

# Modelling the effective binaural signal processing in the auditory system

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

**Carsten Zerbs**  
geb. am 18. Aug. 1967  
in Fulda

Erstreferent: Prof. Dr. Dr. Birger Kollmeier  
Korreferent: Prof. Dr. Volker Mellert  
Tag der Disputation: 28. Januar 2000

# Contents

<b>1</b>	<b>General Introduction</b>	<b>1</b>
<b>2</b>	<b>Literature review</b>	<b>5</b>
2.1	Introduction . . . . .	6
2.2	Aspects of binaural signal processing . . . . .	7
2.3	Cross-correlation models . . . . .	11
2.4	Conclusion . . . . .	14
<b>3</b>	<b>“Effective” binaural signal processing</b>	<b>15</b>
3.1	Introduction . . . . .	16
3.2	Description of the signal processing in the model . . . . .	19
3.2.1	The original model - monaural preprocessing unit . . . . .	19
3.2.2	Optimal detector . . . . .	22
3.2.3	The binaural processing unit . . . . .	23
3.3	Experimental method . . . . .	29
3.3.1	Procedure and subjects . . . . .	29
3.3.2	Apparatus and stimuli . . . . .	29
3.4	Results . . . . .	30
3.4.1	Frequency dependence of binaural thresholds and BMLD . . . . .	30
3.4.2	Interaural time delay of the masker . . . . .	31
3.4.3	Level dependence of binaural thresholds . . . . .	34
3.4.4	Processing of interaural intensity differences . . . . .	34
3.4.5	Forward masking in binaural and monaural configurations . . . . .	37
3.5	Model structure . . . . .	39
3.5.1	Influence of jitter vs. role of neural transduction . . . . .	39
3.5.2	Processing of level and IIDs . . . . .	43
3.6	Discussion . . . . .	47
3.7	Summary . . . . .	50
<b>4</b>	<b>Spectral and temporal effects</b>	<b>51</b>
4.1	Introduction . . . . .	52
4.2	Method . . . . .	54

4.2.1	Procedure and subjects . . . . .	54
4.2.2	Apparatus and stimuli . . . . .	54
4.3	Results: effects of spectral resolution . . . . .	55
4.3.1	Combined interaural time and phase differences . . . . .	55
4.3.2	Frequency dependent interaural phase . . . . .	56
4.3.3	Spectral integration and binaural detection . . . . .	58
4.4	Effects of temporal processing . . . . .	63
4.5	Discussion . . . . .	65
4.5.1	Information processing within one critical band . . . . .	65
4.5.2	External and internal noise statistics in multichannel processing	66
4.6	Summary . . . . .	71
<b>5</b>	<b>Binaural sluggishness</b>	<b>72</b>
5.1	Introduction . . . . .	73
5.2	Results . . . . .	74
5.3	Strategy sluggishness . . . . .	76
5.4	Fast fluctuating interaural parameters . . . . .	79
<b>6</b>	<b>Summary and General Conclusions</b>	<b>82</b>
<b>A</b>	<b>Internal adjustment of external parameters</b>	<b>85</b>
A.1	Factor $q$ for internal amplitude adjustment . . . . .	85
<b>B</b>	<b>Examples of binaural signal processing</b>	<b>86</b>
<b>C</b>	<b>Level-dependent peripheral noise</b>	<b>92</b>
	<b>References</b>	<b>96</b>
	<b>Danksagung</b>	<b>106</b>

# Chapter 1

## General Introduction

The ability of the auditory system to receive and to process acoustic information is an important prerequisite for orientation in surroundings as well as for participation in human communication. The auditory system provides us with abundant information about our environment by performing a transformation of the sound field at our ears into percepts. This local sound field can be quite complex, the affiliated information can be manifold. For instance there might be overlapping sound fields of various spatial and temporal distributed sound sources with different significance and importance for the listener. This superposition of different signals reaches the information processing of one ear as a one dimensional temporal signal. Additionally, each ear not only receives signals from many sources, but it receives the signal from one source on many paths (i. e., with temporal overlapping delays) due to multiple reflections in the environment.

To organize this complex auditory world the auditory system is able to separate various simultaneous sound stimuli into discrete sound sources: auditory objects are assigned to auditory events (Blauert, 1997) to identify these sources and to obtain information about them. The obtained information can contain very different aspects about the sound sources. It can concern the direction and distance of a source (*localization*) or its speed relative to the listener. It can also deal with the inner state of an object like the ringing of an alarm clock. A further important aspect of acoustical information is the meaning of a verbal message. To make acoustic information transfer more efficient, the auditory system seems also to be able to suppress sound from competing sound sources or background noise in favour of one single source to receive a better input. This process of filtering out the portions of information relevant to the listener is denoted as “cocktail party effect” (Cherry, 1953), which is most important for improving speech intelligibility in a noisy environment.

These different aspects of auditory performance are partly made possible due to *monaural* analysis of different temporal and spectral properties of the stimuli, i. e. the properties of signals received with one ear are evaluated. However, a large amount of these tasks can only be performed due to *binaural* information processing which means that *both* ear signals are processed and compared in common on a more

central level. It enables the auditory system to extract additional information from differences between both ear signals that are not available from monaural processing.

Despite the great achievements in information processing, there is a reduction in information content when sound signals are processed in the auditory system due to several physiological limitations in the way how acoustical information is transduced, passed on and evaluated. The loss in information results in a limited capacity to perceive sounds and can be analyzed by detection or discrimination experiments. Detection experiments reveal that sounds can be masked by other sounds. Discrimination experiments show that not every change in a stimulus can be detected and hence demonstrate the degree to which information processing is limited.

Since there are various features of binaural signal processing, it has been extensively investigated from different points of view in many different experimental configurations. We can classify binaural performance according to several *perceptual* attributes or according to *sensitivity* phenomena. Typical absolute perceptual attributes are space (in terms of acoustical localization or lateralization<sup>1</sup>), loudness (Chouard, 1997) or pitch. Typical sensitivity phenomena are spatial resolution, detectability of differences in both ear signals (binaural just noticeable differences, JNDs) or improvement in binaural signal detection when masker and signal are presented with different instead of identical interaural properties (e. g., Durlach and Colburn, 1978).

The latter feature is also called “binaural noise reduction” and is an impressive feature of the auditory system which can easily be observed in daily-life situations: A listener is able focus attention on one speaker in a background of voices. The speaker becomes difficult to understand if the listener plugs one ear. This phenomenon (the above mentioned “cocktail party effect”) arises from the fact that a “signal” (the speaker) can be separated from a background “noise” by a listener if signal and noise have different directions of sound incidence. The improvement in detection (or release from masking) is measured and quantified in binaural psychophysical experiments and compared to an appropriate reference condition where no binaural information is available. It is referred to as the “binaural masking level difference” (BMLD). For speech in noise the improvement is called “binaural intelligibility level difference” (BILD). Binaural speech intelligibility of speech has been investigated e. .g. by Kollmeier (1990) or von Hövel (1984).

The physical quantities on which binaural processing and perception is based on are the sound signals in both ear canals (Blauert, 1997). So the main interests in binaural psychophysics consist on the one hand in the investigation how interaural differences of the ear signals *originate* at a human head in a sound field and on the other hand how they are *evaluated* by the auditory system.

The influence of the human head on the sound field can be approximately investigated by treating it as a rigid sphere of similar size. The first precise mathematical

---

<sup>1</sup>That means the estimation of the direction of a sound source on the basis of signal properties, and, more difficult and less examined, of the distance of a sound source.

description of that problem was given by Lord Rayleigh (Rayleigh, 1907) in context with localization of sound sources. He postulated already the existence of binaural “cues” (features of a stimulus usable for perceptive performance) as basis for localization. Due to orientation of the head relative to the direction of the distant sound source two cues can be derived: interaural level differences (ILD, also abbreviated as interaural intensity differences, IID) and interaural time differences (ITD)<sup>2</sup>. The IIDs of the signal coming from one single sound source in a free-field configuration are produced by diffraction, the ITDs by different lengths of the acoustic pathways reaching both ears. In natural situations with more than one sound source both are time-varying and frequency-dependent magnitudes. For a real head, a complete analytical description is no longer possible. Therefore this simplified description has been refined by considering the complete measured free-field transfer function of the head, the external ears (pinnae) and the ear canals. They can be monitored by placing small probe microphones inside the ear canal presenting binaural stimuli (in the simplest case click pulses). Measurements of such “head related transfer functions” (HRTFs)  $H_{r,\vartheta,\phi}(f)$  on a variety of subjects show that HRTFs show individual characteristics for different subjects (e. g., Mehrgardt and Mellert, 1977; Hartung, 1995). HRTFs can serve to present artificial auditory scenes or “virtual acoustics”, i. e. simulations of free-field conditions via headphones with individual IIDs and ITDs.

In most psychophysical detection experiments synthetic stimuli are presented via headphones. The main advantage of this method is that it permits to control the stimuli and the interaural disparities at the two ears very precisely. This is not the case in free-field presentations because of varying external-ear transfer functions (HRTFs) within a group of subjects. Furthermore, in free-field presentation not every possible combination of interaural parameters can be presented. It can be advantageous to explore features of the auditory system by presenting ear signals via headphone with non-natural stimulus parameters. So the influence of single stimulus attributes like IIDs, ITDs and interaural correlation can be measured independently. Nevertheless there have been studies where stimuli were presented via loudspeakers (Ebata et al., 1968; Burgtorf, 1963; Burgtorf and Wagener, 1968).

In our efforts to combine these different aspects of hearing and to understand how information in the auditory system is processed and evaluated models can be a useful tool. This thesis is concerned with modeling the “effective” binaural signal processing in the auditory system. To this end a well evaluated functional model of monaural signal processing and signal detection which can predict thresholds in a variety of monaural configurations is expanded by a binaural processing unit. This binaural processor works on input signals received from two corresponding monaural frequency channels of the two ears and adopts the concept of the equalization and cancellation theory of binaural signal processing (Durlach, 1972).

---

<sup>2</sup>For the “overall” perception these differences may be only one cue among others. For example in localization of a sound source also cues like the position, orientation and movement of the head yield important cues for localization.

The current work aims at clarifying how successfully such an approach can be applied to various psychophysical paradigms. Simulation of experiments covering various aspects of binaural detection are performed. The corresponding measurements partly have been performed in the framework of the present thesis, partly have been taken from the literature.

The first part of this thesis (Chapter 2) gives a short overview over various ideas, how binaural signal processing takes place in the auditory system and how they are realized in binaural models. The chapter has the intention to motivate the necessity of a suitable monaural preprocessor for a binaural model, which is able to account for a wide range of spectral and temporal aspects of signal processing in the auditory system.

In Chapter 3 the structure of the model is explained, especially the binaural processing unit. After that simulations of critical experiments are performed in order to investigate the effect of the various stages and properties of monaural and binaural signal processing. Some of these simulations are chosen to fix the value of parameters of the binaural processor. All these values are kept constant in all following simulations. The performed simulations cover the topics how different center frequencies are processed, how and up to which degree interaural delays are evaluated. Investigating the influence of masker level and interaural intensity differences on binaural detection results in the observation, that the compression of the stationary part of input signals in the monaural preprocessor is an important prerequisite to predict the outcome of these experiments correctly. From the simulation of binaural and monaural forward masking experiments it can be verified that binaural interaction follows monaural adaptation.

In the following part (Chapter 4) the model developed in the preceding chapter is tested in complex configurations, where the spectral and temporal properties of the model and the auditory system are compared. An important finding in this context is that in many configurations detection takes place in multiple frequency channels.

In Chapter 5 deals with the concept and implementation of a binaural “sluggishness” occurring in detection experiments with time variant interaural parameters. It turns out that the model of binaural signal processing as described in Chapter 3 does not account for such phenomena. A modified strategy of binaural noise reduction is proposed that leads to a sluggish performance of the binaural model. The inertia of the modified processor is based on a sluggishness in steering the mechanism of binaural noise reduction. The concept is compared with the assumption that the binaural system is not able to evaluate fast fluctuating interaural differences.

# Chapter 2

## Binaural models: a literature review

### Abstract

*This chapter gives a rough overview over general aspects and concepts of binaural processing in the auditory system. It describes briefly the main classes of binaural models and compares common and different properties and features of the signal processing in models of binaural interaction. An important reason for the inconveniences of binaural models may be the lack of an adequate monaural preprocessing unit for the binaural processor, which is able to account for spectral and temporal features of the auditory system.*

## 2.1 Introduction

Binaural signal processing is able to overcome some of the restrictions immanent in the auditory processing of monaurally received signals. For instance, since the human cochlea has a restricted frequency resolution, spectrally adjacent signal components excite the same place on the basilar membrane and stimulate the same group of hair cells (in functional terms they are processed within the same frequency channel). So a noise jammer with similar frequency components like the signal reduces the detectability of the latter. On the basis of interaural differences that may differ between signal and noise, however, binaural processing can reduce the influence of the noise. The sensitivity of a human observer to detect a target signal in the background of a masking signal can therefore be substantially better in situations where the observer receives binaural input. One way to characterize the ability of binaural processing in the human auditory system are binaural detection experiments<sup>1</sup>.

In general thresholds from a configuration without interaural differences for signal and masker are chosen as a reference. For headphone presentation this is a monotic or diotic situation<sup>2</sup>. The BMLD for a special binaural configuration is defined as the masked threshold for this configurations subtracted from the reference threshold (both in dB).

Research about BMLDs started in 1948 (Hirsh, 1948; Licklider, 1948). Since then the gain in detection due to binaural processing has been investigated under various experimental conditions in a variety of studies, mainly utilizing earphones (cf. Durlach and Colburn (1978) or Bernstein (1997) for reviews) and a large class of masker and target signals (noise, tones, speech). Most what is known about the binaural system's sensitivity to interaural temporal disparities and interaural intensity disparities is known from such studies. For detection of a low-frequency tonal signal in a broadband noise masker BMLDs up to 15 dB occur. At high frequencies BMLDs are much smaller, e. g. above about 2 kHz they are only 3 to 4 dB. BMLDs are even larger when narrow band noise maskers are used instead of broad band maskers. This is valid for low and high frequencies (Metz et al., 1968; Zurek and Durlach, 1987): BMLDs can amount up to 25 dB at low frequencies and about

---

<sup>1</sup>In detection experiments the threshold of a (target) signal  $S$  in the presence of a (noise) masker  $N$  is determined. To note a binaural configuration the interaural relations of masker and signal are indicated as expansion of the symbols  $N$  and  $S$  with indices which mark the respective interaural configuration. Common symbols are  $m$  for a monotic or  $0$  for diotic presentation. The latter is chosen since it describes a zero interaural phase difference. The interaural phase is described by the phase angle  $\phi$ . A special case is an angle of  $\pi$  (antiphasic condition). Further interaural parameters are described by their common symbols:  $\rho$  for the interaural correlation (coefficient) when statistical independent noise sources are used to present decorrelated noise. The special case of uncorrelated noise is denoted by  $u$ , an interaural time difference by  $\tau$ . The "classical" example is  $N_0S_\pi$ , the configuration with the largest improvement.

<sup>2</sup>In signal presentation via headphone monotic, diotic and dichotic presentation are distinguished. Monotic: A signal is present in only one headphone. Diotic: The same signal is present in both headphones. Dichotic: Different signals are present in each headphone (according to Stumpf (1905)).

12 dB at high frequencies.

