noise reduction algorithm is low for very favourable S/N conditions (due to a ceiling effect both in the unprocessed and processed versions), but increases at decreasing S/N, because the performance in the unprocessed condition decreases more rapidly than that in the processed conditions. For very low signal-to-noise ratios, the benefit is again expected to decrease until both the performance in the unprocessed and the processed versions approaches zero. Thus, the size of the benefit obtainable from the respective algorithms can hardly be compared across studies. This motivated the direct comparison across algorithms performed here because such a comparison can only be performed by a study design as employed here.

In addition, the differences in the benefit achieved for normal-hearing listeners and hearing-impaired listeners as well as the difference between the first experiment (Section 2.3.1) and the second experiment (Section 2.3.2) can be interpreted in terms of the different signal-to-noise ratios employed. The highest signal-to-noise ratios were employed for the hearing-impaired listeners KJ and JK, respectively, who also received a significant benefit from the algorithms at least for configuration hs2. A lower input signal-to-noise ratio was employed for the experiment described in Section 2.3.1 with normal-hearing listeners, where again a considerable benefit was obtained for configuration hs2, but less benefit for the other acoustical situations. The lowest signal-to-noise ratio was employed for the experiment described in Section 2.3.2 for both the normal-hearing listeners and the hearing-impaired subject GM, which resulted in no significant benefit from the algorithms. Since both the experimental paradigm and the subjects differ between these different signal-to-noise ratios employed, a systematic investigation of the effect of the signal-to-noise ratio on the benefit from each of the respective algorithms is still warranted.

With respect to finding c) it should be noted that all the values of the frequency resolution Δf employed here do not exceed 100 Hz, i.e., the ear's critical bandwidth for low frequencies. Also, the grouping and combination of frequency bands was always performed such that an effective bandwidth is achieved which corresponds to one critical band irrespective of the value of Δf employed. Hence, a variation in Δf primarily influences the duration of the impulse responses and the steepness of the filter slopes rather than the bandwidth of the frequency analysis employed. Although some improvement in performance is observed with decreasing values of Δf (especially for the conditions with very unfavourable signal-to-noise ratios where the least deterioration of performance is observed for the highest frequency resolution), a larger effect of the frequency resolution would be expected if larger values of Δf would have been employed. Such values were not included in this study, because for auditory preprocessing it is advisable that the frequency resolution of the technical system prior to the ear should not be less than the ear's frequency resolution. Since the computational effort increases drastically with increasing frequency resolution, the data found here suggest that the increase in algorithmic performance achieved with increasing frequency resolution is very limited. Thus, it seems advisable to use an algorithmic frequency resolution comparable to the ear's critical bands (at least at low frequencies).

With respect to finding d) it is surprising that the increased computational complexity of algorithm M does not yield an advantage over the simpler algorithm F which would justify the additional computation for most situations. However, this finding may also be influenced by the special choice of the signal-to-noise ratios employed here. For example, for the lowest S/N employed in the situation hr1 (cf. Figure 2.11), algorithm M yields less detrimental effects than algorithm F indicating that its processing appears to be more robust at very low signal-to-noise ratios than algorithm F. On the other hand, the maximum effect of algorithm M seems to be smaller than the effect achievable by algorithm F at more favourable signal-to-noise ratios (cf. configuration hs2 in Figure 2.9). Therefore, the maximum benefit obtainable with this algorithm might be at signal-to-noise ratios somewhere between those employed in Figure 2.9 and Figure 2.11 where the benefit obtainable from algorithm F decreases more rapidly than that obtainable from algorithm M. Thus, a closer investigation of the difference between the performance of algorithm F and algorithm M at low signal-to-noise ratio would be desirable. If the better robustness of algorithm M over algorithm F still holds, it might be worthwhile to incorporate some of the properties from algorithm M into algorithm F at least for low signal-to-noise ratios.

2.4 Systems for developing and testing hearing aid algorithms

The types of noise-reduction algorithms to be investigated for use in digital hearing aids and other applications are not independent of the hardware available to perform these algorithms. For the comparison between algorithms described in Section 2.3, for example, an off-line implementation was employed which required from 120 to 675 times real-time for the processing of the test material. Thus, the possibility of modifying any parameters of the algorithms on-line and testing the respective algorithms under a variety of conditions is very restricted. Hence, a real-time implementation that allows an interactive adjustment of processing parameters is highly desirable. Such an implementation may be performed on a wearable or a stationary processing platform. Ideally, the algorithms described in Section 2.2 should be implemented on wearable in-the-canal hearing aids that communicate with each other by some wireless link. However, due to the computational complexity and, subsequently, the hardware requirements for real-time performance of the algorithms described here, such a solution is not feasible with today's technology. In order to outline the current state-of-the-art, this Section describes the current hardware approaches to implement and test these algorithms.

2.4.1 Wearable devices

Given the current technology in integrated circuit design, the power consumption associated with a certain degree of computational complexity imposes severe constraints on the type of processing possible with wearable DSP (digital signal processing) hearing aids. Although a few DSP BTE (behind-the-ear) hearing aids with certain, restricted algorithmic functions are already on the market, more flexible and generally programmable wearable solutions currently only exist as body-worn hearing aids (Faulkner *et al.*, 1990; Arlinger *et al.*, 1994; Grim *et al.*, 1995; Dillier, 1996; Rass, 1996, an overview is given by Steeger, 1996). These devices typically house one multiple-purpose low-power DSP connected to interfaces and supporting hardware for the hearing instrument use (cf., A/D-D/A converter, programming interface, control elements, connection to an earpiece). The algorithms that can be performed with this kind of processor typically include some type of dynamic compression and/or noise reduction such as adaptive beam forming (Dillier, 1996) or directional filtering (Grim *et al.*, 1995; Rass, 1996). However, the computational power of these devices is not sufficient to perform, in real time, the more complex algorithms described in Sections 2.2.2 to 2.2.4. Therefore, multiple-DSP solutions have to be employed that are currently only available as stationary devices.

2.4.2 Stationary devices

Several approaches for real-time simulation of digital hearing aids have been described in the literature that employ stationary equipment ranging from a personal computer (including additional DSP hardware) to a dedicated mainframe computer. Levitt *et al.* (1990), for example, used an array processor as coprocessor for a general-purpose computer to perform real-time simulations of digital hearing aids. Other solutions employ a combination of several DSPs that are housed by a host computer. Each DSP is dedicated to a certain task (e.g., Kollmeier *et al.*, 1993). In the system developed by Peissig (1993), for example, three multiple-purpose floating point DSPs are connected consecutively with serial links. The first DSP performs a stereophonic A/D conversion and a fast Fourier transform on overlapping time segments. The second DSP performs the processing in the frequency domain, and the third DSP performs the transformation back to the time domain and transmitts the data to stereophonic D/A converters. Although the algorithms described in Section 2.2.1 are successfully implemented and tested on this system, the more computationally complex algorithms described in Sections 2.2.2 to 2.2.4 cannot be computed with such a system. The limitations of the system are the processing power of the DSPs involved as well as the limitation of the data transfer between the respective DSPs. Therefore, a new hardware system was developed and tested which is described below.

2.4.3 Stationary apparatus with five signal processors

A block diagram of the real-time hardware set-up is shown in Figure 2.13. The signal processing part of the equipment consists of an ADC (analogue-to-digital converter) and DAC (digital-to-analogue converter) board, a single-DSP VMEbus board and a four-DSP VMEbus board, all in a VMEbus card cage. The ADC and DAC are two-channel 18-Bit converters with built-in, programmable input and output low pass filters. The single-DSP is a 40 MHz Texas Instruments TMS320C40 floating-point signal processor, and the other DSPs are 50 MHz TMS320C40 versions. The VMEbus DSP boards are controlled by a VMEbus SPARC workstation host.

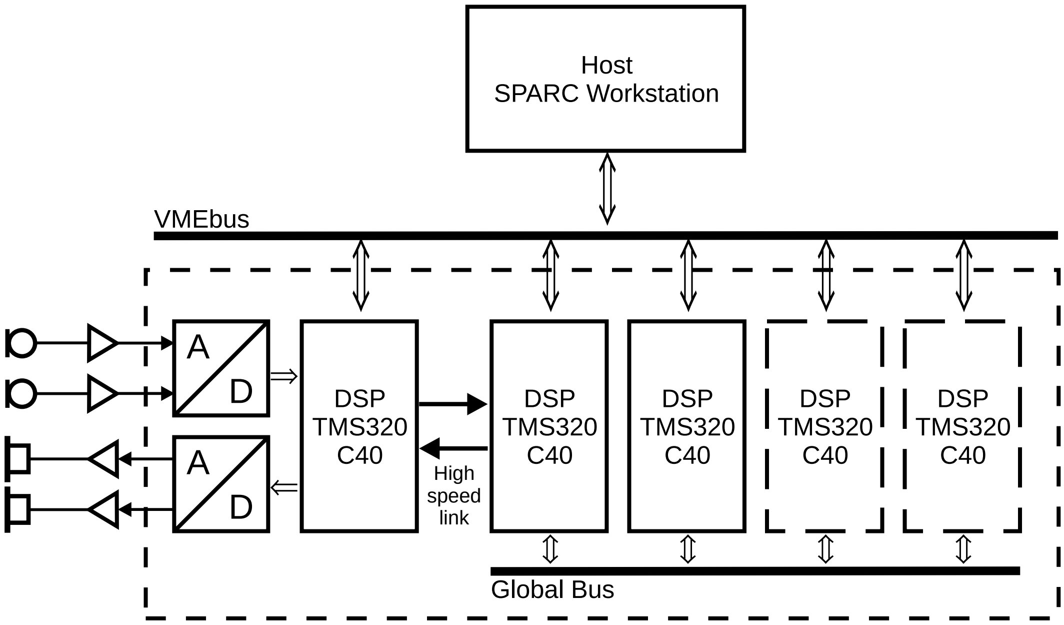


Figure 2.13: Block diagram of the hardware set-up employing VMEbus DSP (digital signal processor) boards, connected to AD/DA converters.

Two microphones are connected via amplifiers to the two-channel ADC. The single-DSP reads the converted time signals, computes the overlapping FFT (Fast Fourier Transform) analysis and sends the complex spectra via a high speed communication port to another signal processor. The single-DSP also receives processed spectra via another communication port and computes the inverse FFT. The time signals reconstructed by an overlap-add technique (Allen and Rabiner, 1977) are then written to the two-channel DAC, which is connected to the receivers via amplifiers. This single-DSP FFT analysis/synthesis system operates up to sampling frequencies of at least 48 kHz.

The hearing instrument algorithms are computed by the four-DSP board connected to the high speed communication ports. For the processing implemented so far, only two of the four DSP's are employed in the computations. The first receives the complex spectra, performs the algorithmic processing and retransmits the processed spectra. The second DSP performs all additional processing of parameters and data required, including data transfer from and to the controlling host workstation.

The signal-processing setup described above is connected to a pair of custom-built ITE (in-the-ear) hearing instruments. This is shown in Figure 2.14 for one ear. For the use in combination with the algorithms described here, ITE instruments are preferable over BTE (behind-the-ear) instruments, because the microphone signals to be recorded by the hearing instruments have to be affected by the head related transfer functions of the head and the outer ear. In applications with real subjects, the individual head-related transfer functions can be recorded with these devices and can be used in the algorithm for reference values. One problem with the ITE instrument in the present application, however, is the large expenditure for individually fabricating the devices, especially if a large number of subjects is involved. Thus, a pair of module ITE hearing aids is used (Siemens Cosmea M) which can be used with the individual earmolds of the subject.

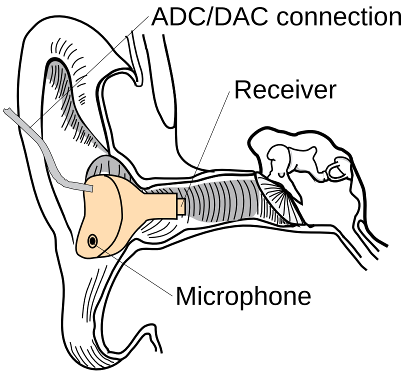


Figure 2.14: Application of an ITE (in-the-ear) hearing instrument to measurements with ADC and DAC connections. The individual earmold is not shown.

The general outline of the signal processing performed with the stationary multiple-DSP apparatus described above is given in Figure 2.15. In addition to a noise reduction technique like those described in Section 2.2, a loudness model-based approach (Hohmann and Kollmeier, 1996) is employed for dynamic compression. Such a combination of dynamic compression and noise reduction is necessary for the application as a (simulated) complete hearing aid. To test any component of the combination, each part of the combined algorithm may be switched on independently, although the dynamic compression part (multiple band gain control) is dependent on the loudness model. In addition, certain interaction effects between noise reduction and dynamic compression have to be considered (see discussion below). Although not all of the algorithms described here have yet been implemented and tested with the hardware and general software layout described above, first informal experiments have been performed with the system using the algorithm described in Section 2.2.1 and dynamic compression algorithms described by Hohmann and Kollmeier (1996) and Marzinzik *et al.* (1996). Besides demonstrating the noise-reduction and dynamic compression capabilities of the respective algorithms that have already been formally evaluated (see Section 2.2.1 and Marzinzik *et al.*, 1996, respectively), this real-time implementation also exhibits the potential of the directional filtering and dereverberation algorithm to suppress unwanted feedback: Any ringing of the hearing aid on one side is uncorrelated with the other side and also exhibits an "unwanted" interaural level and phase difference. Hence the respective frequency band is attenuated. This yields an enhanced feedback margin which was observed with the current implementation. Since these first results are very encouraging, it is assumed that future work will employ the system described here as a general hardware and software framework for the simulation, evaluation and optimization of future "intelligent" digital hearing aids.

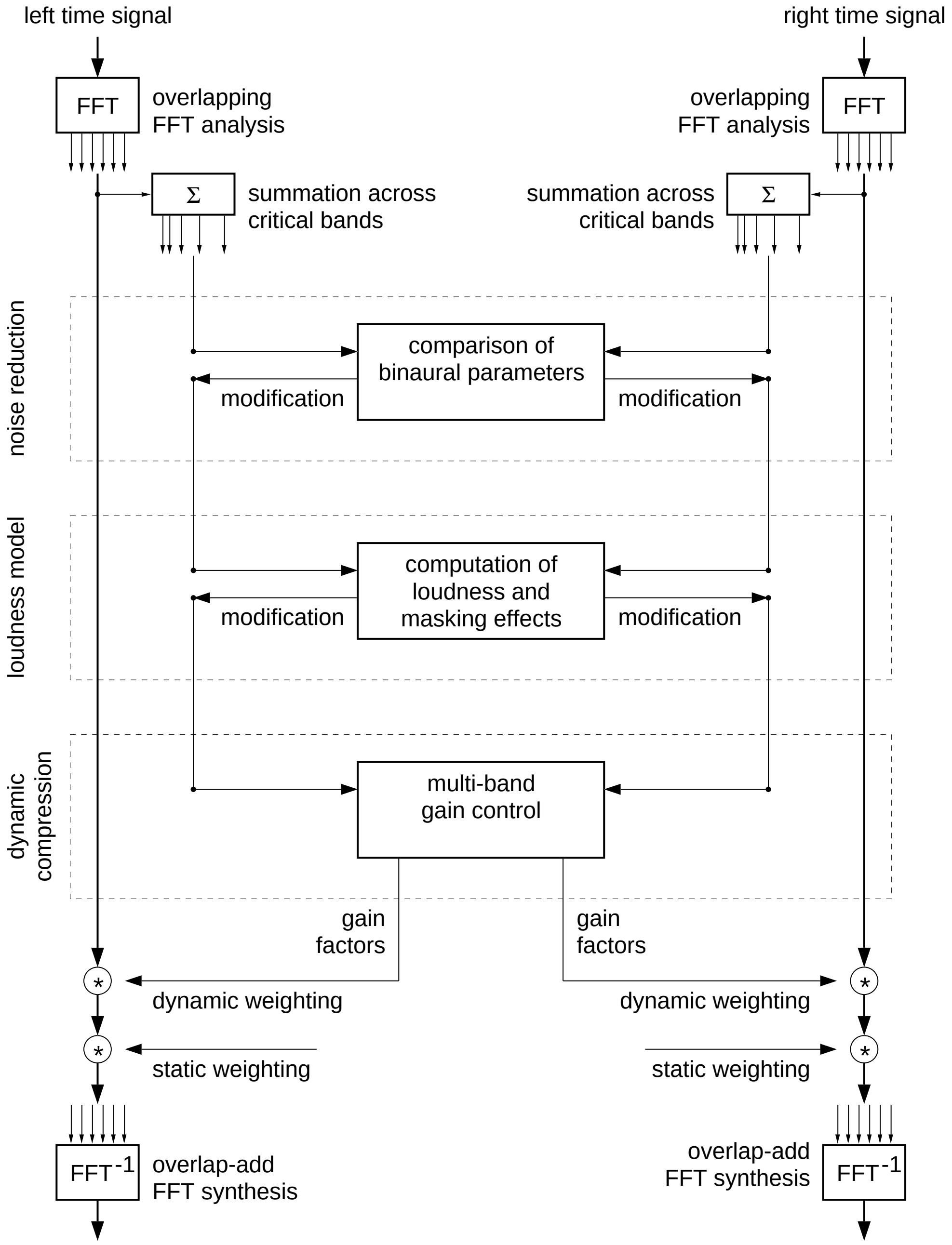


Figure 2.15: Processing scheme of the real-time hearing instrument algorithm.

2.5 Discussion

2.5.1 Relation between models and algorithms

Although the algorithms described here are motivated by psychoacoustical and physiological evidence of the "effective" signal processing for noise reduction in the human auditory system, none of the algorithms claims to be a model of human binaural processing (see Colburn, 1996, for a recent review of such models). Instead, the algorithms described here primarily incorporate the basic signal processing behind these models, such as, e.g., cross-correlating both input channels in each frequency band, exploiting interaural intensity and interaural time differences for gaining localization information about different sound sources, and using organizational or "scene analysis" properties such as analyzing modulation frequencies and combining different cues. An adequate combination of these algorithms might therefore be able to mimic certain aspects of the "effective" processing performed by the binaural system in everyday communication situations without explicitly trying to simulate these properties on a more detailed level. In this respect the approach presented here differs from the model described by Bodden (1993), who based his signal processing on a more detailed model of the binaural system that was originally developed for predicting localization and lateralization experiments. On the other hand, the computational effort required to perform the whole binaural noise reduction process is greatly reduced if not every psychoacoustical and physiological detail of the process is properly accounted for. In this respect, the current algorithms are a type of compromise between pragmatic aspects of signal processing on the one hand and theoretical models of binaural interaction on the other.

2.5.2 Advantages and drawbacks of the current algorithms

Noise reduction algorithms in general have to find a compromise between the magnitude of the noise reduction they yield and the artifacts they produce or the drawbacks that they might produce in situations where their underlying assumptions are not fulfilled. For the directional filtering algorithm (Sections 2.2.1 and 2.3), for example, the performance for a single interfering talker in a non-reverberant environment can be optimized to an attenuation of the jammer of tens of dB. However, the improved performance in this situation is accompanied by processing artifacts as soon as the sound field is more reverberant or if more than one interfering talker is present. Thus, the performance in the "ideal" situation has to be reduced in order to improve the performance in the "non-ideal" situation. Similar arguments hold when comparing the directional filter algorithm with the modulation filter algorithm in Section 2.3: Although the modulation filtering algorithm provides less benefit in the more favourable situations tested than the directional filtering algorithm, it still provides more benefit in the non-ideal situations with an increased number of interfering sound sources and with reverberation. The performance of the noise suppression system based on the localization model is not yet tested.

The fact that each algorithm appears to be optimal only for a certain range of signal-to-noise ratios, amount of reverberation and spatial distribution of target sound and jammers enhances the demand for an optimum combination of different algorithms. Clearly, such a combination can only be performed under the control of an algorithm that supervises which strategy should be employed in the current acoustical situation. The multiple-algorithm approach presented here (Section 2.2.4) appears to be a viable solution to the problem, although a formal test of this framework in combination with several different "binaural" algorithms has not yet been performed.

A severe problem of the algorithms described here is the computational power required for performing them. While the directional filtering algorithm and a simplified version of the localization model-based algorithm have been implemented to run in real time on a system with three multiple-purpose DSP's, the modulation filter algorithm and the combination of directional filtering with cepstral filtering (Sections 2.2.2 and 2.2.4) have only been implemented off line and require a few hundred times as long as real-time on a Sparc station 10. However, more efficient coding of the algorithm will already be sufficient to allow for an implementation on the multiple DSP system described in Section 2.4.3. Each algorithm might be assigned to a dedicated signal processor, so that the structure of this hardware system might already provide a good basis for implementing the "optimum" combination of these different algorithms. Since the computational speed of DSP's is constantly increasing and the power consumption for a given computational task is constantly decreasing, a real-time implementation of such an optimum combination of various algorithms for a wearable hearing aid appears to be feasible within the next few years. Such an implementation might not only be useful for hearing aids, but might as well be useful for, e.g., automatic speech recognition systems and telecommunication applications.

2.5.3 Combination with dynamic compression algorithms

The algorithms presented here only form one essential part of a hearing aid that has to be supplemented with additional features and algorithms to form a complete hearing instrument (such as, e.g., dynamic compression algorithm, algorithms for fitting the hearing aid to the individual patient, feedback cancellation algorithms, user control for different parameters or program options). While most of these additional components of hearing aids appear to be independent of the type of noise reduction employed, at least the interaction with possible types of dynamic compression has to be considered. To a first approximation, the dynamic compression algorithms can operate independently on the noise-reduced output signal of the binaural signal processing scheme. However, the noise reduction scheme might introduce low-level artifacts that will be increased in audibility by the subsequent dynamic compression system. In addition, the noise reduction scheme might try to restore components from the input signal as a part of the "desired" signal that will be regarded as inaudible by a perceptual masking model incorporated in the dynamic compression scheme (for example the scheme proposed by Hohmann and Kollmeier, 1996). Also, certain signal processing operations in the dynamic compression algorithms are the same as in the noise reduction schemes (for example, the signal analysis in auditory critical bands). Therefore, a close interaction between both types of signal processing and even the combined noise reduction/dynamic compression algorithm appears to be desirable. However, such an interacting combination of both types of algorithms might create new kinds of artifacts and unwanted side effects that have not been encountered with any of the noise reduction and dynamic compression algorithms performed in isolation. Therefore, more research and development have to be invested in this area. The hardware platform described in Section 2.4.3 appears to be a good basis for performing this work.

2.5.4 Future developments

Although most of the algorithms described here have been shown to work well in laboratory conditions and on stationary systems, the "effective" benefit they provide for hearing-impaired listeners have yet to be investigated. For this purpose, either a wearable unit has to be employed or a wireless communication link between the patient and a stationary processing apparatus in order for the patient to evaluate the signal processing scheme in as realistic conditions as possible. Hence, field tests using the directional filtering algorithm with a portable device described in Section 2.4.1 are currently performed in the framework of a joint research project with partners in Nürnberg, Giessen and Oldenburg. Grim *et al.* (1995) reported of field tests which had been performed with a device utilizing a directional filtering algorithm similar to that described in Section 2.2.1. Since the processing power required for the more complex algorithms described in Sections 2.2.2 - 2.2.4 is very high, it will take some time until wearable devices will be available that are capable of performing this type of processing. However, since the laboratory results look very promising and the current status of noise reduction hearing aids is relatively poor, it appears necessary to further develop these algorithms into a workable solution that should be available as soon as the hardware prerequisites are fulfilled for wearable hearing aids.

Chapter 3

Directional filtering and dereverberation: Parameter optimisation by paired comparison of sound quality

Abstract

A method is described to systematically investigate the influence of different parameters of a noise reduction algorithm on the subjectively perceived sound quality. The described method is based in a complete paired comparison of all different versions under different acoustical conditions. Each parameter or correlated pair of parameters is systematically varied while keeping the other parameters unchanged. The sample of particular values employed for each parameter is determined in advance with taking into account the influence of the values on audible differences. The algorithm investigated is based on two different strategies which employ the interaural differences in level and phase or the interaural coherence, respectively, of a binaural input signal. In some acoustical conditions, the combination of both strategies yielded poorer results than the single strategies alone in former quality assessment studies. However, the results of the quality assessment presented here indicate that a common set of parameters can be found which is appropriate for a variety of acoustical conditions. The results also indicate that an adaptation of the signal processing strategies to the actual acoustical situation might be useful to increase the overall quality.

3.1 Introduction

Noise reduction techniques, i.e. signal processing algorithms that aim at enhancing speech and at suppressing unwanted noise components in a given acoustical situation or recording, respectively, generally exhibit a trade-off between the reduction of the unwanted noise components and the preservation of the original speech. The former is usually related to the maximum attenuation produced by the employed filter function and the amount of its relative change across time. The preservation of the original speech on the other hand can be described as the absence of audible processing artefacts and is usually inversely related to the previously mentioned maximum amount of filtering. Because the human auditory system is not only very sensitive to processing artefacts produced by noise reduction processing but the subjectively perceived quality of the processed signal is also of crucial importance for, e.g., hearing aid users, the various parameters of a noise reduction algorithm that may influence this quality have to be selected in a careful way. Although there are some objective methods available that predict the perceived quality of speech processed by noise reduction algorithms (cf. Hansen and Pellom, 1998; Marzinzik and Kollmeier, 2000), these methods are not designed to work with any type of target signals or any type of interfering noise signals. Moreover, no common standards exist for the comparison of algorithm performance. Thus, an important issue of algorithm performance evaluation is the assessment of perceived quality with human subjects, which makes the optimisation of a great number of different parameters a very time-consuming task. The situation is even further complicated by the fact that the optimum parameter setting of a given algorithm may vary across listening conditions and listeners and that no common or unambiguous solution may be found.

The current study is therefore concerned with the derivation of an "optimum" parameter set for a given noise reduction algorithm in a variety of acoustical situations and with various listeners. The approach proposed here is

a) to only perform paired comparison judgements of the perceived quality because they are more sensitive to subtle quality differences than absolute quality judgements,

b) to perform a systematic variation of algorithmic parameters by always keeping one subset of parameters fixed and varying the other subset in a systematic way and

c) to perform the same experiments with both different acoustical situations and different individual subjects.

The algorithm employed for this purpose is the algorithm for directional filtering and dereverberation introduced by Peissig (1993) and Kollmeier *et al.* (1993). This algorithm has been suggested and tested for the use in binaural hearing aids, i.e. in an arrangement with two microphones placed near to or inside the right and the left ear, a central processing unit that receives inputs from both sides and that performs the noise suppression, and the output being supplied again to both ears. The advantage of such a setup is the "natural" sampling of the sound field at two points in space that are comparatively far away from each other without applying bulky devices to the patient's head, such as, e.g., broadside or endfire oriented microphone-arrays (cf. Soede *et al.*, 1993; Zurek *et al.*, 1996; see also Chapter 1). The algorithm investigated here is a combination of two different noise reduction strategies that are aimed at two different types of acoustical situations. The directional filtering strategy attenuates signals in each frequency band that exhibits different values of interaural phase difference and level difference than those expected from a certain range of reference directions (usually in front of the subject). This strategy has primarily been shown to be advantageous in anechoic or acoustically rather "dry" situations (Peissig, 1993). It generally fails in highly reverberant or comparatively diffuse acoustical situations. In such situations, however, the dereverberation strategy provides some (subjective) benefit by attenuating those frequency bands which exhibit a small value of interaural coherence, which is an indicator of a reverberant sound field. Hence, a combined algorithm appears promising. Such an algorithm has been implemented in real-time on a laboratory master hearing aid (Wittkop *et al.*, 1997) as well as on a wearable prototype signal processing hearing aid (Rass, 1996). Both systems allow for a quality assessment of the algorithm while simultaneously controling and changing its parameters in real-time.

Unfortunately, the experiences made within the first field tests using the wearable prototype hearing device showed that in comparison to the results obtained for each of the processing strategies alone, the signal quality of the combined directional filtering and dereverberation algorithm was perceived by the subjects as being relatively poor (cf. Albani *et al.*, 1998; Pastoors *et al.*, 1998). Along with this poor signal quality, the results of speech intelligibility measurements exhibited no improvement in speech intelligibility. These results contrasts with laboratory measurements performed by Kollmeier *et al.* (1993) using basically the same strategies. Possible reasons for the poor signal quality delivered in the field tests and the deviations from the laboratory test conditions are the different acoustical conditions, different devices and implementations, respectively, and perhaps inappropriately adjusted processing parameters employed in the field test. Under laboratory conditions, for instance, a processing strategy is tested for each acoustical situation with a particular set of parameters that is specially adapted for that particular situation. In field tests, however, a wider range of acoustical situations occur but the subject can choose only between a limited number of different parameter sets (which are supplied as different "processing programs" on the device). Moreover, a subject might even not select the most appropriate parameter set for a particular situation. Hence, it is important to know how different the "optimum" processing parameters should be for different acoustical situations and whether a parameter set can be found that fulfills the requirements of different acoustical situations. In addition, the subjective preferences may differ across subjects so that the influence of the subject on the rated quality of the hearing aid should be investigated. The current study therefore addresses these questions by performing subjective preference tests with several subjects using a variety of acoustical test situations.

3.2 Description of the algorithm

3.2.1 Original

Peissig (1993) and Kollmeier *et al.* (1993) described an algorithm which performs a noise reduction on binaural microphone input signals by combining a suppression of lateral sound sources and dereverberation. The algorithm described by Peissig (1993), however, directly employs the interaural differences in level and phase. Figure 3.1 sketches the algorithm and its application to binaural microphone signals.

In the original implementation, an overlap-add technique was used for digital frequency analysis and resynthesis (Allen and Rabiner, 1977) using an FFT of 512 samples with Hanning-windowed segments of 408 samples and an overlap rate of 0.5 at a sample rate of 25 kHz. In each frequency channel the average across time of the interaural differences in phase and level, and of the interaural coherence function at zero lag, are determined from the short-term autocorrelation in both channels and the cross-correlation between them.

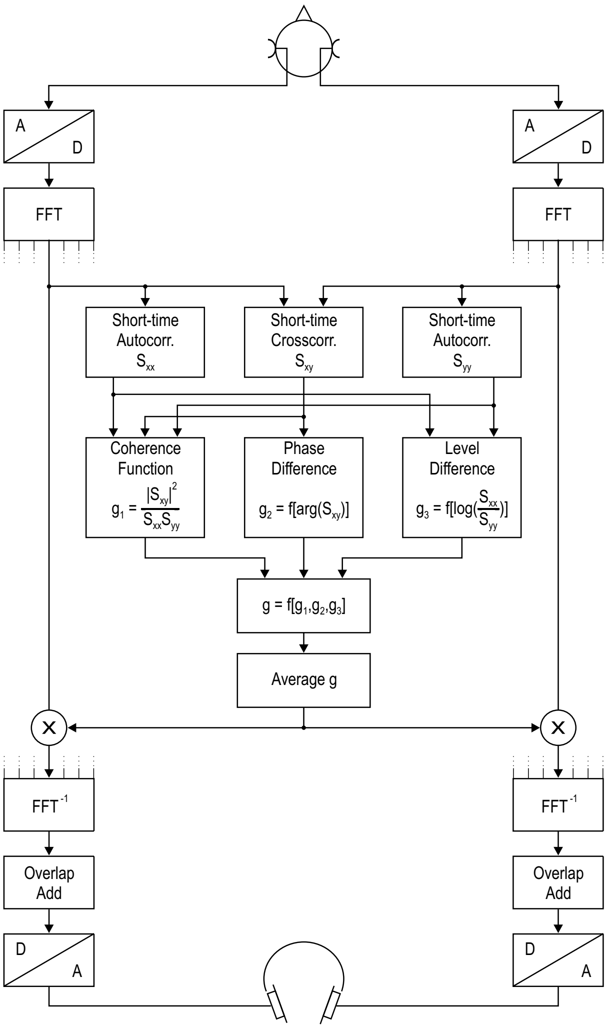

A
D
A
D
FFT
FFT
Short-time Autocorr. S_{xx}
Short-time Crosscorr. S_{xy}
Short-time Autocorr. S_{yy}
Coherence Function
$g_1 = \frac{|S_{xy}|^2}{S_{xx}S_{yy}}$
Phase Difference
$g_2 = f[\arg(S_{xy})]$
Level Difference
$g_3 = f[\log(\frac{S_{xx}}{S_{yy}})]$
$g = f[g_1, g_2, g_3]$
Average g
×
×
FFT^{-1}
FFT^{-1}
Overlap Add
Overlap Add
D
A
D
A

Figure 3.1: Block diagram of the algorithm for directional filtering and dereverberation after Peissig (1993).

In a successive stage, particular weighting factors g_1, g_2 and g_3 are derived for each frequency band. The factors g_2 and g_3, respectively, are calculated from the deviation of the current phase and level differences, respectively, from appropriate reference values. The reference values are obtained from recorded signals of noise sound sources with a particular reference incidence direction, e.g., from the front (i.e., the target is assumed to be in front of the listener). The amount of this deviation is expected to be correlated with the amount of the deviation of the current incidence direction from the reference incidence direction. Thus, phase and level differences that are in a certain range close to the reference values result in factors g_2 and g_3 near to unity. The resulting gain factors decrease with an increasing deviation of the current interaural differences from the reference values. Applying the gain factors g_2 and g_3 to the spectra yields a (frequency dependent) directionality and thus a suppression of lateral noise sources, which is described in detail by Peissig (1993). A similar concept was described by Gaik and Lindemann (1986).

Especially in diffuse sound fields, the interaural phase and intensity differences are not unambiguous indicators of a signal's direction of incidence. Thus, the interaural coherence function is used in addition as proposed by Allen *et al.* (1977) to decide whether the respective frequency channel contains a component of a direct incident sound (high interaural correlation) or diffuse sound or reverberation, respectively. A low interaural correlation coefficient is obtained for the diffuse (reverberant) part of a signal, which results in a low value of g_1, and thus an attenuation of the respective frequency channel.

Finally, the weighting factors g_1, g_2 and g_3 are combined to a total weighting factor g that is used to either suppress the respective frequency band or to leave it unchanged. A temporal and spectral average of the final factor g is performed to reduce the artifacts (see Peissig, 1993, for more details). The weighting factor is then applied to the spectra. Subsequently, the binaural time signals are reconstructed and presented via headphones.

3.2.2 Modifications and extensions

In the implementation employed in this study, an FFT of 512 samples is used with Hanning-windowed segments of 400 samples and an overlap rate of 0.5 at a sample rate of 25 kHz. In contrast to the original algorithm, the power spectra and the complex cross-power spectrum are then summed up across frequency within each critical band. This yields a non-linear frequency scale with 23 bands of 1 Bark bandwidth (cf. Zwicker, 1961). The sum across a critical band of a power spectrum simply yields the total energy, while the respective sum of a complex cross-power spectrum yields the magnitude-weighted mean phase difference as resulting phase and the cross-power sum as resulting magnitude. Both values revealed to be consistent and applicable quantities for further usage in the algorithm.

The evaluation of appropriate, frequency dependent reference values and the calculation of the phase and level gain factors g_2 and g_3 is depicted in Figure 3.2. The reference values are given by the mean interaural level differences $\Delta L(f)$ and phase differences $\Delta\varphi(f)$. They are calculated for the azimuthal angles $0°$, α_{pass} and α_{stop} and denoted as $\langle\Delta L(f)\rangle_{\text{angle}}$ and $\langle\Delta\varphi(f)\rangle_{\text{angle}}$ for the respective angle. These reference values define a certain pass, transition and stop range for the actual values of $\Delta L(f)$ and $\Delta\varphi(f)$. The gain factors are then given by the function $f(x, a, b)$. For the level gain factors, x is $|\Delta L(f) - \langle\Delta L(f)\rangle_{0°}|$

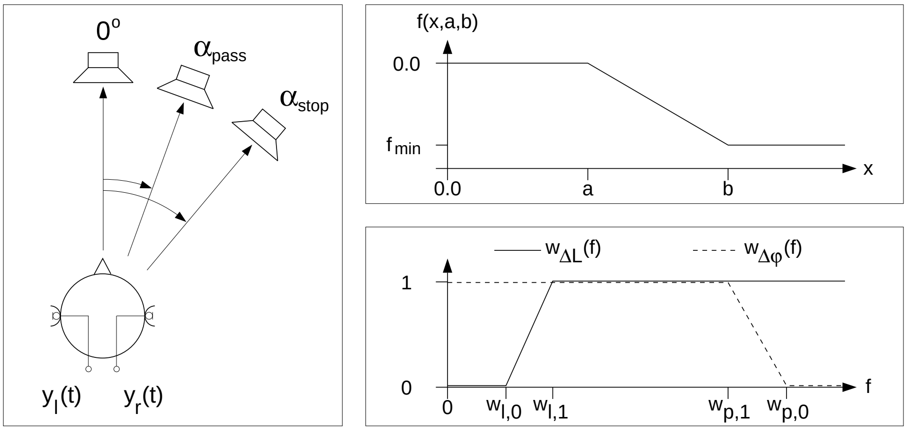


Figure 3.2: In the left panel, the spatial configuration for obtaining the reference values is shown. The sound sources are located at the azimuthal angle of 0°, α_{pass} and α_{stop} at zero degrees elevation. In the upper right panel, the gain function $f(x, a, b)$ employed for computing the level and phase gain factors is depicted. f_{min} denotes the adjustable minimum gain factor. $f(x, a, b)$ is calculated in dB, see text for details. In the lower right panel, the frequency dependent weighting functions $w_{\Delta L}$ and $w_{\Delta\varphi}$ of the level and the phase gains are depicted, where $w_{l,0}$, $w_{l,1}$, $w_{p,1}$ and $w_{p,0}$ are the adjustable frequency limits.

and a and b are $\left|\langle\Delta L(f)\rangle_{\alpha_{\text{pass}}} - \langle\Delta L(f)\rangle_{0^\circ}\right|$ and $\left|\langle\Delta L(f)\rangle_{\alpha_{\text{stop}}} - \langle\Delta L(f)\rangle_{0^\circ}\right|$, respectively. The phase gain factors are calculated in the same way employing $\Delta\varphi(f)$ instead. Since at very low frequencies, the interaural level differences are negligible, and at high frequencies, the interaural phase differences comprise the whole range of $[-\pi; +\pi]$ (due to interaural time differences greater than half a wave cycle and phase wrapping), a frequency dependent weighting of the gain factors g_2 and g_3 is employed. This results in a combined gain factor $g_{2,3}(f) \equiv \frac{w_{\Delta\varphi}(f)\cdot g_2(f) + w_{\Delta L}(f)\cdot g_3(f)}{w_{\Delta\varphi}(f) + w_{\Delta L}(f)}$.

This implementation yields a directionality pattern which is well defined and very similar across frequency (with some exceptions which are discussed below). The directionality is depicted in Figure 3.3 for selected frequency bands of 4 Bark distance[1].