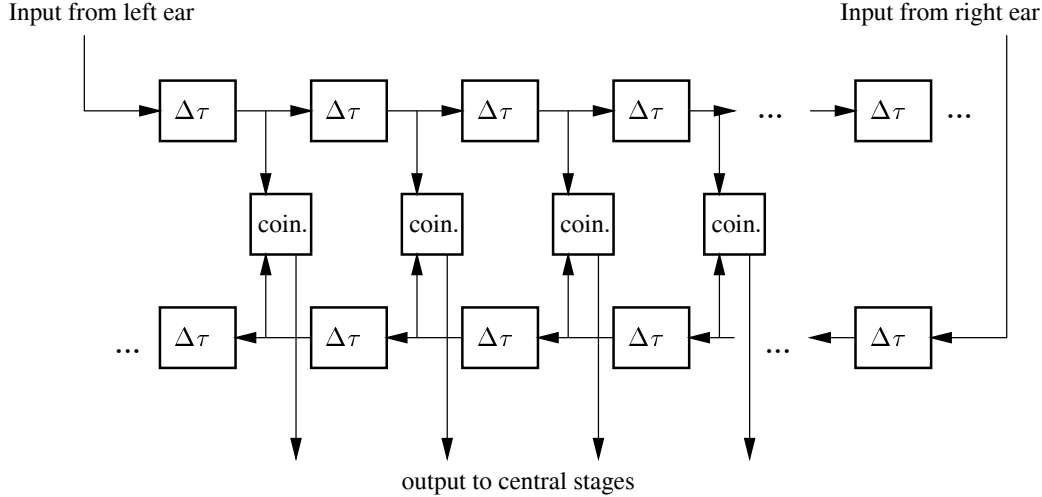
Since the enormous advantages of binaural detection relative to monaural detection were discovered much effort has been spent in understanding the processing of binaural information in the auditory system. Results from detection experiments have inspired the development of theoretical and functional models of binaural interaction, especially of models describing binaural noise reduction. Within the task to understand binaural interaction, models are a useful tool to describe how a listener takes advantage of interaural differences. They can serve to test assumptions about the mechanisms of signal processing or to estimate the size of parameters relevant in binaural listening. Different binaural models are constructed to describe and predict quantitatively certain phenomena related to binaural hearing. They can be classified according to their special subject, e. g., localization, speech recognition or binaural detection experiments. In addition, these models differ in the way how predictions are derived from binaural parameters. This may be achieved by estimating the information available to the binaural system with regard to information theory (for instance the description of interaural signal statistics) or by specifying explicit mechanisms and algorithms of signal processing. Descriptions of binaural processing might be based on both psychophysical and physiological studies and models can be constructed to predict both types of data. Also psychophysical models can include physiological components explicitly. Some of the models of binaural psychophysics include physiological aspects merely in a functional way to both describe the basic aspects of acoustic signal processing and as well to keep explicit computation as simple as possible. In all cases a model has to be described so detailed that it permits quantitative descriptions on the psychoacoustical performance in the concerned task.

The following section describes briefly the main classes of models of binaural signal processing to give a background for a rough classification of the model of the present study. The common features and main differences in signal processing are outlined, some extensions and more detailed descriptions of model features are mentioned.

## 2.2 Aspects of signal processing in binaural models

A primary aspect of all models of binaural interaction is the consideration of interaural time delays, which can be included either explicitly or implicitly. A neural “place mechanism” for extraction of interaural timing information on the basis of a neural network model was first outlined by Jeffress (1948, 1958). He suggested that external interaural delays could be recognized by neural units that record coincidences of neural impulses received from more peripheral nerve fibers from both ears. Each unit receives its input from both sides after a series of internal time delays, i. e., it is most sensitive for exactly one external time delay (and is so assigned to one internal time delay). Such a coincidence network translates external delays into in-

ternal positions of the maximum excitation of the network. One possible schematic realization of the Jeffress place mechanisms is shown in Figure 2.1. It contains two opposing lines of delay elements. Each pair of opposite delay elements feeds input into one cross-coincidence detector element.



**Figure 2.1:** Block diagram of a schematic realization of the Jeffress model.  $\Delta\tau$  labels a block describing short internal delays, “coin.” labels coincidence detector elements.

The elements constructing the original Jeffress model and their operations are not completely specified. Considerable efforts have been made in explicitly describing the neural internal delay mechanism, the coincidence mechanism and the neural input patterns into the process (Colburn and Durlach, 1978) based on physiological data. A comprehensive overview of models giving quantitative formulations of the Jeffress model have been given in a review chapter of Colburn (1996b, Sec.3).

From a functional viewpoint, the model structure proposed by Jeffress can be considered as a stage calculating the interaural short-term running cross-correlation function. A mathematical expression for the short-term correlation function of the ear signals from the right and left ear,  $R$  and  $L$  is given by

$$\Phi(t, \tau) = \int_{-\infty}^t L(t')R(t' - \tau)w(t - t') dt' \quad (2.1)$$

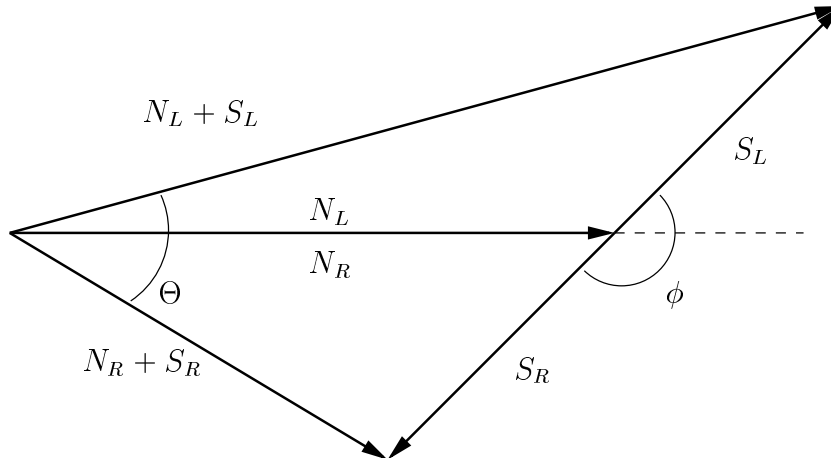
with  $t$  representing running time and  $\tau$  a delay in the signal. The weighting function  $w$  emphasizes values of the product  $L(t')R(t' - \tau)$  that have an argument close to the actual value of  $t$ . Equation 2.1 is only used if the signal  $R$  from the right ear is earlier in time, otherwise the roles of  $R$  and  $L$  are exchanged. In the Jeffress-type models the interaural cross-correlation function (ICF) is calculated from input signals

(more or less “physiological”, depending on the specific model) derived from both ears (Licklider, 1959). This is achieved by describing the operation of a coincidence element as multiplication, specifying the temporal range of inputs that results in an output of a coincidence element and a short-time integration stage behind each unit. The range of interaural delays evaluated has to be restricted since the auditory system does not account for arbitrarily large delays. To include the processing of interaural intensity differences, Jeffress assumed for the original structure of the model that the relative time delay is dependent on intensity in the more peripheral stages. According to this latency hypothesis, more intense stimuli cause the generation of an earlier input into the coincidence network compared to the less intense stimuli. In this way IIDs are translated into ITDs. Other descriptions of binaural processing reject the latency hypothesis and require an extra stage, that accounts for IIDs. In addition, for all models based on cross-correlation a separate stage that accounts for the effect of interaural intensity differences is needed.

An important aspect of ICFs is that under certain assumptions from the short-time ICF the ITD can be estimated. When the ITD is constant for a total binaural stimulus (for example masker plus signal in detection experiments) its value corresponds to the position of the maximum. To that end the ICF must be known over the whole range of possible ITDs. When the stimulus ITD is varying in time, the form of the ICF and consequently the position of its maximum depends on the form and duration of the temporal window used to average the running ICF. Zurek (1981) showed that for the standard binaural detection situations  $N_0S_\pi$  or  $N_0S_m$  the total stimulus interaural phase difference and as a consequence also the ITD is a random variable with a mean value of 0 and a variance depending on the stimulus signal-to-noise ratio. Other binaural stimulus configurations have a corresponding behavior.

The size and the generation of interaural parameters resulting from different interaural stimulus configurations can be geometrically illustrated by vector diagrams suggested by (Jeffress et al., 1956; Jeffress, 1972). Vectors are used to illustrate the amplitude and phase of masker and signal for both ears at an instant in time (from a mathematical point of view they represent instantaneous amplitude and phase of the respective analytic signal). This is illustrated for an  $N_0S_\pi$  configuration in Figure 2.2. The vector representations are only precise for one moment or when tonal maskers are employed.

Since a noise masker varies stochastically in amplitude and phase, the interaural parameters become also stochastic variables. Zurek (1981) derived analytical expressions for the probability distributions of the interaural phase and level difference for different binaural stimuli from detection experiments employing Gaussian noise maskers. Unfortunately, the statistics of these interaural parameters cannot be directly translated into acoustic percepts like lateralization or detection for several reasons. For example, the distribution of interaural parameters allows no conclusion about the rate at which these parameters fluctuate and how it relates to the temporal resolution of the binaural auditory system. The fluctuation rate can be



**Figure 2.2:** Vector diagram representing instantaneous signal amplitude and phase. The dichotic noise ( $N_0$ ) is represented by one vector for both  $N_R$  and  $N_L$ . The  $S_\phi$  signal is presented with an interaural phase  $\phi$  ( $\phi = \pi$  in this illustration). The instantaneous phase difference for the total signal is  $\Theta$ , and the instantaneous amplitude ratio (which is an estimate for the interaural intensity difference) corresponds to the quotient of the vector lengths for the total right and left ear signal,  $|N_R + S_R|/|N_L + S_L|$ .

very different depending on the frequency of the target signal or the noise bandwidth (narrowband noise shows slow internal envelope fluctuations which also affect binaural detection). Following these arguments, interaural differences can not be directly connected with percepts since in auditory signal processing signal features are recoded or lost.

Nevertheless the consideration of interaural parameters can explain several binaural effects. For example Hafter (1971) in his “lateralization model” used the average of a weighted combination of interaural parameters. He could describe BMLDs for binaural configuration with interaural phase differences of the test tone and IIDs of the masker with this single trading variable.

Another binaural model which successfully describes BMLDs from a large class of binaural detection experiments is the *equalization and cancellation* (EC) model (Durlach, 1963, 1972). Since it is the basis for much of the current modeling this model is discussed in more detail here. The EC theory is a relative simple algebraic model which only uses a two parameters. Since it can be easily applied in many situations it is often used to estimate binaural detection data. Calculations are relatively simple since they are performed for signal parameters (like signal energies or signal-to-noise ratio) of unprocessed ear signals. The basic concept is to eliminate as much as possible of the noise masker energy in a binaural configuration in which masker and signal have different interaural relations. This happens by matching the

noise signal and subsequent subtraction of both ear signals. The target signal has to be less affected in the course of this operation than the masker which results in a gain in signal-to-noise ratio in the binaural difference signal relative to the monaural signal-to-noise ratio. This gain denotes the predicted BMLD.

In general the set of available equalization operations is restricted to internal time delays and amplitude adjustments to obtain reasonable predictions for as much as possible interaural configurations. The range of internal delays has to be limited. Two parameters of the model describe the internal noise in the equalization procedure which limits binaural performance. They are specified as random time jitter and random amplitude jitter and follow Gaussian distributions with zero means. They are assumed to be statistical independent from each other, statistical independent for each ear and independent of frequency. Both error types are applied to both ear signals before calculating their difference. In addition to binaural detection data, the model has been applied to describe data from interaural discrimination tasks and binaural “creation-of-pitch”.

The large advantage of the EC-model is that it can be applied in a straightforward manner without large computational expense. However, since the model in its classical form deals only with average signal magnitudes or statistical moments like signal-to-noise ratios or signal energy it does not predict more complex functions of the binaural system that play a role for time varying signal parameters (for instance occurring in forward masking experiments, test tone integration, time dependent interaural correlation). Since no assumptions are made about monaural signal processing (apart from the decomposition of the incoming signals into critical bands), there is neither an absolute monaural threshold included into the model nor a description of level dependence of the BMLD.

To account for the masker level within the model without assumptions about a threshold, Yost (1988) described the level dependence of the BMLD as an effect of decorrelation of the ear signals by internal, only partly correlated noise added from two independent noise sources. He started from the equation

$$\text{BMLD}(N_\rho S_\pi) = 10 \log \left( \frac{k+1}{k-\rho} \right)$$

from EC-theory describing the BMLD in dB as a function of interaural masker correlation  $\rho$ . In this equation  $k$  is the EC-factor, the interaural masker correlation  $\rho$  is expressed as a function of external *and* internal noise intensities.

## 2.3 Cross-correlation models

Several models based on cross-correlation functions have been described to specify features of auditory signal processing.

A first analytical model was developed by Sayers and Cherry (1957) to predict the lateralization of an auditory event and was based on the interaural cross-correlation

of the signal *waveforms*. It used the expression

$$U(t, \tau) = \int_{-\infty}^t L(t')R(t' - \tau)w(t - t') dt' \quad (2.2)$$

for the interaural correlation function. In this equation  $R$  and  $L$  are the *unprocessed* waveforms from right and left ear (no neural or internal representation). The function  $w$  describes the temporal weighting of calculating the correlation (short-term average). To treat IIDs Sayers and Cherry added a constant proportional to the intensity of each ear signal. Afterwards the integrals over the negative and positive values of  $\tau$  in Equation 2.2,  $I_L$  and  $I_R$  respectively are evaluated. They serve as input of a judgment mechanism that extracts the subjective lateral position of a binaural stimulus from

$$\hat{P} = \frac{I_L - I_R}{I_L + I_R}.$$

Blauert and Cobben (1978) used band-pass filtered, rectified and smoothed stimulus waveforms (which they assumed to be reflect the transduction properties of the peripheral system) as input to a model generating the running ICF. If recordings of ear canal signals are used as input into the model from the ICF the direction of a sound source could be inferred. Lindemann (1986a,b) included a mechanism for contralateral inhibition into the model of running cross correlation. This expanded model was intended to account for lateralization experiments in “time-intensity trading” experiments. It results in a sharpening of the peaks of the cross-correlation function compared to the output of a pure Jeffress mechanism. IIDs and ITDs result in a shift of the inhibited ICF. The model shows a kind of precedence effect if the contralateral inhibition is allowed to persist with a short time constant after the first triggering. One disadvantage of the Lindemann model is the large number of parameters describing the amount and persistence of inhibition. Gaik (1990, 1993) added attenuating elements to each step of the delay lines influencing the activity patterns traveling along each line. Through this the latency hypothesis is modeled since IIDs result in a shift of the temporal pattern away from the center. The weights of the damping elements are adjusted in such a way that, for natural combinations of IIDs and ITDs of recorded ear signals within one critical band, the processing results in zero “internal level difference” at the point with the internal delay matching the external delay. Bodden (1993) applied this model as a “cocktail-party processor” that segregates different sound sources and improves the signal-to-noise ratio in situations with different spatially distributed talkers.

A more physiological expansion of the Jeffress network and quantitative approach in describing binaural phenomena concerning detection and discrimination was formulated in a series of publications by Colburn (1973, 1977b,a) and Colburn and Latimer (1978). He included a quantitative model that generates descriptions of patterns of auditory-nerve activity which are the inputs to a central binaural processor. The transduction process from acoustical waveform to neural firing realized important aspects of signal transformation in the auditory system: information is

passed as firing times of each neural fiber, the transformation is stochastic (resulting in restricted performance of the central processor), fibers are frequency selective, firing times are synchronous to the temporal structure of stimulus and, with increasing stimulus intensity, the firing rate increases and saturates for high intensities. The model of auditory nerve activity consist of a bandpass and a lowpass filter, an exponential rectifier and a mechanism generating firing times. The firing times are the output of a non-homogeneous Poisson process whose rate is proportional to the input of the process.

The central processor in Colburn's auditory-nerve-based model acts as a coincidence counter which compares timing information from two fibers (one from each ear) with the same characteristic frequency  $f$  and a single, fixed internal delay  $\tau$ . A coincidence is only detected if two firing events occur nearly simultaneously within a short temporal window of length  $T_w$  (about  $50 \mu\text{s}$ ). With this assumption the output of the coincidence counter may also be treated as a Poisson process with the expectation value

$$E[L(\tau, f)] = T_w \int_0^T r_L(t') r_R(t' - \tau) dt'. \quad (2.3)$$

$L(\tau, f)$  is the number of coincidences for the unit which receives input as rate functions  $r_L$  and  $r_R$  of the stimulus length  $T$ . The expression for  $E[L(\tau, f)]$  in Equation 2.3 as a function of the internal delay  $\tau$  is the cross-correlation function of the neural representations of the ear signals. For small  $T$  it is also a special realization of calculating the running cross correlation function of the Poisson rate functions,

$$E[L(t, \tau, f)] = \int_{-\infty}^t r_L(t') r_R(t' - \tau) w(t - t') p(\tau) dt' \quad (2.4)$$

which depends additionally from running time  $t$  and corresponds to Equation 2.2 apart from the function  $p$ . This function is centered around a small range around  $\tau = 0$  and serves to restrict the range of effective internal delays that contributes to the output of the binaural processing ("centrality").

In its original form the model predicted that overall thresholds are lower than those observed. Colburn (1996a) related these differences to the fact that the model neglects the variability inherent in masker waveforms relative to the variability due to auditory noise statistics.

Modifications of the auditory-nerve-based model were implemented for example by Stern and Colburn (1978) to provide predictions for subjective lateral position of binaural stimuli.

It has been pointed out (Osman, 1971) that for class of correlation models and the common form of the EC model, which are based on the energy as binaural detection variable, it can be written in the same form (ignoring internal time delays)

$$A \int_0^T L^2(t) dt + B \int_0^T R^2(t) dt - 2C \int_0^T L(t) R(t) dt,$$

where  $L$ , and  $R$  are the (preprocessed) ear signals and  $A$ ,  $B$  and  $C$  are weighting factors. In the EC theory, the detection variable results from setting  $A = B = C = 1$ . In correlation models, these constants may be set  $A = B = 1$  and  $C = -1$ .

## 2.4 Conclusion

The examples in the previous sections show that all of the classical models or concepts of binaural signal processing suffer from several inconveniences in the description of psychophysical data. A general binaural model should also provide aspects of monaural auditory signal processing relevant for binaural processing. It is important to have monaural masked thresholds available as “fixpoints” in order to describe the thresholds in masked detection experiments correctly. Also, a correct absolute thresholds is required to restrict the dynamic range in which binaural unmasking can take place. A further important aspect is the temporal dimension of auditory signal processing. All input waveforms should be processed individually. This enables a model to account for inherent noise fluctuations and individual waveforms.

## Chapter 3

# Model of the “effective” binaural signal processing in the auditory system

### Abstract

*This chapter presents a quantitative functional model describing binaural detection experiments. The binaural model is the expansion of a model of the effective temporal signal processing in monaural configurations [Dau et al., J. Acoust. Soc. Am., 99, 3615–3622 (1996)]. It combines several stages of monaural preprocessing with a binaural processing unit and a decision device having properties of an optimal detector. The binaural processing unit follows concepts of the Equalization and Cancellation (EC)-Theory since it performs interaural temporal and amplitude adjustments on internal representations of the ear signals. The model was developed to overcome several restrictions and deficiencies of current models of binaural interaction. In contrast to the EC-model the present model operates on the monaurally preprocessed acoustical waveforms to profit from the properties of temporal and spectral processing of the monaural model.*

*“Critical experiments” employing conditions with broadband noise maskers have been chosen which permit an estimation and adjustment of model parameters and a statement about the structure of the model. The model accounts for the frequency dependence of thresholds in diotic and dichotic configurations with stationary interaural phase differences of signal and masker. In addition, it gives good predictions for thresholds and BMLDs for configurations where a dichotic noise masker with stationary interaural time differences is presented. To deal with the large input range of masker levels and masker IIDs, the compressive behavior of the model is important which has been adapted from the monaural model. Using these elements, the dependence of binaural masked thresholds on masker level and interaural intensity differences is predicted as well as the behavior of thresholds in binaural forward masking experiments.*

### 3.1 Introduction

Binaural processing cannot be treated strictly separated from monaural processing. Physiologically it takes place on a more central level and follows at least some stages of peripheral monaural processing. Therefore a binaural model should incorporate some basic features of monaural signal processing (frequency decomposition, transduction, adaptation, compression and internal noise). However, despite their ability to predict the results for large classes of binaural experiments, almost all of the established binaural models disregard basic monaural features (e. g. they do not predict absolute thresholds or temporal effects in detection experiments). Therefore a combination of principles from monaural *and* binaural signal processing promises an improvement in modeling and understanding of binaural phenomena. This is the aim of the present study.

Two important basic limitations in binaural performance directly result from properties of the monaural auditory processing: (a) the restriction of binaural unmasking to frequencies below approximately 2 to 4 kHz (according to studies using a broadband masker, Hirsh (1948); Hirsh and Burgeat (1958); Webster (1951); Schenkel (1964); Rabiner et al. (1966); Kohlrausch (1984)), and (b) the ability to compensate for interaural time delays of ear signals for not more than 8 or 10 milliseconds. These limitations have to be considered in the description of the binaural signal processing by any model.

The first limitation (i. e. larger BMLDs for low than for high frequencies) originates both from limitations in binaural processing itself (like it is described and employed in the EC theory, Durlach (1972)) and from several possible monaural mechanisms (e. g. Zurek and Durlach, 1987). The first is the *transduction process* from sound signals to neural excitation by the inner hair cells in the inner ear. The transduction process is performed with a decline in the strength of phase locking with increasing stimulus frequency (Ruggero, 1992). Above frequencies of roughly 1 kHz, single inner hair cells cannot follow the phase of the acoustic stimulus. Neural assemblies can follow the fine structure at frequencies above 1 kHz. Hence, neural coding turns gradually from coding of signal waveform to signal envelope. At the level of the auditory nerve, this loss of information in fine structure coding removes interaural time differences which are present in the waveform. The binaural system therefore can make use only of interaural intensity differences in the envelope of the stimulus.

The second factor responsible for the decrease of BMLD with frequency results from the *frequency dependence of the critical bandwidth*: with increasing frequency, the bandwidth of the critical bands increases which causes the effective bandwidth of a broadband masking noise to increase with the center frequency of the respective critical band. Since the maximum rate of fluctuation of a noise band is proportional to its bandwidth, the maximum rate of fluctuation of the output of the auditory filters are increasing with increasing frequency. Subsequently also interaural time and intensity differences for that frequency are fluctuating. Assuming a reduced

sensitivity of the auditory system (or at least of the binaural system) for fast fluctuations these fast varying interaural differences would reduce binaural unmasking (Zurek and Durlach, 1987).

Similar to the restriction on the BMLD imposed by “monaural” and “binaural” properties, the ability of the auditory system to account for interaural time differences (ITDs) is a consequence of monaural and binaural restrictions. For the binaural system, it is assumed that its ability to perform an internal delay of the ear signals is restricted to a certain range. External ITDs beyond this range can still be handled by performing a suboptimal internal delay too small to match the actual external delay. However, the effective correlation time of a noise masker for a tonal signal is restricted by the critical bandwidth of the peripheral filter centered at the signal frequency<sup>1</sup>. Performing an interaural time shift of the noise masker results in a gradual loss of interaural correlation. Small ITDs of noise maskers behave like ITDs for sinusoids since bandpass filtered noise signals can be treated approximately as an amplitude modulated sinusoid. Applying an interaural time delay of the inverse of the center frequency of an  $N_0$  narrow band signal results again almost perfectly in an  $N_0$  configuration. This is reflected in the shape of the autocorrelation function which shows the same periodicity as the bandpass signal and is damped with increasing  $|\tau|$ .

One further important cue in binaural auditory perception are interaural level or intensity differences (IIDs), that, together with interaural time differences (ITDs), form the physical basis of binaural interaction and tasks like localization and binaural noise reduction. This robustness is also reflected in masking configurations like  $N_{\Delta L}S_m$ . In these type of detection experiments, the role of IIDs is destructive, that means that large level differences reduce BMLDs, but the ear is able to deal with a large range of level differences (up to 30 or 40 dB). To describe effects of IIDs in binaural detection, it is necessary to understand monaural or diotic effects.

Fassel (1989) and Kohlrausch and Fassel (1997) discussed results from forward masking experiments in  $N_0S_0$  and  $N_0S_\pi$  configurations in terms of the relative order of monaural adaptation and binaural processing in the auditory system.

A monaural model of temporal auditory signal processing which is able to predict most of the temporal and spectral aspects of monaural auditory signal processing (such as e. g. simultaneous and forward masking and temporal integration) has been proposed by Dau et al. (1996a). It is described in the first part of the following section. The combination of this psychoacoustically motivated preprocessing unit with successful ideas of binaural signal processing (EC-theory) promises synergistic effects and will remove some of the shortcomings of common binaural models. The separation of the total model into different units permits to distinguish the influence

---

<sup>1</sup>The shape of the autocorrelation function for each band limited signal is related to the form of the respective bandpass filter (i. e. the spectrum level of the signal) by the Wiener-Khinchin theorem. It states that the Fourier transform of the autocorrelation function of a signal equals the spectrum level of the signal. Following the uncertainty relation for the Fourier transformation, a signal with broader spectrum has a shorter correlation time.

of monaural and binaural processing on the performance in signal detection. The expansion of the monaural model is tested by simulating its effects on binaural processing and testing the limits that such an approach has.

To do so, simulations of several “critical” experiments are performed and presented in the present chapter. This permits to test the influence of single features of the model structure on the signal processing in the model and to adjust the model parameters associated with these properties. The “optimal” parameters are kept constant in further simulations. These critical experiments concern the influence of basic properties of monaural spectral processing and internal adjustment of external time and level differences. To preclude the use of off-frequency cues in binaural detection, only simulations and corresponding experiments with stochastic broadband noise maskers (noise bandwidth chosen larger than the critical bandwidth) were considered in this chapter.

The following experiments and related aspects of signal processing, concerning topics already mentioned above, have been chosen as suitable for that purpose:

- The frequency dependence of binaural masked thresholds and of the corresponding BMLDs allows an estimation of the internal error immanent in binaural noise reduction of the model
- The dependence of thresholds in configurations with an interaural time shift of the noise masker is used to restrict the range of internal time delays, which are available to the binaural processor
- The dependence of masked thresholds from the level permits to estimate the level of the internal (peripheral) noise and to investigate the influence of compression before binaural processing in the auditory system on BMLDs
- Simulating experiments with stationary IIDs reveal the importance of compression on processing the interaural level differences
- Simulations of binaural and monaural forward masking experiments allow to test the structure of the model in configurations where the role of the order of different stages of signal processing reveals. It confirms experimental findings that adaptation and compression in the auditory system take place before or after binaural processing.

Finally, possible alternative realizations of the signal processing are considered and investigated.

## 3.2 Description of the signal processing in the model

### 3.2.1 The original model - monaural preprocessing unit

In the work of Dau and coworkers (Dau et al., 1996a,b), a quantitative functional model for the “effective” signal processing in the auditory system was described. This computational model allows the prediction of masked thresholds for various monotic and diotic stimulus configurations<sup>2</sup>. The model was originally designed to describe temporal aspects of masking in normal hearing subjects. Recently it has been expanded to describe experiments in modulation detection and modulation discrimination (Dau, 1996; Dau et al., 1997a,b). It has been used as a preprocessing unit for speech intelligibility and speech quality prediction (Hansen, 1998) as well as a front end for robust speech recognition (Tchorz and Kollmeier, 1999). It is also able to describe temporal and compressive properties in the impaired auditory system (Derleth, 1999), where it is used to simulate the performance of subjects with a sensorineural hearing impairment in detection and discrimination experiments. Predictions can be made for a large class of signal properties, such as temporal, spectral and statistical characteristics.

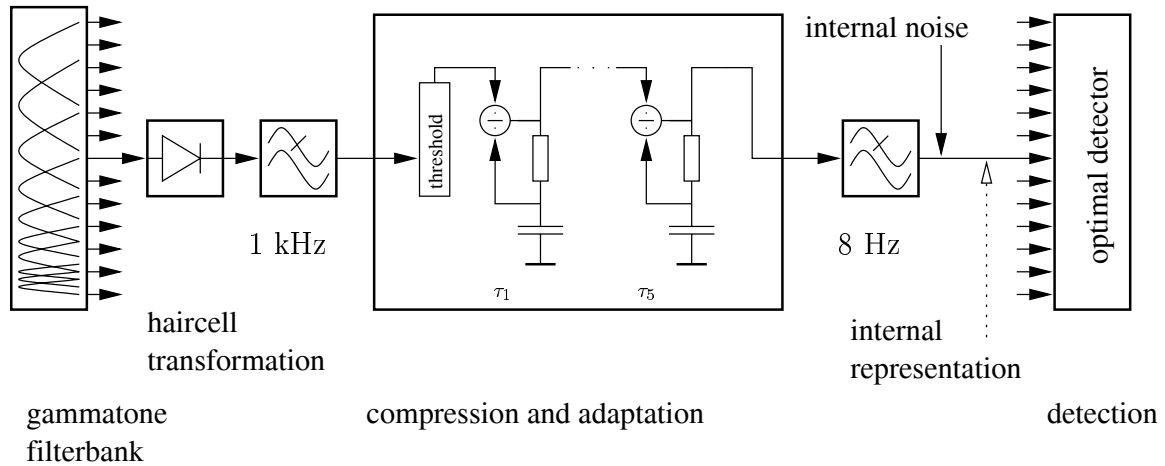
The model combines several stages to simulate aspects of signal transformation in the auditory periphery and “higher” stages with an optimal decision device to describe the decision process within the framework of signal detection theory. Signal processing transforms input signals to an “internal representation”, i. e. a two dimensional time-frequency representation of acoustic stimuli. It mimics the transformation of acoustic stimuli to neural excitation which can be interpreted as input for higher cognitive processes (for example detection or speech perception). A block diagram of the model is given in Figure 3.1.

The first stage in auditory processing consists of a gammatone filterbank of Patterson et al. (1987) describing the bandpass characteristic of the basilar membrane. The 4th order linear bandpass filters have been optimized to fit the data derived from masking experiments using a notched-noise paradigm (Patterson and Moore, 1986; Glasberg and Moore, 1990). The *equivalent rectangular bandwidth* (ERB) of the filters for a give center frequency is chosen according to the equation of Glasberg and Moore (1990) which relates the ERB  $b$  of an auditory filter and the center frequency  $f_0$  (both in Hz):

$$b = l + \frac{f_0}{Q}, \text{ with } l = 24.7 \text{ Hz and } Q = 9.265. \quad (3.1)$$

---

<sup>2</sup>In general the stimuli in corresponding experiments have been presented diotically. The configurations are often referred as monaural experiments, the model will be referred as a monaural model. As a rule of thumb one can state that in detection experiments differences between monaural and diotic presentation are only observed when the stimulus time structure is reproducible (for instance for frozen noise maskers (Langhans and Kohlrausch, 1992)) or for low masker levels (for example observed in Kohlrausch (1984)).



**Figure 3.1:** Block diagram of the monaural psychoacoustical model of auditory processing as described in Dau et al. (1996a,b). The signals are processed with a gammatone filterbank, subjected to adaptation and lowpass-filtered in each peripheral filter. Finally, internal noise is added. The decision device is implemented as an optimal detector.

For simulations of masking experiments with broadband maskers, it is sufficient to consider only the on-frequency filter since off-frequency information is not advantageous for detection: Signal energy which is spread out to adjacent auditory filters is masked by the broad-band noise. Some exceptions occurring in binaural configurations are discussed in the next chapters. If the output of multiple auditory filters has to be considered in simulations, the gammatone filters are arranged along the frequency axis such that their respective amplitude transfer functions overlap at the 3 dB-points.