The directionality in general shows a front-backward symmetry, which is due to the ambiguities of interaural level and phase differences for frontal and backward directions. The above mentioned dissimilarities across frequency are also due to such ambiguities. For example, low attenuation is observed for directions about 90° and 270° at 10 Bark and 90° at 14 Bark. Additionally, a broader directionality is observed for backward directions at frequencies between 14 and 18 Bark. The ambiguities in interaural level differences that produce the respective, low attenuation can be seen in Figure 3.4 which shows the different levels of the left and right ear signal as a function of the incidence direction. For the respective lateral incidence directions and frequencies, these level differences exhibit values which are similar to those for the frontal incidence directions.

[1] The 22 Bark band is already above the cutoff frequencies of the employed microphones and provides no significant data.

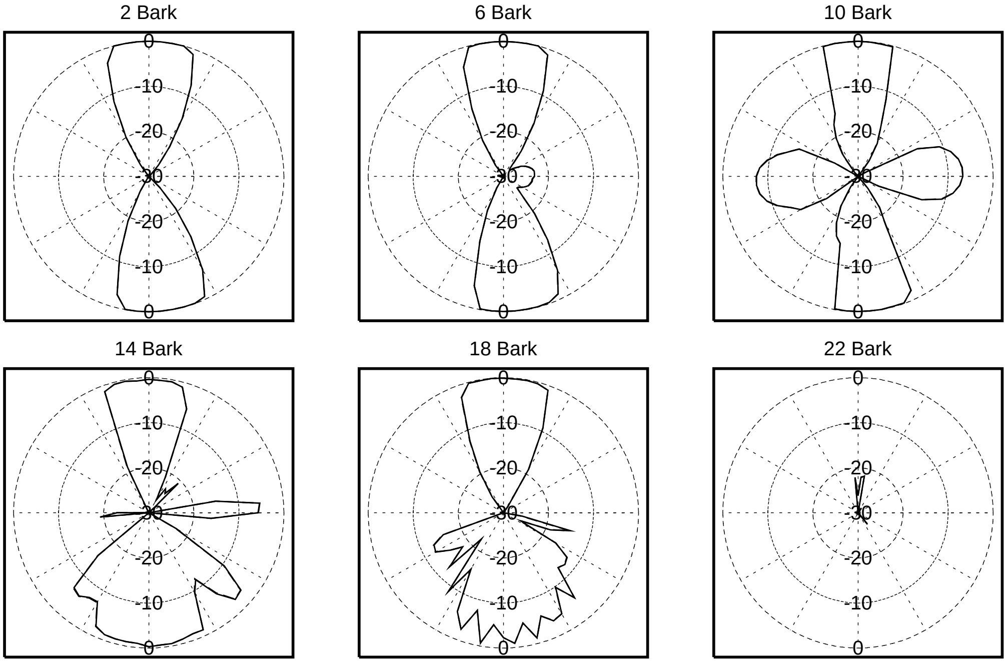


Figure 3.3: Narrow-band directionality of the modified algorithm for directional filtering and dereverberation obtained with the Göttingen dummy head in an anechoic chamber. The center frequencies of the bands with a bandwidth of 1 Bark are given above each panel. The resulting attenuations produced by the algorithm are given as a function of the direction of sound incidence. The radius, i.e., the distance from the outermost circle, gives the attenuation in dB. The numbers placed at the grid circles denote the respective attenuation represented by that particular radius. The azimuth angle is counted clockwise starting with 0° (frontal) at the top, 90° at the right, 180° (backward) at the bottom and 270° at the left side of the plot (think of the head depicted in the left panel of Figure 3.2 placed in the center). The employed reference pass and stop range angles α_{pass} and α_{stop}, respectively (see Figure 3.2), were 20 and 40 degrees, respectively.

The modified algorithm is sketched in Figure 3.5. The parameters which have been systematically varied and tested during the experiments are shown in Table 3.1.

3.3 Method

3.3.1 Procedure and subjects

The subjectively perceived quality of processed speech signals was compared across different processing conditions for the algorithm described above. A complete paired comparison of all different versions was performed by each subject, i.e., the subject compared each version with each other version exactly once. In each trial, the subject was forced to decide which version was perceived to be of better quality, i.e., which version was preferred to listen to.

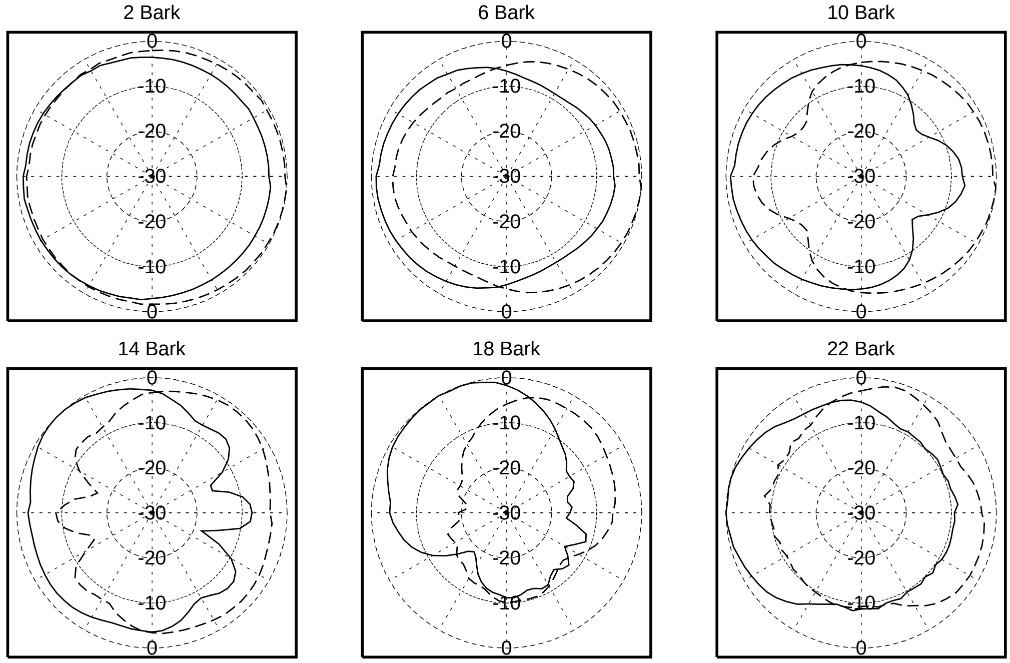


Figure 3.4: Narrow-band levels of the left (solid lines) and right ear (dashed lines) signals obtained with the Göttingen dummy head in an anechoic chamber. The center frequencies of the bands with a bandwidth of 1 Bark are given above each panel. The deviations of the levels (intensities) from the maximum level across all directions and both sides are given as a function of the direction of sound incidence. The radius, i.e., the distance from the outermost circle, gives the deviation in dB. The numbers placed at the grid circles denote the respective difference represented by that particular radius. The interaural level difference is the radial distance between dashed and solid line. For the spatial configuration, see Figure 3.3.

A judgement of equal quality was not allowed. Each subject was instructed in the same way. The order of presentation for all paired versions and within each pair was randomized independently for each subject. The aim was to vary a particular parameter or a set of correlated parameters of the algorithm within each measurement.

In this study, the tested parameter sets were $(\alpha_{\text{pass}}, \alpha_{\text{stop}})$, (τ_1, τ_2) and (a_1, a_2). Since the number N_c of paired comparisons increases rapidly with an increasing number N of different versions $\left(N_c = \frac{1}{2}N(N-1)\right)$, an informal selection of appropriate values was performed in advance for each parameter in order to eliminate values of low evidence and to restrict the total number of different versions. This resulted in a range of 14 to 19 different versions and thus 91 to 171 comparisons per measurement.

Between 10 and 11 clinically normal hearing subjects participated voluntarily in each measurement. Some subjects received an expenditure compensation on an hourly basis. They were aged between 20 and 30 years and all had experience in psychoacoustical measurements.

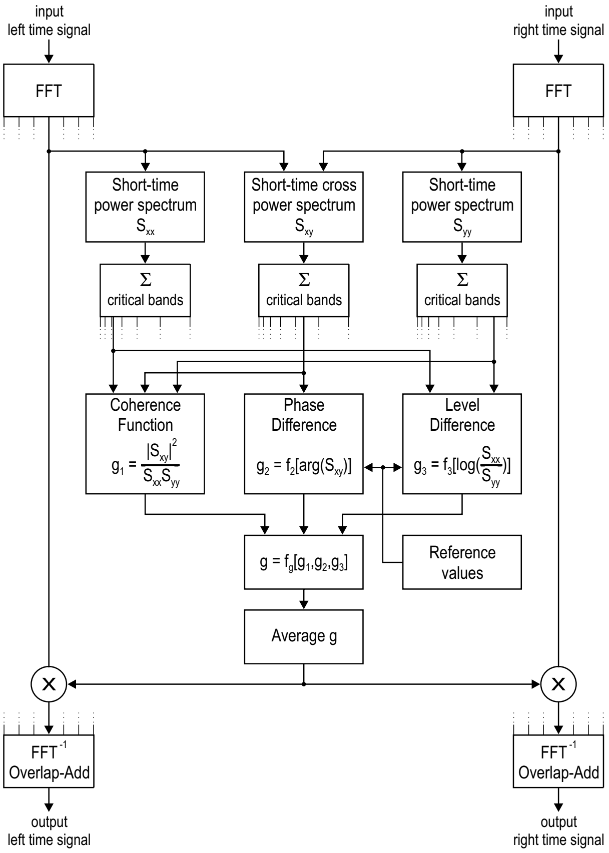


Figure 3.5: Block diagram of the modified algorithm for directional filtering and dereverberation.

Parameter	Description
α_{pass}	Azimuthal angle where pass range ends and transition range starts
α_{stop}	Azimuthal angle where transition range ends and stop range starts
τ_1	Time constant of the first order recursive low pass filter used for a temporal smoothing of the spectral data S_{xx}, S_{yy} and S_{xy}
τ_2	Time constant of the first order recursive low pass filter used for a temporal smoothing of the total gain g
a_1	Maximum attenuation due to the interaural coherence function, i.e., the minimum value of g_1 (negative values in dB denote an attenuation)
a_2	Maximum attenuation due to the interaural phase and intensity differences, i.e., the minimum value of g_2 and g_3 (negative values in dB denote an attenuation)

Table 3.1: Algorithm parameters varied in the experiments.

3.3.2 Apparatus and stimuli

Figure 3.6 shows the spatial configuration of the three employed stimuli **s1**, **s5** and **s6**. All stimuli were recorded dichotically using ITE (In-The-Ear) hearing instruments, worn by a male subject ("central talker") who participated in a conversation. The employed module hearing instruments were Siemens Cosmea M devices with normal microphones and mounted to common ear moulds. The microphones of the hearing instruments were connected to a DAT recorder during the recording. Each stimulus consists of a conversation between the central talker and a male target talker sitting in front at a distance of about one meter. In stimulus **s1**, the conversation takes place in quiet inside a regular seminar room (reverberation time about 0.5 seconds). In stimulus **s5**, an additional interfering female talker utters text passages from a book and is moving slowly within an azimuthal angle range of 45 to 90 degrees with respect to the central talker. In stimulus **s6**, the conversation takes place in a cafeteria during lunch time with a loud and mainly diffuse background noise.

For the assessment, the original signals were loaded from hard disk during the measurement and processed in realtime by a DSP subsystem with five TI TMS320C40 digital signal processors. The processed signals were presented dichotically via amplifier and headphones (Sennheiser HD25) in a sound-insulated booth. For the stimuli **s1** and **s5**, the presentation level was in the range of 65 to 70 dB SPL (coupler measurements) with a signal-to-noise ratio (SNR) of about 0 dB. For stimulus **s6**, the presentation level was up to 78 dB SPL with an SNR of about -10 dB.

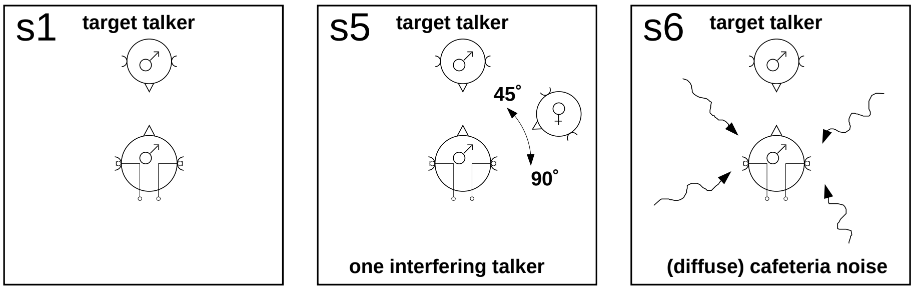
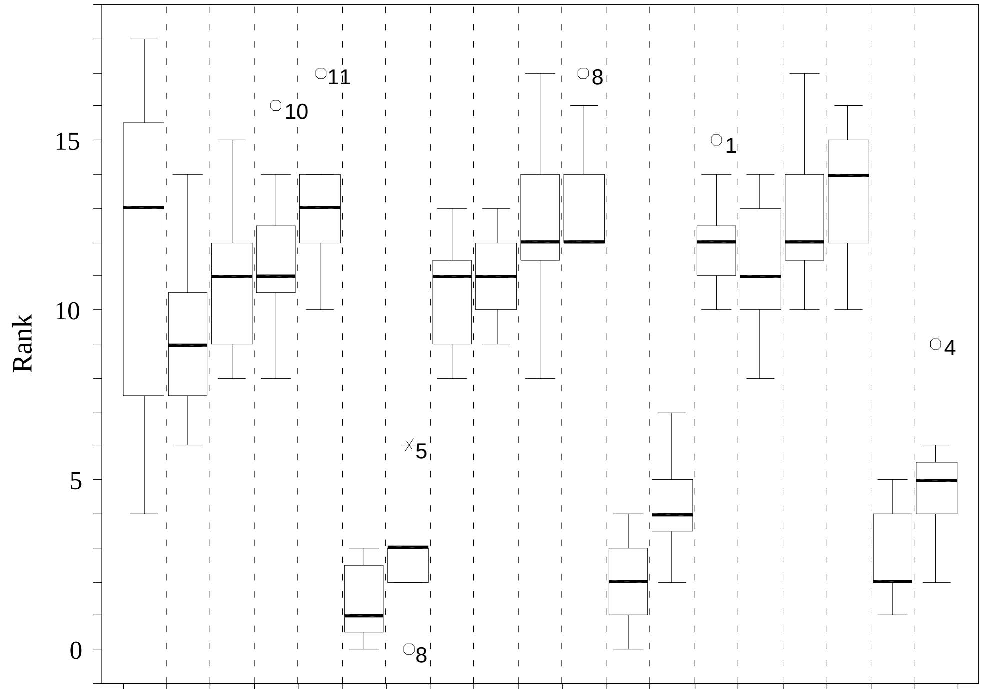


Figure 3.6: Spatial configuration of stimuli **s1**, **s5** and **s6**.

For the binaural processing within the algorithm, appropriate reference values for level and phase differences of different sound incidence directions were required. They were obtained in an anechoic chamber with the same central talker and ITE hearing instruments employed for all signal recordings. All signals were processed using these reference values and presented without further frequency shaping. It was assumed that the frequency response of the whole system, being the same for all presentations, did not affect the relations of the paired comparison judgements. At the beginning of each paired comparison, the signal (about 1 minute of running speech) was presented in an endless loop, starting with the first type of processing switched on. The subject was able to switch the processing type whenever she or he liked to using a handheld touchscreen response box (EPSON EHT-10S), selecting choice "1" or "2". This switching was put into effect without a considerable delay or an interruption of the stimulus presentation. With the processing judged to be of higher quality switched on, the selection of the third choice "better" ended the comparison task.

3.4 Results

3.4.1 Measurement A: Directionality



#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
a_1		-99	-30	-99	-30	-99	-30	-99	-30	-99	-30	-99	-30	-99	-30	-99	-30	-99	-30
a_2		-30																	
τ_1/τ_2		1/60		1/100		60/1		1/60		1/100		60/1		1/60		1/100		60/1	
$\alpha_{\text{pass}}/\alpha_{\text{stop}}$		10/30						20/40						30/50					

Figure 3.7: Relative ranks of versions obtained from measurement **A** (stimulus **s6**). On the abscissa, the 19 different versions of the stimulus are shown. The number of the version is given directly below the axis in the row denoted with a # on the left. Version number 1 is the unprocessed stimulus. For the other versions, the value of the processing parameters a_1 and a_2 and the parameter pairs τ_1/τ_2 and $\alpha_{\text{pass}}/\alpha_{\text{stop}}$ are given in the accordingly denoted rows below. The vertical lines in the rows separate different values. The ordinate gives the rank, i.e., the number of "better" judgements with a maximum possible value of 18. The thick horizontal lines denote the median values, the boxes the range from the first to the third quartile and the outer bars the total range. Circles and asterisks represent outlyers (with the number of the respective subject specified).

This measurement was performed mainly to compare different values of the directionality parameters α_{pass} and α_{stop}. The stimulus was **s6**. Tested values of the parameter set $(\alpha_{\text{pass}}, \alpha_{\text{stop}})$ in degree were $(10, 30)$, $(20, 40)$ and $(30, 40)$, respectively. In addition, a predetermined set of other parameters was employed. The tested values of the parameter set (τ_1, τ_2) in milliseconds were $(1, 60)$, $(1, 100)$ and $(60, 1)$, respectively, and of the parameter set (a_1, a_2) in dB were $(-99, -30)$ and $(-30, -30)$, respectively. All combinations plus the unprocessed version resulted in 19 different versions and thus 171 paired comparisons.

11 subjects participated in this measurement. First, the consistence of the results, i.e., the consistence of all "better" judgements with each other was calculated for each subject. For this, the method of Kendall (1975), described by Bortz *et al.* (1990) was employed. The results of all subjects exhibited significantly consistent results ($\alpha < 0.005$) and were included in the further evaluation. Additionally, the agreement of the judgements across the subjects was evaluated with the method described by the above authors. The resulting coefficient of agreement was $A = 0.43$ with significance level $\alpha < 0.001$, i.e., the agreement of the judgements from different subjects is significantly higher than for judgements obtained at random. A Friedman test revealed a significant influence of the processing version on the results ($\alpha < 0.001$). The results of measurement **A** are depicted in Figure 3.7.

There might be a small tendency that broader directionalities were ranked higher than narrower ones, and that lower maximum attenuations were ranked higher than higher ones, but these tendencies were not significant. A Wilcoxon test was performed for each pair of versions to determine significant differences in the results, and no significant influence of the directionality parameters α_{pass} and α_{stop} was found ($\alpha > 0.05$). For the time constants (τ_1, τ_2), however, the values $(60, 1)$ were significantly ranked lower than any other values ($\alpha < 0.005$).

These finding indicate that for the reverberant stimulus **s6**, the selection of the reference directions is not critical with respect to the quality. On the other hand, small values of τ_2 should be avoided in this situation because they result in audible processing artefacts which have a negative effect on the perceived quality.

3.4.2 Measurement B,C,D: Time constants

These measurements were performed to compare different values of the time constants τ_1 and τ_2. The employed stimuli were **s6** (measurement **B**), **s5** (measurement **C**) and **s1** (measurement **D**). Tested values of the parameter set (τ_1, τ_2) in milliseconds were $(1, 60)$, $(1, 100)$, $(1, 500L)$, $(8, 60)$, $(8, 100)$, $(8, 500L)$, $(20, 60)$, $(20, 100)$, $(20, 500L)$, $(60, 20)$, $(60, 100)$, $(60, 500L)$ and $(100, 20)$. The value of $500L$ for τ_2 denotes a level dependent time constant with a maximum value of 500 milliseconds at an input level in the range of the microphone and speaker noise floor. The value decreases with increasing input level, down to a minimum value of 100 milliseconds at 40 dB above the noise floor. For the other parameters, fixed values were employed. The tested values of the parameter set $(\alpha_{\text{pass}}, \alpha_{\text{stop}})$ in degree were $(20, 40)$, and for the parameter set (a_1, a_2) in dB $(-30, -30)$, respectively. All combinations plus the unprocessed version resulted in 14 different versions and thus 91 paired comparisons per stimulus.

10 subjects participated in each measurement. 12 subjects were involved in total, while 8 subjects participated in all three measurements. Again, the consistence of the results, i.e., the consistence of all "better" judgements with each other was calculated first for each subject. Only 7 (**B**), 8 (**C**) and 9 subjects (**D**), respectively, exhibited significantly consistent results ($\alpha < 0.05$) and were included in the further evaluation. The resulting coefficients of agreement for these subjects were $A = 0.33$ (**B**, $\alpha < 0.001$), $A = 0.08$ (**C**, $\alpha < 0.001$) and $A = 0.03$ (**D**, $\alpha > 0.05$). Although the agreement is low for stimulus **s5**

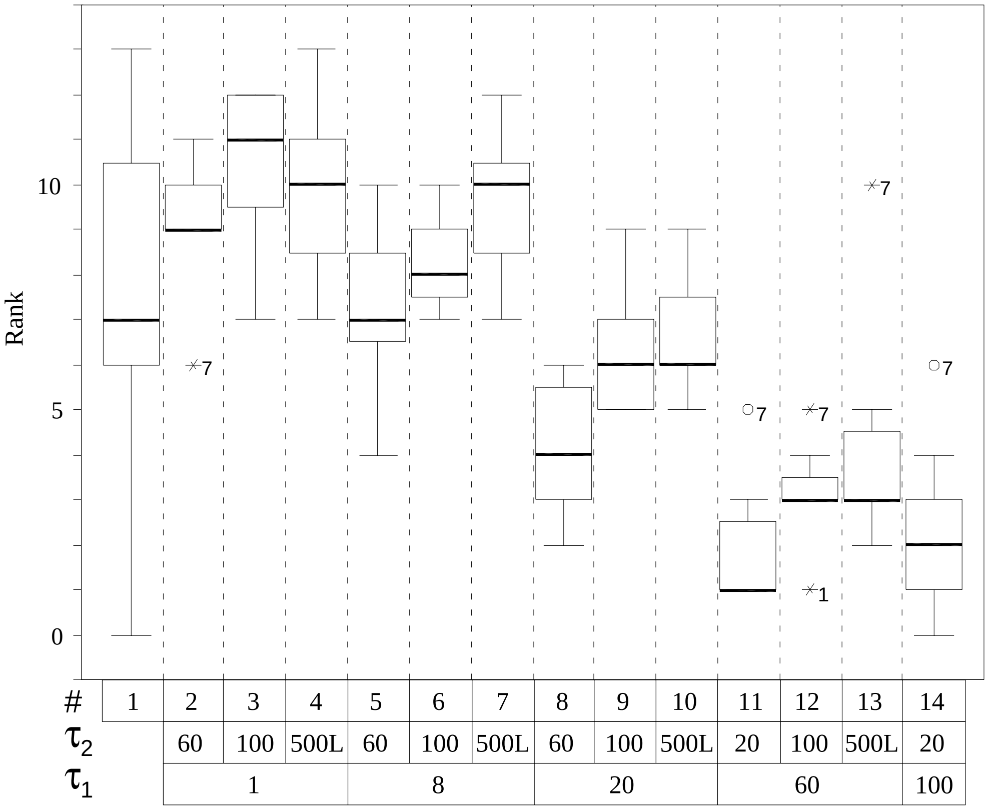
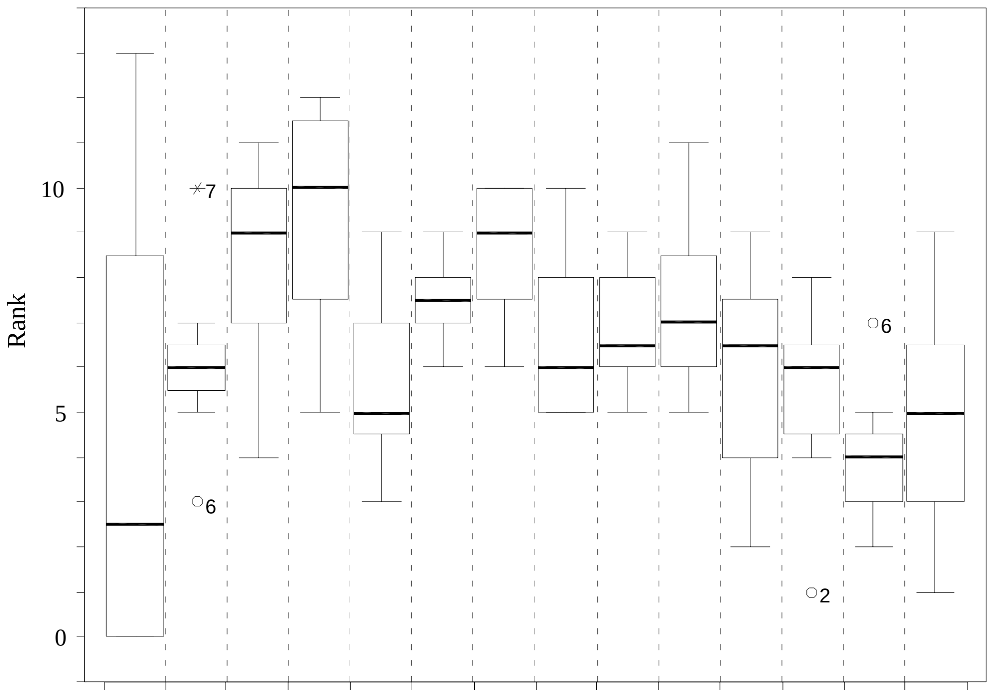


Figure 3.8: Relative ranks of versions obtained from measurement **B** (stimulus **s6**). On the abscissa, the 14 different versions of the stimulus are shown. The meaning of the rows below the axis is similar to Figure 3.7 with τ_2 and τ_1 being the varied parameters here. The ordinate gives the rank in the same way as in Figure 3.7 with a maximum possible rank of 13 in this case.

(**C**), it is still significantly higher than for random judgements. The results of measurement **B**, **C** and **D** are depicted in Figure 3.8, 3.9 and 3.10, respectively. Friedman tests revealed a significant influence of the processing version on the results for stimuli **s6** (**B**) and **s5** (**C**) ($\alpha < 0.005$), but no significant influence for stimulus **s1** (**D**) ($\alpha > 0.05$).

There is a tendency in the results for stimuli **s6** and **s5** that small values of τ_1 were ranked higher than larger ones and that larger values of τ_2 were ranked higher than small ones. A Wilcoxon test was performed for each pair of versions to determine significant differences in the results. For stimulus **s6** (**B**), no significant difference was found between all settings of $(1, \tau_2)$ and $(8, \tau_2)$ ($\alpha > 0.05$ for version numbers 2 - 7, except for number $5 = (8, 60)$). They were all ranked high, and version 3 $(1, 100)$ was significantly ranked higher than the unprocessed version ($\alpha < 0.05$). The settings $(20, \tau_2)$ were ranked inhomogeneously, but rather lower, while the settings $(60, \tau_2)$ and $(100, \tau_2)$ were significantly ranked lower than the settings of $(1, \tau_2)$ and $(8, \tau_2)$ (α at least < 0.05). Hence, the value of τ_1 should not exceed 8 ms in this diffuse situation. For stimulus **s5** (**C**), only few significant differences were found. All processed versions were ranked higher than the unprocessed



#	1	2	3	4	5	6	7	8	9	10	11	12	13	14
τ_2		60	100	500L	60	100	500L	60	100	500L	20	100	500L	20
τ_1		1			8			20			60			100

Figure 3.9: Relative ranks of versions obtained from measurement **C** (stimulus **s5**). See Figure 3.8 for details.

version, with a significant difference for versions 4 $(1, 500L)$ and 7 $(8, 500L)$ $(\alpha < 0.05)$. For stimulus **s1** (**D**), no significant influence at all was found for the different versions. Thus, the results of measurement **D** will not be considered further. However, all processed versions except for one were ranked higher than the unprocessed version.

Taken together, it appears that combinations of a small value of τ_1 ($\leq$ 8 ms) and a comparatively large value of τ_2 ($\geq$ 100 ms) are appropriate for complex acoustical situations.

3.4.3 Measurement E,F: Maximum attenuation

These measurements were performed mainly to compare different values of the maximum attenuations a_1 and a_2. The employed stimuli were **s6** (measurement **E**) and **s5** (measurement **F**). Tested values of the parameter set (a_1, a_2) in dB were $(-99, -30)$, $(-30, -30)$, $(-20, -20)$ and $(-10, -10)$, respectively. In addition, selected values of other parameters were employed. The tested values of the parameter set (τ_1, τ_2) in milliseconds were $(1, 60)$ and $(1, 100)$, respectively, and of the parameter set $(\alpha_{\text{pass}}, \alpha_{\text{stop}})$ in degree were $(10, 30)$ and $(30, 50)$, respectively. All combinations plus the unprocessed version resulted in 17 different versions and thus 136 paired comparisons per stimulus.

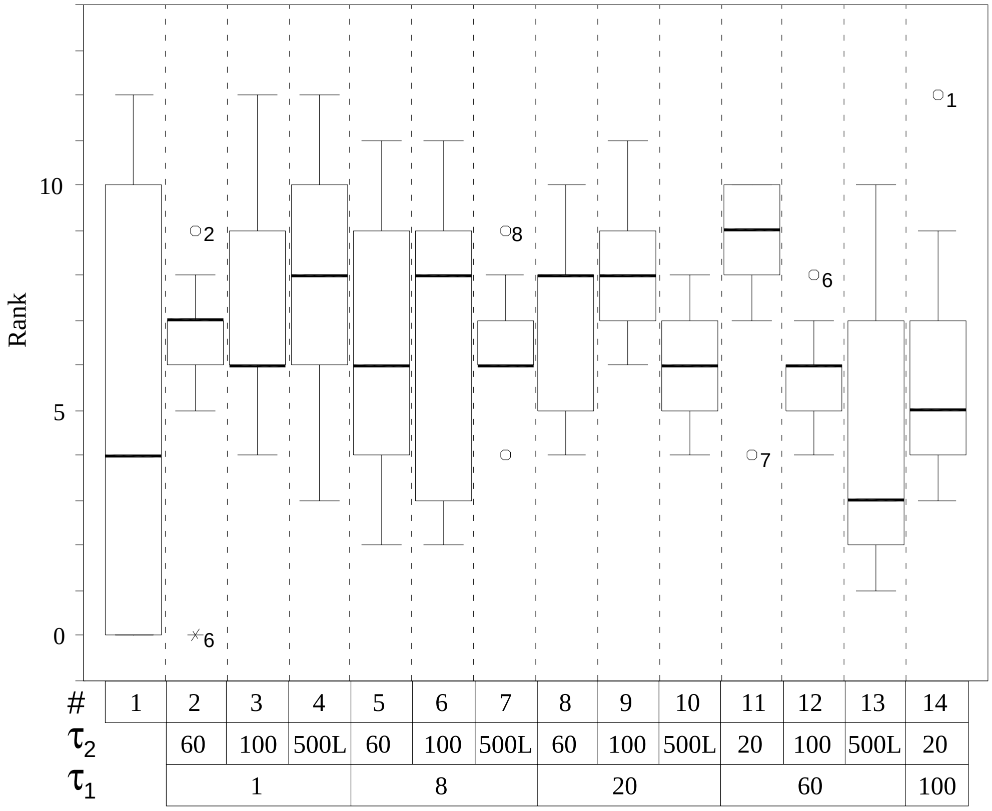
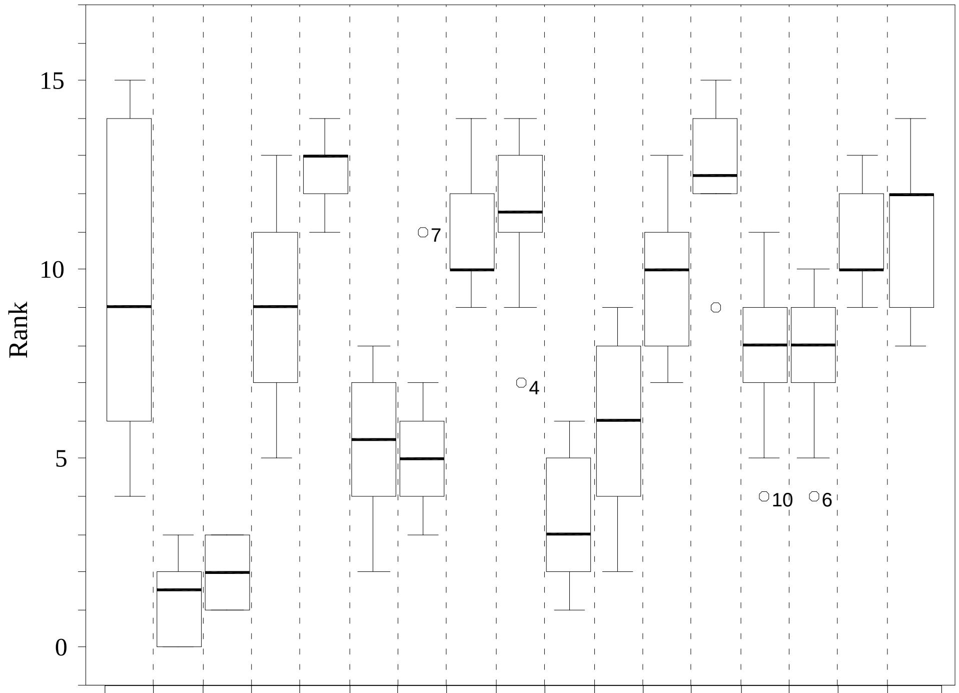


Figure 3.10: Relative ranks of versions obtained from measurement **D** (stimulus **s1**). See Figure 3.8 for details.

10 subjects participated in each measurement. 11 subjects were involved in total, while 9 subjects participated in both measurements. Again, the consistence of the results, i.e., the consistence of all "better" judgements with each other was calculated first for each subject. The results of all subjects exhibited significantly consistent results ($\alpha < 0.005$) and were included in the further evaluation. The resulting coefficients of agreement for the subjects were $A = 0.36$ (**E**, $\alpha < 0.001$) and $A = 0.13$ (**F**, $\alpha < 0.001$). Again, the agreement is low for stimulus **s5** (**F**), but still sigificantly higher than for random judgements. The results of measurement **E** and **F** are depicted in Figure 3.11 and 3.12, respectively. Friedman tests revealed a significant influence of the processing version on the results for both stimuli **s6** (**E**) and **s5** (**F**) ($\alpha < 0.001$).

A Wilcoxon test was performed for each pair of versions to determine significant differences in the results. For stimulus **s6** (**E**), any other parameter setting combined with the greater maximum attenuations $(-99, -30)$ or $(-30, -30)$, respectively, was significantly ranked lower than the respective parameter setting combined with the less maximum attenuation $(-20, -20)$ or $(-10, -10)$, respectively ($\alpha < 0.01$, for versions 15 and 17 is $\alpha < 0.05$). In some cases (versions 4 and 5, 10 and 11, 12 and 13), even $(-10, -10)$ was significantly ranked higher than $(-20, -20)$ ($\alpha < 0.05$). This is a clear indication that perceived quality increases with a decreasing maximum attenuation. Except for the versions



#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
a_1		-99	-30	-20	-10	-99	-30	-20	-10	-99	-30	-20	-10	-99	-30	-20	-10
a_2		-30	-30	-20	-10	-30	-30	-20	-10	-30	-30	-20	-10	-30	-30	-20	-10
τ_1/τ_2		1/60				1/100				1/60				1/100			
$\alpha_{pass}/\alpha_{stop}$		10/30								30/50							

Figure 3.11: Relative ranks of versions obtained from measurement **E** (stimulus **s6**). On the abscissa, the 17 different versions of the stimulus are shown. The meaning of the rows below the axis is similar to Figure 3.7 with the same parameters being the varied here. The ordinate gives the rank in the same way as in Figure 3.7 with a maximum possible rank of 16 in this case.

which were ranked very low (2 and 10, 3 and 11) and versions 7 and 15, no significant difference between the results of different directionality parameters α_{pass} and α_{stop} was found here. In the few cases of significant difference, however, the broader directionality was again ranked higher than the smaller one. There is also again a tendency that the greater value of τ_2 (100 ms) is ranked higher than the smaller one (60 ms). Version 13 was significantly ranked higher than the unprocessed version ($\alpha < 0.05$). For stimulus **s5** (**F**), the results are not as clear as for stimulus **s6**. At least, versions 2 and 3 were significantly ranked lower than any other processed version ($\alpha < 0.05$), which fits into all of the tendencies mentioned above for stimulus **s6**. It should be noted that the versions 5, 9, 13 and 16 were significantly ranked higher than the unprocessed version ($\alpha < 0.05$). This also indicates a high quality of small maximum attenuations. The unprocessed version was again ranked very low.

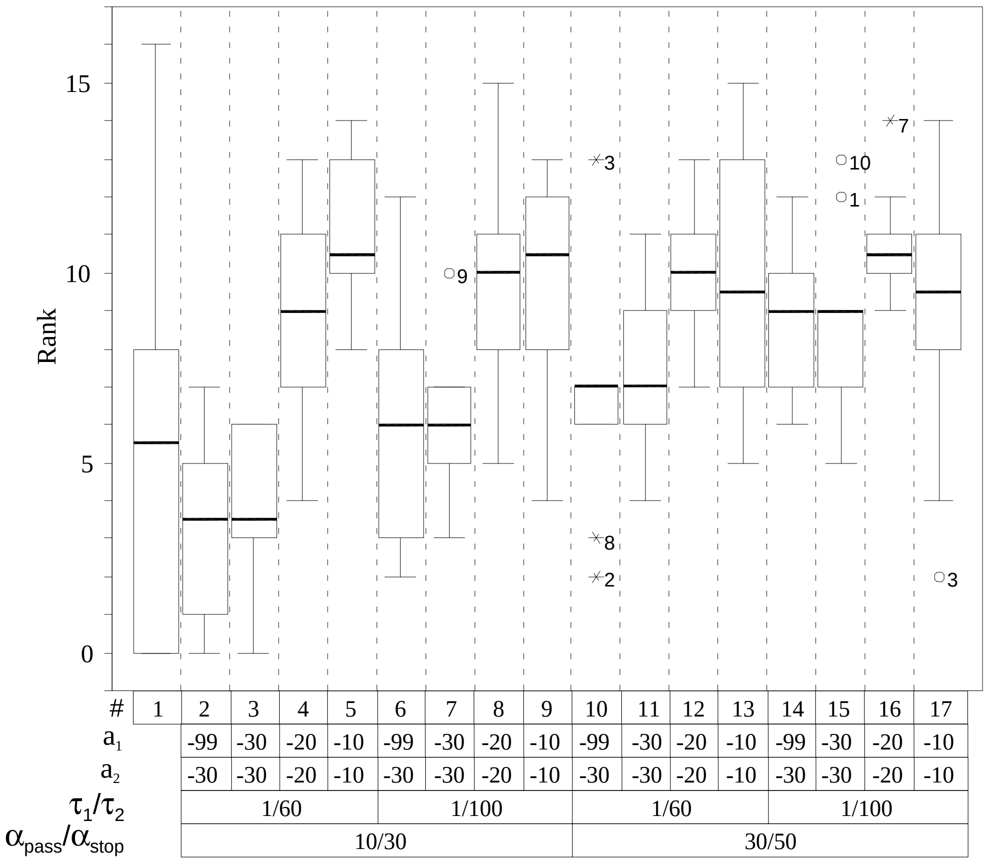


Figure 3.12: Relative ranks of versions obtained from measurement **F** (stimulus **s5**). See Figure 3.11 for details.