Subsequently the output signal of each filterbank channel is half-wave rectified and low-pass filtered with a cut-off frequency of 1 kHz. This stage accounts roughly for the transduction process of the organ of corti from mechanical motion of the basilar membrane to neural excitation of the auditory nerve. A relative shear movement between the tectorial membrane and the organ of corti caused by basilar membrane motions results in bending the tops of the inner hair cells (IHCs). The IHCs are sensitive for motions in one direction, which is modeled by the half-wave rectification. For frequencies larger than about 1 kHz the firing process of the auditory nerve fibers can not follow the fine structure of the acoustic stimulus. So with increasing frequency, the representation in the auditory nerve turns from coding the fine structure of the waveform to envelope of the stimulus (Gulick et al., 1989; Pickles, 1988). This is described by a low-pass filter with a cut-off frequency of 1 kHz. Altering properties of this low-pass filter seems to have little effects on the description of

monaural thresholds. The influence on binaural detection will be discussed later.

In a subsequent stage of the model, five adaptation loops connected in series describe effects of nonlinear adaptation and dynamic compression. They have been initially introduced by Püschel (1988) to describe forward masking data. The input signal to the first adaptation loop is restricted to a fixed minimum value. This operation is performed in order to keep the dynamic system numerical stable and is interpreted as the absolute detection threshold for normal hearing listeners. Derleth et al. (1996) have shown that with an appropriate selection of that value absolute thresholds for impaired listeners can be described as well. Each of the five consecutive adaptation loops consists of a low-pass filter and a dividing element. The time constants of the low-pass filter range from 5 to 500 ms. The input of each loop is used as dividend to the dividing element. Its output is low-pass filtered and fed back as divisor. The output signal in a stationary condition for one single loop is the square root of the input signal. For the chain of five loops the output  $O$  for a stationary input  $I$  is  $O = \sqrt[5]{I} = \sqrt[32]{I}$ , which is close to a logarithmic compression of the input. The logarithmic compression is important to predict a variety of psychoacoustic data with the same model. Hence, in combination with an additive internal noise limiting detection it provides that the model accounts for Weber's law, i. e., a constant resolution of signal amplitude is provided on a logarithmic scale in discrimination experiments.

For non-stationary signals, fast signal fluctuations (compared with the time constants of the low-pass filters which constitute the adaptation loops) are processed approximately linearly, while slowly fluctuating input signals are compressed. The amount of compression depends on the time constants and the charging state of the adaptation loops, i. e., the input signal at previous times.

To map the level range of stationary input signals (from 0 to 100 dB) into an equal range of output signals (from 0 to 100 "model units", MU) the output of the fifth adaptation loop is scaled linearly. Due to adaptation effects (such as, e. g. overshoot for signal onsets or offsets) the value of output signals may be clearly beyond 100 MU if the input signal is non-stationary.

After the adaptation stage, the signal is low-pass filtered with a time constant of 20 ms that corresponds to a cut-off frequency of 8 Hz. In combination with the subsequent detection unit this stage accounts for effects of temporal integration of the model. It also limits of temporal resolution for signal and envelope fluctuations respectively that is available for the subsequent "optimal" processor. The time constant was optimized to account for temporal resolution and integration in detection and masking experiments (Dau et al., 1996b).

## Modifications

The idea of the present study is to preserve the ability of predicting monaural thresholds by not changing the fundamental principles of the monaural model. However, slight changes had to be introduced to improve the performance of the model for

signal levels close to absolute threshold:

The transfer behavior of the ear canal and the middle ear for sound signals at low frequencies (before basilar membrane filtering takes place) shows a high-pass characteristic (e. g. (e. g. Hudde, 1998; Hudde and Engel, 1998)). To account for absolute thresholds in the low-frequency region, a first-order high-pass filter with a cut-off frequency of 1 kHz is introduced to the model before the gammatone filterbank.

Furthermore, uncorrelated noise of a low level (0 dB in each frequency channel) is added in each frequency channel of the left and right part of the model after the auditory filters in order to obtain the appropriate binaural performance at low levels. As a consequence, the monaural absolute thresholds will be slightly increased (about up to 5 dB) due to this additive noise. This “peripheral” noise can be interpreted as physiological internal noise and may originate from mechanical, neural or physiological sources (Shaw and Piercy, 1962; Soderquist and Lindsey, 1972). The noise can be neglected as long as the masker energy within one critical band is relatively high (about more than 40 dB). When the noise level is decreasing, the peripheral noise starts to influence the monaural and binaural thresholds. The uncorrelated noise from two corresponding auditory filters of both ears cannot be removed by the subsequent binaural processor (“internal decorrelation”) and restricts its efficiency at low masker levels. For monaural signals, it is an additional simultaneous masker. The absolute threshold in the model is no longer determined by the minimum input value to the adaptation loops, but by the middle ear model and the peripheral noise. By taking into account both modifications, the behavior of the model for low external masker levels is more appropriate than the original model version (Dau et al., 1996a) which does not account for middle ear effects and includes a less accurate description of the absolute threshold in quiet.

### 3.2.2 Optimal detector

The decision device realizes concepts from signal detection theory (SDT) (Green and Swets, 1988) and is implemented as an optimal detector or cross correlation template matching process. In Dau (1996) the principles of optimal detection processing in the model and the formulation of the decision problem in an  $m$ -Interval Forced-Choice ( $m$ AFC) task are discussed in detail. Here only a short outline of the implementation of the detector is given.

The detector in the present study has been constructed to detect changes in a presented signal (caused by the signal in the masker background). The normalized difference between the internal representations of masker plus suprathreshold signal and the masker alone is calculated and serves as a template for the pattern to be detected. Within the simulation of the measurement procedure, this pattern is compared with differences derived in the same way from the actual signals. To do so, for each presentation the difference is calculated between the internal representations of the actual signal in the masker and a representation of the masker alone signal which

is averaged over several realizations. The comparison is performed by calculating the *unnormalized correlation product* of detector signal and the actual signals. This is the implementation of a matched filter according to statistical decision theory. It is constructed such that it is optimally fitted to the signal in Gaussian distributed noise in a sense that it yields the best signal-to-noise ratio at the output of the filter.

From the correlation product the correct response probability is derived. This is performed according to SDT assuming an additive internal noise of constant level and a Gaussian amplitude distribution of the representations of the acoustical stimuli in the model<sup>3</sup>. The level of the internal noise is chosen such that it restricts the resolution of the model according to a 1-dB-criterion in amplitude discrimination.

### 3.2.3 The binaural processing unit

To expand the monaural model to binaural input, two equal channels describing monaural signal processing in the left and right ear are required. The binaural processing stage receives input from the two monaural frequency channels (peripheral filters) from the left and right ear with the same characteristic frequency. No interaural across-frequency interaction is considered here. The binaural processor facilitates interaction between both preprocessed signals and is located between the both adaptation and compression stages and the temporal integration low pass filters of the two monaural channels. A block diagram of the arrangement of the different signal processing modules is given in Figure 3.2.

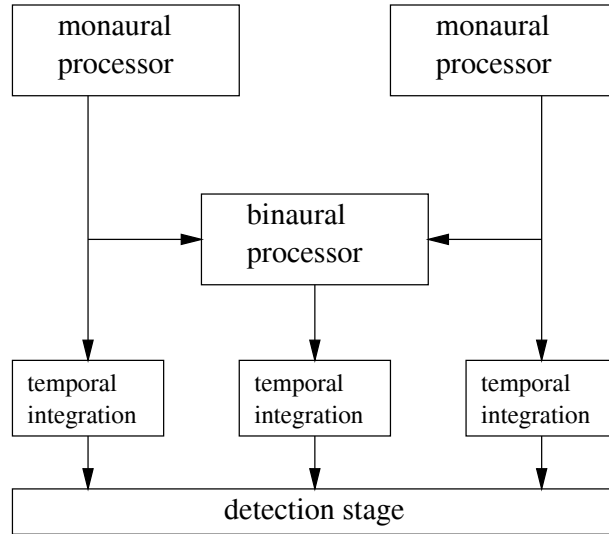
Binaural processing itself takes place in three steps after the adaptation loops and before temporal integration (a detailed block diagram of the signal processing within the complete model is given in Figure 3.3):

1. If there exists a (stationary) interaural level difference of the masker signals in the left and right channel, the signals are adjusted by internal interaural gain factors.
2. The signal in one channel is delayed or advanced relative to the signal of the respective other channel. The advance of the signal corresponds to a delay of the other channel and assumes a symmetric behavior of the model with respect to the left and right ear. This internal time shift is denoted as “internal delay”.
3. Both signals are subtracted from each other and the resulting difference signal is full-wave rectified.

The first two steps correspond to the equalization process in the EC-theory (Durlach, 1972).

---

<sup>3</sup>Two reasons may justify a Gaussian distribution of the noise. First, mathematical, the central limit theorem and second, for practical reasons, that Gaussian distributed variables are easy to handle in calculations.



**Figure 3.2:** Block diagram of the modular arrangement of signal processing units in the complete binaural model.

For the adaptive steering of binaural signal processing, realizations of the interaural cross correlation functions (ICFs) of monaurally preprocessed signals are calculated from the output the adaptation loops, i. e. the input for the binaural process. Multiple realizations are averaged to obtain a representation of the ICF which allows a correct estimate of the external ITD. This averaging is necessary because of the fluctuations of the external noise. Alternatively, only a long-time average over a representation longer than the test signal would result in a correct estimate for the external delay. The calculation is performed for the masker alone signals and for the masker plus suprathreshold test signals. Only the temporal segments of both ear signals are selected that contain the test signal (if present). Both ICFs are restricted to a limited temporal range of  $\pm 4$  ms. The patterns are flanked with 1 ms Hanning ramps to exclude edge effects in the later use of the ICFs. These both representations of interaural cross correlation functions are stored for later use.

In the first step of binaural noise reduction, for each auditory filter  $j$ , the preprocessed signals  $R_j$  and  $L_j$  from both ears are multiplied with the internal gain factors  $r_q, l_q$ . These factors are chosen in such a way that the root mean square (RMS) of the masker alone signals within both ears are equalized. This is done under the constraint that the geometric mean of the RMS values of the scaled signals remains unchanged (see Appendix A.1). In such a way the adjustment does not track the fast fluctuating interaural differences in signal level. The calculation results in

$$r_q = \sqrt{\frac{\text{rms}(L_j)}{\text{rms}(R_j)}} =: q, \quad l_q = \sqrt{\frac{\text{rms}(R_j)}{\text{rms}(L_j)}} = \frac{1}{q}, \quad (3.2)$$

where the RMS value of left and right ear signal is calculated over the entire length

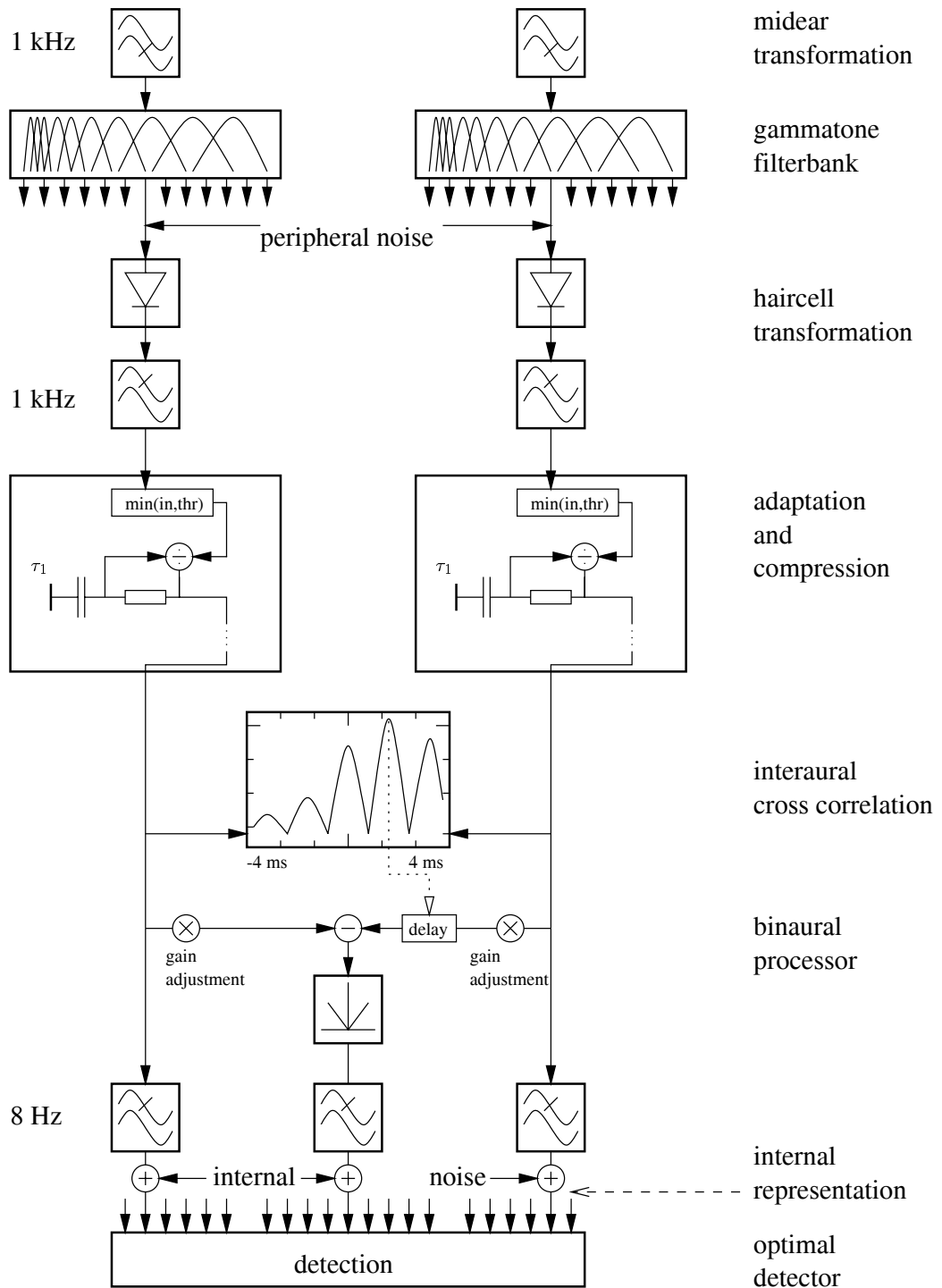


Figure 3.3: Block diagram of the signal processing in the binaural perception model

of a preprocessed noise alone interval. In the present model, due to the compression performed in the adaptation stages, the dynamic range of both ear signals becomes relatively small compared to the dynamic range of input signal (approximately logarithmic compression of stationary signals). For that reason the quotient of the RMS values of left and right ear signal (and its square root) does not show large deviations from unity compared to the IID of the input signals. Examples of occurring values of  $q$ -factors will be given later.

The internal interaural delay  $\tau$  between right and left ear signal is determined in the second step. The size of this internal delay is determined from the stored interaural cross correlation functions. Note that the range of possible correlations of input signals (acoustical waveforms)  $\rho_{\text{in}} \in [-1, 1]$  is mapped to a different range of correlations of output signals  $\rho_{\text{out}} \in [0, 1]$ . This happens mainly due to half-wave rectification in peripheral transduction which leaves only non-negative signals<sup>4</sup>.

The model can choose between two alternative strategies to determine the internal delay  $\tau_{\text{max}}$  between the right and left ear signals. The delay is chosen according to the temporal position of the largest (local) maximum of the correlation function for the noise alone signals if this maximum is large (close to 1). In this situation, the noise in the binaural channel will be canceled as completely as possible with one fixed internal delay (compare Section 2.2, page 9). If, on the other hand, the coherence of the masker waveforms is small, the maximum value of the cross correlation function of the preprocessed noise alone signals will be small and this strategy will fail to cancel a significant portion of the noise. Therefore, a strategy to maximize the test signal in the binaural channel is used in this case. This is achieved by choosing the temporal position  $\tau_{\text{min}}$  of the minimum of the cross correlation function of masker plus suprathreshold signal to determine the internal delay in binaural processing. Since the signal is presented at a relatively high level it dominates the ICF. Therefore one can assume that *after* performing the internal delay the desired signal is approximately antiphase. If such a strategy is assumed, the (sinusoidal) test signal from both ear channels is preserved during the cancellation process, since the half-waves transduced in peripheral processing do not interfere destructively<sup>5</sup>. Since the temporal range of the ICF is restricted to  $\pm 4$  ms, this is also valid for the range of available delays. The local maximum of the restricted interaural cross correlation function can differ from the global optimum for frequencies of roughly below 160 Hz (e. g. in  $N_\pi$  configurations). For these frequencies the first maximum (minimum) of the ICF may be out of the range of available internal delays.

If the equalization process described above would operate in a perfect manner this would result in the following output signal of the binaural processor after the

---

<sup>4</sup>Due to a small amount of decorrelating noise, which is always present in both channels, the “perfect” values 0 and 1 for interaural correlation will not occur for any input signals. Furthermore, the value 0 cannot be reached because of the hard threshold criterion which restricts input to the adaptation stages to a minimum value (DC-offset).

<sup>5</sup>This is exactly true as long as effects of nonlinear adaptation are ignored. Due to nonlinearity and time constants of the adaptation loops this will only be valid approximately.

third (“cancellation”) step:

$$B_j(\tau, t) = |r_q R_j(t) - l_q L_j(t - \tau)| = |q R_j(t) - \frac{1}{q} L_j(t - \tau)|. \quad (3.3)$$

To restrict the effectiveness of the equalization process, the resolution of the temporal fine structure in the binaural channel is reduced. The inexactness in processing is introduced as an uncertainty in the internal temporal alignment of the two ear signals. This uncertainty is implemented similar to the temporal random jitter error in the EC-theory. To calculate the output of the binaural processor, the values of the time delay error caused by “jitter” are taken into account by Gaussian weighted averaging over all uncertainty:

$$B_j(\tau, t) = \exp\left(-\frac{\tau^2}{2\sigma_{\text{eff}}^2}\right) \int \exp\left(-\frac{\tau_{\text{err}}^2}{2\sigma_{\text{err}}^2}\right) |(r_q R_j(t) - l_q L_j(t - \tau - \tau_{\text{err}}))| d\tau_{\text{err}} \quad (3.4)$$

with the optimal internal delay  $\tau$  and the corresponding uncertainty value  $\tau_{\text{err}}$  from a temporal environment around  $\tau$ . The value of the standard deviation  $\sigma_{\text{err}}$  of the uncertainty of temporal adjustment from EC-theory of  $105 \mu\text{s}$  turned out to be a reasonable choice also within the present study (see Section 3.4.1). After calculating the absolute value of the difference signal, the resulting binaural signal is scaled according to the size of the performed internal delay by a Gaussian weighting factor with a standard deviation of  $\sigma_{\text{eff}} = 1 \text{ ms}$ , that means that large delay are weighted less. Since before the detection stage an additive internal noise of constant level is assumed to limit detection, this makes large internal delays less effective (or more noisy). The delay-dependent scaling factor results in a decreasing signal-to-noise ratio with increasing internal delay.

The binaural mechanisms works independently from frequency, because there are no frequency-dependent parameters in its formulation. Nevertheless the results can be frequency-dependent.

Altogether the following features of signal processing limit binaural noise reduction so that no perfect noise reduction is performed:

1. In monaural processing preceding binaural processing, the sum of masker and signal within each ear is processed in a nonlinear way (i. e. half-wave rectification, adaptive compression). On the other hand the difference operation performed in the binaural processing stage is simply linear. Therefore it can no longer cancel out the noise in the input signal completely, even if this would have been possible for the unprocessed input signals (acoustical waveforms), e. g., in an  $N_0 S_\pi$  condition.
2. For any given interaural relation of masker and signal the set of equalization operations or internal interaural transformations available to the binaural system is limited to internal interaural gain and delay. These are suboptimal operations for configurations when it is required to compensate for other

interaural differences than fixed ITDs and IIDs. Examples of such configurations are those with an uncorrelated noise masker ( $N_{\rho=0}$ ) or a masker in an antiphasic condition ( $N_{\pi}$ ) in which an addition of both signals would be optimal.

3. Imperfections in the processing caused by the jitter error leave a larger portion of the noise masker in the binaural signal than an exact, “optimal” delay like given in Equation 3.3.

Equations 3.3 and 3.4 describe functions of running time  $t$  and the internal time delay  $\tau$ . Equation 3.3 additionally depends on the internal gain factor  $q$ , the function in Equation 3.4 furthermore on the parameter describing the standard deviation of the uncertainty error,  $\sigma_{\text{err}}$ .

A generalized model of binaural detection should operate on a two-dimensional time varying pattern of all admissible  $q$ 's and  $\tau$ 's. This could be realized as a combination of all binaural signals formed according to *all* maxima and minima of the ICF. The model in the presented form is restricted to one special combination primarily in order to save computation time. Similar to the simulation of monaural experiments (where a preselection between different frequency channels has to be made) it is assumed that only the channel with the best signal-to-noise ratio has to be considered. For a low-frequency situation there are only one or two relevant maxima or minima with a position close to zero so that selecting only the most prominent peaks (or valleys) is certainly justified. For high-frequency configurations this assumption has to be verified by simulations of corresponding experiments. However, in all configurations due to compression the dependence of the signal in the binaural channel on a special gain factor  $q$  across different conditions should not be very strong. In addition, the evaluation of interaural time differences (represented by the appropriate choice of  $\tau$ ) is primarily performed in the low-frequency region. Hence, we can assume that the current strategy of selection only one combination of  $q$  and  $\tau$  for each frequency is a sufficient approximation also at high frequencies.

For the binaural processed signals a pattern of masker alone and masker plus suprathreshold test signal is calculated in the same way as for the monaural channels and the difference is stored as detector signal.

In comparison to Dau et al. (1996a) the optimal detection stage is slightly changed to work within the expanded model. For both monaural channels and the binaural channel the correct response probability is calculated separately. The detector uses the largest of the probabilities to decide whether the signal in the actual trial has been detected or not<sup>6</sup>.

---

<sup>6</sup>It is also imaginable to construct a total decision variable from a combination of monaural and binaural channels. This possibility has not been tested since the implementation using separate decision variables for monaural and binaural channels already provided sufficient predictive power.

## 3.3 Experimental method

### 3.3.1 Procedure and subjects

Data from binaural detection experiments and from corresponding simulations are presented. Experimental data is partly from own measurements described below, partly from the literature.

Binaural masked tone thresholds were measured (when measurements were performed) and simulated using an adaptive, three-interval forced-choice (3IFC) procedure. The noise masker was presented with defined interaural parameters in three consecutive intervals separated by pauses of 400 ms. In one randomly chosen interval the test-tone was added. The other intervals were leaved unchanged. The subject's task was to specify the interval containing the test tone. During the course of a threshold run, the level of the test tone for each trial was determined according to a 1-up 2-down algorithm (Levitt, 1971) which results in an estimate of the test-tone level necessary for 70.7% correct responses. The step size was 8 dB at the start of the run and divided by two after every two reversals of the test-tone level until the step size reached a minimum of 1 dB, which was held constant through the remainder of the track. Thresholds were defined as the median value of the last eight reversals obtained with this minimum step size. Trial-by-trial feedback was presented during the measurements. The procedure was repeated at least three times for each signal configuration and each subject. All figures show the median and interquartile ranges based on these measurements. All subjects had experience in psychoacoustic measurements and had normal hearing and were aged between 24 and 30 years.

The data of the first binaural detection experiment has been taken from literature (Holube et al., 1995). The second experiment was performed within the framework of this study. Data from the remaining experiments is also from the literature (Kohlrausch, 1984; McFadden, 1968; Fassel, 1989).

### 3.3.2 Apparatus and stimuli

The acoustic stimuli were digitally generated using a sampling frequency of 32 kHz and transformed to analog signals with the aid of a two-channel 16-bit D/A converter. The stimuli were low-pass filtered and dichotically presented via headphones (Stax SR-A headphones using a Stax headphones preamplifier) in a soundproof booth. The generation and presentation of the signals was performed and controlled by a SUN-Workstation using a signal-processing software package developed at the Drittes Physikalisches Institut in Göttingen.

The masker samples were obtained by randomly selecting a segment from a 2-s bandpass-noise buffer. The buffer containing band-limited noise was created in the frequency domain by generating a flat amplitude spectrum within the passband and randomizing the phases. The noise buffer of 2 s was obtained after an discrete inverse Fourier transformation.

## 3.4 Results: Basic measurements and simulations on binaural processing

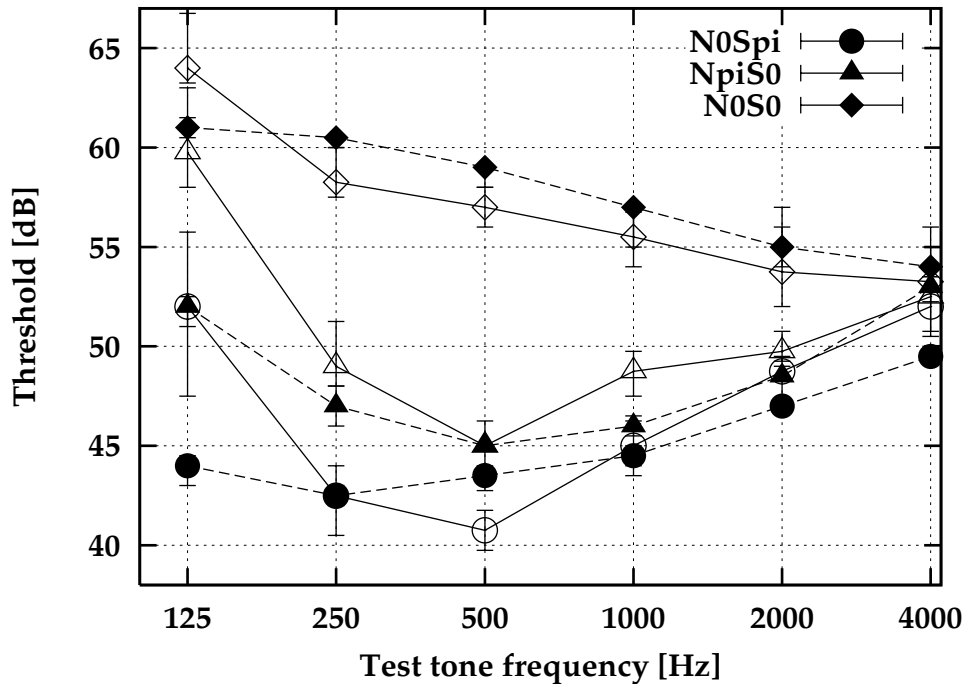
### 3.4.1 Frequency dependence of binaural thresholds and BMLD

An experiment performed by van de Par and Kohlrausch (experimental data published in Holube et al. (1995)) was simulated with the approach described above. The duration of the test tone was 300 ms and it was temporally centered in the 400 ms noise masker. The test tone was presented at 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, and 4 kHz, respectively. The overall level of the noise masker in both measurements and simulations was 70 dB SPL. Its center frequency was equal to the signal frequency and its bandwidth was twice the test signal frequency. So the spectrum level and bandwidth of the masker were dependent of the test-tone frequency. For all signal frequencies the masker bandwidth was broader than the critical bandwidth of the test-tone frequency. Four normal-hearing subjects participated in the experiment. The figures show the central values and interquartiles range of the averaged data of 4 subjects (open symbols, each measurement was repeated 4 times per subject) and 9 simulations (filled symbols), respectively.

Figure 3.4 shows the measured as well as the simulated test-tone levels at threshold for the configurations  $N_0S_0$ ,  $N_\pi S_0$ , and  $N_0S_\pi$  as a function of target signal frequency. The corresponding BMLDs for the two dichotic configurations are plotted in Figure 3.5. Since the overall level of the noise masker was kept constant, the spectral level of the noise decreases with increasing center frequency. The critical bandwidth at the test-tone frequency is increasing with center frequency but this compensates only partly for the decreasing spectral level of the noise masker. Hence, the energy passing the auditory filter at test signal frequency decreases with increasing frequency. As a result, the  $N_0S_0$  thresholds decrease with increasing frequency with a rate of roughly 1.25 dB/Octave for frequencies higher than 250 Hz. This is well in line with predictions from signal detection theory, which predicts a rate of 1.5 dB/Octave (e. g. de Boer, 1966) assuming that peripheral filters have a constant quality factor.

At low frequencies the respective thresholds in both configurations  $N_\pi S_0$  and  $N_0S_\pi$  decrease up to 500 Hz and increase again for higher frequencies. The difference between thresholds in  $N_0S_\pi$  and  $N_\pi S_0$  decreases with increasing frequency. For all conditions (monaural and binaural) approximately the same threshold was measured at a signal frequency of 4 kHz. This results in a diminishing BMLD at this frequency (cf. Figure 3.5).

Except for a slight threshold shift of 2 to 4 dB,  $N_0S_0$  thresholds in simulations follow the behavior of experimental data. The predicted binaural thresholds and corresponding BMLDs follow qualitatively the experimental data, but with some exceptions. Starting from the minimum of measured binaural thresholds curves



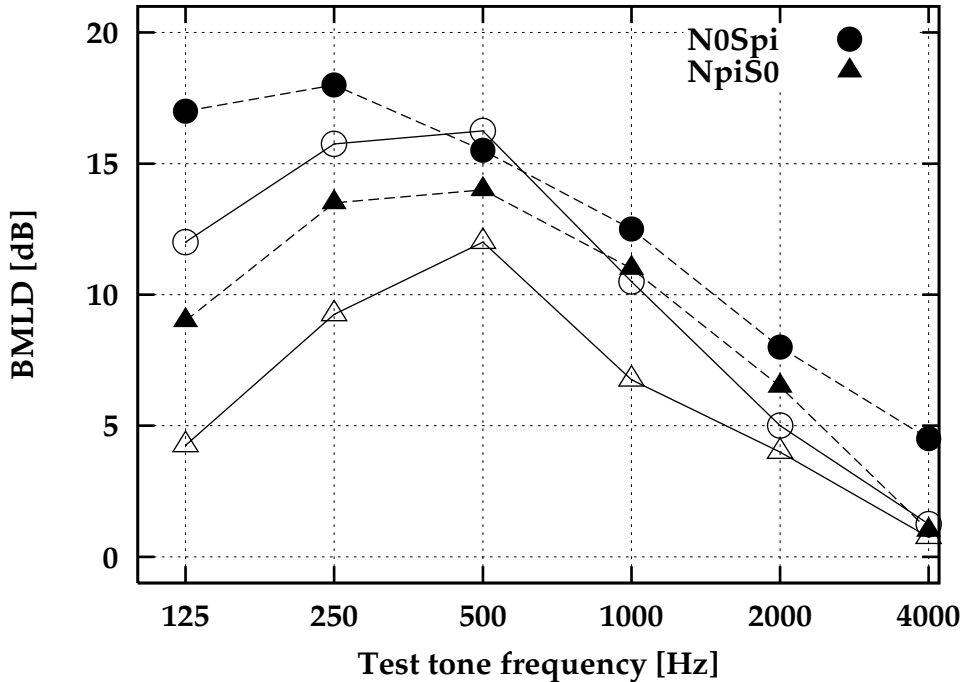
**Figure 3.4:** Thresholds for the configurations  $N_0S_0$  ( $\diamond$ ),  $N_\pi S_0$  ( $\Delta$ ) and  $N_0S_\pi$  ( $\circ$ ) as a function of the test-tone frequency. Open symbols mark data from measurements obtained from Holube et al. (1995), filled symbols data from simulations.

at 500 Hz, predicted thresholds increase slower than measured thresholds both for increasing and decreasing signal frequencies as well. In simulations the threshold for the  $N_0S_\pi$  configuration still decreases for frequencies lower than 500 Hz, and reaches its minimum at 250 Hz. This holds also for the corresponding BMLD which reaches the maximum. For 125 Hz it remains nearly constant at that level. For the  $N_\pi S_0$  configuration, the thresholds increase for lower frequencies but are still about 5 dB lower than experimental results.