3.5 Discussion

The main results of the current study can be summarized as follows:

- For most subjects and conditions, the paired comparison judgements were consistent both within subjects and across subjects. Hence, rank order scales could be constructed that help to identify the most preferable and the least preferable situations according to the subjects judgements.

- Only little interaction across the different parameters was found with respect to the shape of the preference function. This indicates that the optimum value of each parameter is comparatively independent from the respective values of the other parameters.

- The optimum set of parameters is virtually independent from the acoustical situation and the listener. This does not mean that this optimum parameter set yields similar absolute results for e.g. different acoustical situations.

- While the selection of the reference direction is not very crucial for the subjective assessment of the algorithm, the optimum values of the other parameters appeared to be $\tau_1 \leq 8$ ms and $\tau_2 \geq 100$ ms for the time constants and $a_1 \leq -20$ dB and $a_2 \leq -20$ dB for the maximum attenuations.

The results clearly demonstrate that the measurement method employed here (i.e. paired comparison with perceived quality judgements) is a consistent and reliable way to derive optimum parameter combinations for a given algorithm. Of course, this optimisation process relies strongly on the appropriate choice of parameter combinations to be included into the test setup. In the experiments reported here, the parameters were systematically varied in a way that they comprised the whole reasonable range of values. On the other hand, the total number of different values had to be limited because the total measurement time increases with that number in a quadratic way. Thus, the test values were sampled by using results from a pilot experiment with informal listening, where the range of usable parameter values was intersected according to noticeable, i.e. audible differences in the processed signals.

In the experiments, a complete paired comparison of all differently processed versions was performed. This allows for a consistence check of the results of each subject. The results of some subjects indeed turned out to be inconsistent in some particular measurements. But since most subjects rated consistently even in these particular measurements, the rating in general was a task the subjects were able to perform. The statistical evaluation of the results described here gives a relative rank of the different versions with respect to the test criterion, in this case the "better" judgement or personal preference, respectively. The distance of the ranks on a particular "quality scale" was not determined. Such a distance evaluation would be possible if the statistical distribution of the results was known and a sufficient number of subjects was involved. In this study, however, only differences in the relative ranks were considered. A consistent and significant higher rank of a particular version shows at least that there is a noticable improvement and allows for the selection (or the exclusion) of particular parameter values.

Three different stimuli were employed for the comparisons which cover the spatial conditions of speech in quiet (stimulus **s1**), speech with one interfering noise source (stimulus **s5**) and speech in diffuse noise (stimulus **s6**). All signals were recorded in real rooms and exhibit realistic conversational situations. The first situation is important because any signal processing strategy has to preserve a very good signal quality of undistorted speech. The other situations represent the most important spatial noise conditions. The results obtained for stimulus **s1** (measurement **D**) show no significant influence of the processing. This is the desired effect for undistorted speech in quiet. Moreover, most processed versions are ranked higher than the unprocessed signal. This can be explained by a reduced amount of reverberation in the processed signals. For the two other stimuli **s5** and **s6**, the tendencies for particular parameters are similar, e.g., ranks tend to increase with an increasing value of τ_2 in both conditions. In particular, the following effects were found:

- For the directionality parameters ($\alpha_{\text{pass}}, \alpha_{\text{stop}}$), only little significant influence on the ranking was found. In some cases, the broadest directionality was ranked significantly higher than the narrowest (measurements **A**, **E** and **F**). This is consistent with the subjectively found tendency that quality increases slightly with an increasing width of the pass range.
- For the time constants (τ_1, τ_2), considerable and significant influence was found (measurements **A**, **B** and **C**). Quality generally increases with decreasing τ_1 and increasing τ_2, and poor results are obtained for large values of τ_1 and small values of τ_2.

- There was also significant influence of the maximum attenuations (a_1, a_2) (measurements **A**, **E** and **F**). The smaller values $(-10, -10)$, $(-20, -20)$ yielded consistently higher ranks than the greater values $(-30, -30)$, $(-99, -30)$. The tendency is that quality increases with decreasing maximum attenuation.

In general, if the influence of a particular parameter or pair of parameters, respectively, is significant, the tendency of the influence seems to be independent of the other parameters. Also, the observed tendencies may lead to the general impression that quality generally increases with a decreasing influence of the processing, i.e., the less modification is introduced to the original signal, the better is the quality. This would lead to the conclusion that the unprocessed signal always or usually exhibits the highest quality. But the results show that there are indeed processed versions which were significantly ranked higher than the unprocessed version (measurements **B**, **C**, **E** and **F**).

It is striking that the rank of the unprocessed version is varying strongly among the subjects for all measurements. Not only that the total range of ranks is always greater than for any processed version, in some cases (measurements **B**, **C** and **F**) the total range comprises rank 0 up to maximum rank, while some processed versions are ranked very similar by all subjects with a total range of only a few ranks. It might be that subjects are able to distinguish the unprocessed version from a processed version by its naturalness or other properties, and then judge individually prejudiced, depending on what they expect from the processing.

3.6 Conclusions

1. For a conversation (central talker with a single target talker) in quiet, the processing yields a very high signal quality.

2. For a conversation with either a single interfering talker or diffuse noise, respectively, parameter settings were found that were significantly ranked higher than the unprocessed version.

3. The tendencies of parameter influences (if there are any) are consistent across the employed stimuli. Thus, it can be assumed that a parameter setting can be found that is appropriate for various acoustical situations. A reasonable proposal for the parameter values would be $\alpha_{\text{pass}} = 20$, $\alpha_{\text{stop}} = 40$, $\tau_1 = 1$, $\tau_2 = 100$, $a_1 = -20$, $a_2 = -20$, for instance. Nevertheless, there are differences in the ranking of the processed versions in comparison to the unprocessed version for different situations. In particular, the processed versions of stimulus **s6** are generally ranked lower than the processed versions of stimulus **s5** in comparison to the rank of the unprocessed version (measurements **B**, **C** and **E**, **F**). This indicates that the combination of processing strategies used in the algorithm yields better results and is generally more suitable for some conditions than for others.

A main issue of the further development of the algorithm has to be the subjective sound quality and listening comfort, respectively. The above conclusions indicate that it might be useful to adapt the signal processing to the actual acoustical situation. Obviously, the different processing strategies base on different assumptions on the underlying noise condition and can not work equally effective in conditions which considerably differ from the assumptions. The next chapter will focus on the development of a method which allows for a classification of the acoustical situation. Such a method would allow for the selection of processing strategies appropriate for the situation or the adjustment of processing strategies to the situation.

Apart from the above, further investigations are recommended concerning possible general differences between results obtained with the stationary laboratory master hearing aid and the wearable device employed for the field tests. Such differences might be caused by different arithmetics (different FFT time/frequency resolution, floating point vs. fixed point) or different experimental environments and have not been investigated in direct comparison yet.

Chapter 4

Diffusiveness of an acoustical situation: An approach to strategy-selective signal processing

Abstract

In this paper, a measure of the diffusiveness of an acoustical situation is described. This measure is based on the coherence function, but is intended to give a long-term rating of the general diffusiveness rather than a short-time ratio of coherent and incoherent signal energy. It may be obtained from binaural, two-microphone input signals, e.g. dummy head or real-ear recordings, and is suitable for application in binaural hearing aid algorithms. It may be used to either select an appropriate processing strategy or to adjust the processing parameters to values suitable for the situation. It is shown that this measure depends monotoneously on both the amount of reverberation and the number of interfering sound sources present in the signal.

4.1 Introduction

All noise reduction strategies do make one or more assumptions on the acoustical properties of the "target" sound source, i.e. the desired speaker or signal, respectively, and the interfering noise signals and/or the acoustical situation, e.g. the spatial configuration. Based on these assumptions, the algorithm aims at reducing the noise part as much as possible while preserving the target signal as accurately as possible. Hence, the respective applicability is limited to a certain range of acoustical situations that meet the underlying assumption. In applications like digital hearing aids, however, a robust and versatile noise reduction processing is desired that operates in a variety of acoustical conditions, for example by automatically selecting the most appropriate noise reduction technique for the respective situation.

One approach to noise reduction is a directional filtering of two input signals based on the evaluation of binaural parameters. In this approach, assumptions are made about certain interaural parameters of the target signal, e.g. interaural phase differences and level differences. Then, the current signal-to-noise ratio is somehow estimated from the actual interaural parameters. Different techniques and algorithms following this approach have been described in Chapter 1. The described algorithms have been reported to improve speech intelligibility under certain conditions. With an increasing number of competing sound sources and amount of reverberation, however, the actual properties of the interaural parameters more and more deviate from the assumptions. This results in a decline of the processing performance, i.e. a decreasing speech intelligibility and a poor signal quality of the processed signals. The achievable benefit and also the resulting signal quality obviously depends on the match between the acoustical situation which the processing strategy actually is applied to and the situation which was considered when developing the processing strategy. Hence, for noise reduction strategies which employ interaural parameters of binaural input signals, the number of interfering sound sources and the general diffusiveness of the sound field is an important parameter of the acoustical situation. Thus, an automatic classification, i.e., a measure of the diffusiveness of the actual acoustical situation is required that can be used to steer the processing in practical applications. This measure should track the (comparatively slowly changing) general acoustic conditions of the ambient sound field. This would allow for selecting an processing strategy appropriate for the condition and also for optimizing global parameters of the respective strategy. In the following, such a measure is developed and analysed.

4.2 Coherence Function

A classical approach for estimating the incoherent or reverberant part of the signal energy is the coherence function as proposed by Allen *et al.* (1977). They define the following time averages:

$$\begin{aligned}\Phi_{xx}(f,n) &\equiv \left\langle |X(f,n)|^2 \right\rangle &(4.1)\\ \Phi_{yy}(f,n) &\equiv \left\langle |Y(f,n)|^2 \right\rangle &(4.2)\\ \Phi_{xy}(f,n) &\equiv \left\langle X(f,n)Y^*(f,n) \right\rangle, &(4.3)\end{aligned}$$

where X and Y are the short-term spectra of two input signals. f denotes the frequency index, n the time index and $\cdot^*$ is the complex conjugate operator. $\langle\cdot\rangle$ denotes an average across time, which is calculated as a simple first order low-pass filter and denoted as "spectral low-pass" in the following. For an arbitrary time series $Q(n)$, a first order low-pass filter is computed as

$$\langle Q(n)\rangle = \beta \cdot \langle Q(n-1)\rangle + (1-\beta)\cdot Q(n). \tag{4.4}$$

The low-pass filter is characterised by its time constant τ, and the corresponding coefficient β is calculated as

$$\beta = e^{\left(-\frac{1}{\tau\cdot f_{\mathrm{STFT}}}\right)}, \tag{4.5}$$

where f_{STFT} is the sampling frequency of the time series of the spectra (frame rate).

The coherence function $\rho(f, n)$ is eventually given by

$$\rho(f, n) \equiv \frac{|\Phi_{xy}(f, n)|}{\sqrt{\Phi_{xx}(f, n)\Phi_{yy}(f, n)}}. \tag{4.6}$$

Another value commonly used is the Magnitude Squared Coherence $|\rho(f, n)|^2$ (MSC). Allen *et al.* (1977) proposed the coherence function as a direct weighting factor applied to the spectra to reduce room reverberation in two-microphone recordings. This concept was implemented in a noise reduction algorithm for binaural hearing aids by Peissig (1993), for instance, and was reported to considerably reduce reverberation. The time constants employed by Peissig (1993) for calculating the spectral low-pass were in the range of 50 ms to 300 ms. The coherence function itself thus gives information about the coherent signal energy part on a short time scale appropriate for noise reduction rather than a long time scale. However, other authors recently reported to sucessfully employing the coherence function or the MSC, respectively, not only for noise suppression, but also in decision units for different noise suppression strategies.

Bouquin-Jeannès and Faucon (1995) employed the MSC in a decision unit of a voice activity detector to decide whether or not speech was present in the input of two microphone signals in a car. They used a time constant of about 70 ms for calculating the spectral low-pass and then averaged the MSC across frequency. If this averaged MSC was lower than a particular threshold for a time period of about 50 ms, then the average noise spectrum was "learned" by the algorithm. If the averaged MSC was higher than the threshold, a spectral subtraction noise reduction technique was applied employing the previously learned noise spectrum. This application is also based on a rather short time scale in order to detect even short periods of speech in the signal. The authors reported voice activity detection results quite similar to manually labeling the speech periods of the signal. It should be noted that the coherence in principle does not distinguish between speech and non-speech signals, but it can be reasonably assumed that highly correlated signals from two microphones in front of a person represent target speech. The average MSC can thus be considered as effective sound source activity detector rather than a voice activity detector. In a hearing aid algorithm, the average MSC might also be employed to distinguish between the listeners own voice and other signals, since in reverberant environments, the MSC usually is higher for the own voice than for external voices (cf. Section 4.4).

Hussain *et al.* (1997) used the value of the MSC for (manually labeled) noise alone periods to rate the noise and to switch between two different strategies for the reduction of coherent and incoherent noise, respectively. This was done separately for different frequency bands. A time constant of about 60 ms was employed for calculating the spectral low-pass, and an additional time averaging was performed across the noise alone period, which had a fixed length of 100 ms. This is a rather short period, and the authors do not state how to handle shorter or longer noise alone periods (which may occur in real signals). The authors reported significantly better performance of the MSC controlled noise reduction in comparison to fixed noise reduction strategies for anechoic and simulated reverberant signals.

4.3 Long-term coherence

As discussed in the previous section, the coherence function allows for an effective rating and processing of two-microphone inputs for noise reduction purposes. However, the appropriate time scale and averaging method strongly depends on the particular application. The algorithm described in Chapter 3 evaluates interaural phase differences and other parameters to distinguish between a target signal and an interfering signal on a short time scale. This distinction is performed continuously and also effects the signal processing continuously. There are situations, however, in which parameters like interaural phase differences are considerably deteriorated, e.g. with strong reverberation or a large number of spatially separated interfering sound sources (like a cafeteria situation with a loud and rather diffuse background noise). In these situations, a signal processing strategy which relies on these parameters will deteriorate the signal quality or even decrease speech intelligibility. The measure described in the following is intended be used to rate the situation on a rather long time scale in order to decide whether parameters like interaural phase differences can generally be assumed to be reliable or not and thus whether certain signal processing strategies should be applied or not.

For binaural hearing aids and similar applications, the left and right time signals $y_l(t)$ and $y_r(t)$, respectively, are recorded at or in the left and right ear. The equations given in the following are assumed to be applied to a series of short-time Fourier transforms (STFTs) of the signals, calculated as overlapping Fast Fourier Transforms (FFTs), for instance. The time index of the series is denoted as n, and the spectra of the time signals are denoted as $Y_l(f,n)$ and $Y_r(f,n)$, respectively.

In order to reduce statistical fluctuations, the number of frequency channels is reduced from the FFT resolution to a lower number of frequency bands, e.g. critical or third-octave bands. Hence, the following time averages are used instead of equations (4.1) - (4.3):

$$\hat{\Phi}_{xx}(m,n) \equiv \left\langle \sum_{f\in\mathcal{F}_m} |Y_l(f,n)|^2 \right\rangle \tag{4.7}$$

$$\hat{\Phi}_{yy}(m,n) \equiv \left\langle \sum_{f\in\mathcal{F}_m} |Y_r(f,n)|^2 \right\rangle \tag{4.8}$$

$$\hat{\Phi}_{xy}(m,n) \equiv \left\langle \sum_{f\in\mathcal{F}_m} Y_l(f,n)Y_r^*(f,n) \right\rangle, \tag{4.9}$$

where m is the number of the frequency band and $f \in \mathcal{F}_m$ represent all (adjacent) FFT frequency bins within that band. The employed time constant is denoted as $\tau_{\hat{\Phi}}$.

A long-term, i.e. slowly changing function d of the coherence, which will be referred to either as degree of coherence or degree of diffusiveness, respectively, is now defined as

$$d(n) \equiv \left\langle \sum_m w_m h\left(\mathrm{MSC}(m,n)\right) \right\rangle_{\tau_d(R)} . \tag{4.10}$$

h is an appropriate transformation function. With $h(x) = x$, $d(n)$ can be considered as degree of coherence and will be denoted as d_c. In this case, the values of d are in the range of $[0;1]$ (for appropriate values of w_m, see below). A value of 1 means complete coherence and 0 means no coherence or complete diffusiveness, respectively. Alternatively with $h(x) = 1 - \sqrt{x}$, $d(n)$ can be considered as degree of diffusiveness and will be denoted as d_d. In this case, a value of 1 means complete diffusiveness and 0 means complete coherence.

w_m denotes a frequency dependent weighting factor with $\sum_m w_m = 1$. Since for low frequencies, the coherence function is very high, i.e., close to 1 and thus of low evidence, it is advantageous to set $w_m = 0$ for frequency bands below a certain cutoff frequency $m_{d,\min}$ (this effect depends on the distance of the microphones, cf. Dörbecker, 1998). A simple, constant weight is given by

$$w_{m,N} = \begin{cases} 0 & : \ m < m_{d,\min} \\ \frac{1}{N_m} & : \ m \geq m_{d,\min} \end{cases} . \tag{4.11}$$

where N_m is the total number of frequency bands which are summed up. Another possible weight is the energy of each frequency band, for instance given by

$$w_{m,E} = \begin{cases} 0 & : \ m < m_{d,\min} \\ \frac{\max\left\{\hat{\Phi}_{xx}(m,n),\hat{\Phi}_{yy}(m,n)\right\}}{\sum_{m \geq m_{d,\min}} \max\left\{\hat{\Phi}_{xx}(m,n),\hat{\Phi}_{yy}(m,n)\right\}} & : \ m \geq m_{d,\min} \end{cases} . \tag{4.12}$$

This weighting allows bands with more energy to have more influence on $d(n)$ than bands with less energy. The influence of different weightings is shown below.

The operator $\langle\cdot\rangle_{\tau_d(R)}$ denotes an average across time. Again, this may be calculated as a simple first order low-pass filter with a time constant τ_d of at least a few seconds. However, the aim is to rate the acoustical situation based on actual sound incidence and not on pauses (which usually are not coherent anyway due to microphone noise). A level dependent time constant thus yields much more stable results, especially for speech signals including pauses (cf. Fig. 4.5). For this, the maximum total energy $I_{\max}$ and its moving average $\hat{I}_{\max}$ are defined as

$$I_{\max}(n) \equiv \max\left\{ \sum_{m \geq m_{d,\min}} \hat{\Phi}_{xx}(m,n), \sum_{m \geq m_{d,\min}} \hat{\Phi}_{yy}(m,n) \right\} \tag{4.13}$$

$$\hat{I}_{\max}(n) \equiv \langle I_{\max}(n) \rangle_{0,\tau_{\hat{I}}} \quad , \tag{4.14}$$

which means that $\hat{I}_{\max}(n)$ is the low-pass filtered value of $I_{\max}(n)$ using an instantaneous attack and a release time constant $\tau_{\hat{I}}$. Finally, the time constant τ_d is adjusted between its minimum and maximum value $\tau_{d,\min}$ and $\tau_{d,\max}$, respectively, using the ratio R of current energy to moving maximum energy in dB:

$$R \equiv 10 \log_{10} \left(\frac{I_{\max}(n)}{\hat{I}_{\max}(n)} \right) . \tag{4.15}$$

Considering the signal energy in dB, it is quite reasonable to give values a high weight when R is in a range near to 0 dB and thus using a low time constant τ_d. If the value of R

decreases and falls below a certain limit, the time constant τ_d should considerably increase. The function $\tau_d(R)$ is thus defined as

$$\tau_d(R) = \begin{cases} \tau_{d,\min} & : \quad R \geq R_{\max} \\ \tau_{d,\max} & : \quad R \leq R_{\min} \\ \tau_{d,\min} + \frac{(R - R_{\max})(\tau_{d,\max} - \tau_{d,\min})}{R_{\min} - R_{\max}} & : \quad R_{\max} > R > R_{\min} \end{cases} . \tag{4.16}$$

This gives a trapeziform function which is depicted in Fig. 4.1.

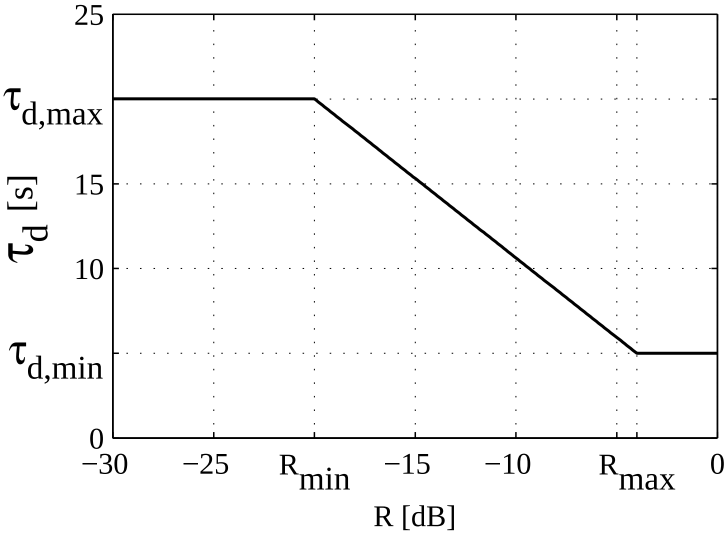


Figure 4.1: τ_d as a function of the ratio R of current energy to maximum energy (see text for details). The particular values shown here have been used for the example calculations.

4.4 Examples

The resulting value of d calculated for a particular signal using equation (4.10) depends on a variety of different parameters. All calculated values given in the examples below have the following parameters in common: The initial frequency analysis has been performed using a 512 point FFT with a 400 point hanning window and 200 point window shift at a sampling frequency of 25 kHz. Subsequently, a summation following Eq. (4.7) - (4.9) across 23 critical bands, each with a bandwidth of 1 Bark was performed (cf. Zwicker, 1961). The Bark frequency scale was preferred over the more recent ERB scale (cf. Moore and Glasberg, 1983) simply because the frequency resolution was limited to a fixed distance at lower frequencies due to the FFT analysis. The values of the time constants were $\tau_{\hat{\Phi}} = 40$ ms, $\tau_{\hat{I}} = 20$ s, $\tau_{d,\min} = 5$ s and $\tau_{d,\max} = 20$ s. Other parameters were $m_{d,\min} = 4$ Bark (approx. 400 Hz), $R_{\max} = -4$ dB and $R_{\min} = -20$ dB. Differing or additional parameters will be specified for each example, if necessary.

4.4.1 Stationary signals

First, stationary two-channel signals were used as input signals to calculate values of d_c and d_d. The correlation of both channels was systematically varied by adding a correlated and a diffuse part at a certain level ratio. The correlated part of the signal consisted of identical white noise in both channels with a mean energy I_c, while the diffuse part consisted of uncorrelated white noise in both channels, each with a mean energy I_d. The ratio $\frac{I_c}{I_d}$ of both energies is called correlation-to-diffusiveness ratio (CDR).

Fig. 4.2 depicts d_c and d_d, respectively, as functions of the CDR, calculated with the frequency weights $w_{m,N}$ and $w_{m,E}$ according to equations (4.11) and (4.12). The shown values are mean values calculated for each particular CDR from a signal of 30 s duration. The deviation from the mean value within one signal is very small (the maximum deviation was 0.3 dB absolute and 6 % relative, respectively, but typically smaller). Compared to the constant frequency weights $w_{m,N}$, the energy dependent weights $w_{m,E}$ seem to enlarge the range of values of d for low CDRs, but have small effects for medium and higher CDRs.

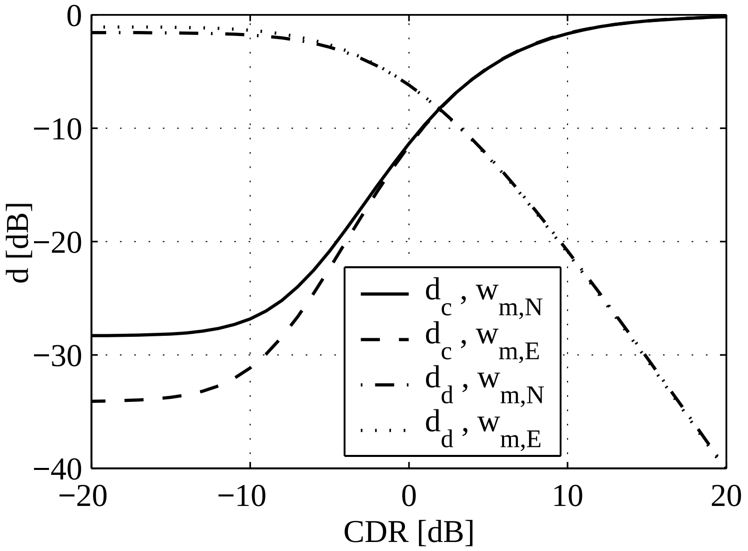


Figure 4.2: Degree of coherence d_c and degree of diffusiveness d_d as functions of the correlation-to-diffusiveness ratio (CDR) of a noise signal (see text for details). $w_{m,N}$ and $w_{m,E}$ denote the frequency weights given by equations (4.11) and (4.12).

4.4.2 Binaural recordings

Values of d_c and d_d, respectively, have also been calculated from binaural recordings of various acoustical situations. All recordings were made with the same pair of hearing instruments (Siemens Cosmea M) worn by either a subject or a dummy head. The recorded signals had durations between 30 and 120 seconds. Three different rooms were employed, a non-reverberant (anechoic) chamber, a moderately reverberant, small room ($T_{60} \approx 0.5$ s) and a highly reverberant, large room ($T_{60} \approx 2$ s). A single target speaker was presented by a loudspeaker in front of the subject or dummy head. Additionally, up to a maximum of 4 interfering speakers from different spatial locations (45, 135, 225 and 315 degrees azimuth) were presented by loudspeaker. The distance of the loudspeakers to the microphones was approx. 1.5 m, if not specified differently. In the highly reverberant room, there was also a cafeteria noise situation recorded with diffuse noise present (no interfering speaker).

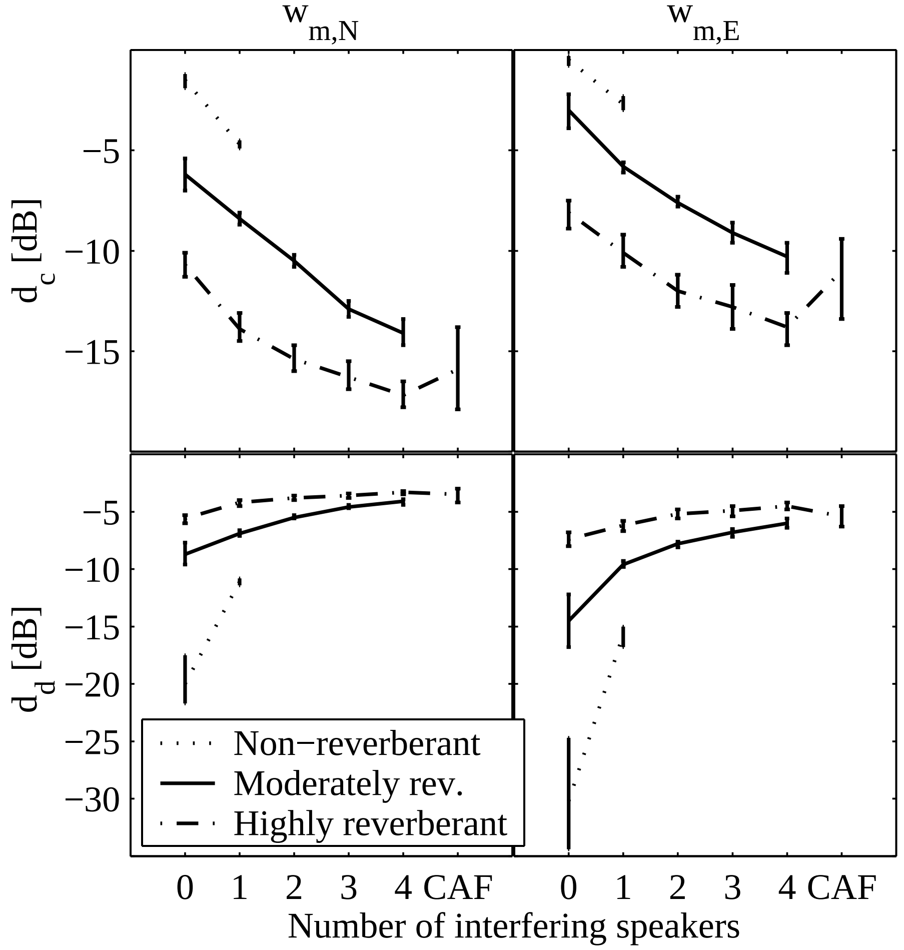


Figure 4.3: Degree of coherence d_c and degree of diffusiveness d_d calculated from binaural recordings. The abscissa denotes the number of present interfering speakers (0-4) and the cafeteria situation (CAF), respectively. A target speaker was present in all situations. The lines denote the mean values of d_c (upper panels) and d_d (lower panels), while the errorbars give the total range of values calculated from the particular signal. The results are given separately for the non-reverberant room (dotted line), the moderately reverberant room (solid line) and the highly reverberant room (dash-dotted line) in each panel. The left panels were calculated employing the weights $w_{m,N}$, the right panels employing the weights $w_{m,E}$ according to equations (4.11) and (4.12).

Fig. 4.3 shows d_c and d_d, respectively, as a function of the number of interfering speakers and the cafeteria noise situation, respectively. Left and right panels depict values for different weighting factors w_m, while the different lines denote different rooms. Obviously,

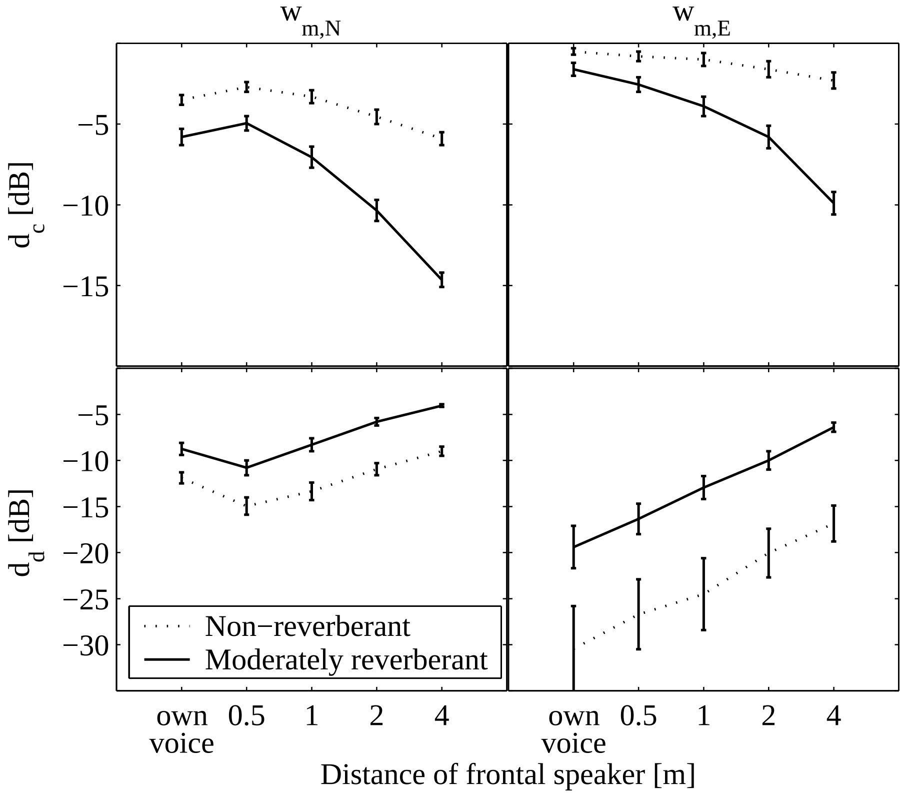


Figure 4.4: Degree of coherence d_c and degree of diffusiveness d_d as a function of the distance of a frontal speaker, compared to the value calculated for the own voice. See Fig. 4.3.

d_c decreases and d_d increases, respectively, with the number of interfering speakers in the same room and also with the amount of reverberation for a fixed number of interfering speakers. Fig. 4.4, shows d_c and d_d, respectively, as a function of the distance of a frontal target speaker and the own voice of the hearing instrument wearer, respectively. In this case, the employed weighting factors have considerable influence on the values. For the weighting factors $w_{m,E}$, d_c and d_d are monotonous functions of the distance with the highest (d_c) and lowest (d_d) value, respectively, for the own voice. Hence, if the amount of reverberation is known (or can be estimated at least to some extent), d_c and d_d may be used to distinguish between the own voice and distant (external) talkers. This monotonous dependency is not observed if the weighting factors $w_{m,N}$ are used. Apparently, the weighting with the spectral signal energy is useful for an estimation of the distance of a single talker.

The upper panel of Fig. 4.5 depicts the total level of a signal with 11 s target speech, 10 s pause (ambient noise) and another 11 s target speech recorded in the moderately reverberant room. The lower panel shows time courses of d_c calculated from that signal using different values for the time constant τ_d. The fast acting version of d_c (thin line) is considerably correlated to the signal level (correlation coefficient 0.46). The 5 s low-pass filtered version (thick solid line) is already quite stable, but decreases considerably during the 10 s speech pause. The version calculated with the level dependent low-pass filter (thick dashed line) is the least fluctuating stable version.

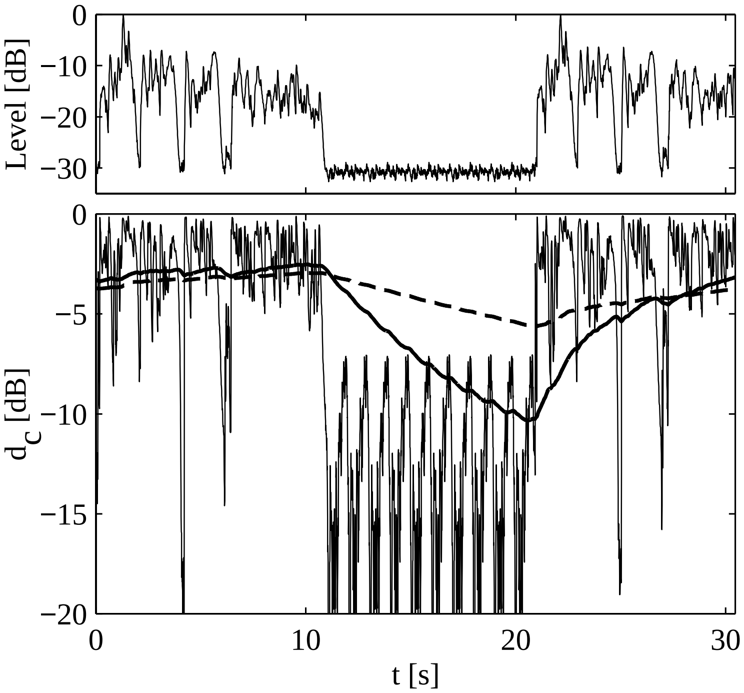


Figure 4.5: Time courses of d_c for a signal with a single target speaker. The upper panel shows the total level of the signal as a function of time. The lower panel shows a fast acting d_c calculated with $\tau_{d,\min} = \tau_{d,\max} = 0$ s (thin solid line), d_c calculated with $\tau_{d,\min} = \tau_{d,\max} = 5$ s (thick solid line) and d_c calculated with $\tau_{d,\min} = 5$ s, $\tau_{d,\max} = 20$ s (thick dashed line). The employed frequency weights were $w_{m,E}$ in all cases.

4.5 Discussion

The definition of the degree of coherence and the degree of diffusiveness, respectively as described in this study is quite arbitrarily chosen, but provides a long-term rating of the acoustical situation. Although the resulting values depend on the particular parameter set employed, the values shown in Fig. 4.3 demonstrate that for a fixed set of parameters, the degree of diffusiveness, for instance, i) increases with an increasing amount of reverberation for the same acoustical situation and ii) increases with an increasing number of speakers in the same room. Obviously, it is not possible to conclude from the calculated values whether a high degree of diffusiveness is in particular the result of a high reverberation or a high number of sound sources. As long as the respective effect is the same, it is not necessary to have this information anyway. However, if additional information about the actual amount of reverberation is available, the degree of diffusiveness allows for an estimate of the number of (spatially separated) sound sources. The time courses of the degree of coherence as depicted in Fig. 4.5 demonstrate that the proposed level dependent low-pass filtering provides a more stable measure than a simple low-pass filtering of the Magnitude Squared Coherence, especially in long signal pauses.

The proposed measure allows for a rating of acoustical situations with respect to their complexity. In situations which are assumed to be too complex for a particular processing strategy, this strategy may be switched off. For instance, interaural phase differences can reasonably be assumed to be deteriorated and unreliable in situations with strong reverberation or a lot of interfering sound sources. Thus, a noise reduction strategy which is based on the evaluation of these phase differences should be switched off if the degree of coherence is low. However, the appropriate particular values, i.e. the time constants and the boundary between "simple" and "complex"situations used for such a rating, will surely depend on the particular application. From Fig. 4.4 it may concluded that the degree of coherence might also be suitable to detect an activity of the own voice in contrast to other sound sources (in a reverberant environment). However, smaller time constants $\tau_{d,\max}$ and $\tau_{d,\min}$ would be required to detect the utterance of a single word or sentence.

The proposed method does not allow for a distinction between target and interfering signal or the detection of noise alone periods. For this, additional assumptions about the target and/or the noise are required, e.g. fluctuating speech and stationary noise, which will depend on the application. The proposed large time constants are not appropriate to detect or react on, for instance, a single spoken word or a short speech pause (although smaller time constants may be used for these purposes). For the intended application in a hearing aid, however, it is not desirable to switch noise reduction processing strategies on and off very fast, because this would yield audible and annoying artefacts.

It should be noted that since the degree of coherence is an average across frequency, the transfer functions (e.g. cutoff frequencies) of the recording devices also have influence on the calculated values. A direct comparison of values is thus only possible for signals recorded with the same microphone characteristics and transfer functions. An application of the method proposed here to the selection of the appropriate noise reduction strategy in a hearing aid algorithm is given in the next chapter.

Chapter 5

Strategy-selective noise reduction algorithm: Technical description and evaluation

Abstract

Different binaural signal processing strategies for noise reduction are derived which are based on particular assumptions on the properties of the target signal and the undesired interfering signals. The processing strategies are evaluated with respect to their technical performance using artificial signals. They are shown to function if the underlying assumptions are met. All processing strategies are combined within a single, strategy-selective algorithm which automatically selects appropriate processing strategies depending on the acoustical situation. For this, the previously introduced degree of diffusiveness is employed to classify the situation and to switch off particular processing strategies if necessary. The time constants of the processing are optimized employing the results of subjective preference measurements. Using these optimized parameter values, the processing in general exhibits a very high sound quality.

5.1 Introduction

Noise reduction systems or algorithms, respectively, for hearing aids have received considerable interest in the past, primarily because the reduced speech intelligibility under noisy conditions is one of the major complaints in hearing impaired subjects. A promising algorithm that uses binaural information to suppress lateral noise sources and reverberation has been introduced by Peissig (1993) and was described and further developed in Chapters 2 and 3. Laboratory studies with hearing impaired subjects proved that this algorithm is capable of improving speech perception in noise under certain conditions (cf. Peissig, 1993). However, field studies which were performed employing a wearable digital signal processor hearing aid revealed that the subjective sound quality of the processing is rated rather poor by hearing impaired subjects in real-life conditions (cf. Pastoors *et al.*, 1998; Albani *et al.*, 1998). Hence, in order to make such a binaural noise reduction algorithm more acceptable for the hearing aid users, the resulting subjective sound quality of the processing has to be improved considerably. The current study therefore attempts to optimize the sound quality of the binaural noise reduction by optimizing a set of strategies for particular acoustical situations they are suitable for and by allowing the whole algorithm to adapt to the respective acoustical situation.

The noise reduction algorithm described so far (cf. Chapters 2 and 3) is based on two different processing techniques or strategies, respectively, i.e. dereverberation processing and directional filtering. Experiences and listening tests show that although the dereverberation processing can cause some audible processing artefacts, the directional filtering is much more critical with respect to the sound quality. This strategy thus limits the maximum achievable sound quality of the whole algorithm. One reason for this is that the interaural signal parameters that are used to calculate the directional filtering gain factors are unreliable and unstable in the presence of multiple noise sound sources or diffuse noise. However, if the sound quality is optimized in such critical situations, the noise reduction capabilities will considerably decrease in acoustically “easy” situations.

The aim of the algorithm described here is the best possible restoration of the original short-time spectral target signal intensity. This should result in lowest possible processing artefacts and thus in high sound quality. Parts of the algorithm after Peissig (1993) have been taken over unchanged in the new algorithm and other parts have been modified. In addition, a new processing strategy for cancelling out a single (lateral) noise source has been added which preserves a high signal quality. Also, a decision unit has been included which is capable of automatically rating the diffusiveness, i.e., the complexity of the current acoustical situation. The rating is used to switch on or off particular processing strategies of the new algorithm, depending on the expected deterioration caused by the processing in that situation. This allows for the optimization of processing strategies for situations in which they can achieve a benefit and to switch them off if they can not. This avoids using the same parameter setting in all situations which neither causes much deterioration nor achieves much benefit in any case. The parameters of the new algorithm have been investigated with respect to the sound quality by paired comparisons and subjective judgement of relative sound quality with normal hearing subjects. The results will be described and compared to results obtained with the algorithm after Peissig (1993).

5.2 Algorithm

The following considerations concerning the reduction of noise in a noisy signal will focus on the estimation of the short-time spectral amplitude (STSA) of the original, undegraded target signal. Thus, only magnitude gain factors will be derived and the phase of the degraded signal will be preserved in the processed signal. The STSA has been found to be of major importance in speech perception, and the method of preserving the degraded phase is often used (cf. Lim and Oppenheim, 1979; Boll, 1979; McAulay and Malpass, 1980; Wang and Lim, 1982). The ratio of the estimated magnitude of the signal alone over the magnitude of the degraded signal is applied to the spectra as magnitude gain factors, if such an estimation is possible. This method can be considered as a parametric Wiener filtering (cf. Lim and Oppenheim, 1979).

It is assumed that the acoustical environments in which the algorithm is applied are mainly characterized by a combination of the following "model" situations. The first situation, which is depicted in the left panel of Fig. 5.1, is a situation with one target sound source and some additional, diffuse noise. The second situation, which is depicted in the right panel of Fig. 5.1, consists of one target sound source and one interfering sound source, which are clearly separated in their spatial localization, i.e., mainly in their azimuthal localization. Additionally, an extension of the second situation to more than one interfering sound source will be considered in section 5.2.3.

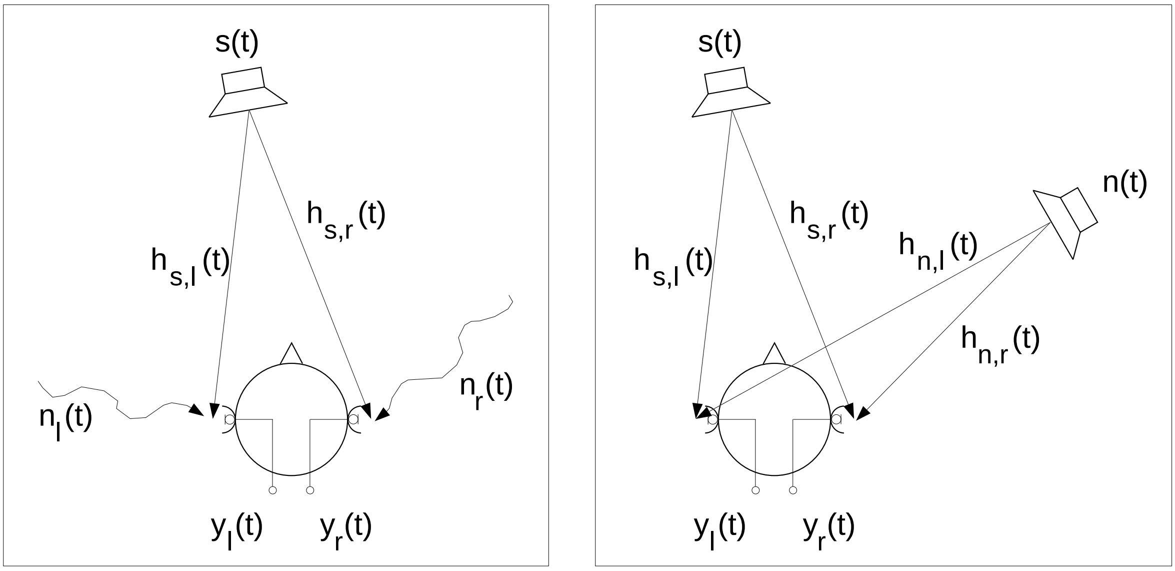


Figure 5.1: Acoustical "model" situations considered for the development of the noise reduction algorithm. The left panel shows one target sound source $s(t)$ and diffuse noise $n_l(t)$ and $n_r(t)$ at the left and right ear, respectively. The right panel shows one target sound source $s(t)$ and one interfering sound source $n(t)$. The impulse responses $h_{s,l}(t)$, $h_{s,r}(t)$ and $h_{n,l}(t)$, $h_{n,r}(t)$, respectively, of the transfer systems from the sound sources to the left and right ear affect the target signal $s(t)$ and the interfering signal $n(t)$, respectively. The left and right microphone signals $y_l(t)$ and $y_r(t)$ are recorded at the locations of the left and right ear.

The time signals and their corresponding FFT spectra, respectively, are denoted using lowercase and uppercase letters, respectively. While a Fourier transform is not time dependent any more, a series of short-time Fourier transforms (STFTs), as often used in digital signal processing, still is, and the equations presented here are assumed to be applied to STFT series. However, the time indices of a series of the functions defined in the following are usually not specified in order to achieve more clearness of the equations. The following definitions will be referred to in general (STFTs of time signals depicted in Figure 5.1):

$$Q_Y \equiv \frac{Y_l(f)}{Y_r(f)}, \qquad Q_S \equiv \frac{H_{s,l}(f)}{H_{s,r}(f)}, \qquad Q_N \equiv \frac{H_{n,l}(f)}{H_{n,r}(f)}. \tag{5.1}$$

5.2.1 Strategy 1: One target, diffuse noise

The starting point for the following considerations is the dereverberation processing of the algorithm after Peissig (1993). Since there are no theoretical limitations to specific acoustical situations, this stage is of major importance for the algorithm especially in complex environments. It was based on the empirical approach of employing the Magnitude Squared Coherence (MSC) for dereverberation, proposed by Allen *et al.* (1977). However, the implementation of Peissig (1993) exhibited considerable processing artefacts when applying the respective gain factors directly to the spectra without further low pass filtering. The underlying acoustical situation is thus analytically investigated in the following in order to obtain gain factors which restore the original target signal intensity more accurately.

In the model situation with one target sound source and some additional, diffuse noise (as depicted in the left panel of Fig. 5.1), the Fourier transforms of the left and right microphone signals are given as

$$\begin{array}{rclcl} Y_l(f) & = & S(f) \cdot H_{s,l}(f) + N_l(f) & \equiv & X_l(f) + N_l(f) \\ Y_r(f) & = & S(f) \cdot H_{s,r}(f) + N_r(f) & \equiv & X_r(f) + N_r(f), \end{array} \tag{5.2}$$

where $X_l(f)$ and $X_r(f)$ represent the target signal parts of the spectra.

It is assumed that $H_{s,l}(f)$ and $H_{s,r}(f)$ and therefore Q_S are not or very slowly changing with time. With the prerequisite that $s(t)$, $n_l(t)$ and $n_r(t)$ are not correlated and with $\langle\cdot\rangle$ as expectation value operator (i.e., average across time) and $\cdot^*$ as the complex conjugate operator, the expected values of the power spectra $Y_l(f)Y_l^*(f)$, $Y_r(f)Y_r^*(f)$ and the cross power spectrum $Y_l(f)Y_r^*(f)$, respectively, are given as

$$\langle Y_l(f)Y_l^*(f)\rangle = \left\langle |X_l(f) + N_l(f)|^2 \right\rangle = \left\langle |X_l(f)|^2 \right\rangle + \left\langle |N_l(f)|^2 \right\rangle, \tag{5.3}$$

$$\langle Y_r(f)Y_r^*(f)\rangle = \left\langle |X_r(f) + N_r(f)|^2 \right\rangle = \left\langle |X_r(f)|^2 \right\rangle + \left\langle |N_r(f)|^2 \right\rangle, \tag{5.4}$$

$$\langle Y_l(f)Y_r^*(f)\rangle = \left\langle |S(f)|^2 \right\rangle \cdot H_{s,l}(f)H_{s,r}^*(f). \tag{5.5}$$

Using these quantities, the aim of the algorithm would be, for instance, to restore $X_l(f)$ from $Y_l(f)$ by applying appropriate gain factors $G_l(f)$, i.e. $\widehat{X_l}(f) = G_l(f) \cdot Y_l(f)$ with $\widehat{X_l}(f)$ representing an estimate of $X_l(f)$. Appropriate definitions of magnitude gain factors $G_{\mathrm{corr},l}$ and $G_{\mathrm{corr},r}$ for application to the left and right microphone signal spectrum, referred to as correlation gain factors, are given by

$$G_{\mathrm{corr},l}(f) \equiv \sqrt{\frac{|\langle Y_l(f)Y_r^*(f)\rangle \cdot Q_S^*|}{\langle Y_l(f)Y_l^*(f)\rangle}} = \sqrt{\frac{\langle |X_l(f)|\rangle^2}{\langle |X_l(f) + N_l(f)|^2\rangle}}, \tag{5.6}$$

$$G_{\mathrm{corr},r}(f) \equiv \sqrt{\frac{\left|\langle Y_l(f)Y_r^*(f)\rangle \cdot \frac{1}{Q_S}\right|}{\langle Y_r(f)Y_r^*(f)\rangle}} = \sqrt{\frac{\langle |X_r(f)|\rangle^2}{\langle |X_r(f) + N_r(f)|^2\rangle}}. \tag{5.7}$$

The expectation value operator $\langle\cdot\rangle$ may be realized as an approximation by a simple first order low-pass filter with the time constant τ_Y. Instead of directly employing the MSC, the correlation gain factors employ the magnitude square root of the coherence function, which results in less attenuation and less processing artefacts. Additionally, the impulse responses $h_{s,l}(t)$ and $h_{s,r}(t)$ are taken into account.

In order to calculate the above correlation gain factors, the ratio Q_S has to be estimated. This can be done by using the ratio Q_Y and an appropriate decision unit which decides whether the current Q_Y is assumed to represent a target signal activity or not. Once this decision is made, the ratio Q_S can be obtained by averaging the target-representing Q_Y across time. A decision unit for this purpose and an estimator of Q_S is described in Section 5.2.4. As an alternative approach, a constant value of Q_S, calculated in advance as mean value of Q_Y for a noise signal with frontal sound incidence direction in a non-reverberant environment, can be used as a fixed estimate of Q_S.

In order to evaluate the technical performance of the processing strategy, the correlation gain factors were calculated from a binaural signal mixed with diffuse noise. The target signal and the diffuse noise were recorded separately, i.e. $x_l(t)$ and $x_r(t)$ were first recorded using a dummy head with 10 seconds of white noise as target signal $s(t)$ present only (cf. Figure 5.1). From this recording, the mean ratio $\langle Q_S \rangle$ was calculated as an average across time. The signals $y_l(t)$ and $y_r(t)$ were then calculated by adding uncorrelated (diffuse) noise signals $n_l(t)$ and $n_r(t)$ to the recorded signals $x_l(t)$ and $x_r(t)$. This was done for the three different signal-to-noise ratios of -6 dB, 0 dB and +6 dB (mean SNR across the 10 seconds signal). Hence, the gains were evaluated for a total of 30 seconds signal. The theoretically required gain $G_{\text{theo},x}(f)$ was calculated from the known separate signals as

$$G_{\text{theo},x}(f) = \sqrt{\frac{\langle |X_x(f)| \rangle^2}{\langle |X_x(f) + N_x(f)|^2 \rangle}}, \tag{5.8}$$

where x is l or r, respectively. All signal spectra $X_x(f)$ and $X_x(f) + N_x(f)$ and thus the theoretically required gain factors were calculated in the same way as within the noise reduction processing, i.e. with the same time and frequency resolution. Finally, the correlation gain factors were calculated from the noisy signals $y_l(t)$ and $y_r(t)$ using both the fixed, known ratio $\langle Q_S \rangle$ and the running estimate of Q_S (cf. Section 5.2.4).

Figure 5.2 shows $G_{\text{corr},l}(f)$ plotted against the theoretical gain $G_{\text{theo},l}(f)$ for three different frequency bands. The center frequency f_c is specified in each panel, the bandwidth is 1 Bark (critical band). The values of $G_{\text{corr},l}(f)$ shown in the upper panels were calculated using the fixed, known ratio $\langle Q_S \rangle$, while the lower panels were calculated using the running estimate of Q_S. The time constant τ_Y was 40 ms for all panels. The correlation coefficient r of both gains is given in each panel. The figure shows that r is high especially for higher frequencies[1]. Comparing the upper and lower panels shows that employing the estimate of Q_S instead of its known, fixed value does not considerably decrease r.

In most cases, the correlation of the values is quite high. For the low frequency band, the correlation gain seems to be overestimated at low SNRs. This does not deteriorate the sound quality, since the target signal is not attenuated more than necessary and the information is preserved in the processed signal. The noise reduction performance, however, is not at optimum in that situation. This effect is not suprising because the amount of inter-microphone coherence depends on the frequency and on the distance of the microphones (cf. Dörbecker, 1998). It is higher at lower frequencies and thus yields higher correlation

[1] For the right correlation gain factors $G_{\text{corr},r}(f)$, which are not shown here, the values of r are slightly smaller probably due to the nonlinear influence of Q_S^* and $\frac{1}{Q_S}$.

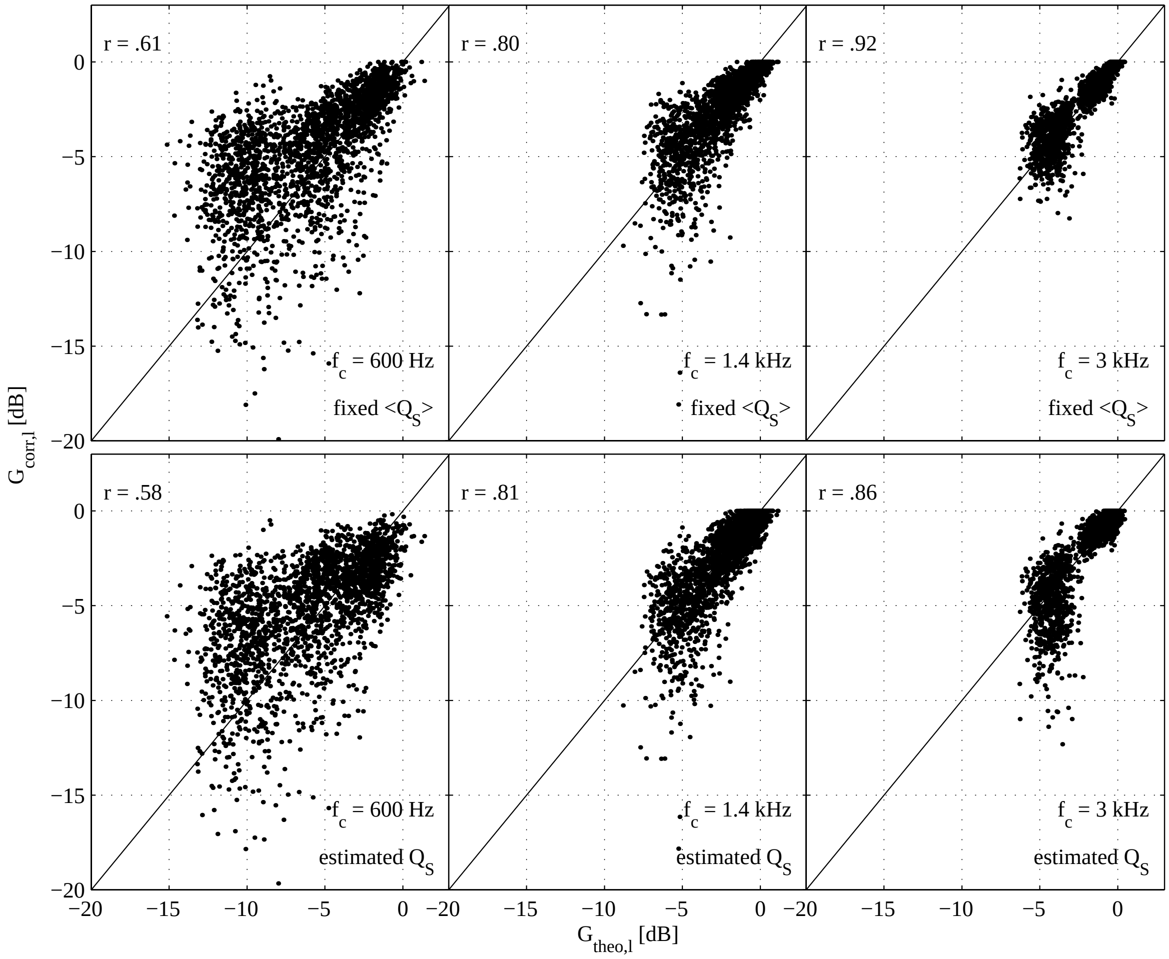


Figure 5.2: Correlation gain factors $G_{\mathrm{corr},l}(f)$ plotted against the theoretically required gain $G_{\mathrm{theo},l}(f)$ for different center frequencies f_c. The correlation coefficient r between both gains is given in the upper left corner of each panel.

gains. For the high and mid frequency bands, Figure 5.2 shows the general tendency of an underestimation of G_{corr} particularly for the signals with the lowest SNRs (mean SNR -6 dB). Apparently, for small values of $\langle|X_l(f)|\rangle^2$, the numerator $|\langle Y_l(f)Y_r^*(f)\rangle \cdot Q_S^*|$ of Equation (5.6) is no longer an accurate and stable estimate of $\langle|X_l(f)|\rangle^2$. In general, the deviation between $G_{\mathrm{corr},x}(f)$ and $G_{\mathrm{theo},x}(f)$ is found to increase with a decreasing SNR (and thus with a decreasing $G_{\mathrm{corr},x}(f)$). In addition, a larger time constant τ_Y results in a smaller deviation and a higher value of the correlation coefficient. This can be explained by the fact that larger time constants yield a better approximation of the theoretically assumed expectation operator. On the other hand, if the time constant is large, the calculated gain will not follow the real fluctuations of the SNR in an appropriate way. There is a contrary dependency between the accuracy of the G_{corr} estimate and its adaptation rate, which make an optimization of the time constant τ_Y necessary that is based on sound quality considerations.

5.2.2 Strategy 2: One target, one interfering sound source

In the model situation with one target sound source and one interfering sound source (as depicted in the right panel of Fig. 5.1), the Fourier transforms of the left and right microphone signals are given as

$$\begin{aligned} Y_l(f) &= S(f) \cdot H_{s,l}(f) + N(f) \cdot H_{n,l}(f) \\ Y_r(f) &= S(f) \cdot H_{s,r}(f) + N(f) \cdot H_{n,r}(f). \end{aligned} \tag{5.9}$$

In the following, the left and right target signal spectra $S(f) \cdot H_{s,l}(f)$ and $S(f) \cdot H_{s,r}(f)$ and the left and right interfering signal spectra $N(f) \cdot H_{n,l}(f)$ and $N(f) \cdot H_{n,r}(f)$ will be denoted as

$$\begin{aligned} S_l &\equiv S(f) \cdot H_{s,l}(f) \quad , \quad & S_r &\equiv S(f) \cdot H_{s,r}(f) \ , \\ N_l &\equiv N(f) \cdot H_{n,l}(f) \quad , \quad & N_r &\equiv N(f) \cdot H_{n,r}(f) \ . \end{aligned} \tag{5.10}$$

The fractions $\frac{S_l}{N_l}$ and $\frac{S_r}{N_r}$ are referred to as left and right signal-to-noise ratio (SNR), respectively. As can be easily shown (see Appendix A), the left and right SNR equal to

$$\frac{S_l}{N_l} = -\frac{Q_Y - Q_N}{Q_Y - Q_S} \cdot \frac{Q_S}{Q_N}, \qquad \frac{S_r}{N_r} = -\frac{Q_Y - Q_N}{Q_Y - Q_S}. \tag{5.11}$$

Hence, magnitude gain factors $G_{\text{int},l}$ and $G_{\text{int},r}$ for application to the signal spectra, referred to as interfering gain factors, can be obtained from the signal-to-noise ratios as:

$$G_{\text{int},l}(f) \equiv \left| \frac{S_l}{S_l + N_l} \right| = \left| \frac{1}{1 + \frac{N_l}{S_l}} \right|, \qquad G_{\text{int},r}(f) \equiv \left| \frac{S_r}{S_r + N_r} \right| = \left| \frac{1}{1 + \frac{N_r}{S_r}} \right|. \tag{5.12}$$

To obtain the left and right SNR, the ratios Q_S and Q_N have to be estimated. Again, this can be done by using the ratio Q_Y and an appropriate decision unit which decides whether the current Q_Y is assumed to represent the target or the interfering signal. Once this decision is made, the estimated ratios Q_S and Q_N, respectively, can be obtained by averaging the target-representing and the interfering-representing Q_Y, respectively, across time. A decision unit for this purpose and an estimator of Q_S and Q_N is described in Section 5.2.4.

For the evaluation of the performance of the SNR estimation, a target signal at 0 degree azimuth and an interfering signal at 60 degrees azimuth were employed. Both binaural signals were 10 seconds of CCITT speech-shaped noise (cf. CCITT G.227), recorded separately using a dummy head in a non-reverberant room. The target signal was repeated three times at the same level, representing the signals $x_l(t)$ and $x_r(t)$. The interfering noise was repeated three times with mean levels of -6 dB, 0 dB and +6 dB relative to the target signal, giving the signals $n_l(t)$ and $n_r(t)$. The signals $y_l(t)$ and $y_r(t)$ were then calculated by adding the respective target and interfering signal. The real SNR was calculated from both separate signals, while the estimated SNR was calculated from the mixture alone within the algorithm. All signal spectra and thus the estimated SNR as well as the real

SNR were calculated in the same way as within the noise reduction processing, i.e. with the same time and frequency resolution. Both the fixed, known ratios $\langle Q_S \rangle$ and $\langle Q_N \rangle$ and the running estimates of Q_S and Q_N were used to calculate the estimated SNR values (cf. Section 5.2.4).

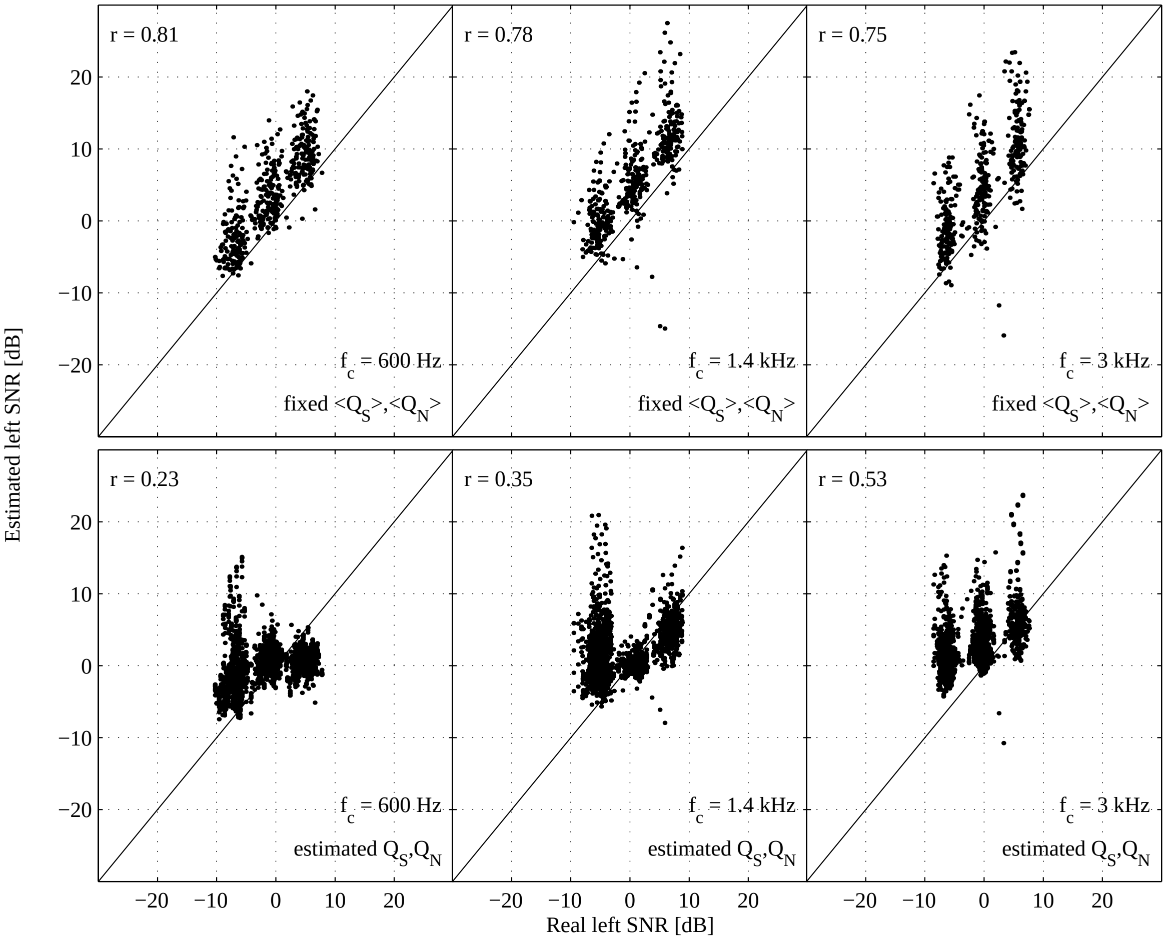


Figure 5.3: Left SNR (signal-to-noise ratio) values estimated according to Eq. (5.11) plotted against the actual left SNR calculated from the known target and interfering signal. Again, the upper panels show values for different center frequencies f_c calculated using the known mean values $\langle Q_S \rangle$ and $\langle Q_N \rangle$, respectively, while the lower panels were calculated using the running estimates of Q_S and Q_N. The correlation coefficient r between both SNRs is given in the upper left corner of each panel.

Figure 5.3 shows the estimated SNR values plotted against the real SNR for three different frequency bands of the left channel. The center frequency f_c is specified in each panel, the bandwidth is 1 Bark (critical band). The SNR values shown in the upper panels were calculated using the fixed, known ratios $\langle Q_S \rangle$ and $\langle Q_N \rangle$, while the lower panels were calculated using the running estimates of Q_S and Q_N (cf. Section 5.2.4). The time constant τ_Y was 40 ms for all panels. The correlation coefficient r between both gains is given in each panel. For the right channel, which is not shown here, r is higher for high frequencies, but lower for other frequencies. The upper panels shows that r is high when the known ratios $\langle Q_S \rangle$ and $\langle Q_N \rangle$ are used to calculate the estimated SNR. There is a systematic overestimation of the SNR, i.e. the estimated values are usually greater than or equal to the actual SNR. This effect obviously reduces the noise reduction performance. However, the target signal information is not reduced in the processed signal, because the actual SNR appears to be a quite precise lower limit for the estimated values and the signal is thus not overattenuated by the respective gain factors.

The lower panels shows that the correlation coefficient r is considerably lower especially for the lower frequencies if the running estimates of Q_S and Q_N are used instead of their known, real values. Obviously, the running estimates of Q_S and Q_N are not to accurate enough yet to yield SNR estimates with approximatly the same accuracy as for the known values $\langle Q_S \rangle$ and $\langle Q_N \rangle$. This contrasts with the correlation gain factors described in the previous section where the performance only slightly decreased when using the running estimates. Hence, the running estimates of Q_S and Q_N should be improved in the future. However, an advantage of estimating the SNR using Eq. (5.11) is that the estimated values Q_S and Q_N are expected to change rather slowly in time, while the fast changing value Q_Y is computed directly from the short-time spectra. An application of gain factors given by (5.12) should thus result in small processing artefacts due to fast changing gain factors. This strategy is of course not expected to work in situations with multiple interfering signals. It is, however, expected to work well in situations without a target signal or an interfering signal, respectively.

5.2.3 Strategy 3: Directional filter

As an extension of the situation with one target sound source and one interfering sound source (right panel of Fig. 5.1), one target sound source at the front and one or more lateral interfering sound sources are considered now in an empirical approach. This approach is based on techniques for the suppression of lateral sound sources originally described by Gaik and Lindemann (1986) and Peissig (1993). The interaural level differences ΔL and interaural phase differences $\Delta\varphi$ are employed to separate parts of the signal which origin from sound sources with different azimuthal locations. These interaural differences are defined as:

$$\Delta L(f) \equiv 10 \cdot \log_{10} \left[\frac{Y_l(f) Y_l^*(f)}{Y_r(f) Y_r^*(f)} \right] \tag{5.13}$$

$$\Delta\varphi(f) \equiv \arg \left[Y_l(f) Y_r^*(f) \right], \tag{5.14}$$

where ΔL is calculated in dB and $\Delta\varphi$ is in the range of $[-\pi; +\pi]$.

For the further computations, reference values of interaural level and phase differences of particular, azimuthal sound incidence directions are required. These particular incidence directions $\alpha_{\text{front}} = 0°$, α_{pass} and α_{stop} will be referred to as frontal direction, pass-range direction and stop-range direction, respectively. These directions denote the azimuthal angle of the sound source at zero degree elevation, as depicted in the left panel of Fig. 5.4.

The reference values of the interaural differences are obtained as:

$$\left.\begin{array}{rcl}\Delta L_x(f) & \equiv & \langle \Delta L(f)\rangle \\ \Delta\varphi_x(f) & \equiv & \langle \Delta\varphi(f)\rangle\end{array}\right\}\text{with the sound source at } \alpha_x \text{ azimuth,} \tag{5.15}$$

where x denotes the front, pass or stop index, respectively and the expectation operator $\langle\cdot\rangle$ denotes the arithmetic mean value of a certain period, usually a few seconds. The mean values are calculated within the algorithm with appropriate signals, e.g., pink or white noise from the respective incidence direction in a non-reverberant environment.

The gain factors $G_{\Delta L}$ and $G_{\Delta\varphi}$, referred to as level and phase gain factors, are then computed from the actual interaural differences and the reference values. For this, the following definitions are used:

$$\delta X_{\text{pass}}(f) \equiv \min\left\{\left|\Delta X_{\text{pass}}(f) - \Delta X_{\text{front}}(f)\right|, \left|\Delta X_{\text{stop}}(f) - \Delta X_{\text{front}}(f)\right|\right\} \tag{5.16}$$

$$\delta X_{\text{stop}}(f) \equiv \max\left\{\left|\Delta X_{\text{pass}}(f) - \Delta X_{\text{front}}(f)\right|, \left|\Delta X_{\text{stop}}(f) - \Delta X_{\text{front}}(f)\right|\right\} \tag{5.17}$$

$$\delta X(f) \equiv \left|\Delta X(f) - \Delta X_{\text{front}}(f)\right|, \tag{5.18}$$

where X denotes the level L or the phase φ, respectively. The values of δL_{pass} and $\delta\varphi_{\text{pass}}$ represent the deviations of the mean level and phase differences at the pass-range direction from the respective mean values at the frontal direction. The same holds for δL_{stop}, $\delta\varphi_{\text{stop}}$ at the stop-range direction, while $\delta L(f)$ and $\delta\varphi(f)$ are the deviations of the actual level and phase differences from the respective mean values at the frontal direction.

The level and phase gain factors are defined as:

$$G_{\Delta L}(f) \equiv f(\delta L(f), \delta L_{\text{pass}}(f), \delta L_{\text{stop}}(f)) \tag{5.19}$$

$$G_{\Delta\varphi}(f) \equiv f(\delta\varphi(f), \delta\varphi_{\text{pass}}(f), \delta\varphi_{\text{stop}}(f)), \tag{5.20}$$

where $f(x, a, b)$ denotes the gain function depicted in the upper right panel of Fig. 5.4, which yields gain factors in dB. The gain factors $G_{\Delta L}$ and $G_{\Delta\varphi}$, respectively, are similar to the gain factors g_3 and g_2, respectively, described by Peissig (1993), with some modifications. The first modification is the frequency dependency of all reference values, which yields an approximately frequency independent directionality of the processing. Other modifications concern the frequency-specific combination and the application of the gain factors, which is described below.

The level and phase gain factors are of maximum value, if the actual level and phase differences exhibit values within the range of $\Delta L_{\text{front}} \pm \delta L_{\text{pass}}$ and $\Delta\varphi_{\text{front}} \pm \delta\varphi_{\text{pass}}$, and they are of minimum value, if the actual level and phase differences exhibit values outside the range of $\Delta L_{\text{front}} \pm \delta L_{\text{stop}}$ and $\Delta\varphi_{\text{front}} \pm \delta\varphi_{\text{stop}}$. This is expected to coincide with sound incidence directions within the azimuthal pass-range of $0^\circ \pm \alpha_{\text{pass}}$ and outside the azimuthal stop-range of $0^\circ \pm \alpha_{\text{stop}}$, respectively. Thus, the level and phase gain factors are expected to be high for signals from sound sources within the pass-range and low for other, lateral sound sources outside the stop-range with a smooth transition inbetween.

Finally, the level and phase gain factors are combined to obtain the gain factors G_{lat}, referred to as lateral gain factors:

$$G_{\text{lat}}(f) \equiv \frac{w_{\Delta L}(f) \cdot G_{\Delta L}(f) + w_{\Delta\varphi}(f) \cdot G_{\Delta\varphi}(f)}{w_{\Delta L}(f) + w_{\Delta\varphi}(f)}, \tag{5.21}$$

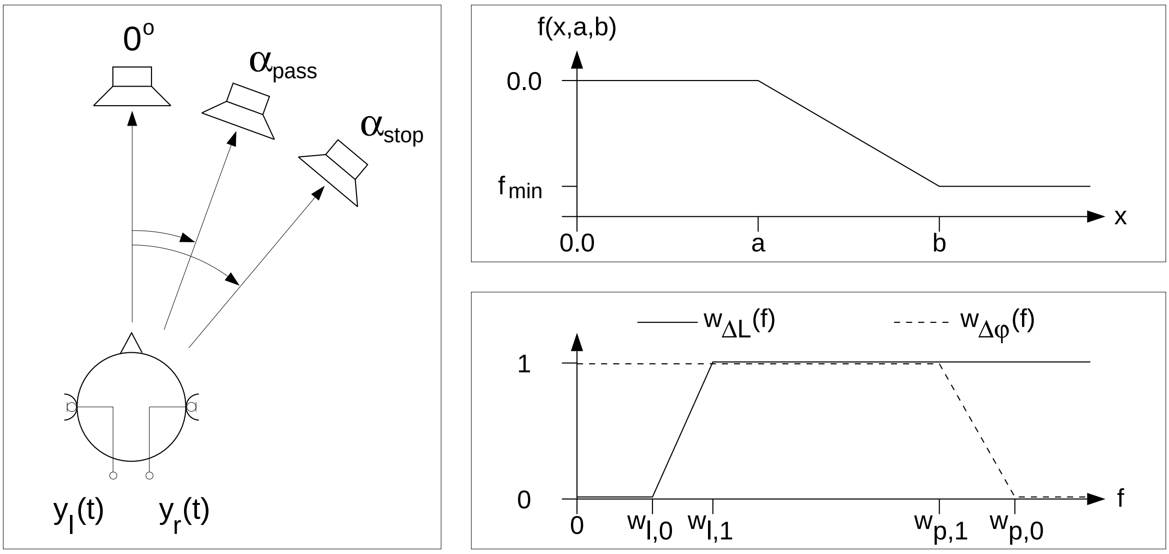


Figure 5.4: In the left panel, the spatial configuration for obtaining the reference values is shown. The sound sources are located at the azimuthal angle of 0°, α_{pass} and α_{stop} at zero degree elevation. In the upper right panel, the gain function $f(x, a, b)$ employed for computing the level and phase gain factors is depicted, where f_{min} is the adjustable minimum gain factor. $f(x, a, b)$ is calculated in dB. In the lower right panel, the frequency dependent weighting functions $w_{\Delta L}$ and $w_{\Delta\varphi}$ of the level and the phase gains are depicted, where $w_{l,0}$, $w_{l,1}$, $w_{p,1}$ and $w_{p,0}$ are the adjustable frequency limits.

where $w_{\Delta L}$ and $w_{\Delta\varphi}$ denote frequency dependent weighting functions for the level and the phase gain factors. Since the interaural level differences are negligible at very low frequencies, the level gains are useless there. On the other hand, at high frequencies the interaural phase differences comprise the whole range of $[-\pi; +\pi]$ (due to interaural time differences greater than half a wave cycle and phase wrapping) and thus the phase gains are useless there. The frequency dependent weighting functions, which are depicted in the lower right panel of Fig. 5.4, take this into account. The values of the employed frequency limits $w_{l,0}$, $w_{l,1}$, $w_{p,1}$ and $w_{p,0}$ have to be obtained from appropriate noise signals of different incidence directions in non-reverberant environment.