For high frequencies, the predicted BMLDs stay above the experimental ones. For 4 kHz the  $N_0S_\pi$  BMLD is still at a level of 4.5 dB. In addition, the predicted difference of the  $N_0S_\pi$  and the  $N_\pi S_0$  configuration between 500 Hz and 2 kHz is smaller than for the experimental data.

### 3.4.2 Interaural time delay of the masker

In the next experiment, the effect of external interaural time differences of the masker on binaural detection were studied. As in the studies of Jeffress et al. (1952) and Langford and Jeffress (1964), simultaneous thresholds for a diotic tonal signal ( $S_0$ ) were measured as a function of the ITD of the noise masker. Sometimes such configurations are referred to with the misleading term decorrelation (by an



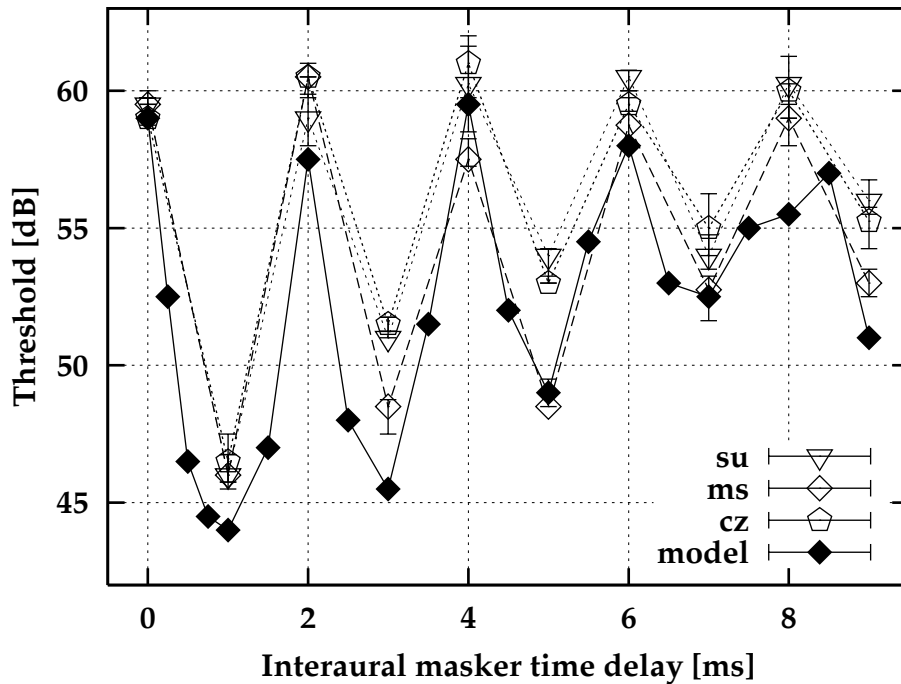
**Figure 3.5:** Binaural masking level differences for the configurations  $N_\pi S_0$  ( $\Delta$ ) and  $N_0 S_\pi$  ( $\circ$ ) from 3.4 in measurements obtained from Holube et al. (1995) (open symbols) and simulations (filled symbols).

external delay) which should not be mixed up with decorrelation by mixing noise from independent noise sources. The purpose was to determine to which extent the binaural system is able to compensate for ITDs in detection experiments. In the experimental configuration, the masker ITD was varied in steps of 1 ms from 0 to 9 ms. In the corresponding simulations, some values in between the integer multiples of 1 ms have been considered. The test tone had a frequency of 500 Hz and a duration of 200 ms and was both spectrally and temporally centered in a noise masker with a bandwidth of 1 kHz and a duration of 500 ms. The noise masker was presented with an overall level of 70 dB SPL. In order to avoid spectral splatter both the test signal and noise maskers were gated with 20 ms Hanning ramps. Thresholds are expressed as level of the test tone.

To produce the  $N_\tau$ -masker the first noise sample was randomly selected from a 2000-ms noise buffer described above. A second sample of the same length was selected with a temporal shift of the size of the desired ITD (delayed noise). Both samples were presented dichotically and were gated with the same Hanning ramps so that the ITD was coded in the fine structure of both waveforms and not in the onset of the presented signals.

Figure 3.6 shows the experimental data for three subjects (open symbols) and results from corresponding simulations (filled diamonds). The test-tone level at

threshold is plotted as a function of the interaural time difference of the masker. The point  $N_{\tau=0}S_0$  for no internal delay is the reference point for calculating the BMLD (a plot would result in the same curve as in 3.6 mirrored at the  $\tau$ -axis). There is a good agreement in performance between the subjects. Thresholds of subject ms are for several delays some dB lower than those of the other subjects, in particular for the minima of the threshold curves which occur in the odd multiples of 1 ms. The interquartile ranges for all subjects are smaller than 2.5 dB (exception: subject su for  $\tau = 4$  ms with an interquartile range of 3.125 dB).



**Figure 3.6:** Experimental masked thresholds for three subjects (open symbols) and the corresponding model predictions (filled symbols) for a configuration with an external interaural time difference ( $N_{\tau}S_0$ ) as a function of the interaural delay of the noise masker.

All threshold graphs show a periodic behavior with a general damping in each cycle. In the simulated thresholds, the structure within each cycle tends to be asymmetric: the (negative) slope for small ITDs is larger than the (positive) slope for larger ITD-values. This behavior is consistent with experimental data from Rabiner et al. (1966), Langford and Jeffress (1964) or van der Heijden and Trahiotis (1999). Thresholds in simulations for values of the ITD which result in a binaural advantage in detection (ITDs different from even integer multiples of 1 ms) are some dB below experimental thresholds. This is the result of the scaling of the signal in the binaural channel which has been chosen in such a way that binaural thresholds and not the BMLDs are predicted optimally. Except for this difference, the agreement between

simulations and experimental data is very good up to masker ITDs of 7 ms. The reason why predictions deviate from measured thresholds for larger masker ITDs is unclear. It may result from the manner how the internal delay is restricted which produces an inadequate internal compensation for large external delays.

### 3.4.3 Level dependence of binaural thresholds

An experiment to establish the relation between BMLD and signal frequency by Kohlrausch (1984) was repeated in simulations. The test tone frequency was 300 Hz, its duration 250 ms including 20 ms Hanning ramps (linear ramps in the experiment)<sup>7</sup>. The noise masker was digitally generated, low-pass filtered at 1 kHz and attenuated to the desired overall level which was varied from 20 to 80 dB (corresponding to a spectrum level between -10 and 50 dB/Hz for the low frequencies). Thresholds have been simulated for the configurations  $N_0S_\pi$  and  $N_0S_0$  as reference. The reference in the experiment was  $N_\pi S_\pi$ .

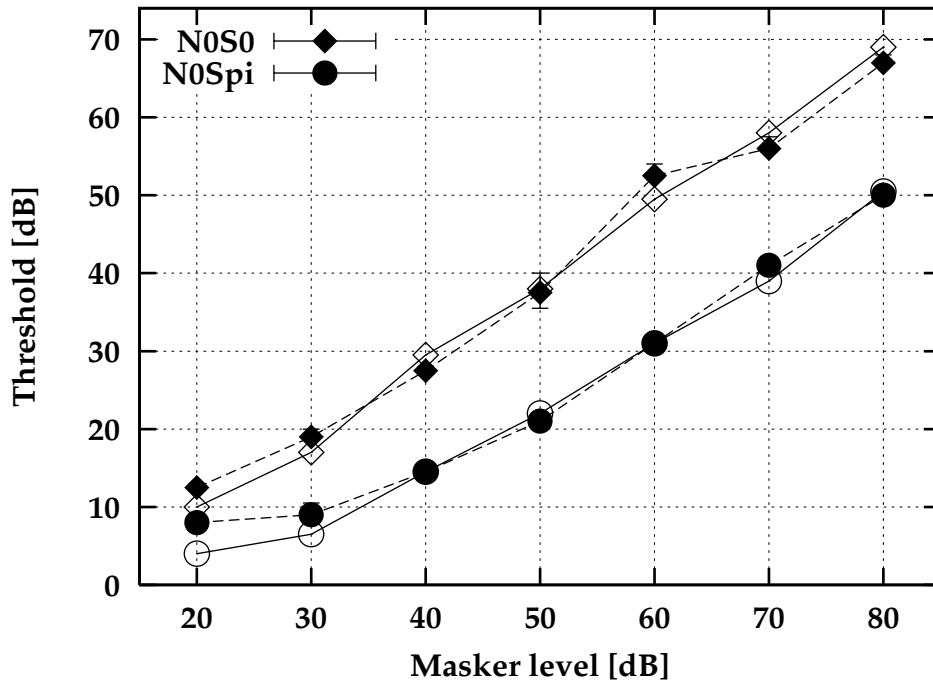
Figure 3.7 shows results for both experiments (open symbols) and corresponding simulations (filled symbols). The level of the test tone at threshold is plotted as a function of masker level. In general the correspondence between experimental data and thresholds from simulations is quite good. Larger differences occur only for low masker levels and for one outlier for the  $N_0S_0$  threshold at 60 dB masker level. The BMLDs from simulations are slightly smaller than in the experiment. Note that simulations were performed without the highpass filter describing the ear canal and the middle ear. Taking it into account would lead to slightly higher thresholds for the lowest masker levels.

### 3.4.4 Processing of interaural intensity differences

McFadden (1968) measured binaural masked thresholds for the configurations  $N_mS_m$  and  $N_0S_\pi$  as function of the masker level. Moreover he measured thresholds in an  $N_0S_\pi$  configuration where the sum signal consisting of masker and test signal is attenuated at one side. This configuration is denoted as  $N'_0S'_\pi$ . Both signals are scaled with the same factor, i. e. the signal-to-noise ratios for each ear as well as the interaural phase differences are the same as in the  $N_0S_\pi$  configuration. Only overall amplitude differences across both ears were introduced. The masker was a white Gaussian noise. In the experimental configuration, the masker spectrum level was varied from -15 to 45 dB/Hz with a step size of 10 dB/Hz. For the  $N'_0S'_\pi$ -configuration the masker spectrum level was kept constant at 45 dB/Hz in one ear and varied in the other ear. The signal was a 400 Hz tone with a duration of 250 ms and 10 ms ramps. Psychometric functions were determined for all three configurations with a two-interval forced choice method and calculating the proportion of correct

---

<sup>7</sup>A configuration at a medium test-tone frequency (300 Hz) is considered here in the simulations because the model does not account well for the binaural thresholds below about 250 Hz, c. f. see Section 3.4.1.



**Figure 3.7:** Binaural and diotic thresholds as a function of the overall masker level. Experimental data from Kohlrausch (1984), with  $N_\pi S_\pi$  ( $\diamond$ ) instead of  $N_0 S_0$  as reference configuration. Test-tone frequency was 300 Hz.

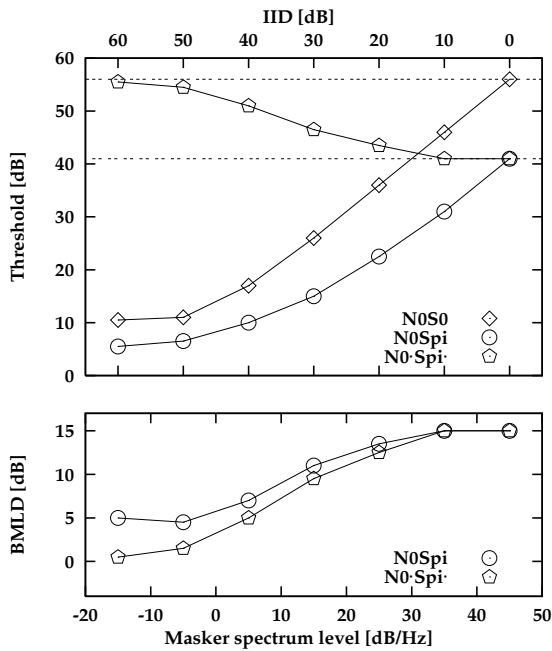
responses for each configuration and several signal-to-noise ratios. The signal level which resulted in 80% correct detections was considered as threshold.

In contrast to the experimental setup described by McFadden, thresholds in the simulations were obtained by an adaptive, three-interval forced choice method as described in Section 3.3.1 resulting in an estimate of signal level necessary for 70.7% correct responses. Unlike to the simulations in the previous section, the ear canal and middle ear highpass filter was taken into consideration.

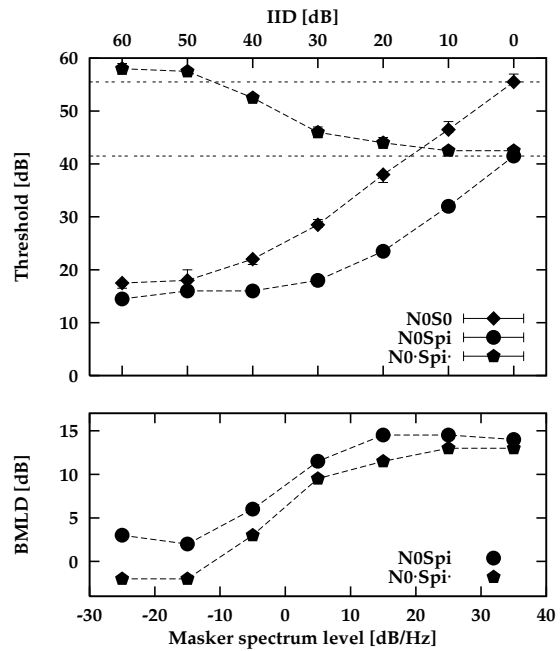
For high masker levels and a diotic masker ( $N_0 S_0$  and  $N_0 S_\pi$ ) the simulated thresholds are approximately 10 dB higher than thresholds from the McFadden data. Since predicted thresholds were in good agreement with experimental data in similar conditions (see above) the differences between experiment and simulations are probably due to a different level definition, threshold criterion, and/or calibration procedure than in the McFadden data. Therefore in the simulations the range of masker spectrum levels was chosen 10 dB lower (spectrum level ranging from -25 dB/Hz to 35 dB/Hz) to fit the thresholds.

Figure 3.8 shows the data from McFadden (1968), Figure 3.9 the corresponding results from simulations. The upper panels in both figures show the masked thresholds as a function of the masker spectrum level. For the  $N'_0 S'_\pi$  configuration (pentagons) the signal level at threshold in the ear with the higher masker level is

plotted as function of the lower masker spectrum level presented in the other ear. Since the masker spectrum level at one ear was kept constant, the x-axis can as the shifted IID (upper x-axis). This way of plotting is chosen since thresholds can be better compared with the  $N_0S_0$  reference condition at the highest masker level. The horizontal dashed lines indicate the reference thresholds for the  $N'_0S'_\pi$  configuration: the upper line for the  $N_0S_0$  at the highest masker level, which is the reference point to calculate the BMLD, and for  $N_0S_\pi$  which is the lower limit for the threshold curve. The lower panels show the corresponding BMLDs for  $N_0S_\pi$  and  $N'_0S'_\pi$ .



**Figure 3.8:** Experimental data from McFadden (1968). Level dependence of masked thresholds for the configurations  $N_0S_0$ ,  $N_0S_\pi$  and  $N'_0S'_\pi$  (upper panel) and corresponding BMLDs (lower panel). See text.