Fig. 5.5 shows directionality patterns, i.e., polar plots of the gain $G_{\text{lat}}(f)$ as a function of α and the respective Azimuth-plane directivity index for particular frequency bands and for the broadband condition. CCITT speech-shaped noise was used as acoustical signal in all conditions. The curves denote the mean frequency band gain for the particular frequency bands and the mean total gain for the broadband condition. All values shown were measured with the left-ear signals, the right-ear signals produced similar values. To measure the values, microphone signals of In-The-Ear (ITE) hearing instruments plugged into the ears of the Göttingen dummy head were recorded in an anechoic room. The

Azimuth-plane directivity indices were calculated as

$$D(f_k) \equiv \frac{|G_{\mathrm{lat}}(f_k, \alpha = 0)|^2}{\frac{1}{2\pi}\int_{\alpha=0}^{2\pi} |G_{\mathrm{lat}}(f_k, \alpha)|^2 d\alpha} \approx \frac{|G_{\mathrm{lat}}(f, \alpha = 0)|^2}{\frac{1}{N_l}\sum_{l=0}^{N_l-1} |G_{\mathrm{lat}}(f, \alpha_l)|^2} \quad (5.22)$$

$$DI \equiv \sum_k I_k 10 \log_{10} D(f_k), \quad (5.23)$$

where k denotes the index of the frequency band, f_k its respective center frequency, l the index of the azimuth angle, α_l its respective value and I_k the importance function (cf. Desloge *et al.*, 1997, for the directivity indices[2] and Pavlovic, 1987, for the importance function I_k for average speech). The directionality patterns exhibit the characteristic forward/backward ambiguity of the level and phase differences and also an increase in gain for $\alpha \approx 90$ and $\alpha \approx 270$ degrees for particular frequencies due to an increase in level at the opposite side of the head caused by constructive interference in the diffraction pattern.

The lateral gain factors have empirically been found to function as expected and to be able to considerably attenuate signal parts of lateral sound sources, if not too many lateral sound sources are present and/or the acoustical situation is not dominated by diffuse noise signals. Otherwise, the level and phase differences are deteriorated and unreliable and lead to deteriorated and thus useless lateral gain factors.

5.2.4 $\langle Q_S \rangle$ and $\langle Q_N \rangle$ estimator

The ratios $\langle Q_S \rangle$ and $\langle Q_N \rangle$, i.e., the expected values of Q_S and Q_N are employed for the computation of gain factors within the algorithm. To estimate $\langle Q_S \rangle$ and $\langle Q_N \rangle$, respectively, the lateral gain factors G_{lat} introduced in section 5.2.3 are used as follows:

If the value of $G_{\mathrm{lat}}(f)$ is greater than the limit $G_{\mathrm{lat,S}}$, the actual ratio Q_Y is assumed to represent Q_S at this particular frequency. On the other hand, if the value of $G_{\mathrm{lat}}(f)$ is less than the limit $G_{\mathrm{lat,N}}$, the actual ratio Q_Y is assumed to represent Q_N at this particular frequency. In one of these two cases, the expected value $\langle Q_S \rangle$ or $\langle Q_N \rangle$, respectively, is recalculated using Q_Y as Q_S or Q_N, respectively. Otherwise, nothing is done for this particular frequency, i.e., the values of $\langle Q_S \rangle$ and $\langle Q_N \rangle$ are kept unchanged. The limits $G_{\mathrm{lat,S}}$ and $G_{\mathrm{lat,N}}$ have to be adjusted to about 0.0 dB and a few dB less, respectively (cf. function $f(x, a, b)$ in Figure 5.4).

The employed expectation value operator $\langle \cdot \rangle$ of $\langle Q_S \rangle$ and $\langle Q_N \rangle$ is an intensity weighted first order low-pass filter with the minimum time constant τ_{Q_x}. For the intensity weighting, the maximum intensities $I_{\max}$ are calculated and low-pass filtered with a 0 ms attack time constant and a usually large release time constant $\tau_{I_{\max}}$:

$$I_{\max}(f) \equiv \max\{Y_l(f)Y_l^*(f), Y_r(f)Y_r^*(f)\} \quad (5.24)$$

$$\overline{I}_{\max}(f) \equiv \langle I_{\max}(f) \rangle \quad \text{with 0 ms attack, } \tau_{I_{\max}} \text{ release time constant.} \quad (5.25)$$

With this, the intensity weights w_I and the weighted filter factor γ_{Q_x} are defined as:

$$w_I(f) \equiv \frac{I_{\max}(f)}{\overline{I}_{\max}(f)} \quad \text{with } w_I(f) \in [0; 1] \quad (5.26)$$

[2] Note that Desloge *et al.* (1997) denote the Azimuth-plane directivity indices as $D_{az}(f_k)$ and $D_{az,I}$, respectively.

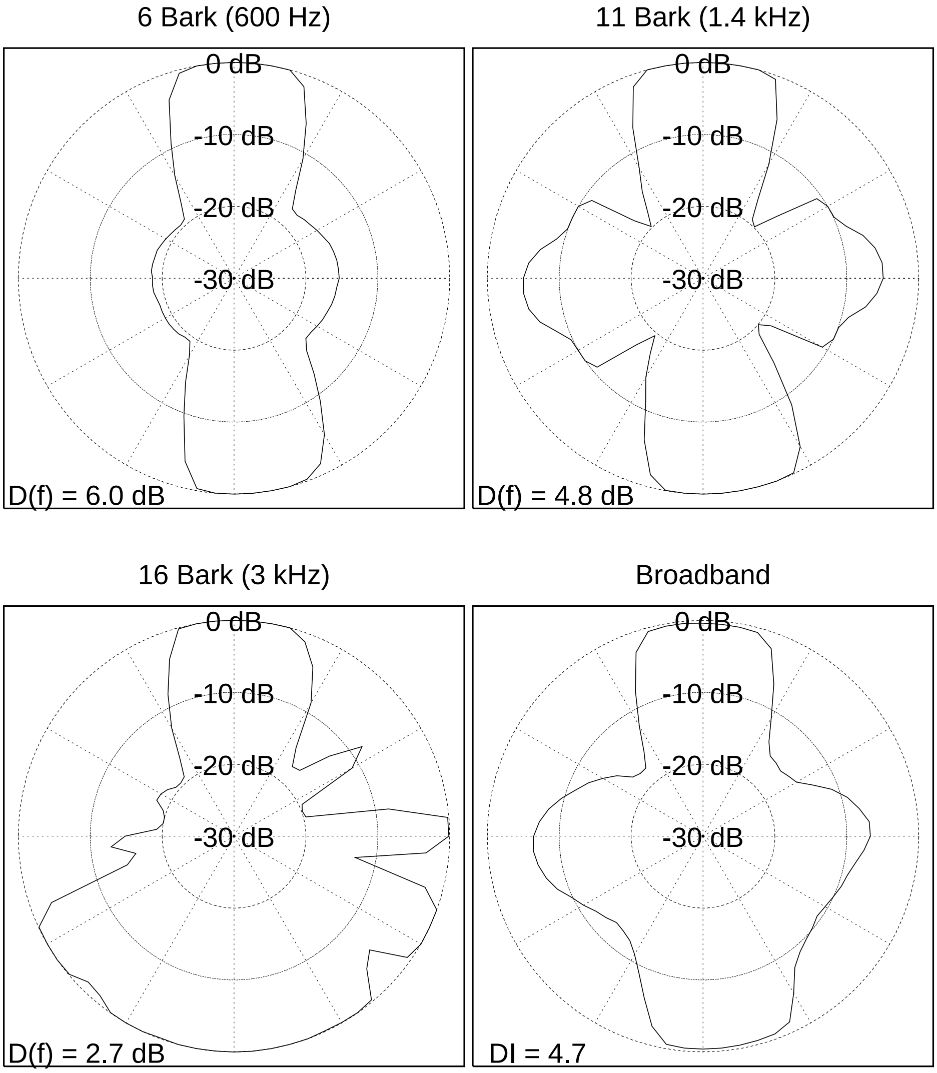


Figure 5.5: Measured gain $G_{\text{lat}}(f)$ as a function of the azimuthal incidence direction α (polar plots with top view). The respective Azimuth-plane directivity index $D(f)$ and DI, respectively, is given in each panel. All values are given in dB (dB is not specified for DI due to its definition). At the top of each panel is $\alpha = 0$, to the right 90, at the bottom 180 and to the left 270 degrees. The radius gives the gain as denoted at the circular grid lines. The upper panels and the lower left panel show the values for particular frequency bands with center frequencies of 6 Bark ≈ 600 Hz, 11 Bark ≈ 1.4 kHz and 16 Bark ≈ 3 kHz (bandwidth 1 Bark each). The lower right panel show the values for the broadband condition.

$$\gamma_{Q_x}(f) \;\equiv\; 1 - w_I(f) \cdot \left[1 - e^{-\left(\tau_{Q_x} \cdot f_r\right)^{-1}}\right], \tag{5.27}$$

where f_r denotes the frame rate of the short-time spectra. With n as time index of the STFT series, the expected value of Q_S or Q_N is then calculated as:

$$\langle Q_x \rangle_n \equiv \gamma_{Q_x}(f) \cdot \langle Q_x \rangle_{n-1} + (1 - \gamma_{Q_x}(f)) \cdot Q_x, \tag{5.28}$$

where x denotes S or N, respectively. The estimated values of $\langle Q_S \rangle$ and $\langle Q_N \rangle$ are initially set at the beginning of the processing using the reference values (which are originally obtained as complex values, anyway):

$$\langle Q_S \rangle_0 \equiv 10^{\frac{1}{20}\Delta L_{\text{front}}(f)} \cdot e^{i\Delta\varphi_{\text{front}}(f)} \tag{5.29}$$

$$\langle Q_N \rangle_0 \equiv 10^{\frac{1}{20}\Delta L_{\text{stop}}(f)} \cdot e^{i\Delta\varphi_{\text{stop}}(f)}. \tag{5.30}$$

In Figure 5.6, real, i.e., calculated mean values $\langle Q_S \rangle$ and $\langle Q_N \rangle$ are shown in comparison to their respective estimate, calculated from speech signals. The curves show that the estimates follow quite closely the calculated mean values. In this case, the phase was estimated more accurately than the magnitude.

As already noted in section 5.2.3, the lateral gain factors are deteriorated if too many lateral sound sources are present or the acoustical situation is dominated by diffuse noise signals. Hence, the estimated value of $\langle Q_S \rangle$ is also deteriorated in these cases, while $\langle Q_N \rangle$ is not a valid ratio due to the absence of a single interfering noise source. In both cases mentioned above, the cross-correlation of the left and right microphone signals is low or, in other words, the diffusiveness of the acoustical situation is high. In the next section, a method will be described to employ the degree of diffusiveness for the determination of the validity of the lateral gain factors and $\langle Q_S \rangle$ and $\langle Q_N \rangle$, respectively.

5.2.5 Degree of diffusiveness

The degree of diffusiveness d_d has been introduced and described in Chapter 4 as a measure of the general diffusiveness of an acoustical situation. It is a reasonable assumption that the lateral gain factors and $\langle Q_S \rangle$ and $\langle Q_N \rangle$, respectively, are the more deteriorated and unreliable, the more diffuse the current acoustical situation is, i.e., the higher the value of d_d is. Thus, the influence of the lateral gain factors and $\langle Q_S \rangle$ and $\langle Q_N \rangle$ on the particular gain factors of the algorithm will be controlled depending on the value of the degree of diffusiveness.

d_d is defined by Equation 4.10 and the respective transformation function $h(x)$, cf. Section 4.3. The calculated value of d_d is in the range of $[0;1]$, whereas 0 means complete correlation between left and right signal and 1 means complete diffusiveness, i.e., no correlation between the left and right signal. The value is initially set to 1 to reduce processing artefacts. The actual diffusiveness of an acoustical situation is then subdivided by means of the limits $d_{\text{corr,max}}$ and $d_{\text{diff,min}}$, which have to be empirically obtained from appropriate signals:

$$d_{\text{corr,max}} \equiv \text{upper limit of } d_d \text{ for mainly correlated signals} \tag{5.31}$$

$$d_{\text{diff,min}} \equiv \text{lower limit of } d_d \text{ for mainly diffuse signals,} \tag{5.32}$$

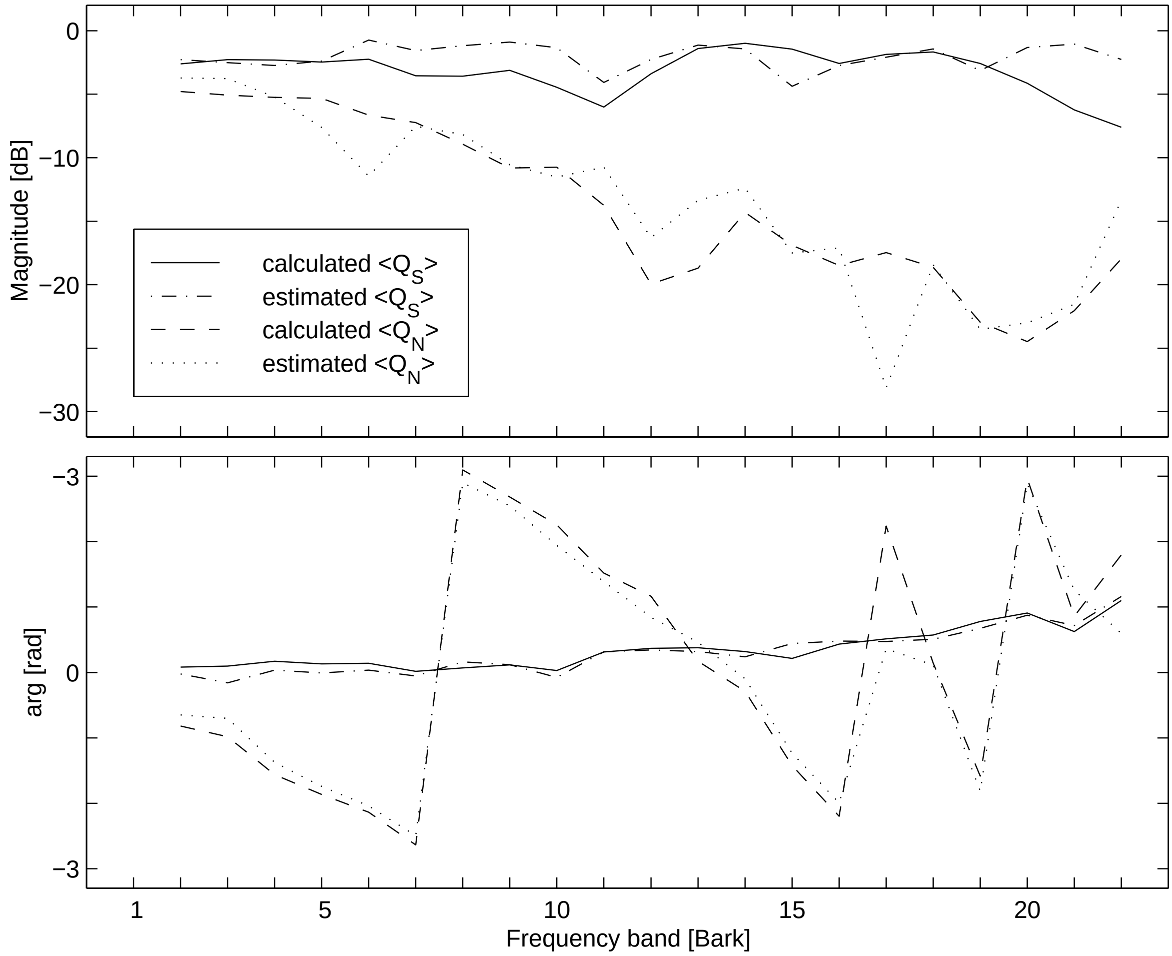


Figure 5.6: Calculated mean values $\langle Q_S \rangle$ and $\langle Q_N \rangle$ in comparison to the respective estimated values, depicted as magnitude and phase values in each frequency band. Anechoic dummy head recordings of running speech uttered by a male talker alone at the front and a female talker alone at 50 degrees azimuth, respectively, were used to calculate the mean values. The mixture of both speech signals was employed as input for the estimation of both values. The estimated values were taken as a snapshot from the algorithm after the processing of a few seconds signal (all estimates were initialized to zero at the start of the processing).

with $d_{\mathrm{corr,max}} \leq d_{\mathrm{diff,min}}$. For the use as a smooth transition function between the ratings "mainly correlated" and "mainly diffuse", t_d is defined as:

$$t_d \equiv \begin{cases} 1 & : \; d_d \leq d_{\mathrm{corr,max}} \\ \frac{d_d - d_{\mathrm{diff,min}}}{d_{\mathrm{corr,max}} - d_{\mathrm{diff,min}}} & : \; d_{\mathrm{corr,max}} < d_d < d_{\mathrm{diff,min}} \\ 0 & : \; d_d \geq d_{\mathrm{diff,min}} \end{cases} . \tag{5.33}$$

t_d may be used as an exponent with gain factors which are to be controlled by the diffusiveness. The application of t_d to particular gain factors will be described in the next section.

5.2.6 Gains

In the first stage of the noise reduction, the uncorrelated, i.e., diffuse parts of the left and right microphone signal spectra are attenuated. For this, the correlation gain factors $G_{\text{corr},l}$ and $G_{\text{corr},r}$ introduced in section 5.2.1 are employed with a modification concerning the influence of $\langle Q_S \rangle$. Taking into account the considerations made in section 5.2.4 and 5.2.5, the degree of diffusiveness and the transition function t_{diff}, respectively, is employed to determine the influence of $\langle Q_S \rangle$ when computing the correlation gains. With the definition of

$$|Q'_S| \;\equiv\; |\langle Q_S \rangle|^{t_{\text{diff}}} \tag{5.34}$$

as an "effective" estimate of Q_S which assumes a value of unity if the acoustical situation is too diffuse (i.e. $t_{\text{diff}} = 0$), the appropriate gain factors are obtained from (5.6) as:

$$G'_{\text{corr},l}(f) \;\equiv\; \sqrt{\frac{|\langle Y_l(f) Y_r^*(f) \rangle| \cdot |Q'_S|}{\langle Y_l(f) Y_l^*(f) \rangle}} \;\equiv\; \sqrt{\text{stcc}_l(f) \cdot |Q'_S|} \tag{5.35}$$

$$G'_{\text{corr},r}(f) \;\equiv\; \sqrt{\frac{|\langle Y_l(f) Y_r^*(f) \rangle| \cdot \frac{1}{|Q'_S|}}{\langle Y_r(f) Y_r^*(f) \rangle}} \;\equiv\; \sqrt{\text{stcc}_r(f) \cdot \frac{1}{|Q'_S|}}. \tag{5.36}$$

The correlation usually is low at high frequencies, if reverberation or diffuse noise is present. This results in a general low-pass characteristic of the correlation gains. Applying the square root to the gain factors above a certain frequency $f_{\text{corr},\sqrt{}}$ was empirically found to reduce this effect and thus increase the signal quality in reverberant or diffuse situations without affecting the signal quality in non-reverberant situations. The accordingly modified gain factors are defined as

$$\widehat{G}_{\text{corr},X}(f) \;\equiv\; \begin{cases} G'_{\text{corr},X}(f) & : \; f < f_{\text{corr},\sqrt{}} \\ \sqrt{G'_{\text{corr},X}(f)} & : \; f \geq f_{\text{corr},\sqrt{}} \end{cases}, \tag{5.37}$$

where X denotes l or r, respectively.

These correlation gains have been found to considerably attenuate diffuse noise and reverberation while preserving a high quality of the target sound in a wide range of acoustical situations, not only the one originally assumed in section 5.2.1. Additionally, the correlation gains achieve an effective feedback suppression, since a feedback howl usually does not occur simultaneously and correlated at both ears. Because of the latter, it is also recommended to apply the correlation gains in any acoustical situation for any binaural hearing aid signal processing application.

In the second stage of the noise reduction, the components of a single interfering sound source in the signal spectra are attenuated. For this, the interfering gain factors $G_{\text{int},l}$ and $G_{\text{int},r}$ introduced in section 5.2.2 are employed with a slight modification. Tests yielded that in reverberant situations, the interfering gain factors tend to overestimate the signal-to-noise ratios and thus the interfering gain factors (i.e. the gain factors are closer to unity as necessary). The original interfering gain factors given by (5.12) employ $|S_l + N_l|$ and $|S_r + N_r|$, respectively, as signal magnitude. The expression $|S_l|^2 + |N_l|^2$, for instance, is always greater than or equal to $|S_l + N_l|^2$, although the less correlated S_l and N_l are and the longer the employed short-time spectra are, the smaller is the difference between the two expressions. Informal listening tests revealed that employing $\sqrt{|S_l|^2 + |N_l|^2}$ and $\sqrt{|S_r|^2 + |N_r|^2}$, respectively, yields better results in reverberant situations than employing $|S_l + N_l|$ and $|S_r + N_r|$, respectively, without deteriorating the signal in non-reverberant situations. Thus, appropriate interfering gain factors are now defined as:

$$G'_{\mathrm{int},l}(f) \;\equiv\; \sqrt{\frac{\left|\frac{S_l}{N_l}\right|^2}{\left|\frac{S_l}{N_l}\right|^2 + 1}}\,, \qquad G'_{\mathrm{int},r}(f) \;\equiv\; \sqrt{\frac{\left|\frac{S_r}{N_r}\right|^2}{\left|\frac{S_r}{N_r}\right|^2 + 1}}, \tag{5.38}$$

where $\frac{S_l}{N_l}$ and $\frac{S_r}{N_r}$ are obtained employing (5.11) with the respective estimates of $\langle Q_S \rangle$ and $\langle Q_N \rangle$.

As already noted in section 5.2.4, the estimated value of $\langle Q_S \rangle$ is deteriorated and $\langle Q_N \rangle$ is not a valid ratio, if too many lateral sound sources are present or the acoustical situation is dominated by diffuse noise signals. Tests confirmed that in situations with diffuse noise, the interfering gain factors yield an indefinite attenuation of about 3 dB. Thus, the interfering gain factors are applied depending on the actual degree of diffusiveness. Again, the transition function t_{diff} is suitable for this. The appropriate interfering gain factors $\widehat{G}_{\mathrm{int},l}$ and $\widehat{G}_{\mathrm{int},r}$ for application to the left and right spectra, respectively, are defined as:

$$\widehat{G}_{\mathrm{int},l}(f) \;\equiv\; \left[G'_{\mathrm{int},l}(f)\right]^{t_{\mathrm{diff}}}\,, \qquad \widehat{G}_{\mathrm{int},r}(f) \;\equiv\; \left[G'_{\mathrm{int},r}(f)\right]^{t_{\mathrm{diff}}}. \tag{5.39}$$

These interfering gain factors have been found to considerably attenuate a single interfering sound source in non-reverberant and moderately reverberant situations while preserving a very high quality of the target sound.

In the third stage of the noise reduction, signal parts of lateral sound sources are attenuated. For this, the lateral gain factors G_{lat} derived in section 5.2.4 are employed. Since the lateral gain factors exhibit strong fluctuations in time, it is necessary to low-pass filter the gain factors given by (5.21) to avoid the so-called musical tone effect. Additionally, the lateral gain factors G_{lat} can not be calculated satisfactory in the presence of too many lateral sound sources or dominant, diffuse noise and thus also have to be applied depending on the actual degree of diffusiveness. Again, the transition function t_{diff} is used. The appropriate lateral gain factors $\widehat{G}_{\mathrm{lat}}$ for application to both the left and right spectra are given by:

$$\widehat{G}_{\mathrm{lat}}(f) \;\equiv\; \left[\langle G_{\mathrm{lat}}(f)\rangle\right]^{t_{\mathrm{diff}}}. \tag{5.40}$$

where the expectation value operator $\langle\cdot\rangle$ denotes a first order low-pass filter with the time constant τ_{lat}. These lateral gains have been found to considerably attenuate lateral sound sources in non-reverberant and reverberant situations.

As a combination of the gain factors derived so far, the total gain factors $\widehat{G}_l$ and $\widehat{G}_r$ defined as

$$\widehat{G}_l(f) \equiv \min\left\{\widehat{G}_{\text{corr},l}(f)\cdot\widehat{G}_{\text{int},l}(f),\widehat{G}_{\text{lat}}(f)\right\} \tag{5.41}$$

$$\widehat{G}_r(f) \equiv \min\left\{\widehat{G}_{\text{corr},r}(f)\cdot\widehat{G}_{\text{int},r}(f),\widehat{G}_{\text{lat}}(f)\right\} \tag{5.42}$$

have been found to be appropriate. They are directly applied to the spectra Y_l and Y_r of the left and right microphone signals to derive the corresponding, weighted spectra $\widehat{Y}_l$ and $\widehat{Y}_r$:

$$\widehat{Y}_l(f) \equiv \widehat{G}_l(f)\cdot Y_l(f)\ , \qquad \widehat{Y}_r(f) \equiv \widehat{G}_r(f)\cdot Y_r(f). \tag{5.43}$$

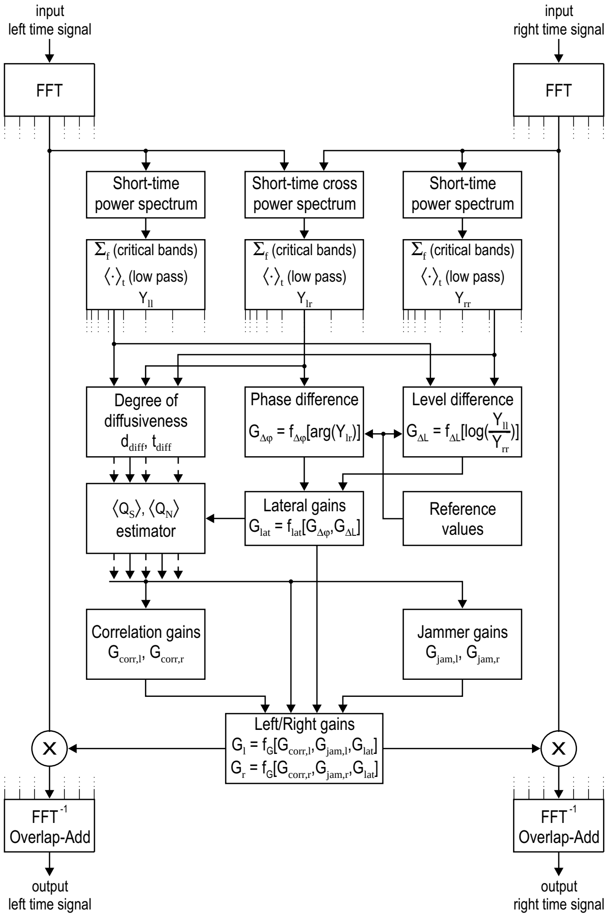

input left time signal
input right time signal
FFT
FFT
Short-time power spectrum
Short-time cross power spectrum
Short-time power spectrum
Σ_f (critical bands) $\langle\cdot\rangle_t$ (low pass) Y_{ll}
Σ_f (critical bands) $\langle\cdot\rangle_t$ (low pass) Y_{lr}
Σ_f (critical bands) $\langle\cdot\rangle_t$ (low pass) Y_{rr}
Degree of diffusiveness d_{diff}, t_{diff}
Phase difference $G_{\Delta\varphi} = f_{\Delta\varphi}[\arg(Y_{lr})]$
Level difference $G_{\Delta L} = f_{\Delta L}[\log(\frac{Y_{ll}}{Y_{rr}})]$
$\langle Q_S\rangle$, $\langle Q_N\rangle$ estimator
Lateral gains $G_{lat} = f_{lat}[G_{\Delta\varphi},G_{\Delta L}]$
Reference values
Correlation gains $G_{corr,l}$, $G_{corr,r}$
Jammer gains $G_{jam,l}$, $G_{jam,r}$
Left/Right gains
$G_l = f_G[G_{corr,l},G_{jam,l},G_{lat}]$
$G_r = f_G[G_{corr,r},G_{jam,r},G_{lat}]$
FFT^{-1} Overlap-Add
FFT^{-1} Overlap-Add
output left time signal
output right time signal

Figure 5.7: Block diagram of the strategy-selective algorithm for dereverberation and suppression of lateral noise sources.

Fig. 5.7 shows a block diagram of the algorithm. In the current implementation, an FFT of 512 samples is used with Hanning-windowed segments of 400 samples and an overlap rate of 0.5 at a sample rate of 25 kHz. The power spectra and complex cross power spectrum are then summed up across frequency within critical bands to yield a non-linear frequency scale with 23 bands each of 1 Bark bandwidth. The sum across a critical band of a power spectrum simply yields the total energy, while the sum across a critical band of a complex cross power spectrum yields the magnitude weighted mean phase difference as resulting phase and the cross power sum as resulting magnitude.

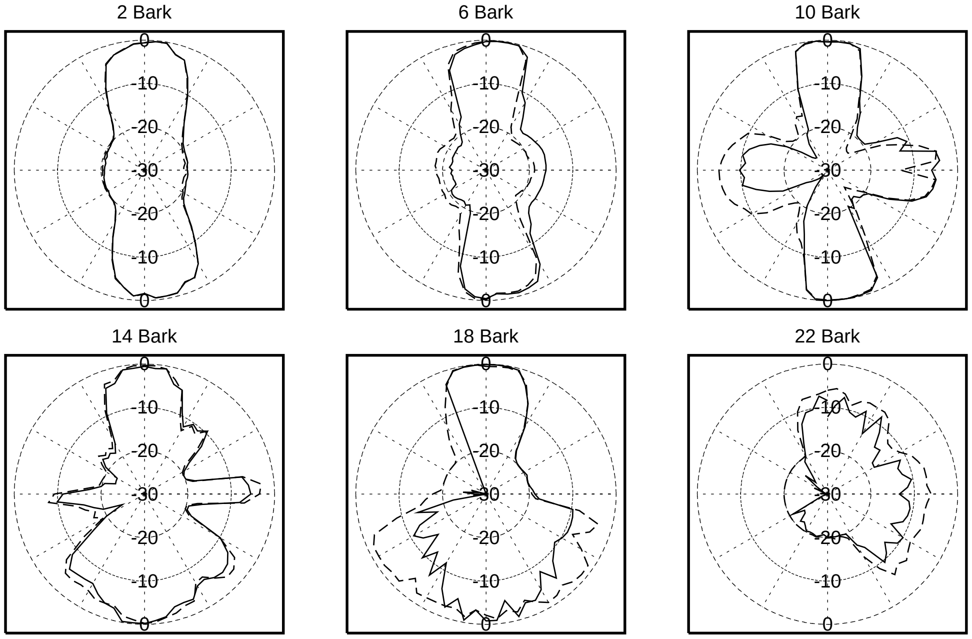


Figure 5.8: Narrow band directionality of the introduced algorithm, obtained with the Göttingen dummy head in an anechoic chamber. The center frequencies of the bands are denoted above each panel, bandwidth is 1 Bark. The resulting attenuations produced by the algorithm are given as a function of the sound incidence direction, where solid lines represent the left channel gains and dashed lines the right channel gains. The radius, i.e., the distance from the outermost circle gives the attenuation in dB. The numbers placed at the grid circles denote the respective attenuation represented by that particular radius. The azimuth angle is counted clockwise starting with 0° (frontal) at the top, 90° at the right, 180° (backward) at the bottom and 270° at the left side of the plot (think of the head depicted in the left panel of Figure 5.4 placed in the center). The employed reference pass and stop range angles α_{pass} and α_{stop}, respectively (see Figure 5.4), were 20 and 40 degrees, respectively.

In order to allow for a direct comparison of the new algorithm with the algorithm described in Chapter 3, Figure 5.8 shows the directionality patterns for selected frequency bands of 4 Bark distance in the same way as Figure 3.3 does for the algorithm in Chapter 3. In contrast to Figure 5.5, the total directionality of the whole algorithm is shown here. The directionality patterns of Figure 3.3 and Figure 5.8 are quite similar with respect to the front-backward symmetry, dissimilarities across frequency due to ambiguities of interaural level and phase differences for frontal and backward directions and also the diffraction effects for incidence directions close to 90 and 270 degrees azimuth at medium frequency bands. From the directionality patterns, no significant differences between the two algorithms can thus be concluded. From the differences in the processing, significant differences are expected to be observed mainly at subjective ratings of sound quality in different acoustical situations, especially in complex acoustic environments.

5.3 Optimisation of time constants based on subjective evaluation

The algorithm described above is characterized by a large number of parameters which all have to be adjusted appropriately. Some parameters like most parameters of the directional filter and the correlation gain factors can either be chosen from former investigations (i.e. α_{pass}, α_{stop}, f_{min}, $w_{l,0}$, $w_{l,1}$, $w_{p,1}$, $w_{p,0}$) or due to simple theoretical considerations or listening tests (i.e. $G_{\text{lat,S}}$, $G_{\text{lat,N}}$, $f_{\text{corr},\sqrt{}}$). The time constants of the different low-pass filters, however, have to be investigated in more detail. τ_{Q_x} and $\tau_{I_{\text{max}}}$, for instance, were found to be not too critical and adjusted according to informal listening tests. τ_Y and τ_{lat}, on the other hand, have a significant influence on the signal quality and also on the effect of the processing, although the signal quality is the major issue of the optimisation process. Thus, formal tests concerning the signal quality are described in the following which have been performed to find appropriate values for these time constants.

5.3.1 Procedure and subjects

The subjectively perceived quality of processed speech signals was compared for different values of the time constants τ_Y and τ_{lat} employed in the algorithm described above (see Sections 5.2.1 and 5.2.6). The values of these time constants were systematically varied during the experiment, and the experiment was repeated in different acoustical situations. For each acoustical situation, the subjects performed a complete paired comparison of all different versions, i.e., a subject compared each version with each other version exactly once. In each trial, the subject was forced to decide which version was perceived to be of better quality, i.e., which version was preferred to listen to. A judgement of equal quality was not allowed. The order of presentation for all paired versions and within each pair was randomized independently for each subject. In order to limit the total number of different versions, the test values were determined by using results from an informal listening test where the range of usable parameter values was sampled according to audible differences. This resulted in 18 versions and thus 153 comparisons for each situation. The particular values are listed below.

10 clinically normal hearing subjects participated voluntarily in each measurement. Some subjects received an expenditure compensation on an hourly basis. They were aged between 20 and 30 years and all had prior experience with psychoacoustical measurements.

5.3.2 Apparatus and stimuli

Figure 5.9 shows the spatial configuration of the three employed stimuli **s1**, **s5** and **s6**. All stimuli were recorded dichotically using ITE (In-The-Ear) hearing instruments, worn by a male subject ("central talker") who participated in a conversation. The employed module hearing instruments were Siemens Cosmea M devices with normal microphones and mounted to common ear moulds. The microphones of the hearing instruments were connected to a DAT recorder during the recording. Each stimulus consists of a conversation between the central talker and a male target talker sitting in front at a distance of about one meter. In stimulus **s1**, the conversation takes place in quiet inside a regular seminar room (reverberation time about 0.5 seconds). In stimulus **s5**, an additional interfering female talker utters text passages from a book and is moving slowly within an azimuthal angle range of 45 to 90 degrees with respect to the central talker. In stimulus **s6**, the conversation takes place in a cafeteria during lunch time with a loud and mainly diffuse background noise.

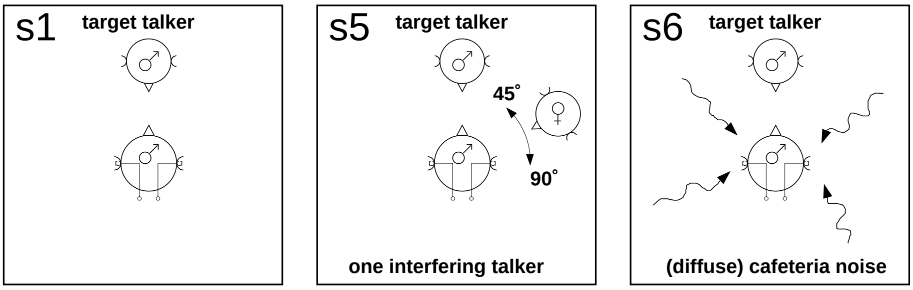


Figure 5.9: Spatial configuration of stimuli **s1**, **s5** and **s6**.

For the assessment, the original signals were loaded from hard disk during the measurement and processed in realtime by a DSP subsystem with five TI TMS320C40 digital signal processors. The processed signals were presented dichotically via amplifier and headphones (Sennheiser HD25) in a sound-insulated booth. For the stimuli **s1** and **s5**, the presentation level was in the range of 65 to 70 dB SPL (coupler measurements) with a signal-to-noise ratio (SNR) of about 0 dB. For stimulus **s6**, the presentation level was up to 78 dB SPL with an SNR of about -10 dB. All signals were presented without further frequency shaping. It was assumed that the frequency response of the whole system, being the same for all presentations, did not affect the relations of the paired comparison judgements. At the beginning of each paired comparison, the signal (about 1 minute of running speech) was presented in an endless loop, starting with the first type of processing switched on. The subject was able to switch the processing type whenever she or he liked to using a handheld touchscreen response box (EPSON EHT-10S), selecting choice "1" or "2". This switching was put into effect without a considerable delay or an interruption of the stimulus presentation. With the processing judged to be of higher quality switched on, the selection of the third choice "better" ended the comparison task.

5.3.3 Parameters

The 17 different test values of the parameter set $(\tau_Y, \tau_{\text{lat}})$ in milliseconds were $(1, 8)$, $(1, 20)$, $(1, 60)$, $(1, 100)$, $(8, 20)$, $(8, 60)$, $(8, 100)$, $(20, 8)$, $(20, 20)$, $(20, 60)$, $(20, 100)$, $(40, 8)$, $(40, 20)$, $(40, 60)$, $(60, 8)$, $(60, 20)$ and $(60, 60)$.

The values of the other parameters were virtually set to the values proposed in the prior description of the algorithm. Concerning the degree of diffusiveness, however, the parameters $d_{\mathrm{diff,min}}$ and $d_{\mathrm{corr,max}}$ had to be adjusted to be appropriate for the setup and stimuli used in this experiment[3]. This was done by evaluating d_d for the employed stimuli and a variety of additional recordings (with the same microphones) and for all values of τ_Y used in the experiment. The limits $d_{\mathrm{diff,min}}$ and $d_{\mathrm{corr,max}}$ where then chosen in a way that the stimuli **s1** and **s5** were rated as being mainly correlated, and stimulus **s6** as being mainly diffuse at nearly any time (with the values being automatically adjusted for different time constants τ_Y). As a result of this procedure, the algorithm selected the respective processing strategies for the stimuli without changing this selection during the presentation of a particular stimulus.

It should be noted that the degree of diffusiveness itself was calculated in a simplified way using a level independent time constant τ_d for the final low-pass filter (see Section 4.3). But since the limits $d_{\mathrm{diff,min}}$ and $d_{\mathrm{corr,max}}$ where appropriately adjusted and the particular stimuli employed here exhibited no considerable speech or noise pauses, this simplified calculation had no considerable influence on the signal processing.

The employed values of the processing parameters are listed in Table 5.1.

Parameter	Value	Section
τ_Y	1 to 60 ms	5.2.1
τ_{lat}	8 to 100 ms	5.2.6
α_{pass}	20 deg.	5.2.3
α_{stop}	40 deg.	
f_{min}	-20 dB	
$w_{l,0}$	5 Bark	
$w_{l,1}$	7 Bark	
$w_{p,1}$	12 Bark	
$w_{p,0}$	14 Bark	
$G_{\mathrm{lat,S}}$	0.01 dB	5.2.4
$G_{\mathrm{lat,N}}$	-2.0 dB	
τ_{Q_x}	200 ms	
$\tau_{I_{\mathrm{max}}}$	2 s	
$f_{\mathrm{corr},\sqrt{\ }}$	20 Bark	5.2.6

Table 5.1: Employed parameters of the algorithm.

Additionally, appropriate reference values for level and phase differences of different sound incidence directions were required for the calculations of the directional filter within the algorithm. These reference values were obtained in an anechoic chamber with the same central talker and ITE hearing instruments employed for all signal recordings. All signals were processed using these reference values.

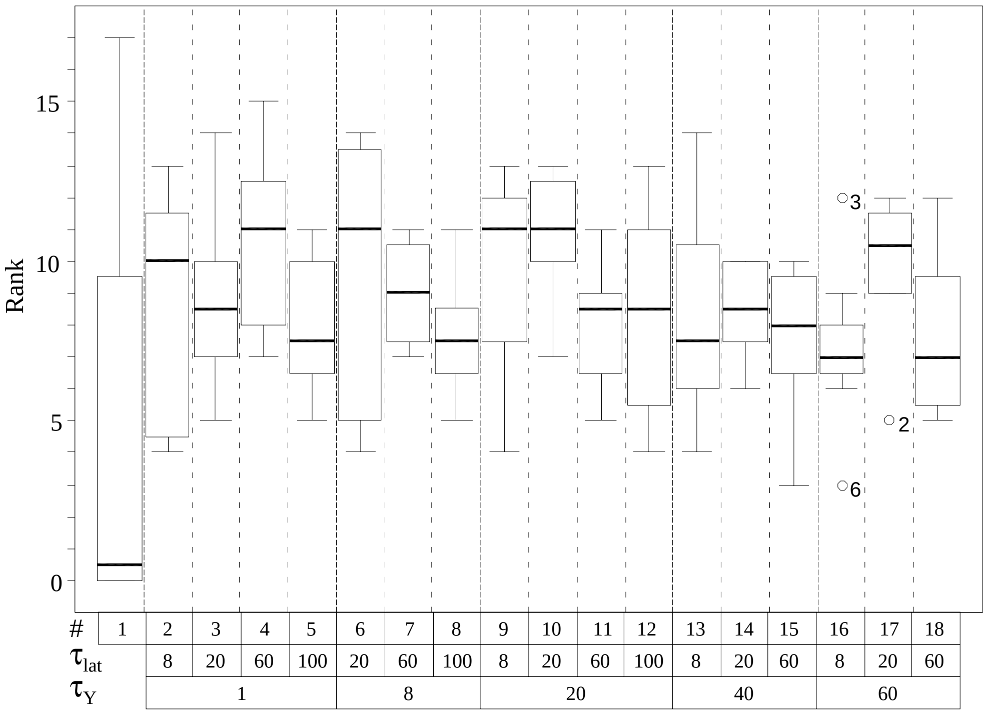


Figure 5.10: Relative ranks of versions for stimulus **s1**. On the abscissa, the 18 different versions of the stimulus are shown. The number of the version is given directly below the axis in the row denoted with a # on the left. Version number 1 is the unprocessed stimulus. For the other versions, the values of the processing parameters τ_Y and τ_{lat} are given in the accordingly denoted rows below. The vertical lines in the rows separate different values. The ordinate gives the rank, i.e., the number of "better" judgements with a maximum possible value of 17. The thick horizontal lines denote the median values, the boxes the range from the first to the third quartile and the outer bars the total range. Circles and asterisks represent outlyers (with the number of the respective subject specified).

5.3.4 Results

The results for stimulus **s1** are depicted in Figure 5.10. First, the consistence of the results, i.e., the consistence of all "better" judgements with each other was calculated for each subject. For this, the method of Kendall (1975), described by Bortz *et al.* (1990) was employed. Only 8 of the 10 subjects exhibited significantly consistent results and were included in the further evaluation ($\alpha < 0.01$). Then, the agreement of the judgements across the subjects was evaluated with the method described by the above authors. The resulting coefficient of agreement was $A = 0.02$ (with $\alpha > 0.05$). The subjects thus exhibited no significant agreement, and a Friedman test also revealed no significant influence of the processing version on the results ($\alpha > 0.05$). Hence, these results will not be discussed in detail here. However, all processed versions were ranked higher than the unprocessed

[3] The calculated value of d_d depends not only on the stimulus, but also on the transfer functions of the microphones and the employed particular value of τ_Y.

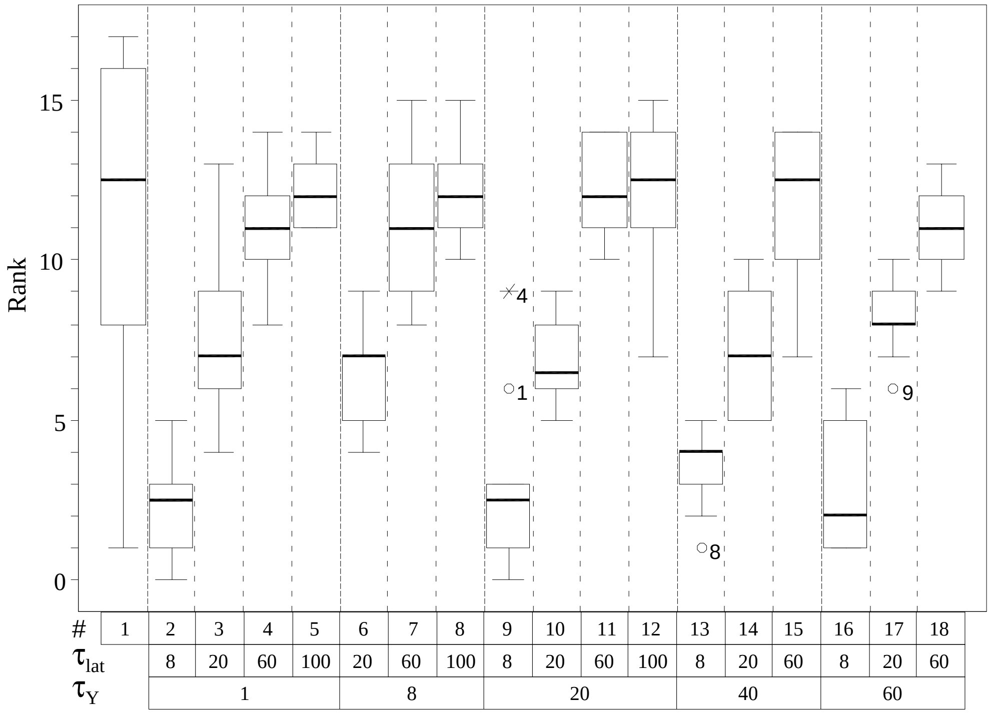


Figure 5.11: Relative ranks of versions for stimulus **s5**. See Figure 5.10 for details.

version in this situation (conversation in quiet). It is striking that the ranks of the unprocessed version comprise the whole range of 0 up to 17, while its median value is about 0.5 and thus extremely low. In contrast, the median values of all processed versions comprise only the small range of 7 up to 11 (i.e., they were all ranked more or less in the middle of the possible range). This led to a closer look at the results obtained for stimulus **s1** which is described later.

The results for stimulus **s5** are shown in Figure 5.11. Again, the consistence of the results, i.e., the consistence of all "better" judgements with each other was calculated first for each subject. Here, the results of all subjects exhibited significantly consistent results and were included in the further evaluation ($\alpha < 0.001$, in two cases $\alpha < 0.05$). The coefficient of agreement exhibited a value of $A = 0.33$ with significance level $\alpha < 0.001$, i.e., the agreement of the judgements across subjects was significantly higher than for judgements obtained at random. A Friedman test additionally revealed a significant influence of the version on the results ($\alpha < 0.001$). A Wilcoxon test was then performed for each pair of versions to determine significant differences in the results. The time constant τ_Y in general shows little influence on the sound quality. For a fixed value of τ_Y, however, the rank increases with an increasing value of τ_{lat}. Moreover, the version with the respective smallest values of τ_{lat} was significantly ranked lower than the versions with larger values of τ_{lat} or than the unprocessed version ($\alpha < 0.01$, in some cases $\alpha < 0.05$). Large values of τ_{lat} thus seem to be appropriate here. It should be noted that again the variability of the ranks of the unprocessed version is extremely high, while the variability is rather small for all processed versions.

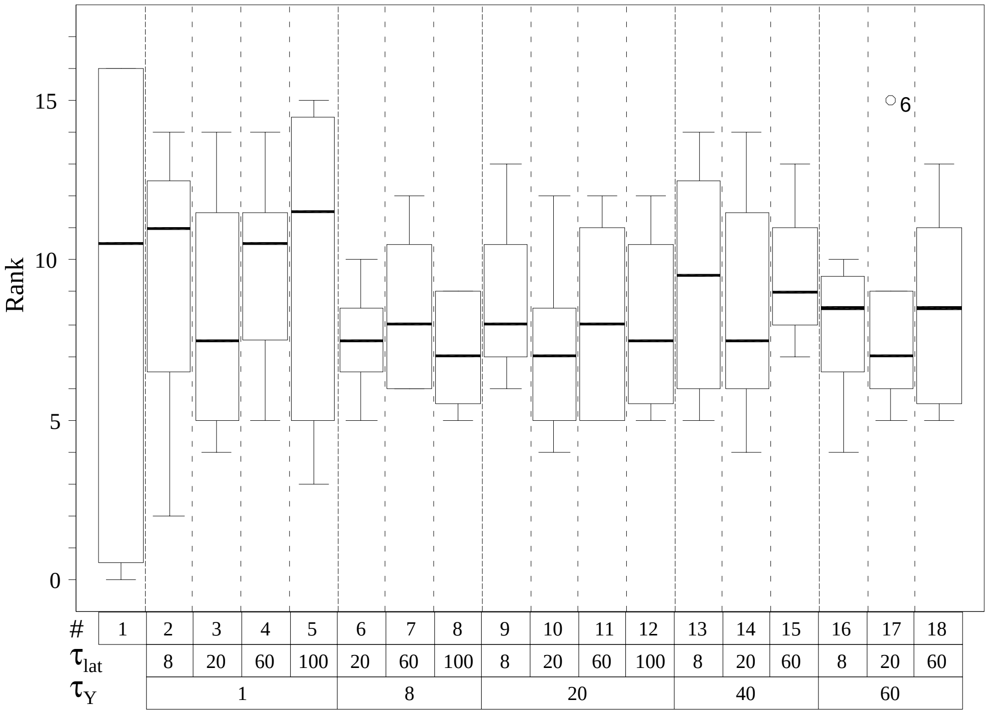


Figure 5.12: Relative ranks of versions for stimulus **s6**. See Figure 5.10 for details.

The results for stimulus **s6** are shown in Figure 5.12. Like for stimulus **s1**, only 8 of the 10 subjects exhibited significantly consistent results and were included in the further evaluation ($\alpha < 0.01$). And again, these subjects exhibited no significant agreement ($A = -0.02$ with $\alpha > 0.05$), and a Friedman test also revealed no significant influence of the processing version on the results ($\alpha > 0.05$). Hence, the results will not be discussed in detail. Obviously, the time constant τ_Y has no significant effect on sound quality in this situation (for normal hearing subjects). The time constant τ_{lat} should have no influence at all in this situation anyway, because the directional filter is switched of in this situation due to the high degree of diffusiveness. However, it should be noted that again the variability across subjects of the ranks of the unprocessed version is extremely high, i.e., the ranks comprise almost the whole range of possible values.

As mentioned above, the results for stimulus **s1** were investigated in more detail because of the totally different rankings of the unprocessed version in comparison to all processed versions. The subjects were split up into two groups. 5 subjects who rated the unprocessed version with a rank less than 8.5 were considered as group **PP**, which means that they mainly preferred the processed versions. 3 subjects rated the unprocessed version with a rank greater than 8.5 and were considered as group **PU**, which means that they mainly preferred the unprocessed version. The separate results of both groups are depicted in Figure 5.13.

Due to the particular classification of the subjects, no further conclusions concerning the rank of the unprocessed version can be drawn from the split results. However, when evaluating these two groups separately, Friedman tests revealed a significant influence of

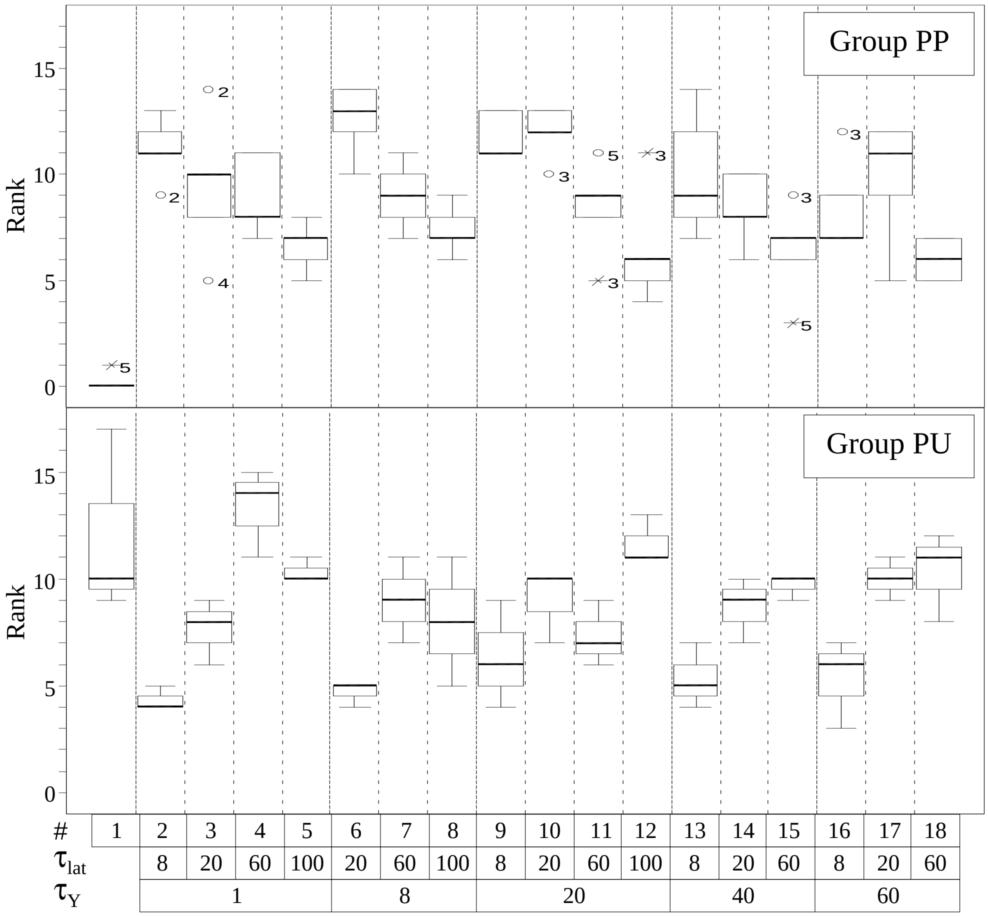


Figure 5.13: Relative ranks of versions for stimulus **s1** as shown in Figure 5.10, here depicted separately for group **PP** (upper panel) and group **PU** (lower panel). Group **PP** represents all subjects who mainly preferred the processed versions, while group **PU** represents all subjects who mainly preferred the unprocessed version.

the processing version for both groups ($\alpha < 0.01$) and thus conclusions can be drawn about the influence of the tested time constants on the ranking of the processed versions within each group. Wilcoxon tests were then performed for each pair of versions to determine significant differences in the results. For group **PP**, the rank decreases with an increasing value of $\tau_{\rm lat}$ (for a fixed value of τ_Y). In some cases, the largest value of $\tau_{\rm lat}$ was ranked significantly lower than smaller values ($\alpha < 0.05$). No significant differences between versions with different values of τ_Y were found. It should be noted that, although the rank of the unprocessed version is surely determined by the criterion of group **PP**, there is almost no variability at all in the ranks of the unprocessed version within this group. For group **PU**, the small number of subjects does not allow for any significant differences to be found. However, in contrast to group **PP**, there is a tendency that the rank increases with an increasing value of $\tau_{\rm lat}$. An increasing value of $\tau_{\rm lat}$ can be considered as an decreasing amount of processing, because a larger time constant results in slower changes of the gains.

Obviously, the subjects of group **PP** not only clearly prefer the processed versions, they also prefer a higher amount of processing to a smaller amount of processing. In contrast to this, the subjects of group **PU** in general prefer the unprocessed version and they also prefer "less processed" versions to "more processed" versions. In order to yield a signal quality similar to or even better than for the unprocessed signal for both groups of subjects, a rather large value of τ_{lat} (about 60 ms) seems to be appropriate for τ_{lat} in this situation.

Taken together, values of about 40 ms for τ_Y and about 60 ms for τ_{lat} are appropriate for all investigated situations. For these values, the quality of all processed stimuli was judged as being similar to or even better than for the unprocessed version.

The effect of the unprocessed version being ranked with an extremely high interindividual variation is consistent with the findings of the experiments described in Chapter 3. Obviously, the subjects assess an unprocessed stimulus somehow different than a processed version of it. This holds for for different spatial noise situations, different processing strategies and a variety of different parameter settings. This indicates that subjects are able to detect even small signal processing artefacts very well. It seems possible that subjects notice whether a signal is processed or not and then judge depending on what they expect from a signal processing in the respective situation. Thus, conclusions about relative differences between unprocessed and processed versions have to be drawn carefully and might be limited to a certain group of subjects.

5.4 Discussion

The binaural noise reduction algorithm described in this chapter was developed with the aim of providing a high subjective sound quality of processed signals for a variety of acoustical situations. Additionally, no particular assumptions were made about a specific kind of target signals, such as, e.g. speech, and a specific kind of interfering signals, such as, e.g. gaussian noise. The main assumption of the algorithm is the spatial (azimuthal) separation of target and interfering signals, and diffuse signals are considered as being undesired noise. Although target and interfering signals are assumed to be not correlated for parts of the processing, other parts of the algorithm are not based on this assumption.

The correlation gain factors described in Section 5.2.1 were demonstrated to exhibit a high correlation between the calculated gain and the theoretically required gain for a totally diffuse noise signal, even if the employed ratio Q_S of left over right target signal spectra is estimated during the processing. Informal listening tests also showed that applying the correlation gain factors yields a high subjective signal quality in easy acoustical situations as well as in very difficult situations, e.g. situations with many different interfering sound sources. Moreover, the employed correlation gain factors exhibit considerable less audible fluctuations in time than the formerly used technique of directly applying the magnitude squared coherence (cf. Peissig, 1993).

The interfering gain factors described in Section 5.2.2 were derived under the assumption of only two different sound sources present. Using this assumption, the SNR can be calculated from the interaural relation, i.e. ratios of the signals (also referred to as binaural parameters), without having to know the target or interfering signal itself. With the knowledge of all binaural parameters, the correlation between the calculated SNR and the real SNR was shown to be quite high. However, under realistic conditions, these binaural parameters have to be estimated from the actual microphone signals, which are in fact a mixture of parameters from both sound sources. An appropriate technique is thus required to extract or estimate the binaural parameters of each sound source alone. The decision unit employed for the estimation (cf. Section 5.2.4) is based on the directional filter strategy that is also a part of the algorithm. Although this decision unit is able to estimate magnitude and phase of the required binaural ratios $\langle Q_S \rangle$ and $\langle Q_N \rangle$, the interaural difference in level is not estimated as accurately as the interaural difference in phase even under non-reverberant conditions (cf. Figure 5.6). A possible explanation might be the influence of fluctuations of the numerator and the denominator on the estimated magnitude (e.g., the linear mean value of -6 dB and +6 dB is not 0 dB, but approx. +1.9 dB). The inaccurate magnitude of the estimated ratios $\langle Q_S \rangle$ and $\langle Q_N \rangle$ might again be the reason that the estimate of the SNR calculated from the estimated ratios $\langle Q_S \rangle$ and $\langle Q_N \rangle$ considerably differs from the SNR calculated with known ratios $\langle Q_S \rangle$ and $\langle Q_N \rangle$ (within the calculation of the interfering gain factors). An improvement of the decision unit or in general of the $\langle Q_S \rangle$ and $\langle Q_N \rangle$ estimator is thus desireable and should be a main topic of further investigation on this processing strategy. An advantage of this processing strategy is the high subjective quality of the processed signal which was observed in informal listening tests. Even if the estimated $\langle Q_S \rangle$ and $\langle Q_N \rangle$ are inaccurate or the acoustical situation is more complex than the assumed presence of two sound sources, the signal is slightly attenuated, but beyond that not considerably deteriorated. The expected low fluctuations in time of the gain factors were also confirmed in informal listening tests.

The lateral gain factors described in Section 5.2.3 represent a modified version of the directional filter described by Peissig (1993). They have empirically been found in the past to function as expected and to be able to considerably attenuate signal parts of lateral sound sources, if not too many lateral sound sources are present and/or the acoustical situation is not dominated by diffuse noise signals. Otherwise, the processed signal exhibits audible processing artefects which considerably decrease the signal quality. However, the integration of the so-called degree of diffusiveness into the algorithm makes, amongst other possible applications, a full usage of the advantages of the lateral gain factors possible while avoiding its well-known drawbacks (cf. Sections 5.2.5 and 5.2.6). This is a great advantage over former algorithms which also integrated different processing strategies, but without having the possibility to automatically switch particular strategies or to adapt them to different acoustical situations. The degree of diffusiveness is an indicator for the complexity of the current acoustical situation with respect to the number of sound sources and the amount of reverberation. The properties of this measure should be investigated in more detail in the future, especially with respect to possible applications in hearing aid algorithms.

The time constants τ_Y and τ_{lat} were investigated in detail in order to find the optimum values for different acoustical situations. For the (diffuse) cafeteria situation, no significant influence of the time constants on the signal quality was found. No optimisation of these parameters is thus required for this situation. For the situation with one target talker and one interfering talker, however, higher values of τ_{lat} yield significantly better results than lower values, while τ_Y again shows no significant influence. For the situation with only one target talker alone, however, the results are not that clear. At least there is some evidence that confirms the above finding concerning τ_{lat}. Since the results exhibit no clear optimum, the value of τ_Y employed for further experiments described in the next chapter was eventually chosen according to former investigations of Wittkop *et al.* (1997), where parts of the algorithm were used.

The experiments described above are quite similar to the experiments described in Chapter 3, where a slightly modified version of the algorithm after Peissig (1993) was investigated. Most results of the previous experiments from Chapter 3 and the experiments described in this study are consistent. In contrast to the previous results, however, no critical influence of the time constant τ_Y was found for the new algorithm in the diffuse cafeteria noise situation. Since in the new algorithm, the directional filter is switched off in this situation while it is not in the previous algorithm, the deterioration of the directional filter gains in this situation is a reasonable explanation of this effect. The experiments described above thus also demonstrate that τ_Y is not critical for the correlation gain factors with respect to the sound quality.

Taken together, for both time constants τ_Y and τ_{lat} of the new algorithm values were found that are appropriate for all investigated acoustical situations. Further experiments can thus be performed for a variety of situations without changing the parameters. Moreover, the described algorithm seems to be indeed suitable for different acoustical situations with a single set of parameters, since the processed versions were rated by the normal hearing subjects as being of equal or even higher quality than the unprocessed version in all tested situations (when using the "optimum" values found for τ_Y and τ_{lat}). This is a promising result, because common single-microphone noise reduction techniques described in the literature are often reported to provide no benefit or even to produce deterioration with respect to speech intelligibility in speech in noise tests under realistic conditions (cf. Dillon and Lovegrove, 1993; Marzinzik, 2000). Since the benefit provided for hearing-impaired listeners is often even larger than for normal-hearing listeners, it can be expected that the described algorithm is appropriate for noise reduction in practical digital binaural hearing aids.

5.5 Conclusions

Binaural recordings or hearing aid arrangements can be used for promising noise reduction processing strategies. Two strategies of this kind that are based on physical binaural signal parameters were derived theoretically in this work. One strategy (correlation gain) exhibits a very high signal quality and is suitable for application in all investigated acoustical situations. The other strategy (interfering gain) was shown to function at least if the underlying assumptions are fullfilled and all required (binaural) information is known. In realistic situations, however, the accuracy of the signal-to-noise ratio estimation decreases propably due to inaccurate estimations of binaural signal parameters. The signal quality, on the other hand, was found to be high in all situations. A more accurate estimation of binaural signal parameters is thus desirable and should be an issue of future investigation.

A total of three different binaural noise reduction strategies were integrated in a complex noise reduction algorithm. The algorithm was especially designed to yield a high quality of the processed output signals. First tests with normal hearing subjects show that the signal quality indeed is high for different acoustical situations with only one single set of processing parameters (time constants).

In general, the theoretical evaluations and informal listening tests of the processing strategies are quite promising. However, all objective measures available so far and the quality assessments made within this study do not yet give evidence about the actual noise reduction performance. Hence, an evaluation with hearing impaired subjects has to be performed in the future.

Chapter 6

Strategy-selective noise reduction algorithm: Evaluation with hearing impaired listeners

Abstract

The previously introduced strategy-selective binaural noise reduction algorithm for hearing aids is evaluated with eight hearing-impaired subjects who exhibit two different types of hearing loss (high frequency hearing loss and flat hearing loss). The subjective preference as well as speech reception thresholds (SRTs) in noise are measured under realistic free-field conditions in a laboratory environment. The subjective preference is assessed with a complete paired comparison paradigm including a consistency evaluation. The SRTs are measured using a sentence test with an adaptive procedure. Additionally, SRTs are measured for a diotic listening condition presented by headphones. The results of the subjective assessment show a high quality of the processed signal especially in the diffuse cafeteria noise situation. The algorithm is shown to be able to improve the SNR under certain conditions, although a significant improvement of the SRT is found only in the case of diotic presentation. The results also suggest that there might be an improvement of speech intelligibility for subjects with a flat hearing loss in the free-field (dichotic) listening situation with interfering speech signals or diffuse cafeteria noise. Since the results exhibit no deterioration of signal quality or speech intelligibility in any investigated situation and some improvement in some of the situations, the strategy-selective algorithm appears very promising for real life conditions.

6.1 Introduction

A variety of binaural hearing aid algorithms have been developed, implemented and evaluated in the past in order to give hearing-impaired listeners support in their deteriorated speech communication in noise. These algorithms generally aim at reducing ambient noise and undesired sounds while preserving or even enhancing target speech signals which usually are assumed to originate from the front of the listener. In laboratory studies, the developed algorithms always appeared to be very promising with respect to sound quality and speech intelligibility (cf. Peissig, 1993; Kollmeier *et al.*, 1993; Kollmeier and Koch, 1994; Wittkop *et al.*, 1997). In several field tests, however, there was no evidence found for any real benefit for hearing impaired people provided by the processing implemented in wearable devices (cf. Albani *et al.*, 1998; Pastoors *et al.*, 1998). Moreover, subjects even complained about the poor sound quality in some acoustical real-life situations. This motivated the development and optimization of the algorithm described in Chapter 5 which aims at providing an acceptable sound quality under different acoustical conditions.

In the study described here, this new algorithm is investigated with respect to sound quality and speech intelligibility with hearing impaired subjects under realistic acoustical free-field conditions. The former algorithm after Peissig (1993) is also evaluated in order to allow for a comparison. Although performed in a laboratory environment, the employed stimuli and acoustical conditions are very similar to every-day life situations. The laboratory setup, on the other hand, allows for controlling certain factors which may have an influence on the perceived sound quality and the speech intelligibility, e.g. absolute presentation level, hearing instrument configuration and acoustic environment. The influence of the type of hearing loss (i.e. flat or high frequency loss) has been investigated in some of the experiments. The experiments include judgements of subjective preference in order to assess which noise reduction processing technique sounds better to the subjects (including the case of no noise reduction processing) and whether the preferences are significant or not. Additionally, speech reception thresholds for sentences were measured for the different processing techniques as a direct measure of speech intelligibility in noise using the Oldenburg sentence test (cf. Wagener *et al.*, 1999a-c).

6.2 Algorithms and parameters

The first algorithm, referred to as Algorithm "Fixed" after Peissig (1993, cf. Kollmeier *et al.*, 1993), has been described in detail in Chapter 3. The values of the processing parameters were chosen according to the specifications and recommendations given in Chapter 3 (with some slight deviations due to an adjustment with respect to the parameters of the new algorithm). In particular, the parameters which have been varied in Chapter 3 were set to $\alpha_{\text{pass}} = 20$, $\alpha_{\text{stop}} = 40$, $\tau_1 = 8$, $\tau_2 = 60$, $a_1 = -20$ and $a_2 = -20$.

The second algorithm investigated is the new algorithm described in Chapter 5, referred to as Algorithm "Selective". The values of the processing parameters of this algorithm were chosen according to the specifications given in Chapter 5. Hence, the time constants were set to $\tau_Y = 40$ ms and $\tau_{\text{lat}} = 60$ ms.

For both algorithms, binaural reference values are required for the signal processing of the directional filter (cf. Sections 3.2.2 and 5.2.3). For the free-field measurements in experiment 1 and 2, these reference values were measured individually for each subject and employed for the processing during the experiments. For the diotic presentation by headphone in experiment 3, all binaural signals were recorded using a dummy head and were processed off-line prior to the measurement. Hence, the reference values were also measured in advance with the same dummy head used for the recordings. The processing then employed these reference values, and the same accordingly processed signals were presented to all subjects.

6.3 Experiment 1: Subjective preferences

6.3.1 Procedure

The subjective preference of different signal processing strategies for noise reduction in hearing aids was measured with hearing impaired listeners. The measurement was performed in free-field conditions with a binaural hearing aid supply. The investigated signal processing conditions were “Linear” (linear amplification alone), “Linear Fixed” (linear amplification plus noise reduction algorithm “Fixed”) and “Linear Selective” (linear amplification plus noise reduction algorithm “Selective”). A complete paired comparison was performed by each subject, i.e., the subject compared each processing condition with each condition once. This resulted in three comparisons for each particular stimulus condition or acoustical situation, respectively. Within each comparison task, the stimulus was presented in an endless loop. The two different processing conditions appeared to the subject as being two different hearing aid programs with the first program switched on at the beginning of the task. The subject was able to arbitrarily switch between the programs whenever she or he liked to and without a particular time limit. For this, a handheld touchscreen response box was employed with the displayed choice of “1” or “2”. The switching was put into effect without a considerable delay or an interruption of the stimulus presentation. With the preferred processing switched on, the selection of the third choice “better” ended the comparison task. A statement of no preference was not allowed. The order of all pairs and presentations was randomized independently for each subject. The procedure was repeated for three different stimulus conditions.

6.3.2 Apparatus and stimuli

All stimuli were presented under free-field conditions in a sound-insulated booth[1] using the measurement setup depicted in Figure 6.1. This setup consisted of two functional parts. The first functional part performed the measurement control, the stimulus presentation and the assessment of the subject’s response. Each stimulus consisted of a speech signal S and an interfering or so-called noise signal N. These two signals were played by the PC using digital-to-analogue converters (DACs) and presented to the subject by the corresponding loudspeakers (depending on the stimulus condition). The presentation level was controlled by a PC controlled audiometer and final amplifiers.

The second functional part of the setup was the hearing aid and signal processing part that allowed for a real-time binaural hearing aid simulation. The subject was placed in the center of the booth, wearing her or his individual right and left ear moulds. The employed modular hearing instruments (Siemens Cosmea M) were plugged into the ear moulds and connected by wire to the signal processing system. The right and left microphones of the hearing instruments were connected to the digital-to-analogue converters (ADCs), while the DACs were connected to the respective receivers via analogue multiband amplifiers (t.c.

[1] IAC 403A, inside extensions 223.5 x 213.5 x 199.5 = width x length x height in cm.

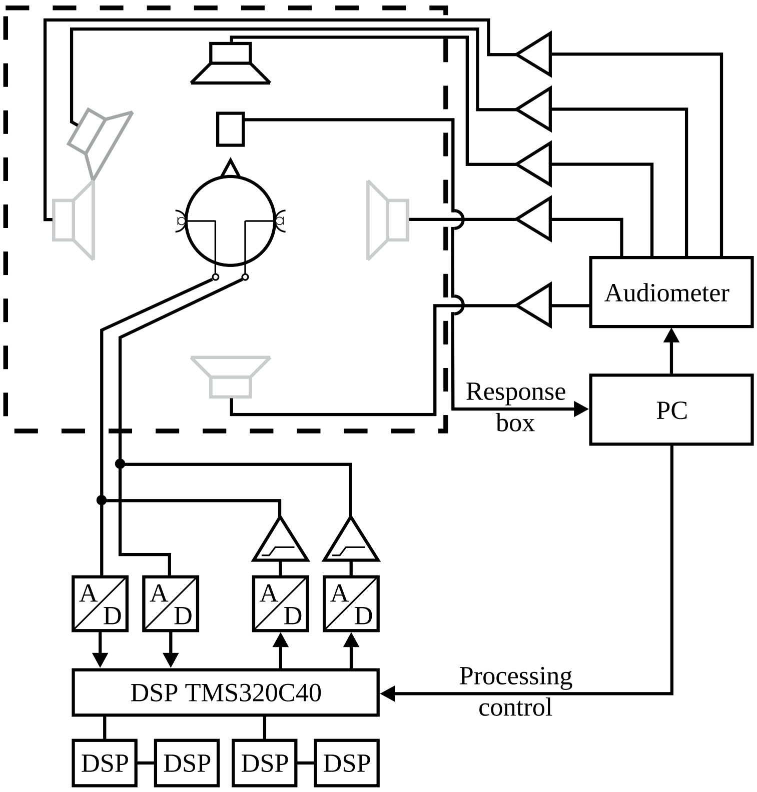


Figure 6.1: Free-field measurement setup and signal processing apparatus. In the top left corner, the sound-insulated booth is shown with the different loudspeakers and the subject placed in the center. To the right, the computer controlled measurement system is shown which performs the presentation of the stimuli and the assessment of the subject's response. In the bottom, the real-time signal processing system is sketched which processes the microphone signals of the hearing aid and plays the processed signals to the hearing aid receivers.

electronics TC1128X 28 Band Graphic Equalizer) used for an individual frequency shaping. The signal processing system included 5 TI TMS320C40 digital signal processors (DSPs). One of the DSPs performed the FFT and inverse FFT, while each of two independent DSP pairs calculated separately one of the two investigated processing strategies. The particular signal processing strategy applied to the hearing instrument signals was selected by the PC used for the measurement control, depending on the current selection made by the subject with the response box. The switching between the different processing strategies was put into effect without a considerable delay or an interruption of the stimulus presentation.

Figure 6.2 shows the spatial configurations of the three employed stimulus conditions $S_0N_{60/\text{noise}}$, $S_0N_{60/\text{speech}}$ and S_0N_{diff}. All stimulus conditions were a mixture of a target speech signal from the front at 0 degrees azimuth and an additional, interfering (noise) signal. The target speech was in all cases a concatenation of five randomly selected sentences from the Oldenburg sentence test, presented in an endless loop for each preference judgement. For stimulus condition $S_0N_{60/\text{noise}}$, the interfering signal was the original speech-

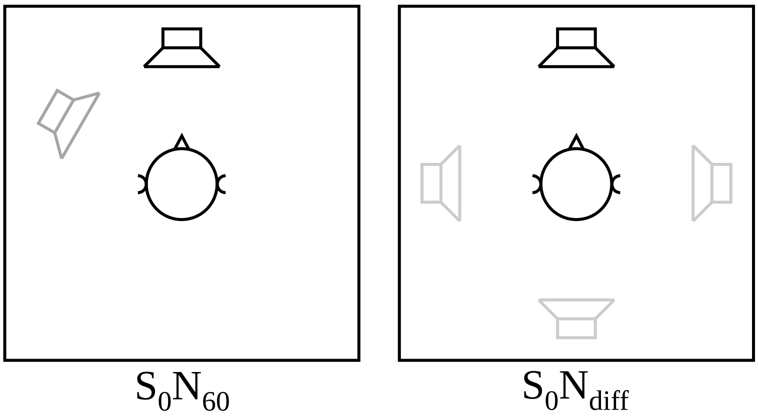


Figure 6.2: Spatial configurations of the employed stimulus conditions. The stimuli consist of target speech from the front (S_0) and interfering signals from 60 degrees to the left (N_{60}) or diffuse noise from different directions (N_{diff}), respectively.

shaped noise from the Oldenburg sentence test (cf. Wagener *et al.*, 1999c). For stimulus condition $S_0N_{60/\text{speech}}$, the interfering signal was a mixture of running speech from a male and a female speaker (reading text passages from two different books). For stimulus condition S_0N_{diff}, the interfering signal was uncorrelated cafeteria noise (recorded in a large cafeteria room during lunch time) from three different loudspeakers at 90, 180 and 270 degrees azimuth. The decorrelation was realized by using delayed versions of the same signal with a delay of about 2 seconds between each loudspeaker. Stimulus condition S_0N_{diff} indeed can be considered as being diffuse, since the degree of diffusiveness (cf. Chapter 4) calculated for the signal was in the range of real binaural cafeteria recordings (see below). Since there was no particular time limit, all interfering (noise) signals were also presented in an endless loop during each comparison task. The repeated interfering signals, however, were much longer than the repeated sentences (a couple of minutes, depending on the signal).

All stimuli were presented at a total level of about 65 dB SPL, measured at the center of the booth. If a subject complained about this level being either uncomfortable high or too low, the total level was accordingly adjusted (the total range of employed levels was 60 to 65 dB, but usually 65 dB was kept). The signal-to-noise ratios (SNRs) of the presented stimuli (or signal-to-jammer ratio, respectively, in the case of the non-noise interfering signals) were individually selected for each subject and stimulus condition. They were chosen as being the SNR corresponding to 50 % speech intelligibility (called speech reception threshold or SRT) in the processing condition “Linear” plus 6 dB. The respective SRT was individually measured (cf. experiment 2, Figure 6.4, light grey bars). The value (SRT + 6 dB) was chosen because at this particular SNR, the expected speech intelligibility for normal hearing listeners approaches approx. 100 % (cf. Wagener *et al.*, 1999b). In this case, the speech should be understandable, but variations in the SNR or in the speech quality, e.g. due to signal processing, should be able to influence the speech intelligibility itself or the listening effort required to fully understand the sentences. For higher SNR values, the intelligibility is so high that small variations in the processing probably would not affect the effort required to understand 100 % of the speech. For smaller SNR values, on the other hand, it might happen that the speech is not understandable at all. In this case, the subjects would not be able to appropriatly compare different processing techniques (and would in general not accept the processing). The chosen SNR thus seemed to be appropriate for a subjective comparison of different processing techniques.

In addition, the values of the degree of diffusiveness d_d for the different stimulus conditions had to be evaluated for the algorithm "Selective" and the respective limits of the decision unit had to be adjusted (cf. Section 5.3.3). In particular, the range of the calculated value of d_d throughout the employed stimulus conditions was assessed. From this, the lower limit $d_{\text{diff,min}}$ of d_d for mainly diffuse situations and the upper limit $d_{\text{corr,max}}$ of d_d for mainly correlated situations where set in a way that stimulus condition S_0N_{diff} was rated as being diffuse during the whole presentation while stimulus conditions $S_0N_{60/\text{noise}}$ and $S_0N_{60/\text{speech}}$ were rated as being correlated. This was possible because the value of d_d was significantly higher throughout the whole signal of condition S_0N_{diff}[2] than for the stimulus conditions $S_0N_{60/\text{noise}}$ and $S_0N_{60/\text{speech}}$.