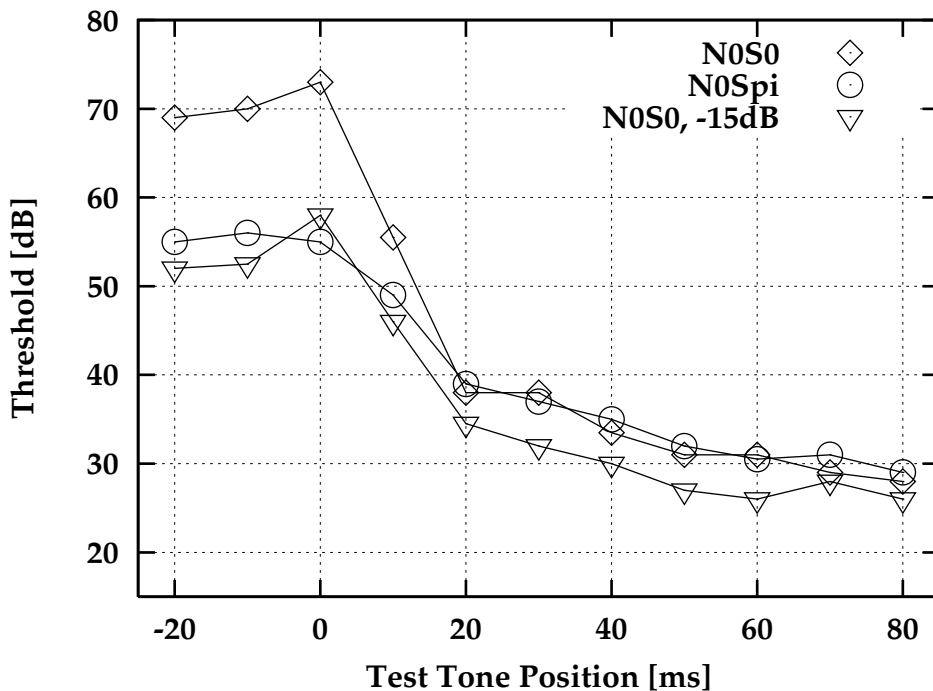
**Figure 3.9:** Simulated thresholds (upper panel) and BMLDs (lower panel) for same configurations as in Figure 3.8. Simulations performed with internal amplitude adjustment for the  $N'_0S'_\pi$  configuration.

Except from the different range of masker levels and apart from the low masker levels there is a good agreement between experimental and simulated thresholds. With increasing masker level, simulated thresholds do not reach the same absolute thresholds in quiet as in the experiment. They stay about 7 dB ( $N_0S_0$ ) and 9 dB ( $N_0S_\pi$ ) above thresholds from human observers. The transition between maximum BMLD (difference between both curves) and minimum BMLD with decreasing masker level is smoother for the experimental data than in the simulated thresholds. The predicted thresholds for  $N'_0S'_\pi$  are in good agreement with the experiment. Both have the same sigmoid shape and show significant BMLDs up to

IIDs of 40 dB. The simulated threshold curve performs a slightly steeper transition between both endpoints of the curve.

### 3.4.5 Forward masking in binaural and monaural configurations

Fassel (1989) and Kohlrausch and Fassel (1997) measured masked diotic and dichotic thresholds in forward masking configurations for several masker durations and levels. The signal was a 10 ms tone pulse 500 Hz shaped with Hanning ramps. In the experiments the masker was a frozen noise sample with a flat bandpass spectrum between 20 and 1000 Hz. Since the original noise sample was not available in the simulations, running noise was used. For masking experiments performed in broadband noise (broader than the critical bandwidth) there is no large difference between frozen and running noise configurations (masked threshold differ less than 2 dB). The noise masker was presented at an overall level of 70 dB SPL in  $N_0S_0$  and  $N_0S_\pi$  configurations as well as 55 dB for the  $N_0S_0$  configurations. This corresponds to a diotic situation in which the simultaneously masked thresholds are at an equal level as the  $N_0S_\pi$  thresholds at the higher masker level.

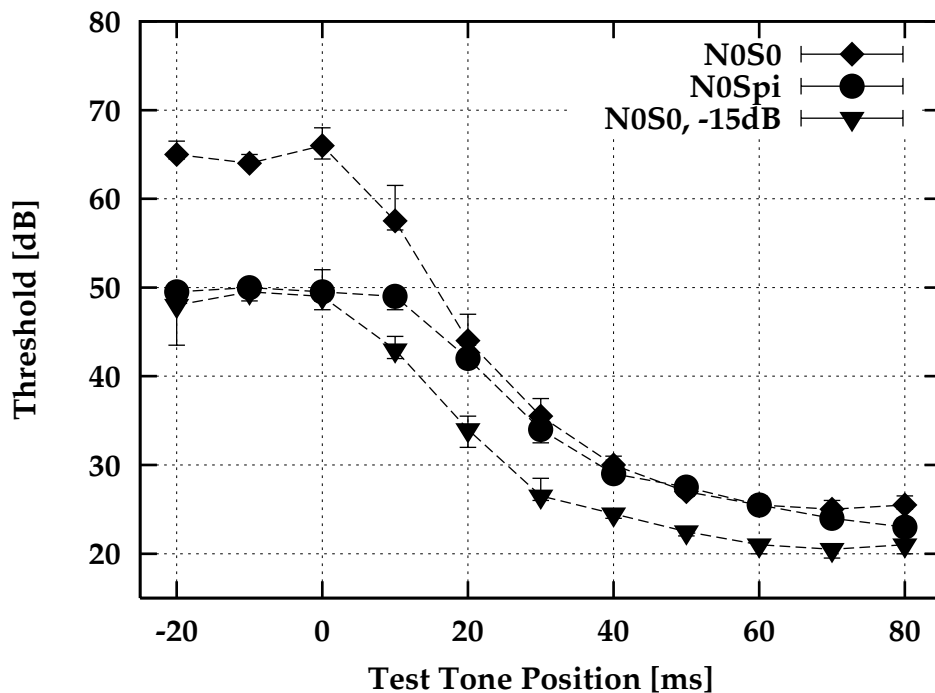


**Figure 3.10:** Experimental thresholds from binaural and monaural forward masking experiments. Results are from Fassel (1989) for one subject. Thresholds are plotted as a function of the relative temporal position of the end of the test tone relative to the end of the noise masker.

Figure 3.10 shows the experimental results from Kohlrausch and Fassel (1997). Forward masking thresholds are plotted as function of the temporal position of the end of the test signal relative to the end of the noise masker. A test signal presented at 0 ms is just still completely within the noise masker.

The  $N_0S_0$  (-15 dB) and  $N_0S_\pi$  threshold curves approximately show the same values in simultaneous masking (temporal positions  $< 0$  ms), but diverge in forward masking. The diotic curve is steeper for delays up to 20 ms, and remains about six dB below dichotic thresholds. The diotic curve for the higher masker level in forward masking starts also with a steeper slope than the  $N_0S_\pi$  curve. Both curves converge at about 20 ms.

Figure 3.11 shows the results from the corresponding simulations of the three configurations. The main effects and the general course of the three curves are the same as for the experimental data:  $N_0S_0$  (-15 dB) and  $N_0S_\pi$  curves starting from an equal level in simultaneous masking diverge in forward masking,  $N_0S_0$  and  $N_0S_\pi$  thresholds for the same masker level converge in forward masking. Over the whole range of signal positions, thresholds are some dB lower than in the experiment. Also the slope of all three simulated curves is flatter for the first 30 ms. In addition, the range over which diotic and dichotic thresholds approach in forward masking for the same masker level is slightly larger (about 30–40 ms) than in the experimental data (20 ms).



*Figure 3.11: Results from the simulations of the forward masking experiments in Figure 3.10.*

## 3.5 Evaluation of the model structure and “critical” model parameters

The binaural processing model proposed here is able to simulate several different binaural detection experiments to a satisfactory degree with a comparatively simple add-on to a pair of models of the “effective” monaural processing. Since the respective role of each processing element in the binaural model can not be identified beforehand, a more detailed analysis is given below. It also will be considered which general conclusions about binaural signal processing in the human auditory system can be drawn from the interaction between experimental findings and theoretical models of the experiments considered here.

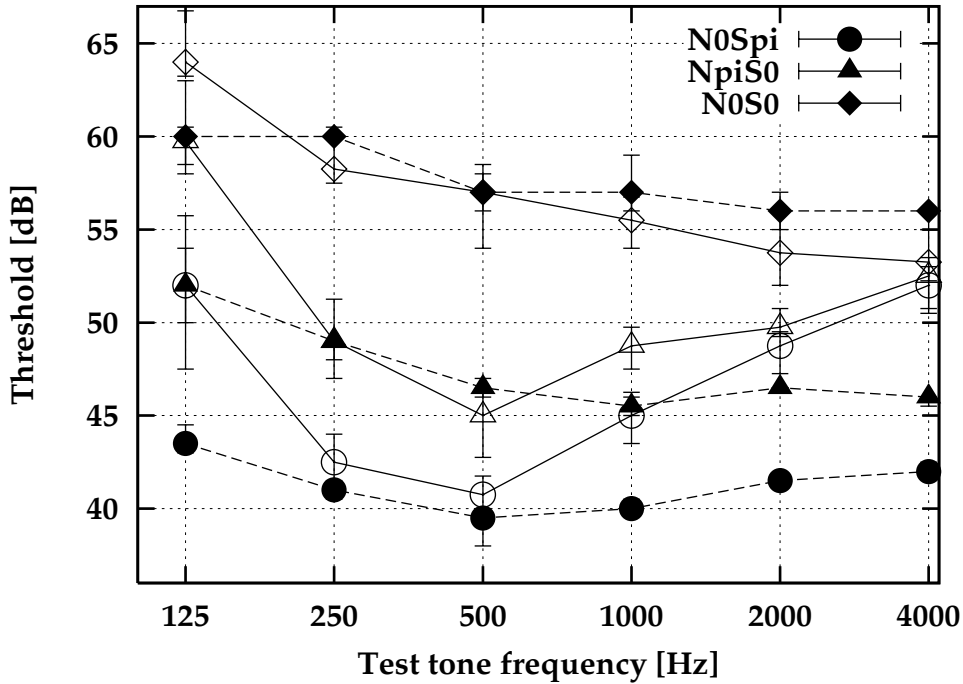
### 3.5.1 Influence of jitter vs. role of neural transduction

The most plausible reasons that may account for the limited efficiency of binaural interaction in high frequency configurations are the loss of neural phase locking, an increasing “critical” bandwidth or a binaural mechanism which is less effective with increasing frequency. The first two properties are already modeled by the respective monaural processing units which include a model of neural transduction and a gammatone filterbank, respectively.

To assess the relative contribution of the different stages of signal processing, a systematic variation of those model parameters and model properties was performed that directly influence the accuracy of the binaural processing unit.

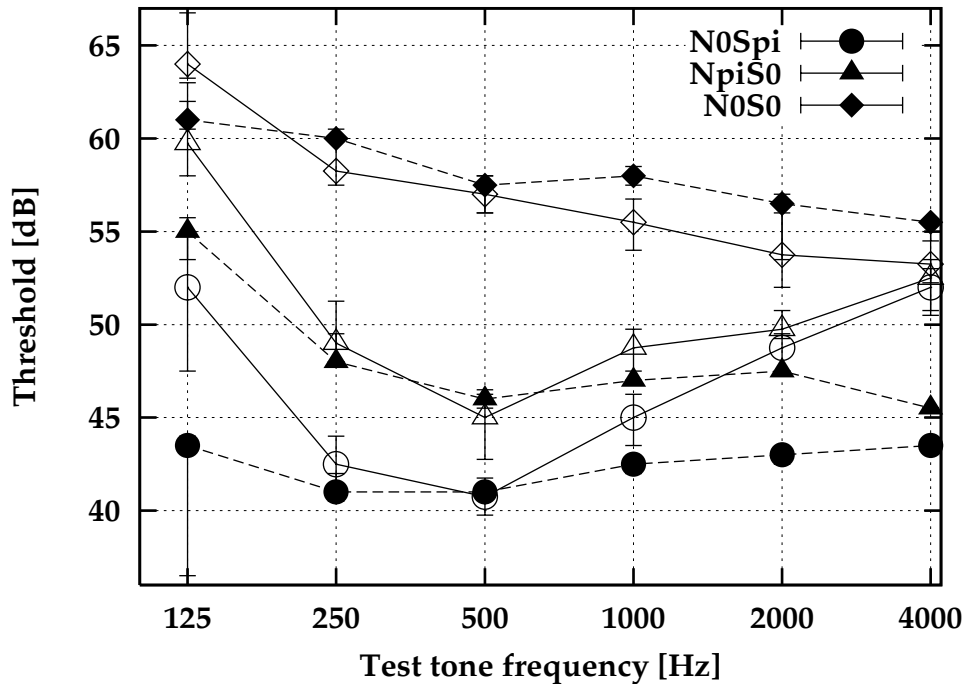
First, the simulations of the frequency dependence of thresholds have been repeated with a modified model in which the binaural processor is assumed to work with a “perfect” mechanism of noise reduction (*without* averaging over a range of internal delays, that means according to Equation 3.3). The remaining processing stages were left unchanged. Figure 3.12 shows results from these simulations with the modified binaural processing unit in comparison with the experimental results (the same as in Section 3.4.1). The results from simulations (filled symbols) are presented with their medium value and interquartile range from five simulations for each point. The signal in the binaural channel is scaled in such a way that the results from simulations in the  $N_0S_\pi$  configuration at 500 Hz fit approximately the corresponding masked threshold for human observers. Thus, the simulated binaural thresholds at this frequency are slightly lower than in Figure 3.4.

In Figure 3.12 the shape of the simulated  $N_\pi S_0$ - and  $N_0 S_\pi$ -threshold curves show much less variation across frequency than the experimental data and the simulations performed with time jitter (cf. Figure 3.4). In particular thresholds are too low at high frequencies indicating a higher binaural efficiency in the model than actually observed in the data. At frequencies below 1 kHz threshold agree quite well with simulations in Figure 3.4. Obviously, the uncertainty in binaural processing has no influence for low frequencies, since the interaural phase of the ear signals does not change significantly in the range covered by the time jitter.



**Figure 3.12:** Frequency dependence of masked thresholds for the same configurations as in Figure 3.4. Experimental data is plotted with open symbols. Simulations (filled symbols) have been performed with a binaural processor which is implemented without random jitter. The peripheral low-pass filter is of 2<sup>nd</sup> order with  $f_{cut-off} = 1\text{ kHz}$  (see text).

The second stage in signal processing which limits binaural detection is the synchronization low-pass filter describing loss of neural phase-locking. To assess the influence of this stage on binaural detection performance, the same simulation as before (binaural processor without processing uncertainty) was repeated with a low-pass filter with a lower cut-off frequency (650 Hz) and a higher order (5<sup>th</sup> order) than in the original simulations. Such a filter would reduce more of the waveform fine-structure in the transduction process and leaves less information available at high frequencies. The results of the simulations are plotted with filled symbols in Figure 3.13 together with the experimental thresholds (open symbols). Again, the medium values and interquartile ranges from five simulations is presented. Compared with the results from Figure 3.12, there are nearly no noticeable differences apart from a slight shift in binaural thresholds. This shift results from the changed characteristic of the low-pass filter since it reduces more of the signal energy available in high frequencies available at the output of the respective peripheral filter and hence deteriorates the performance. Since the observed shift is independent of center frequency, the decreasing efficiency of binaural unmasking at high frequencies can not be explained by the model on the basis of the removal of the signal fine



**Figure 3.13:** Threshold curves from experimental data (open symbols) and simulation data (filled symbols) for the same configurations as in Figure 3.12 (binaural processor without random jitter) but with peripheral low pass-Filter with  $f_{\text{cut-off}} = 650$  Hz of 5<sup>th</sup> order.