6.3.3 Subjects

Eight hearing impaired subjects participated voluntarily in this study. They were aged between 16 and 65. All of them were hearing aid users with usually binaural hearing aid supply. All subjects had some experiences in psychoacoustic measurements and they all received an expenditure compensation on an hourly basis. The subjects were selected to have a symmetric, sensorineural hearing loss with an air-bone gap of not more than 10 dB for all frequencies tested. The audiograms of all subjects are given in Table 6.1. The upper four subjects in the table exhibited a high frequency loss with a steep slope of the hearing threshold between low and high frequencies. The lower four subjects exhibited a rather flat or moderately sloping hearing loss.

For each subject and each ear, the signal processing was fitted in advance to the measurements to yield an individual hearing loss compensation by linear amplification. The amplification was adjusted using both digital attenuation within the digital signal processing and digitally programmable, analogue multiband amplifiers between the DACs and the receivers (see Figure 6.1). The amount of amplification, i.e. the target gain was chosen in order to compensate for the deviation of the hearing impaired's most comfortable loudness level from that of normal hearing subjects. The most comfortable loudness level was determined by a categorical loudness scaling procedure (cf. Hohmann, 1993; Hohmann and Kollmeier, 1995; Launer *et al.*, 1996) for the common audiometric frequencies 500 Hz, 1 kHz, 2 kHz, 3 kHz, 4 kHz and 6 kHz. For lower and higher frequencies, the respective value of the nearest available frequency was used. The calculated target gain values are shown

[2] The range of values of d_d for stimulus condition S_0N_{diff} was also compared to values calculated for real binaural recordings using the same microphones in a cafeteria noise situation (a crowded cafeteria during lunch time with a reverberation time $T_{60} \approx 2$ s). The values were quite similar and the real recordings were also rated as a diffuse situation using the same limits as employed for the measurement stimuli. The particular calculated values of d_d and thus of the limits $d_{\text{diff,min}}$ of d_d and $d_{\text{corr,max}}$ are depending on various parameters like microphones, room acoustics and spatial configuration and are not listed here.

Subject	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	6 kHz	8 kHz
NF	10/10	15/10	35/35	60/65	70/70	70/80	85/90
HH	10/15	15/20	15/15	60/60	70/75	70/70	70/80
WH	10/15	15/15	20/20	40/50	70/70	80/85	65/80
JW	10/ 5	10/ 0	10/ 5	25/40	65/70	70/70	75/75
KM	40/35	50/40	45/40	55/50	75/60	70/75	70/65
BD	45/55	50/55	50/55	30/45	55/70	65/80	70/75
AA	40/35	50/50	65/70	60/60	60/60	85/65	80/75
DD*	20/20	30/30	45/40	55/45	60/60	60/65	60/55

Table 6.1: Pure tone audiograms of the subjects, measured with headphones (Telephonics TDH-39P). The values are the right/left hearing thresholds in dB HL at the specified audiometric frequencies.

in Table 6.2. The fitting procedure was performed with an insertion gain control by In-Situ measurements using a PortaREM 2000 real ear measurement system. For the highest frequencies 6 kHz and 8 kHz, the theoretical target gain usually was not fully provided by the instruments. The insertion gain also had to be adjusted in some cases due to feedback howling or on request of the subjects. All subjects reported the amplification being at least satisfying and comparable to their own hearing aids. Two subjects even prefered the experimental linear amplification to their own aids. This was not investigated further, but it may be explained by the high frequency resolution of 30 third-octave bands used for the frequency shaping which allowed for a very accurate and smooth insertion gain adjustment across frequency.

Subject	500 Hz	1 kHz	2 kHz	4 kHz	6 kHz
NF	0/ 0	3/ 3	19/19	21/24	32/35
HH	5/ 6	5/11	15/17	27/32	37/46
WH	9/ 0	14/ 3	18/15	34/30	39/39
JW	4/ 0	0/ 4	14/ 5	23/29	31/37
KM	11/ 4	11/ 7	12/ 7	15/ 7	18/19
BD	8/ 3	9/ 4	0/ 0	13/13	27/21
AA	13/26	18/26	20/13	19/18	31/28
DD*	13/ 0	19/ 9	20/12	24/25	21/25

Table 6.2: Target gains right/left in dB as calculated from the categorical loudness scaling results.

6.3.4 Results

The results of the preference judgements are shown in Figure 6.3 for all subjects and all stimulus conditions. Rank 1 indicates that during the three comparisons, the processing

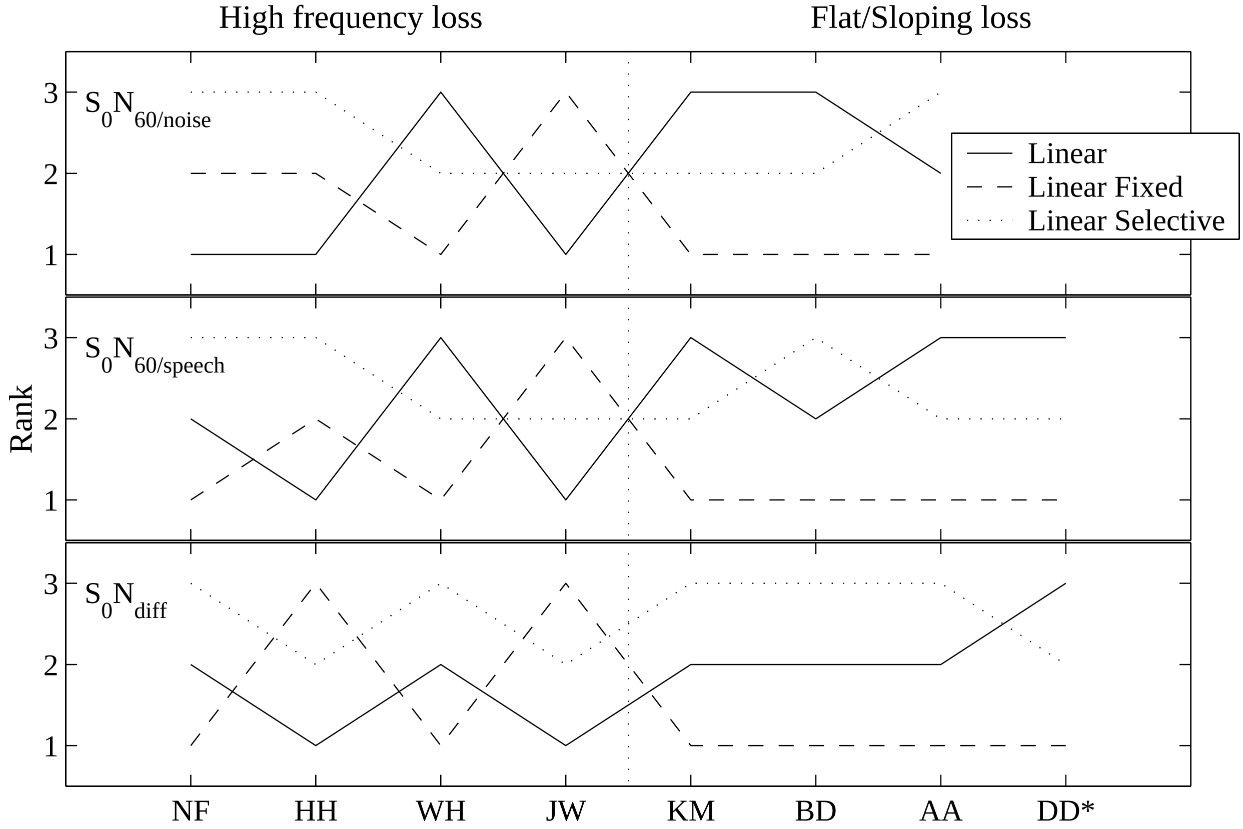


Figure 6.3: Ranks of preferences for the three different processing types "Linear" (solid line), "Linear Fixed" (dashed line) and "Linear Selective" (dotted line). Each panel shows the results for one of the stimulus conditions $S_0N_{60/\mathrm{noise}}$, $S_0N_{60/\mathrm{speech}}$ and S_0N_{diff}, as denoted in the uper left corner of the panel. The ordinate gives the rank obtained by paired comparison of the processing types. A processing type with a higher rank was preferred to all processing types with a lower rank. The abscissa denotes the 8 subjects. The 4 subjects on the left had a high frequency hearing loss, while the 4 subjects on the right had a rather flat or sloping hearing loss.

strategy was not prefered at all. Rank 2 indicates that the processing was prefered once, and rank 3 means two preferences which is the maximum, since one particular strategy was present only in two of the three comparisons. To further evaluate the data, a preference was considered as a "greater than" relation and the consistence of all three relations was defined as being the requirement for further consideration. Hence, the consistence of all preferences with each other was verified for all subjects and all stimulus conditions. Only in one case (subject DD*, condition $S_0N_{60/\mathrm{noise}}$), the relations were not consistent and these ranks were thus not included in the results.

For the stimulus conditions $S_0N_{60/\mathrm{noise}}$ and $S_0N_{60/\mathrm{speech}}$, no particular preference of any algorithm to the "Linear" processing or vice versa can be seen in the results. Since the ranks are quite similar in both acoustical situations for each subject, the preferences seem to be subject specific rather than processing specific here. Except for subject WH, however, all subjects prefered “Linear Selective” to “Linear Fixed”. This clearly indicates that although no consistent preference of a noise reduction to “Linear” or vice versa was found, “Linear Selective” yields a better perceived quality than “Linear Fixed”.

For stimulus condition S_0N_{diff}, 7 subjects prefered a noise reduction processing to the “Linear” processing. In this acoustical situation, the noise reduction thus seems to consistently improve the signal quality. Additionally, 6 subjects prefered “Linear Selective” to “Linear Fixed”, which again is a strong indication that “Linear Selective” yields a better quality than “Linear Fixed”.

6.4 Experiment 2: Dichotic speech intelligibility

6.4.1 Procedure

In this experiment, the speech intelligibility, i.e., the speech reception threshold (SRT) in noise was measured for the same signal processing techniques investigated in experiment 1. The SRT is defined as the signal-to-noise ratio (SNR) where 50 % of the target speech (in this case 50 % of the words of a sentence) is correctly repeated by the subject in a particular acoustical situation. The same measurement setup was used and the same subjects participated in this experiment as in experiment 1. Hence, the speech intelligibility was measured in free-field conditions with binaural, dichotic listening using a real-time hearing aid simulation.

The employed speech intelligibility test was the adaptive Oldenburg sentence test (cf. Wagener *et al.*, 1999a-c). This test allows for an in principle unlimited repetition of SRT measurement tasks, because the test is designed to appear as an open intelligibility test, i.e., a test with an unlimited number of different test sentences. This is maintained by using syntactically fixed, but semantically inpredictable sentences with a limited set of possible words (e.g. “Thomas gets seven red shoes” or “Thomas gets eighteen expensive cars”). The stimulus or acoustical noise conditions, respectively, were the same as used in experiment 1 ($S_0N_{60/\mathrm{noise}}$, $S_0N_{60/\mathrm{speech}}$ and S_0N_{diff}). The test procedure was as follows. A test sentence and the interfering signal were presented to the subject at a certain SNR. The subject had to repeat all words of the sentence she or he understood. The correctly repeated words were marked on a touchscreen response box (EPSON EHT-10S) by the measurement supervisor. An adaptive procedure converging on 50 % intelligibility controlled the SNR during the presentation of a total of 30 sentences. Finally, the SRT was estimated by fitting a logistic function to all responses of the subject using a maximum likelihood procedure. The whole task was repeated for all investigated processing techniques and for all stimulus conditions. For the adjustment of the SNR during the measurement, the target speech level was varied while the noise level was fixed. The absolute presentation level was the same as in experiment 1 for the aided processing conditions (with hearing instrument). For the unaided condition, the presentation level was individually adjusted for each stimulus condition in order to maintain a comfortable loudness level of the interfering signals. On the average (across all subjects and stimulus conditions), the resulting presentation level was +1.3 dB higher for the unaided condition than for the aided conditions.

6.4.2 Apparatus, stimuli and subjects

The subjects, the apparatus and the stimulus conditions (i.e. their spatial configuration and the interfering signals) were the same as for experiment 1 described above in Section 6.3. The employed processing parameters of the algorithms were also identical. The target speech within one measurement task, however, was the respective sentence according to the adaptive Oldenburg sentence test procedure. The SNR of the stimulus was adjusted by the adaptive procedure within each task.

6.4.3 Results

The measured SRTs of all hearing impaired subjects and the respective mean values of normal hearing subjects are shown in Figure 6.4. It can clearly be seen that the subjects with a high frequency hearing loss on the left side of the figure generally perform better than the subjects with a rather flat hearing loss on the right side. However, both groups of subjects perform considerably worse than normal hearing subjects, i.e., they all exhibit a higher SRT than normal hearing subjects for all stimulus and processing conditions. The individual results indicate that some subjects do exhibit an improvement of SRT due to the noise reduction processing (compared to the processing "Linear") and some do not. It should be noted that two subjects (BD and DD*) exhibit a general negative effect of the hearing aid supply, i.e., the SRTs are higher for all conditions with a hearing aid than for the unaided condition. This can be explained by the fact that the unaided condition and the aided conditions were measured in different sessions, i.e. on different days, and the intraindividual variance can thus be rather high. Additionally, the subjects had to use a hearing aid during the measurements they were not accustomed to, which might deteriorate speech intelligibility for some subjects. Another factor is the (presumably limited) influence of the presentation level: Although both for the unaided condition and for the aided conditions the same categorical loudness impression was targeted (comfortable loudness of the interfering signal), the average presentation level (as well as the average spectrum) at the subject's ear was different for the unaided condition than for the aided conditions. The important differences, however, are those between the conditions with the different hearing aid processing schemes, i.e. "Linear", "Linear Fixed" and "Linear Selective". These conditions were measured during a single session.

The median values of the SRTs are shown in Figure 6.5. In general, the results for "Linear" and "Linear Selective" look very similar, while "Linear Fixed" obviously performs worse than the two other conditions. Similarly, a T-test revealed no significant difference between the SRTs of processing "Linear" and "Linear Selective" ($\alpha > 0.1$), but a significant difference between SRTs of processing "Linear" and "Linear Fixed" and of processing "Linear Fixed" and "Linear Selective" ($\alpha < 0.001$). This again demonstrates the better performance of algorithm "Selective" in comparison to algorithm "Fixed", but is somewhat disappointing with respect to the total effect of algorithm "Selective". Finally, Wilcoxon tests were performed separately for each stimulus condition. For stimulus condition $S_0N_{60/\mathrm{noise}}$, there was again no significant difference found between processing "Linear" and "Linear Selective" ($\alpha > 0.1$), but a significant difference between processing "Linear" and "Linear Fixed" and between processing "Linear Fixed" and "Linear Selec-

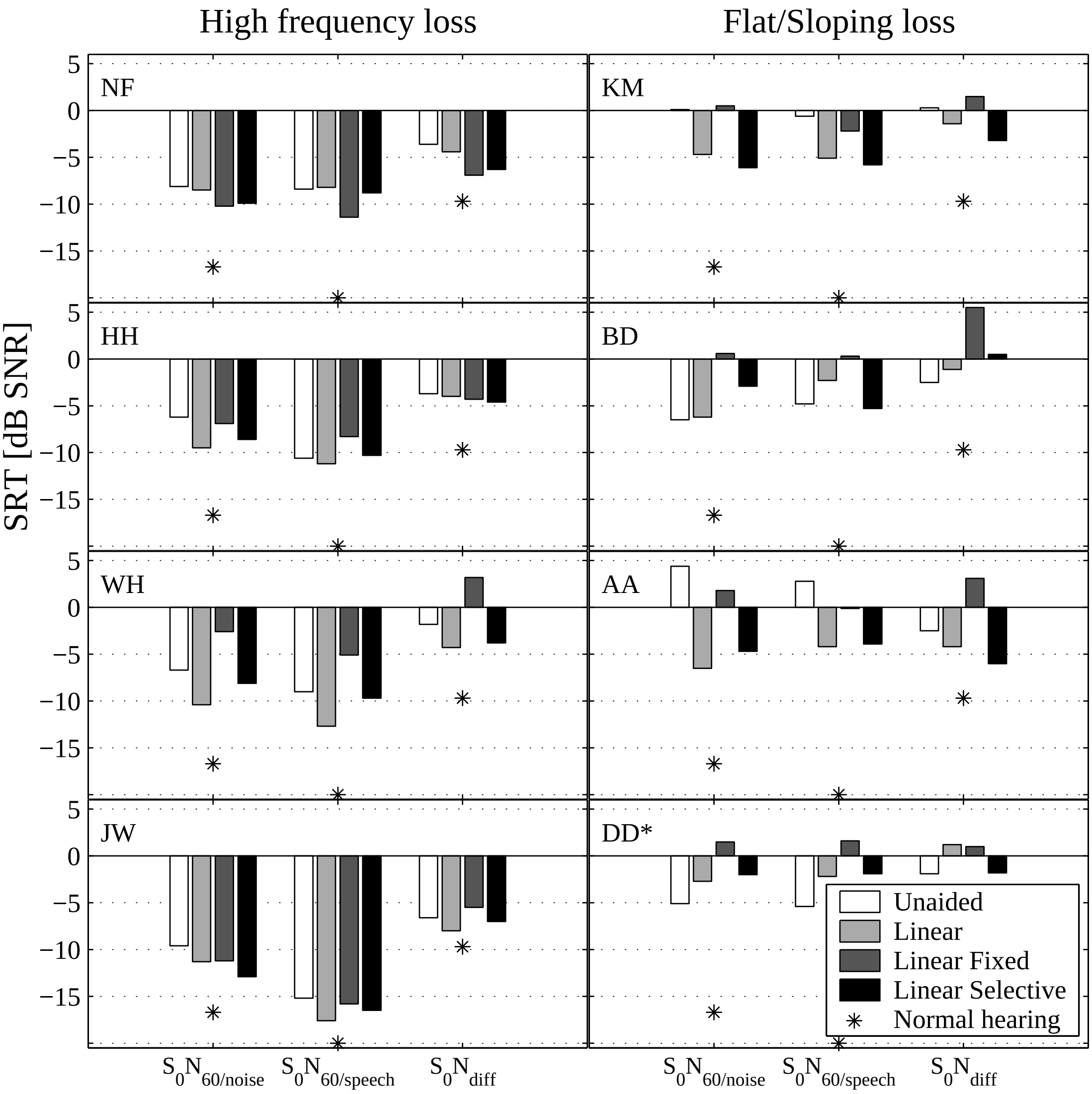


Figure 6.4: Measured SRT of 8 hearing impaired subjects for different conditions of binaural (dichotic) hearing instrument supply. Each panel shows the results of one subject. The 4 subjects on the left had a high frequency hearing loss, while the 4 subjects on the right had a rather flat or sloping hearing loss. The abscissa denotes the three stimulus conditions. For each stimulus condition, the respective bars denote the SRT in the following processing conditions (from left to right): Unaided (without a hearing instrument), aided with linear amplification (frequency shaping), aided with linear amplification and algorithm “Fixed” and aided with linear amplification and algorithm “Selective”. The asterisks represent the respective mean SRT of normal hearing subjects (without hearing instrument).

tive” ($\alpha < 0.05$). This stimulus condition includes the most difficult interfering signal of the three employed stimulus conditions, because the speech shaped noise exhibits the maximum spectral and temporal masking of the target speech. For stimulus conditions $S_0N_{60/\text{speech}}$ and S_0N_{diff}, again there was no significant difference found between the processing “Linear”

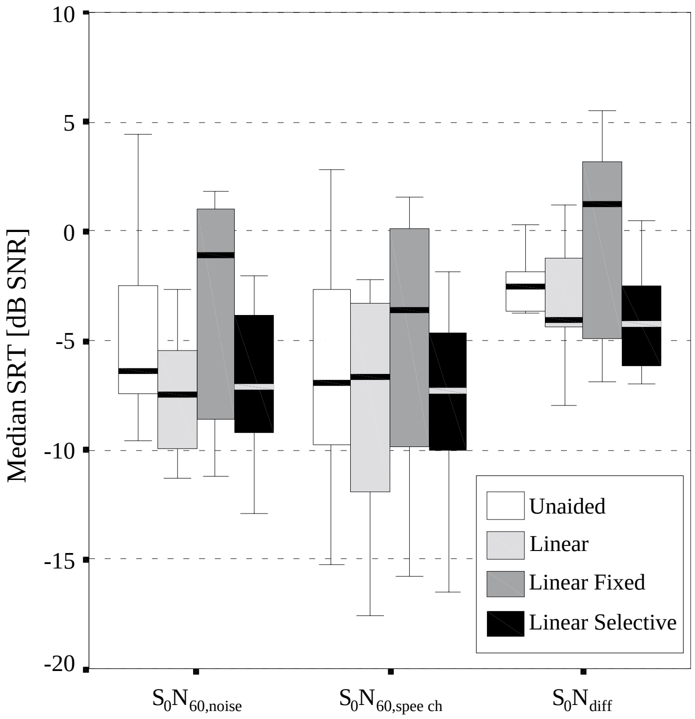


Figure 6.5: Mean values of the SRTs shown in Figure 6.4. The thick horizontal lines denote the median values, the boxes the range from the first to the third quartile and the outer bars the total range. Each median was calculated for 8 subjects in the respective condition, see Figure 6.4 for details.

and "Linear Selective", but also no significant difference between processing "Linear" and "Linear Fixed" ($\alpha > 0.1$ in all cases except for processing "Linear" and "Linear Fixed" and stimulus condition $S_0N_{60/\mathrm{speech}}$ with $\alpha > 0.05$). As above, a significant difference was found for both stimulus conditions between the processing "Linear Fixed" and "Linear Selective" ($\alpha < 0.05$). This means that the advantage of algorithm "Selective" over algorithm "Fixed" is more significant than the advantage of no noise reduction processing over algorithm "Fixed". Due to the small differences and the small number of subjects, however, this is only a hint that algorithm "Selective" might improve the SRT in certain noise conditions. Hence, the experiment was not able to prove this assumption which is based so far only on informal listening and the technical evaluation of the noise reduction processing strategies.

6.5 Experiment 3: Diotic speech intelligibility

In the experiments 1 and 2, the subjects were able to listen binaurally to the unprocessed and the processed dichotic sounds and hence were able to use their own central binaural processing capabilities to suppress undesired, spatially separated noise. However, the performance of the binaural noise suppression abilities varies considerably across hearing impaired listeners (cf. Häusler *et al.*, 1983; Kinkel *et al.*, 1991; Holube, 1993; Gabriel *et al.*, 1992). This interindividual variability may be the reason that experiment 2 was not able to prove a significant improvement of speech intelligibility in noise for the processing "Linear Selective" which aims to replace or at least support the (possibly impaired) binaural noise suppression capabilities in hearing impaired listeners. In order to eliminate the factor "remaining individual binaural processing capabilities" from the individual speech test performance, the benefit of the processing "Linear Selective" was tested in this experiment using a diotic condition, i.e. identical signals were presented to both ears of the subject. Hence, the speech intelligibility was measured for the signal processing conditions "Linear" (linear amplification alone) and "Linear Selective" (linear amplification plus noise reduction algorithm "Selective"). As in experiment 2, the Oldenburg sentence test was employed to measure the SRT for a speech shaped noise as interfering signal (cf. Section 6.4.1). However, a different set of subjects and a different hearing aid fitting procedure was employed as in experiment 2.

6.5.1 Apparatus and signals

The measurement setup of this experiment differs from the setup employed for experiment 2. One major difference was that all signals were presented diotically by headphones. Additionally, the signal processing was performed in advance to the measurement and not in real-time during the measurement. The signals were processed for various different SNRs and stored on the hard disk of the PC. The measurement setup is depicted in the left panel of Figure 6.6. The already processed signals were played back from the PC using digital-to-analogue converters (DACs) and presented to the subject by headphones via analogue multiband amplifiers (used for the individual frequency shaping) and a final amplifier (used for the overall level adjustment).

Only one stimulus condition with a spatial S_0N_{-60} configuration was employed in this experiment. The interfering noise was again the speech-shaped noise from the Oldenburg sentence test, the stimulus condition will thus be referred to as $S_0N_{-60/\mathrm{noise}}$. This condition is similar to stimulus condition $S_0N_{60/\mathrm{noise}}$ employed in experiments 1 and 2 (cf. Sections 6.3.2 and 6.4.2), but the sound incidence direction of the interfering speech shaped noise was 60 degrees to the right instead to the left. For the experiment, the target speech and the interfering noise were recorded dichotically in advance with a Head Acoustics dummy head in a seminar room with a reverberation time of $T_{60} = 0.6s$. Speech and noise were summed up at various SNRs, processed with the different processing techniques and stored on hard disk to be available during the speech intelligibility measurement. For the processing with algorithm "Selective", the binaural reference values measured for the employed dummy head were used. During the measurement, only the right channel of the processed signal (i.e. the channel with the less favourable SNR) was presented diotically to both ears. All specified SNRs are the respective value of the right channel before the processing, calculated from the RMS values of the separately recorded speech and noise signals. Although the processed signal was the same for both ears, the hearing loss compensation was performed individually for each ear.

Like in the experiments 1 and 2, the values of the degree of diffusiveness d_d were calculated for the employed signals (cf. Section 6.3.2). The limits $d_{\mathrm{diff,min}}$ and $d_{\mathrm{corr,max}}$ were

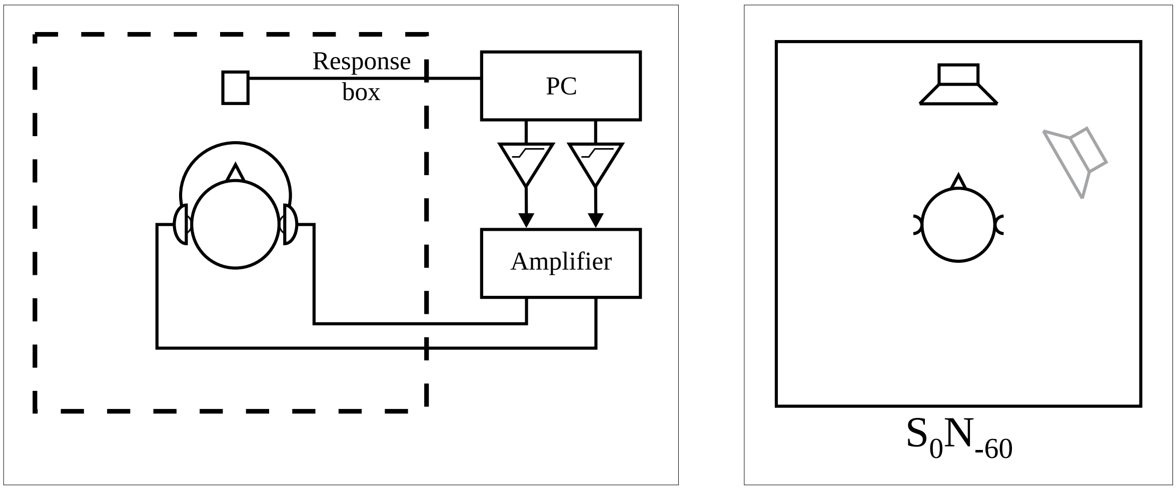


Figure 6.6: Setup and signal configuration used for the diotic speech intelligibilty measurement. The left panel shows the setup with the subject placed in the center of a sound-insulated booth. The PC controls the measurement and the frequency shaping amplifiers and also performs the presentation of the signals and the assessment of the subject's response. The right panel depicts the spatial configuration S_0N_{-60} used for the recording of the signals.

again set in a way that the signals were rated as being mainly correlated and thus all signal processing strategies were switched on. All other parameters of the algorithm were adjusted to the values used for the experiments 1 and 2 (cf. Section 6.2 and Chapter 5).

6.5.2 Subjects

Six sensorineural hearing impaired subjects participated voluntarily in this study. They were aged between 23 and 78 and they were all hearing aid users. All subjects had some experiences in psychoacoustic measurements and they all received an expenditure compensation on an hourly basis. The subjects were selected to have a rather symmetric hearing loss. The audiograms of all subjects are given in Table 6.3. All subject had rather moderate hearing losses sloping in median from 30 dB at 250 Hz to 77 dB at 8 kHz.

During the assessment, the digitally programmable, analogue multiband amplifiers were used to provide an individual hearing loss compensation by linear amplification separately for each ear of the subject. In this experiment, the electrical linear amplification was adjusted to exactly satisfy the "one-half gain rule" (cf., e.g., Dempsey, 1994). No further coupler measurements were performed. The total presentation level was individually adjusted in a preliminary test run in a way that the overall loudness impression was at the top border of the comfortable loudness range.

Subject	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	6 kHz	8 kHz
BD	45/55	50/55	50/55	30/45	55/70	65/80	70/75
GM	35/45	40/45	50/45	65/60	90/90	90/90	90/90
HM	25/20	45/35	55/50	55/50	60/60	65/70	70/80
KF	20/30	30/35	55/50	60/55	60/70	65/50	80/55
KR	25/30	30/20	35/30	40/30	60/75	70/75	65/75
WH	15/50	20/45	30/45	50/45	55/60	50/65	80/80

Table 6.3: Pure tone audiograms of the subjects. measured with headphones (Telephonics TDH-39P). The values are the right/left hearing thresholds in dB HL at the specified audiometric frequencies.

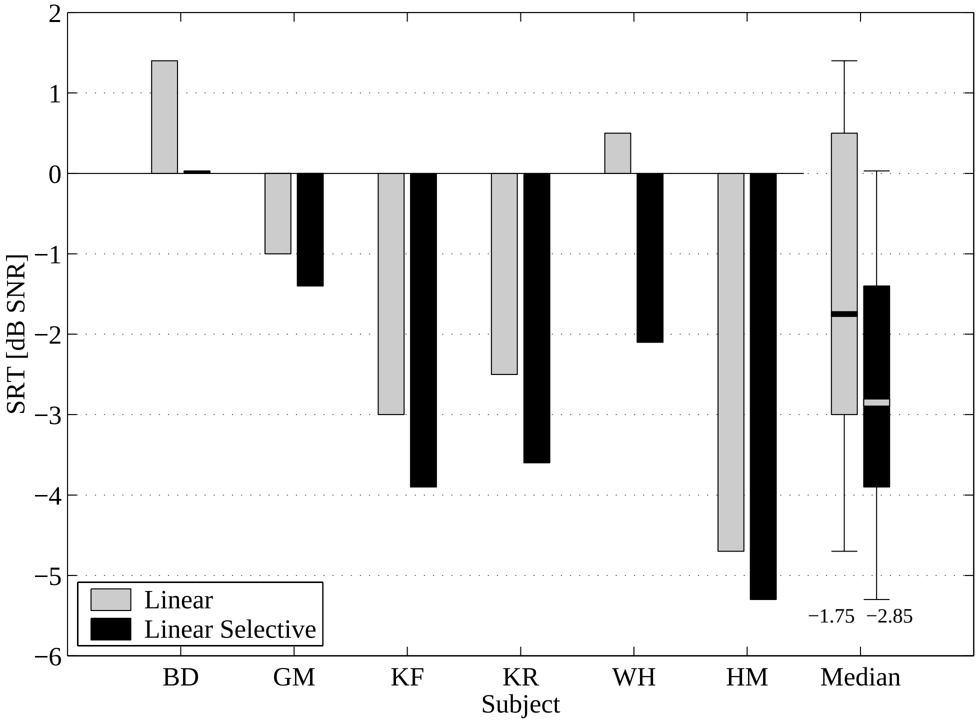


Figure 6.7: Measured SRTs of 6 hearing impaired subjects for diotic presentation. The abscissa denotes the subjects and the mean values, respectively. For the mean values (rightmost bars), the thick horizontal lines denote the median values which are additionally given below the bars, the boxes denote the range from the first to the third quartile and the outer bars the total range. The light grey bars represent the processing with linear amplification alone. The black bars give the results obtained with linear amplification and noise reduction algorithm “Selective”.

6.5.3 Results

The measured SRTs of all subjects and the respective mean values are shown in Figure 6.7. All subjects exhibit an improvement in SRT by the processing of algorithm “Selective”. A

Wilcoxon test revealed that the differences between "Linear" and "Linear Selective" are significant ($\alpha < 0.05$). The median improvement is 1.1 dB, the maximum improvement is 2.6 dB (subject WH). Hence, a significant improvement of the SRT was found in this experiment.

6.6 Discussion

In this study, the efficacy of two binaural noise reduction algorithms for hearing aids with respect to subjective preference and speech intelligibility was investigated under realistic free-field conditions with real hearing instruments. The acoustical conditions included standard situations with a target speaker and a single interfering sound source (either a speech-shaped noise or a mixture of 2 talkers). Additionally, a simulated cafeteria situation with multiple ambient noise sources was tested which was shown to exhibit a diffusiveness of the sound field comparable to a real cafeteria situation. Although this situation is rather difficult to test in laboratory free-field conditions and thus often omitted, it is a very important every-day situation for hearing aid wearers and therefore was also investigated in this study.

The subjective preference was tested in a straight forward way. The hearing impaired subjects simply switched between two different hearing aid programs without any delay or signal disruption and then directly chose the prefered program. The employed paradigm of a complete paired comparison allows not only for an appropriate evaluation of the preference, but also for a consistency verification. The results of the assessment clearly indicate that the subjects prefered the new strategy-selective algorithm to the older algorithm after Peissig (1993) in all situations. However, in the situations with a single interfering sound source, there was no general preference of the strategy-selective algorithm to the unprocessed version or vice versa across the subjects. Although this indicates that the quality of the processed version is comparable to the original signal, the performance of the processing still has to be further improved in these particular situations in order to yield a real benefit for the hearing aid user. In the cafeteria noise situation, the subjects clearly preferred the strategy-selective algorithm in comparison to the unprocessed version. This demonstrates the high signal quality of the processing in this situation and also a real benefit for the hearing impaired, since the diffuse ambient noise was at least subjectively reduced.

The influence of the algorithms on speech intelligibility was first tested with a standard sentence test in the same realistic acoustical free-field conditions. Although the strategy-selective algorithm yields significantly better results than the algorithm after Peissig (1993), no significant improvement of speech intelligibility compared to the unprocessed signal was found in any situation. There is only a slight hint that the strategy-selective algorithm might slightly improve speech intelligibility for the interfering speech and in the cafeteria situation (stimulus conditions $S_0N_{60/\mathrm{speech}}$ and S_0N_{diff}). The results obtained with this algorithm differ more significantly from the (worse) results obtained with the algorithm after Peissig (1993) than the results of the unprocessed version do. With regard to the limited number of subjects, however, this can not be considered as a proof of an effective speech intelligibility improvement. In an additional experiment, the strategy-selective algorithm was tested with the same sentence test and a similar binaural processing of the signals, but a diotic presentation of the signals by headphone in order to avoid effects of central binaural processing in the auditory system of the subjects. In this case, the processing was shown to significantly improve speech intelligibility for the most difficult noise signal, i.e. speech shaped noise.

The strategy-selective processing was thus proved to reduce noise and to effectively improve the signal-to-noise ratio, but the effect was not sufficient to yield an improvement of speech intelligibility in the case of normal dichotic (binaural) listening. Since the processing improves the signal-to-noise ratio, there are two possible explanations for the absence of speech intelligibility improvement: The first is that even in hearing impaired listeners, the binaural performance of the human auditory system itself is superior to the noise reduction processing and obtains no additional information by the processing that would help to further improve intelligibility. The remaining binaural processing ability indeed varies considerably across hearing impaired subjects in a way that is not predictable from the audiogram (cf. Häusler *et al.*, 1983; Kinkel *et al.*, 1991; Holube, 1993; Gabriel *et al.*, 1992). Hence, even if the remaining binaural abilities of the subjects who participated in this study might on the average have been superior to the effect provided by the algorithm, it can be expected that other subjects with less binaural abilities will profit more from the processing. It is possible to classify subjects with respect to their residual binaural processing abilities by measuring the binaural intelligibility difference or BILD. The BILD is the difference in SRT for monaural and binaural signal presentation in a setup where target speech and interfering noise are spatially separated (cf. Hövel, 1984; Holube, 1993; Kollmeier, 1997; Kühnel and Kollmeier, 1997). The influence of the BILD on speech intelligibility in combination with binaural processing techniques should be investigated in future evaluations. Although the remaining binaural abilities are not predictable from the audiogram, the individual results of experiment 2 indicate that the subjects with a rather flat hearing loss derive more benefit from the noise reduction processing in the conditions $S_0N_{60/\text{speech}}$ and S_0N_{diff} than the subjects with a high frequency loss (for condition S_0N_{diff}, the strategy-selective processing yields an improvement of speech intelligibility for 3 of the 4 respective subjects). This effect may be due to binaural abilities, but also due to the more favourable absolute SNR values. The SRTs are in the range of about 0 down to -5 dB for the subjects with a rather flat hearing loss, which is a range where the processing should be able to effectively reduce noise. For the subjects with a high frequency loss, the SRTs are in general lower with values down to less than -15 dB. This already approaches the range of normal hearing. At this very low SNR, the processing can at best be expected to subjectively reduce noise and to improve the perceived signal quality, but not to effectively improve the SNR.

Another possible explanation for the lack of speech intelligibility improvement would be that the processing deteriorates the signal in a way that compensates for the effect of SNR improvement. Since an improvement of speech intelligibility was found for diotic presentation, this deterioration must be assumed to mainly effect binaural signal parameters. In this case, it can be expected that measurement configurations can be found where either signal quality or speech intelligibility is significantly deteriorated due to the processing. Until now, this was not observed.

It should be noted that the usage of hearing instruments instead of headphones in general makes the investigation of small effects in signal quality and speech intelligibility more difficult (but also more realistic). The available frequency range of the receivers is smaller than for headphones, for instance, and hearing instruments also exhibit internal noise which can be audible. These factors can influence, e.g., speech intelligibility (cf. Albani *et al.*, 1998). However, it is very promising that the strategy-selective processing was found to not deteriorate neither quality nor speech intelligibility in any of the investigated conditions even at very low signal-to-noise ratios.

6.7 Conclusions

In this study, the new strategy-selective algorithm introduced in Chapter 5 was shown to exhibit a very high subjectively perceived quality of the processed signal. In the difficult cafeteria noise situation, the hearing impaired subjects clearly preferred this algorithm to the unprocessed condition. The strategy-selective algorithm also performed clearly better than the previous algorithm after Peissig (1993) in all conditions with respect to both signal quality and speech intelligibility. Moreover, the processing did not deteriorate speech intelligibility in comparison to linear amplification alone even in acoustically "difficult" situations. Hence, the method of selecting particular processing strategies depending on the acoustical situation introduced with this algorithm is very promising and can thus be recommended for the use in future developments of hearing aid algorithms.

Although the algorithm exhibits a significant signal quality improvement in the cafeteria situation, the noise reduction performance of the algorithm still requires further development and improvement. Especially the performance with a single interfering sound source is still disappointing, even though this condition can be considered as an "easy" noise situation. An effective speech intelligibility improvement was found only in the case of a diotic signal presentation. Although this proves that the algorithm can reduce noise and improve the SNR, the effect could not be shown to yield a real benefit in terms of SRT under realistic free-field conditions for the hearing impaired subjects. It may, however, yield more benefit if the SNR is more favourable than it was in the SRT measurements especially of the subjects with a high frequency hearing loss. It may also be more beneficial for listeners with a more severe hearing loss as well as for cochlear implant patients.

Chapter 7

Summary and conclusions

In this study, processing strategies for the reduction of undesired noise in binaural input signals were described and investigated. After a review in chapter 2 considering strategies and algorithms that have been published in the past, a modified version of the algorithm introduced by Peissig (1993) (see also Kollmeier *et al.*, 1993; Wittkop *et al.*, 1997) using a fixed processing was described in chapter 3. A complete paired comparison paradigm was described and shown to be a suitable method to investigate the algorithm with respect to the subjectively perceived sound quality. Parameters of the processing were systematically varied, and a parameter combination was found that is appropriate for all of the investigated acoustical conditions.

In chapter 4, a measure was proposed which allows for the long-term rating of the actual acoustical condition with respect to its diffusiveness or “complexity”, respectively. This measure was shown to be a monotonous function of the number of spatially separated sound sources (if the amount of reverberation is constant) and a monotonous function of the amount of reverberation (if the number of spatially separated sound sources is constant) and also of the distance between the recording microphones and a single sound source. The long time scale of several seconds of this measure aims at the automatic classification of the situation and the general selection of appropriate noise reduction strategies. While these strategies themselves act on a very short time scale (in order to follow the fluctuations of the actual SNR), the switching between different strategies should be performed rather slow in order to avoid audible and disturbing artefacts.

In chapter 5, a strategy-selective algorithm was eventually described which utilizes the automatic rating of the situation to switch on or off particular noise reduction strategies. Two different strategies for the reduction of undesired interfering signals in binaural input signals were theoretically derived. A third strategy that is based on an empirical approach of suppressing lateral sound sources was also integrated in the algorithm. The properties of the strategies were first investigated using artificial signals. Then, the processing was optimized with respect to the sound quality. For this, the method described in chapter 3 was used which includes a systematic variation of processing parameters and the assessment of subjective preference judgements in a complete paired comparison paradigm. Finally, the algorithm was evaluated with respect to sound quality and speech intelligibility with hearing impaired subjects in chapter 6.

The performance of the strategy-selective algorithm was found to be superior to or equal to the algorithm with the fixed processing in all investigated acoustical conditions and with respect to both signal quality and speech intelligibility for hearing impaired subjects. Moreover, the strategy-selective algorithm was also found to be superior to or equal to the condition without any noise reduction processing in all of the investigated acoustical conditions. In particular, the hearing impaired subjects clearly prefered the strategy-selective algorithm in the diffuse cafeteria noise situation. In this situation, only the correlation gain factors are switched on which is a processing strategy based on the assumption of diffuse background noise. The other processing strategies which assume one or a few interfering sound sources are not active in this situation. This avoids disturbing processing artefacts that occur in the algorithm with a fixed processing. Since the annoyance of disturbing interfering signals and the related listening effort is an important factor of the benefit a hearing impaired patient can derive from a noise suppression processing (cf. Marzinzik, 2000), the clear preference of the strategy-selective algorithm to the no-processing condition is a very promising result. In the other acoustical situations, however, it cannot be observed that the strategy-selective processing or no noise reduction processing is clearly preferred.

Considering the speech intelligibility, the results are not as clear as for the subjective preference. In the realistic free-field conditions with dichotic listening (which includes some reverberation), no significant improvement of the SRT in noise was found for the strategy-selective algorithm. At least it was found that this algorithm causes no deterioration at all even at very low SNRs (cf. Fig. 6.4). The median SRTs were even slightly (but not significantly) improved for the interfering speech and the cafeteria noise. In these two situations, the individual results are quite promising, especially for the subjects with a flat hearing loss. For these subjects, the resulting SRTs are in general higher and thus more favourable for the signal processing. For a larger number of subjects with an appropriate individual hearing loss, a (probably small, but) significant improvement of the SRT may be found in future investigations.

Although the results of this study with realistic acoustical situations exhibit less noise reduction performance than earlier studies using anechoic conditions (e.g. Peissig, 1993), there still is sufficient evidence that the strategy-selective noise reduction processing is able to effectively improve the SNR under certain conditions: The measurements with a diotic presentation of the binaurally processed signals by headphones show a significant improvement of the SRT in the investigated situation with a single interfering noise source (cf. Fig. 6.7). This performance clearly surpasses state-of-the-art monaural noise reduction techniques (see Marzinzik, 2000, for an overview). Another evidence for the ability of the algorithm to improve the SNR was found by Kleinschmidt *et al.* (2000), who used the strategy-selective noise reduction processing as a preprocessor for robust automatic speech recognition. The noise reduction processing resulted in an improvement of the recognition rate, i.e., the percentage of correctly recognized words in conditions with a single, lateral noise source. The shift of the recognition rate as a function of SNR exhibited an effective improvement of the SNR of up to 6 dB in a reverberant environment and up to 12 dB in an anechoic environment due to the processing in this particular application. It should be noted that a monaural noise reduction technique investigated in the same study also yielded a comparable SNR improvement for the employed stationary interfering signals. In contrast to the monaural technique, however, the binaural processing can be expected to also operate on temporal non-stationary interfering signals.

Taken together, the strategy-selective noise reduction algorithm is superior to a comparable algorithm with fixed processing strategies with respect to both sound quality and speech intelligibility. An effective improvement of sound quality for hearing impaired subjects was found in terms of subjective preference. This is a real benefit for hearing impaired patients, since the listening effort is an important factor for hearing speech in noise. A significant improvement in speech intelligibility in terms of SRT was found only for diotic or monaural applications, respectively. However, the results indicate that an effective, i.e. significant benefit in terms of SRT might be shown in the future also in normal, dichotic listening situations for appropriate hearing impaired patients, i.e. for patients with particular types of hearing losses. The development of the "ideal" hearing aid, i.e., a hearing aid that completely restores all abilities of hearing to that of normal hearing, is a challenge and the ultimate goal of audiological research, from which we are still far away. However, very probably only a binaural hearing aid will be able to really approach this goal. The development and investigation of binaural processing techniques in general thus should be an issue of future research.

Appendix A

SNR equations

In this Section, particular equations concerning the SNR introduced in section 5.2.2 are derived. To achieve more clearness, the frequency dependency (f) of the spectra $S(f)$, $N(f)$, $H_{s,l}(f)$, $H_{s,r}(f)$, $H_{n,l}(f)$ and $H_{n,r}(f)$ is not denoted in the following equations. Using the definitions (5.1) yields

$$\begin{aligned}
-\frac{Q_Y - Q_N}{Q_Y - Q_S} \cdot Q_S \cdot \frac{H_{n,r}}{H_{s,l}} &= -\frac{\frac{S\cdot H_{s,l}+N\cdot H_{n,l}}{S\cdot H_{s,r}+N\cdot H_{n,r}} - \frac{H_{n,l}}{H_{n,r}}}{\frac{S\cdot H_{s,l}+N\cdot H_{n,l}}{S\cdot H_{s,r}+N\cdot H_{n,r}} - \frac{H_{s,l}}{H_{s,r}}} \cdot \frac{H_{s,l}}{H_{s,r}} \cdot \frac{H_{n,r}}{H_{s,l}} \\
&= -\frac{\frac{S\cdot H_{s,l}\cdot H_{n,r}+N\cdot H_{n,l}\cdot H_{n,r}}{S\cdot H_{s,r}\cdot H_{n,r}+N\cdot H_{n,r}\cdot H_{n,r}} - \frac{H_{n,l}\cdot S\cdot H_{s,r}+H_{n,l}\cdot N\cdot H_{n,r}}{S\cdot H_{s,r}\cdot H_{n,r}+N\cdot H_{n,r}\cdot H_{n,r}}}{\frac{S\cdot H_{s,l}\cdot H_{s,r}+N\cdot H_{n,l}\cdot H_{s,r}}{S\cdot H_{s,r}\cdot H_{s,r}+N\cdot H_{n,r}\cdot H_{s,r}} - \frac{S\cdot H_{s,r}\cdot H_{s,l}+N\cdot H_{n,r}\cdot H_{s,l}}{S\cdot H_{s,r}\cdot H_{s,r}+N\cdot H_{n,r}\cdot H_{s,r}}} \cdot \frac{H_{n,r}}{H_{s,r}} \\
&= -\frac{\frac{S\cdot[H_{s,l}\cdot H_{n,r}-H_{n,l}\cdot H_{s,r}]+N\cdot[H_{n,l}\cdot H_{n,r}-H_{n,l}\cdot H_{n,r}]}{S\cdot H_{s,r}\cdot H_{n,r}+N\cdot H_{n,r}\cdot H_{n,r}}}{\frac{S\cdot[H_{s,l}\cdot H_{s,r}-H_{s,l}\cdot H_{s,r}]+N\cdot[H_{n,l}\cdot H_{s,r}-H_{s,l}\cdot H_{n,r}]}{S\cdot H_{s,r}\cdot H_{s,r}+N\cdot H_{n,r}\cdot H_{s,r}}} \cdot \frac{H_{n,r}}{H_{s,r}} \\
&= -\frac{\frac{S\cdot[H_{s,l}\cdot H_{n,r}-H_{n,l}\cdot H_{s,r}]}{S\cdot H_{s,r}+N\cdot H_{n,r}} \cdot \frac{1}{H_{n,r}}}{\frac{N\cdot[H_{n,l}\cdot H_{s,r}-H_{s,l}\cdot H_{n,r}]}{S\cdot H_{s,r}+N\cdot H_{n,r}} \cdot \frac{1}{H_{s,r}}} \cdot \frac{H_{n,r}}{H_{s,r}} \\
&= -\frac{S}{N} \cdot \frac{H_{s,l} \cdot H_{n,r} - H_{n,l} \cdot H_{s,r}}{-H_{s,l} \cdot H_{n,r} + H_{n,l} \cdot H_{s,r}} \\
&= \frac{S}{N}. \qquad \text{(A.1)}
\end{aligned}$$

Equations for the signal-to-noise ratios of the left and right spectrum now can be derived as

$$\frac{S_l}{N_l} \overset{(5.10)}{=} \frac{S}{N} \cdot \frac{H_{s,l}}{H_{n,l}} \overset{(A.1)}{=} -\frac{Q_Y - Q_N}{Q_Y - Q_S} \cdot Q_S \cdot \frac{H_{n,r}}{H_{s,l}} \cdot \frac{H_{s,l}}{H_{n,l}} = -\frac{Q_Y - Q_N}{Q_Y - Q_S} \cdot \frac{Q_S}{Q_N} \qquad \text{(A.2)}$$

$$\frac{S_r}{N_r} \overset{(5.10)}{=} \frac{S}{N} \cdot \frac{H_{s,r}}{H_{n,r}} \overset{(A.1)}{=} -\frac{Q_Y - Q_N}{Q_Y - Q_S} \cdot Q_S \cdot \frac{H_{n,r}}{H_{s,l}} \cdot \frac{H_{s,r}}{H_{n,r}} = -\frac{Q_Y - Q_N}{Q_Y - Q_S}, \qquad \text{(A.3)}$$

which gives the equations (5.11). Additionally and without further proof, the binaural SNR, e.g., the SNR of the sum of the left and right spectrum, can be derived from (A.1) as

$$\frac{S_b}{N_b} \equiv \frac{S \cdot [H_{s,l} + H_{s,r}]}{N \cdot [H_{n,l} + H_{n,r}]} = -\frac{Q_Y - Q_N}{Q_Y - Q_S} \cdot \frac{Q_S + 1}{Q_N + 1}. \qquad \text{(A.4)}$$

References

Albani, S., B. Gabriel, V. Hohmann and B. Kollmeier (**1998**). Konzept und Ergebnis einer Feldstudie eines digitalen Hörgeräte-Algorithmus zur Störgeräuschreduktion. *Zeitschrift für Audiologie*, **Suppl. 1**:100–102. ISSN 1437–8914.

Albani, S., J. Peissig and B. Kollmeier (**1996**). Model of binaural localization resolving multiple sources and spatial ambiguities. In B. Kollmeier (Ed.), *Psychoacoustics, speech and hearing aids*, pp. 227–232. Singapore: World Scientific. ISBN 981022561X.

Allen, J. B. (**1996**). Derecruitment by multi-band compression in hearing aids. In B. Kollmeier (Ed.), *Psychoacoustics, speech and hearing aids*, pp. 141–152. Singapore: World Scientific. ISBN 981022561X.

Allen, J. B., D. A. Berkley and J. Blauert (**1977**). Multimicrophone signal-processing technique to remove room reverberation from speech signals. *J. Acoust. Soc. Am.*, **62(4)**:912–915.

Allen, J. B. and L. R. Rabiner (**1977**). A unified approach to short-time fourier analysis and synthesis. *Proc. of the IEEE*, **65**:1558–1564.

Arlinger, S., J. Hellgren and T. Lunner (**1994**). A wearable digital hearing aid. In *Issues in Advanced Hearing Aid Research*. Lake Arrowhead: House Ear Institute (Los Angeles).

Asano, F., S. Hayamizu, Y. Suzuki, S. Tsukui and T. Sone (**1996**). Array signal processing applicable to hearing aids. In *Proc. of the Joint ASA/ASJ Meeting*. Honolulu, Hawaii.

Berghe, J. V. and J. Wouters (**1998**). An adaptive noise canceller for hearing aids using two nearby microphones. *J. Acoust. Soc. Am.*, **103(6)**:3621–3626.

Bodden, M. (**1993**). Modeling human sound-source localization and the cocktail-party-effect. *Acta Acustica*, **1**:43–56.

Boll, S. F. (**1979**). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust., Speech, Signal Processing*, **24**:113–120. ISSN 1053–587X.

Boll, S. F. (**1991**). Speech enhancement in the 1980s. In S. Furui and M. M. Sondhi (Eds.), *Advances in Speech Signal Processing*, pp. 309–325. New York: Dekker. ISBN 0–8247–8540–1.

Bortz, J., G. A. Lienert and K. Boehnke (**1990**). *Verteilungsfreie Methoden in der Biostatistik*. Berlin Heidelberg New York: Springer-Verlag. ISBN 3–540–50737–X.

Bouquin-Jeannès, R. Le and G. Faucon (**1995**). Study of a voice activity detector and its influence on a noise reduction system. *Speech Communication*, **16**:245–254. ISSN 0167-6393.

Brainard, M. S., E. I. Knudsen and S. D. Esterly (**1992**). Neural derivation of sound source location: Resolution of spatial ambiguities in binaural cues. *J. Acoust. Soc. Am.*, **91**:1015–1027.

Brey, R. H., M. S. Robinette, D. M. Chabries and R. W. Christiansen (**1987**). Improvement in speech intelligibility in noise employing an adaptive filter with normal and hearing impaired subjects. *J. Reh. Res. Dev.*, **24**:75–86.

Bronkhorst, A. W. and R. Plomp (**1989**). Binaural speech intelligibility in noise for hearing-impaired listeners. *J. Acoust. Soc. Am.*, **86(4)**:1374–1383.

Cappé, O. (**1994**). Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. *IEEE Trans. Speech and Audio Processing*, **Vol. 2, No. 2**:345–349.

Casseday, J. H. and E. Covey (**1987**). Central auditory pathways in directional hearing. In W. A. Yost and G. Gourevitch (Eds.), *Directional Hearing*. New York: Springer-Verlag.

Colburn, H. S. (**1996**). Binaural psychoacoustics and models. In B. Kollmeier (Ed.), *Psychoacoustics, speech and hearing aids*, pp. 211–220. Singapore: World Scientific. ISBN 981022561X.

Cox, R. M., G. C. Alexander and I. M. Rivera (**1991**). Comparison of objective and subjective measures of speech intelligibility in elderly hearing-impaired listeners. *J. Speech Hear. Res.*, **34**:904–915.

Damaske, P. and B. Wagener (**1969**). Richtungshörversuche über einen nachgebildeten Kopf. *Acustica*, **21**:30–35.

Dempsey, J. J. (**1994**). Hearing aid fitting and evaluation. In J. Katz (Ed.), *Handbook of clinical audiology*, pp. 723–735. Baltimore: Williams & Wilkins. 4th edition, ISBN 0–683–04548–2.

Desloge, J. G., W. M. Rabinowitz and P. M. Zurek (**1997**). Microphone-array hearing aids with binaural output - Part I: Fixed-processing systems. *IEEE Trans. Speech and Audio Processing*, **Vol. 5, No. 6**:529–542.

Dillier, N. (**1996**). Anpaßverfahren und Evaluationsergebnisse mit neuen Algorithmen für digitale Hörhilfen. In *Fortschritte der Akustik – DAGA '96*, pp. 68–71. Oldenburg: DEGA.

Dillon, H. and R. Lovegrove (**1993**). Single-microphone noise reduction systems for hearing aids: a review and an evaluation. In G. A. Studebaker and I. Hochberg (Eds.), *Acoustical factors affecting hearing aid performance*, pp. 353–372. Boston: Allyn and Bacon. 2nd edition, ISBN 0–205–13778–4.

Dörbecker, M. (**1998**). Sind kohärenzbasierte Störgeräuschreduktionsverfahren für elektronische Hörhilfen geeignet? – Modelle zur Beschreibung der Kohärenzeigenschaften kopfbezogener Mikrofonsignale. In *5. ITG-Fachtagung Sprachkommunikation*, pp. 53–56. Dresden: Hrsg. Rüdiger Hoffmann. ISBN 3–8007–2350–6.

Elberling, C., C. Ludvigsen and G. Keidser (**1993**). The design and testing of a noise reduction algorithm based on spectral subtraction. *Scand. Audiol.*, **Suppl. 38**:39–49.

Ephraim, Y. and D. Malah (**1985**). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP–33, No. 2**:443–445. ISSN 1053–587X.

Faulkner, A., A. J. Fourcin and B. C. J. Moore (**1990**). Psychoacoustic aspects of speech pattern coding for the deaf. *Acta Otolaryngol (Stockh.)*, **Suppl. 469**:172–180.

Gabriel, K. J., J. Koehnke and H. S. Colburn (**1992**). Frequency dependence of binaural performance in listeners with impaired binaural hearing. *J. Acoust. Soc. Am.*, **91**(1):336–347.

Gaik, W. and W. Lindemann (**1986**). Ein digitales Richtungsfilter, basierend auf der Auswertung interauraler Parameter von Kunstkopfsignalen. In *Fortschritte der Akustik – DAGA '86*, pp. 721–724. Bad Honeff: DPG Kongreßgesellschaft. ISSN 0720–2253.

Gelnett, D. J., J. A. Sullivan, M. J. Nilsson and S. D. Soli (**1995**). Field trials of a portable prototype digital hearing aid. *J. Acoust. Soc. Am.*, **97(5, Pt. 2)**:3346.

Gingsjö, A. L. (**1996**). A wearable digital master hearing aid with transient reduction software. *J. Acoust. Soc. Am.*, **100(4, Pt. 2)**:2741.

Goldsworthy, R. (**1998**). Beamforming algorithms used for noise reduction. In *Symposium on Speech and Hearing Sciences Program.* Harvard, MS: MIT.

Graupe, D., J. K. Grosspietsch and S. P. Basseas (**1984**). Self-adaptive filtering of environmental noises from speech. In *Proc. AIAA/IEEE 6th Avionics Sys. Conf.* Baltimore, MD.

Graupe, D., J. K. Grosspietsch and S. P. Basseas (**1987**). A single-microphone-based self-adaptive filter of noise from speech and its performance evaluation. *J. Reh. Res. Dev.*, **24(4)**:119–126.

Greenberg, J. E. and P. M. Zurek (**1992**). Evaluation of an adaptive beamforming method for hearing aids. *J. Acoust. Soc. Am.*, **91(3)**:1662–1676.

Griffiths, L. J. and C. W. Jim (**1982**). An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propag.*, **30**:27–34.

Grim, M., M. Terry and C. Schweitzer (**1995**). Evaluation of a non-linear frequency domain beam-forming algorithm for use in a digital hearing aid. In Mike Newman (Ed.), *Proceedings of the 15th International Congress on Acoustics*, pp. 77–80. Trondheim, Norway. ISBN 82–595–8995–8.

Hansen, J. H. L. and B. L. Pellom (**1998**). An effective quality evaluation protocol for speech enhancement algorithms. In *Int. Conf. on Spoken Language Processing, ICSLP '98*. Sydney, Australia.

Häusler, R., H. S. Colburn and E. Marr (**1983**). Sound localization in subjects with impaired hearing. *Acta Otolaryngol (Stockh.)*, **Suppl. 400**:1–62.

Hoffman, M. W., T. D. Trine, K. M. Buckley and D. J. Van Tasell (**1994**). Robust adaptive microphone array processing for hearing aids: realistic speech enhancement. *J. Acoust. Soc. Am.*, **96(2, Pt. 1)**:759–770.

Hohmann, V. and B. Kollmeier (**1995**). Weiterentwicklung und klinischer Einsatz der Hörfeldskalierung. *Audiologische Akustik*, **34(2)**:48–64.

Hohmann, V. and B. Kollmeier (**1996**). Perceptual models for hearing aid algorithms. In B. Kollmeier (Ed.), *Psychoacoustics, speech and hearing aids*, pp. 193–202. Singapore: World Scientific. ISBN 981022561X.

Hohmann, Volker (**1993**). *Dynamikkompression für Hörgeräte – Psychoakustische Grundlagen und Algorithmen*. Düsseldorf: VDI-Verlag. Reihe 17, Nummer 93, ISBN 3–18–149317–1.

Holube, Inga (**1993**). *Experimente und Modellvorstellungen zur Psychoakustik und zum Sprachverstehen bei Normal- und Schwerhörigen*, (Ph.D. thesis). Universität Göttingen.

Hövel, H. von (**1984**). *Zur Bedeutung der Übertragungseigenschaften des Außenohres sowie des binauralen Hörsystems bei gestörter Sprachübertragung*, (Ph.D. thesis). RWTH Aachen.

Humes, L. E., L. A. Christensen, F. H. Bess and A. Hedley-Williams (**1997**). A comparison of the benefit provided by well-fit linear hearing aids and instruments with automatic reductions of low-frequency gain. *J. Speech Lang. Hear. Res.*, **40(3)**:666–685.

Hussain, A., D. R. Campbell and T. J. Moir (**1997**). A new metric for selecting sub-band processing in adaptive speech enhancement systems. In *Proc. EUROSPEECH '97*, Vol. **Vol. 5**, pp. 2611–2614. ESCA.

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

Kates, J. M. (**1986**). Signal processing for hearing aids. *Hearing Instruments*, **37(2)**:19–22.

Kendall, M. G. (**1975**). *Rank correlation methods.* London: Griffin & Co. ISBN 0 85264 199 0.

Kießling, J. (**1997**). Versorgung mit Hörgeräten. In J. Kießling, B. Kollmeier and G. Diller (Eds.), *Versorgung und Rehabilitation mit Hörgeräten*, pp. 49–110. Stuttgart: Georg Thieme Verlag. ISBN 3–13–106821–3.

Kinkel, M., B. Kollmeier and I. Holube (**1991**). Binaurales Hören bei Normal- und Schwerhörigen I: Methoden und Ergebnisse. *Audiologische Akustik*, **30**:192–201.

Kleinschmidt, M., J. Tchorz and B. Kollmeier (**2000**). Combining speech enhancement and auditory feature extraction for robust speech recognition. *Speech Communication.* (accepted for publication), ISSN 0167-6393.

Kollmeier, B. and R. Koch (**1994**). Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction. *J. Acoust. Soc. Am.*, **95**:1593–1602.

Kollmeier, B., J. Peissig and V. Hohmann (**1993**). Real-time multiband dynamic compression and noise reduction for binaural hearing aids. *J. Reh. Res. Dev.*, **30**(1):82–94.

Kollmeier, B. and M. Wesselkamp (**1996**). Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *J. Acoust. Soc. Am.* submitted.

Kollmeier, Birger (**1997**). Signal processing for hearing aids employing binaural cues. In Gilkey, Robert H., Anderson and Timothy R. (Eds.), *Binaural and Spatial Hearing in Real and Virtual Environments*, pp. 753–775. Mahwah, New Jersey: Lawrence Erlbaum Assoc.

Kompis, M. and N. Dillier (**1994**). Noise reduction for hearing aids: Combining directional microphones with an adaptive beamformer. *J. Acoust. Soc. Am.*, **96(3)**:1910–1913.

Konishi, M., T. T. Takahashi, W. E. Wagner, W. E. Sullivan and C. E. Carr (**1988**). Neurophysiological and anatomical substrates of sound localization in the owl. In W. E. Gall G. M. Edelman and W. M. Cowan (Eds.), *Auditory Function: Neurobiological Basis of Hearing.* John Wiley & Sons.

Kühnel, V. and B. Kollmeier (**1997**). Sprachaudiometrie in Ruhe und Störgeräusch: Göttinger Satztest im Vergleich zu konventionellen Sprachverständlichkeitstests im klinischen Einsatz. In *Fortschritte der Akustik - DAGA97*, pp. 91–92. Kiel: DEGA, Oldenburg.

Langner, G. (**1992**). Periodicity coding in the auditory system. *Hear. Res.*, **60**:115–142.

Launer, S., I. Holube, V. Hohmann and B. Kollmeier (**1996**). Categorical loudness scaling in hearing-impaired listeners – Can loudness growth be predicted from the audiogram? *Audiological Acoustics*, **35 (4)**:156–163. ISSN 0172–8261.

Levitt, H., A. Neumann and J. Sullivan (**1990**). Studies with digital hearing aids. *Acta Otolaryngol (Stockh.)*, **Suppl. 469**:57–69.

Lim, J. S. (**1983**). *Speech Enhancement*. Prentice-Hall. ISBN 0–13–829705–3.

Lim, J. S. and A. V. Oppenheim (**1979**). Enhancement and bandwidth compression of noisy speech. *Proc. of the IEEE*, **67**:1586–1604. ISSN 0018–9219.

Lindemann, E. (**1995**). Two microphone nonlinear frequency domain beamformer for hearing aid noise reduction. In *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*. Mohonk, NY.

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

Lindemann, W. (**1986b**). Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front. *J. Acoust. Soc. Am.*, **80**:1623–1630.

Lu, M. and P. M. Clarkson (**1993**). The performance of adaptive noise cancellation systems in reverberant rooms. *J. Acoust. Soc. Am.*, **93(2)**:1122–1135.

Marzinzik, M. (**2000**). *Noise Reduction Schemes for Digital Hearing Aids and their Use for the Hearing Impaired*, (Ph.D. thesis). Universität Oldenburg.

Marzinzik, M., J. E. Appell, V. Hohmann and B. Kollmeier (**1996**). Evaluation of dynamic compression algorithms using a loudness model for hearing impaired listeners. In B. Kollmeier (Ed.), *Psychoacoustics, speech and hearing aids*, pp. 203–207. Singapore: World Scientific. ISBN 981022561X.

Marzinzik, M. and B. Kollmeier (**1999**). Development and evaluation of single-microphone noise reduction algorithms for digital hearing aids. In T. Dau, V. Hohmann and B. Kollmeier (Eds.), *Psychophysics, physiology and models of hearing*, pp. 279–282. Singapore: World Scientific. ISBN 981–02–3741–3.

Marzinzik, M. and B. Kollmeier (**2000**). Benefits of noise suppression for the hearing impaired. In *Fortschritte der Akustik – DAGA 2000*, p. in preparation. Oldenburg: DEGA.

McAulay, R. J. and M. L. Malpass (**1980**). Speech enhancement using a soft-decision noise suppression filter. *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP–28, No. 2**:137–145. ISSN 1053–587X.

Moore, B. C. J. and B. R. Glasberg (**1983**). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J. Acoust. Soc. Am.*, **74**:750–753.

Nielsen, H. B. and C. Ludvigsen (**1978**). Effect of hearing aids with directional microphones in different acoustic environments. *Scand. Audiol.*, **7(4)**:217–224.

Parsons, T. W. (**1976**). Separation of speech from interfering speech by means of harmonic selection. *J. Acoust. Soc. Am.*, **60(4)**:911–918.

Pastoors, A. D., T. M. Gebhart, B. Kollmeier and J. Kießling (**1998**). Digitales Prototyp-Hörgerät im Feldtest. *Zeitschrift für Audiologie*, **Suppl. 1**:103–105. ISSN 1437–8914.

Pavlovic, C. V. (**1987**). Derivation of primary parameters and procedures for use in speech intelligibility predictions. *J. Acoust. Soc. Am.*, **82(2)**:413–422.

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

Peissig, Jürgen (**1993**). *Binaurale Hörgerätestrategien in komplexen Störschallsituationen.* Düsseldorf: VDI-Verlag. Reihe 17, Nummer 88, ISBN 3–18–148817–8.

Peterson, P. M., N. I. Durlach, W. M. Rabinowitz and P. M. Zurek (**1987**). Multimicrophone adaptive beamforming for interference reduction in hearing aids. *J. Reh. Res. Dev.*, **24(4)**:103–11.

Plomp, R. (**1978**). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J. Acoust. Soc. Am.*, **63(2)**:533–549.

Rass, U. (**1996**). A wearable signal processor system for the evaluation of digital hearing aid algorithms. In B. Kollmeier (Ed.), *Psychoacoustics, speech and hearing aids*, pp. 273–276. Singapore: World Scientific. ISBN 981022561X.

Rass, U. and G. H. Steeger (**1999**). A high performance wearable signal processor system for the evaluation of digital hearing aid algorithms. In *Proc. of the Joint ASA/EAA/DEGA Meeting.* Berlin, Germany.

Schroeder, M. R., B. S. Atal and J. L. Hall (**1979**). Optimizing digital speech decoders by exploiting masking properties of the human ear. *J. Acoust. Soc. Am.*, **66**:1647–1652.

Soede, W. (**1990**). *Improvement of Speech Intelligibility in Noise: Development and evaluation of a new directional hearing instrument based on array technology.* Delft, The Netherlands: Delft University of Technology. ISBN 90–9003763–2.

Soede, W., A. J. Berkhout and F. A. Bilsen (**1993**). Development of a directional hearing instrument based on array technology. *J. Acoust. Soc. Am.*, **94(2, Pt. 1)**:785–798.

Sone, T., Y. Suzuki, F. Asano, T. Takasaka, M. Ohashi and K. Yamaguchi (**1995**). A portable digital hearing aid with narrow-band loudness compensation and the fitting system for it. In *Proc. of the 15th ICA, Vol. 4*, pp. 265–268. ISBN 82–595–8995–8.

Stearns, S. D. (**1991**). *Digitale Verarbeitung analoger Signale.* München Wien: R. Oldenbourg Verlag. ISBN 3–486–21986–3.

Steeger, G. H. (**1996**). Classical solutions and new concepts in hearing aid technology. In B. Kollmeier (Ed.), *Psychoacoustics, speech and hearing aids*, pp. 263–272. Singapore: World Scientific. ISBN 981022561X.

Strube, H. W. (**1981**). Separation of several speakers recorded by two microphones (cocktail-party processing). *Signal Processing*, **3**:355–364.

Stubbs, R. J. and Q. Summerfield (**1988**). Evaluation of two voice-separation algorithms using normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.*, **84(4)**:1236–1249.

Stubbs, R. J. and Q. Summerfield (**1991**). Effects of signal-to-noise ratio, signal periodicity, and degree of hearing impairment on the performance of voice-separation algorithms. *J. Acoust. Soc. Am.*, **89**:1383–1393.

Sullivan, T. M. and R. M. Stern (**1993**). Multi-microphone correlation-based processing for robust speech recognition. In *Proceedings of the ICASSP, Minneapolis.*

Summerfield, Q. and R. J. Stubbs (**1990**). Strength and weakness of procedures for separating simultaneous voices. *Acta Otolaryngol (Stockh.)*, **Suppl. 469**:91–100.

Suzuki, Y., S. Tsukui, F. Asano, R. Nishimura and T. Sone (**1999**). New design method of a binaural microphone array using multiple constraints. *IEICE Trans. Fundamentals*, **Vol. E82-A, No. 4**:588–596.

Terry, M., C. Schweitzer, E. Lindemann and J. Melanson (**1994**). Evaluation of a prototype beamforming binaural hearing aid. *J. Acoust. Soc. Am.*, **95(5, Pt. 2)**:2991.

v. d. Malsburg, C. and J. Buhmann (**1992**). Sensory sementation with coupled neural oscillators. *Biol. Cybern.*, **67**:233–242.

van Campernolle, D. (**1990**). Hearing aids using binaural processing principles. *Acta Otolaryngol (Stockh.)*, **Suppl. 469**:76–84.

Verschuure, J. and W. A. Dreschler (**1996**). Dynamic compression hearing aids. In B. Kollmeier (Ed.), *Psychoacoustics, speech and hearing aids*, pp. 153–164. Singapore: World Scientific. ISBN 981022561X.

Wagener, K., T. Brand and B. Kollmeier (**1999a**). Development and evaluation of a German sentence test II: Optimization of the Oldenburg sentence test. *Zeitschrift für Audiologie*, **38 (2)**:44–56. ISSN 1435–4691.

Wagener, K., T. Brand and B. Kollmeier (**1999b**). Development and evaluation of a German sentence test III: Evaluation of the Oldenburg sentence test. *Zeitschrift für Audiologie*, **38 (3)**:86–95. ISSN 1435–4691.

Wagener, K., V. Kühnel and B. Kollmeier (**1999c**). Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test. *Zeitschrift für Audiologie*, **38 (1)**:4–15. ISSN 1435–4691.

Wang, D. L. and J. S. Lim (**1982**). The unimportance of phase in speech enhancement. *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP–30, No. 4**:679–681. ISSN 1053–587X.

Welker, D. P., J. E. Greenberg, J. G. Desloge and P. M. Zurek (**1997**). Microphone-array hearing aids with binaural output - part ii: A two-microphone adaptive system. *IEEE Trans. Speech and Audio Processing*, **Vol. 5, No. 6**:543–551.

Wesselkamp, M., K. Kliem and B. Kollmeier (**1992**). Erstellung eines optimierten Satztests in deutscher Sprache. In B. Kollmeier (Ed.), *Moderne Verfahren der Sprachaudiometrie, Buchreihe Audiologische Akustik*. Heidelberg: median-verlag. ISBN 3–922766–15–3.

Wesselkamp, Matthias (**1994**). *Messung und Modellierung der Verständlichkeit von Sprache*, (Ph.D. thesis). Universität Göttingen.

Wittkop, T., S. Albani, V. Hohmann, J. Peissig, W. S. Woods and B. Kollmeier (**1997**). Speech processing for hearing aids: Noise reduction motivated by models of binaural interaction. *ACUSTICA · acta acustica*, **83**(4):684–699. ISSN 0001–7884.

Woods, W. S., M. Hansen, T. Wittkop and B. Kollmeier (**1996a**). A simple architecture for using multiple cues in sound separation. In *Proceedings, ICSLP 96, Philadelphia*.

Woods, W. S., M. Hansen, T. Wittkop and B. Kollmeier (**1996b**). Using multiple cues for sound source separation. In B. Kollmeier (Ed.), *Psychoacoustics, speech and hearing aids*, pp. 253–258. Singapore: World Scientific. ISBN 981022561X.

Zurek, P. M., J. E. Greenberg and W. M. Rabinowitz (**1996**). Prospects and limitations of microphone-array hearing aids. In B. Kollmeier (Ed.), *Psychoacoustics, speech and hearing aids*, pp. 233–244. Singapore: World Scientific. ISBN 981022561X.

Zwicker, E. (**1961**). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *J. Acoust. Soc. Am.*, **33**:248.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Oldenburg, den 15. Dezember 2000

Thomas Wittkop

Danksagung

An dieser Stelle möchte ich all denen ganz herzlich danken, die zum Gelingen dieser Arbeit beigetragen haben. Insbesondere gilt mein Dank

Herrn Prof. Dr. Dr. Birger Kollmeier dafür, dass er die Anfertigung dieser Dissertation ermöglicht und unterstützt hat, sowie für die Betreuung, die vielen Anregungen und die Zeit, die er in diese Arbeit investiert hat,

Dr. Volker Hohmann für die Betreuung und vielfältige Unterstützung bei der Erstellung dieser Arbeit,

den Mitarbeiterinnen und Mitarbeitern des Hörzentrum Oldenburg, insbesondere Dr. Birgitta Gabriel und Kerstin Sommer, sowie Anita Gorges und Bajo Meenen für die Mitarbeit und Mithilfe an den Messungen und technische Unterstützung,

Mark Marzinzik für die Mithilfe insbesondere an Kapitel 6,

allen Kolleginnen und Kollegen aus AG Medizinische Physik, GK Psychoakustik, AG Akustik, SFB Neurokognition sowie Euro-GK Neurosensorik für die bereitwillige Mithilfe, technische Unterstützung und gute Arbeitsatmosphäre,

allen Versuchspersonen für ihre Mithilfe,

sowie Stephan Albani, Jens E. Appell, Daniel Berg, Dr. Birgitta Gabriel und Dr. Volker Hohmann für ihre ständige und vielfältige Unterstützung.

Diese Arbeit wurde unterstützt vom Bundesministerium für Bildung und Forschung, von der Europäischen Union und von der Deutschen Forschungsgemeinschaft.

Lebenslauf

Ich wurde am 9.9.1968 als zweites Kind von Rotraut Wittkop, geb. Hamann, und Helmut Wittkop in Hamburg mit deutscher Staatsangehörigkeit geboren.
Ab 1974 besuchte ich die Grundschule Kronstieg in Hamburg, von der ich 1976 zur Grundschule Harksheide Süd in Norderstedt wechselte. Von 1978 an besuchte ich das Gymnasium im Schulzentrum Süd in Norderstedt, das ich im Mai 1987 mit dem Abitur abschloss.
Zum Wintersemester 1988/89 begann ich das Studium der Physik an der Georg-August-Universität in Göttingen und legte dort im April 1991 die Vordiplomsprüfung ab. Im Juli 1993 begann ich am Dritten Physikalischen Institut der Universität Göttingen unter der Anleitung von Prof. Dr. Dr. Birger Kollmeier mit der Diplomarbeit "Vergleich binauraler digitaler Hörgerätestrategien zur Störgeräuschunterdrückung". Die Diplomprüfung im Fach Physik legte ich zum 1. Juli 1994 in Göttingen ab.
Seit dem 1. Juli 1994 arbeite ich in der Arbeitsgruppe Medizinische Physik der Universität Oldenburg als wissenschaftlicher Mitarbeiter. Dort fertigte ich unter der Anleitung von Prof. Dr. Dr. B. Kollmeier diese Dissertation an. Bis zum April 1998 habe ich im vom Bundesministerium für Bildung und Forschung geförderten Projekt "Entwicklung und Bewertung von digitalen Hörgeräte-Algorithmen, Anpassungsverfahren und Prototypen" mitgearbeitet. Vom Mai 1998 bis Juni 2000 arbeitete ich am Projekt "SPACE" zur Entwicklung und Evaluation von Signalverarbeitungsstrategien für Schwerhörende mit, das von der Europäischen Union gefördert wurde. Seit Juli 2000 bin ich im Rahmen des DFG Schwerpunktprogramms "Grundlagen und Verfahren verlustarmer Informationsverarbeitung (VIVA)" tätig.