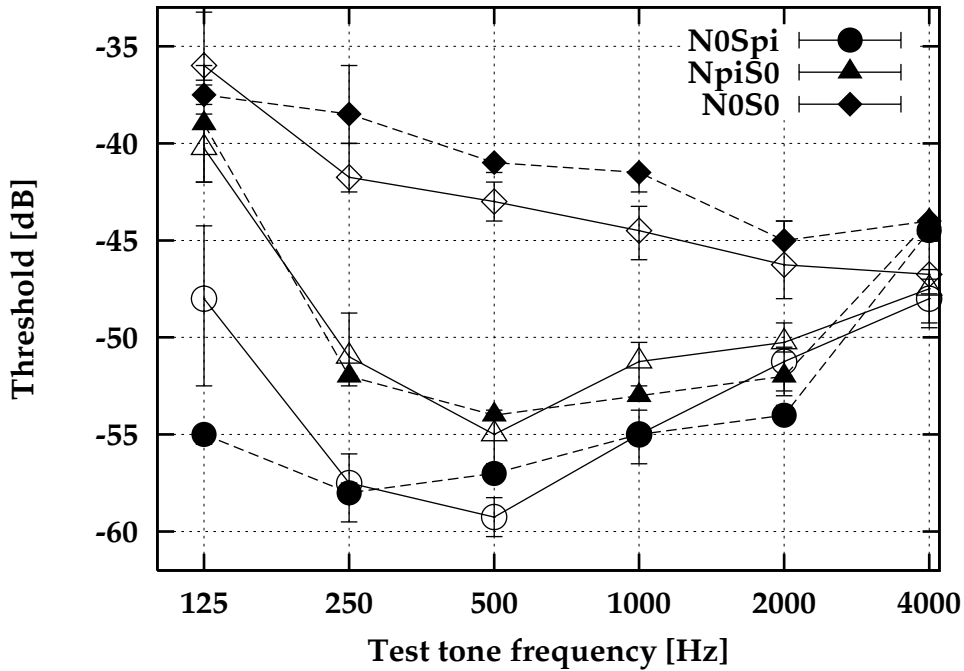
structure alone in combination with an appropriate peripheral bandwidth (which has not been considered here).

A higher order and steeper synchronization low-pass filter was also estimated in physiological and psychophysical studies. Weiss and Rose (1988) measured the frequency dependence of synchronization of cochlear nerve fibers in different auditory receptor organs. The transfer characteristics are described by synchronization low-pass filter functions. The order of these filters was between 4 and 6, the cut-off frequency varied from 340 Hz to 2.5 kHz. These values motivated the parameters employed in Figure 3.13.

Bernstein and Trahiotis (1996) obtained similar results for human listeners describing their experimental data from binaural detection experiments. They measured the percent correct  $[p(c)]$  as a function of signal-to-noise ration for several center frequencies. In order to predict the percent correct they used the normalized interaural *cross correlation coefficient* of a half-wave, square-law rectified and low-pass filtered version of the ear signals. They obtained the best fit of experimental data with a 4<sup>th</sup> order low-pass filter with a cut-off frequency of  $f_{\text{cut-off}} = 450$  Hz.

Figure 3.14 finally illustrates two different effects of changing two more model parameters that show a significant effect only for high or low center frequency, re-

spectively. A steeper low-pass filter after half-wave rectification. 4<sup>th</sup> order instead of 5<sup>th</sup> order as used before combined with a binaural uncertainty in temporal equalization (“time jitter”), which primarily affects the high-frequency range. Furthermore, the range of available internal time delays has been restricted to  $\pm 3$  ms (realized as described in Section 3.2.3), which only affects binaural low-frequency thresholds. The modifications have been employed in one single simulation since their respective influence can be restricted to different frequency ranges. In comparison to Figures



**Figure 3.14:** Influence of the uncertainty in binaural processing in combination with a low-pass filter of 4<sup>th</sup> order after half-wave rectification. The range of available internal delays has been restricted to  $\pm 3$  ms. Interaural configurations are the same as in Figure 3.12.

3.12 and 3.13, the BMLDs at high frequencies are better predicted. At 4 kHz, the BMLD is even too low. This effect may be due to the restriction to only one internal time delay in combination with the 4<sup>th</sup> order low-pass filter. The usage of only one time delay may be too strict for the high frequency configurations, since there are several maxima of the ICF at ITDs close to zero with similar function values available. The optimal combination of binaural signals resulting from different possible internal delays might therefore improve binaural thresholds in high-frequency configurations, when peripheral transduction leaves a sufficient amount of fine-structure. Hence, a low-pass filter of a higher order and with a lower cut-off frequency than applied in the monaural model might yield adequate binaural thresholds.

In the low-frequency range, the predicted thresholds for an  $N_{\pi}S_0$  configuration

can be fitted quite well by restricting the range of available internal delays. However, the thresholds for  $N_0S_\pi$  are not changed. The restricted size of the internal delay cannot be made responsible alone for deviations of predicted and experimental thresholds at low frequencies for the following reason: For a signal frequency of 125 Hz and 250 Hz, the predicted  $N_\pi S_0$ -threshold coincidences with the experimental threshold by restricting the range of possible internal delays. An internal delay of a half of the period of the signal (of the frequency  $f_c$ ) is required for an  $N_\pi S_0$  configuration to perform an optimal noise reduction in the subsequent cancellation stage. A smaller delay would result in a suboptimal noise reduction. This is not the case for the  $N_0S_\pi$  thresholds where a zero internal delay is required which is always available. Hence, the restriction of the delay does only account for parts of the experimental data a further mechanism different from a noisy delay has to account for the discrepancy found at low frequencies.

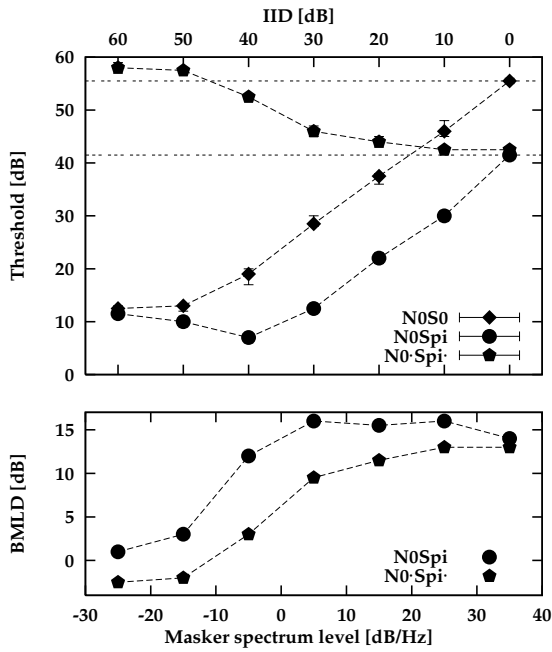
### 3.5.2 Processing of level and IIDs

This section investigates the influence of monaural compression of stationary signals and of the peripheral noise on the performance of the binaural model in situations where threshold are determined as a function of the masker level and of the masker IID. To this end, simulations of the experiment performed by McFadden (1968) (already described in Section 3.4.4) were repeated without employing peripheral noise and without internal gain adjustment.

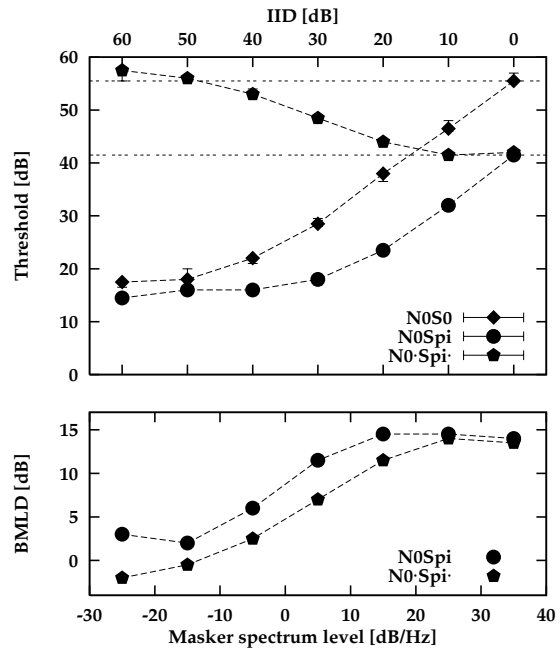
Apart from the low-frequency behavior, the model reproduces experimental performance quite well. Deviations occur when masked thresholds are close to absolute threshold where the model performance of the model is determined by the middle ear highpass filter and the peripheral noise. Several factors may be responsible for this behavior. For instance, the peripheral stages of the model do not take into consideration the level dependence of the auditory filter shape. From masking experiments utilizing a spectrally notched-noise paradigm where the signal is placed asymmetrically in the notch and from physiological tuning curves it is known that auditory filter are sharper tuned in at low levels and getting broader with increasing level. For the model performance this would mean that less of the masker energy falls in the filter and thresholds were lower. Assuming that the peripheral noise is also filtered in the auditory periphery, even the absolute threshold would be reduced. As a second factor, the influence of the peripheral noise has to be verified.

Simulations of the McFadden experiment performed without peripheral noise demonstrate that decorrelating noise is necessary to describe binaural thresholds at low masker levels correctly. Results from such simulations are plotted in Figure 3.15 (masked thresholds in the upper panel and BMLDs in the lower one). Both  $N_0S_0$  and  $N_0S_\pi$  absolute thresholds (the asymptotic value of threshold for low masker levels) are some dB lower than in the simulations in Figure 3.9. The difference between absolute thresholds at low masker levels, which is also present in the experimental data, does not occur. It can be interpreted as binaural threshold in the uncorrelated

peripheral noise. In contrast to experimental results, with increasing masker level the simulated  $N_0S_\pi$  thresholds drop to a minimum lower than absolute threshold at  $-5$  dB/Hz, where the BMLD is already close to its maximum. The binaural processor achieves maximal efficiency shortly after it receives some input. For the “undershoot” of  $N_0S_\pi$  thresholds with increasing masker level a kind of stochastic resonance<sup>8</sup> might be responsible. This effect can only be removed when a sufficiently large amount of (peripheral) internal noise is added to the signal processing in the model<sup>9</sup>. In the experiment, the BMLD increases gradually over a input range of 40 dB.



**Figure 3.15:** Predicted thresholds (upper panel) and corresponding BMLDs (lower panel) for the same configurations as in Figure 3.8. Simulations performed without decorrelating peripheral noise.



**Figure 3.16:** The same predicted thresholds and BMLDs as in Figure 3.15. Simulations for the  $N_0'S_\pi'$  configurations were performed without internal gain adjustment.

Another important and interesting aspect of binaural signal processing in the configuration with stationary interaural level differences is the influence of the in-

<sup>8</sup>Stochastic resonance can be defined as a cooperative effect occurring in nonlinear input-output systems, in which a weak periodic signal interacting with large fluctuations leads to a large amplification of the periodic signal component in the output signal (cf. Wiesenfeld and Moss (1995) or Narins et al. (1997)).

<sup>9</sup>Simulations with different shapes of the input-output-characteristic of the binaural processor (instead of a full-wave rectification) have been tested. This can for instance describe a reduced sensitivity of the binaural system at low signal levels. For tested characteristics the “undershoot” in thresholds with increasing masker level was preserved.

ternal equalization of those differences. As it has been described in the previous chapter, in the internal adjustment only the long-term average of the masker energy is taken into account. Due to the nonlinearity of the adaptation stages the equalized signals become more dissimilar with increasing masker IID such that after the cancellation step a significant portion of the masker still remains.

Because of the almost logarithmic compression of the stationary part of signals their internal representations have already a similar RMS-value. The question is, how much compression alone contributes to the model's behavior. Therefore the simulation of the experiment from McFadden (1968) has been repeated without adjusting the average amplitude of the internal representations of the ear signals. The result are plotted in Figure 3.16. The predicted  $N'_0 S'_\pi$ -threshold curve does not differ very much from the thresholds in Figure 3.9 (with internal gain adjustment), especially the endpoints of both curves are corresponding. The slope of the threshold curve is slightly flatter than for these with internal adjustment. The effect of equalization on predicted thresholds is very small, thresholds in Figure 3.16 correspond even slightly better to the experimental data. Following these results, monaural compression provides a sufficient "normalization" of input signals to enable the model to deal with the large input range and to explain binaural detection data concerning masker input levels and IIDs. In the present model of binaural processing it can even take over the role of the equalization of the signal amplitude in the EC-model.

A quantitative measure how IIDs are represented on an internal level are the  $q$ -factors from Equation 3.2 which define the internal scaling factors  $r_q = q$  and  $l_q = 1/q$  for the right and left ear signal. The  $q$ -factors occurring in the simulations for the McFadden-experiment are listed in Table 3.1 for the IIDs investigated in the simulations. Each factor is the average values calculated for five internal represen-

IID [dB]	$q$	$q_{\text{stat}}$
60	0.8965	0.8977
50	0.8969	0.9140
40	0.9046	0.9306
30	0.9339	0.9475
20	0.9645	0.9647
10	0.9852	0.9821
0	1	1

**Table 3.1:**  $q$ -factors occurring in simulations of the experiment from McFadden (1968). The factors are defined in Equation 3.2 used to adjust the RMS-values of the internal representation of the right and left ear signal (see text).

tations of independent noise samples (middle row). The values for the small and for the large investigated IIDs are close to the value  $q_{\text{stat}}$  one would expect for purely

stationary (constant) input signals (right row). This can be interpreted as follows: when the external masker IIDs are small the internal representations for both ear signals are similar in their temporal fine structure. With increasing IID apart from the deviating RMS-values because of the nonlinear processing also the fine structure becomes different for both ear signals. For large IIDs the ear where the masker with the lower level is presented the representation is a constant determined by the constant value describing the hard threshold criterion before the first adaptation loop. In this case  $q$  becomes exactly equal to  $q_{\text{stat}}$ .

The decrease in BMLD with decreasing overall level of the masker has been investigated in many studies (Hirsh, 1948; Dierks and Jeffress, 1962; Dolan, 1968; McFadden, 1968). A common explanation of the masker level dependency of the BMLD is based on additive, partly uncorrelated internal noise to both ear signals. Robinson and Jeffress (1963) had shown that the size of BMLD as the interaural correlation of the external, experimentally controlled noise masker decreased. As the external noise level is lowered, the interaurally decorrelated internal noise constitutes a larger proportion of the effective masking noise and lowers the size of the BMLD. Dierks and Jeffress (1962) fitted an internal noise with an interaural correlation of 0.25. McFadden (1968) stated a value of 0.35. Yost (1988) used a correlation of 0.30, and he estimated the level of the internal noise as 29 dB with supraaural headphones and 22 dB with insert phones which is in line with other studies, (e. g. Berger, 1983). All studies considered the overall level of the external masker and internal noise, but not the energy within the respective critical band.

The values of the fitted interaural correlation of the internal noise and the high level of the internal decorrelating noise (compared to the level employed in the present model) may be a consequence of the fact that the underlying model description does not incorporate a monaural threshold.

The treatment of signal intensities and the origin of BMLDs within the binaural model of the present study is different. It is a cumulative effect of an absolute threshold, peripheral additive noise, compression and internal noise. Since the model incorporates an absolute threshold, the range of masker levels for which the binaural processor receives input has a lower limit. On the other hand the energy of the internal representation do not show a strong dependence from external masker energy because of the approximately logarithmic compression of the stationary part of the signal. Assuming that a constant portion of the noise masker is canceled by binaural processing, the signal level at threshold shows also no strong dependence on the external masker level. Since the resolution of the model is restricted by a constant internal noise (which is different from the EC theory), also binaural thresholds should show a linear dependence from masker level (both in dB). For low masker levels the decorrelation described in Section 3.2.3 becomes relevant so that the sum of external and internal noise cannot be removed completely.

The description of absolute thresholds in the model by an effective description of the ear canal and the middle ear transduction together with decorrelating internal noise yield predictions for absolute thresholds which are some dB too high in the low-

frequency range. Deviations from experimental data may result from inadequate descriptions of both features. So the utilized highpass filter has a steeper transfer function in the low-frequency range than absolute threshold. Furthermore, additive peripheral noise might be a description which is too simple to describe internal decorrelation.

To improve the predictions for the thresholds as a function of masker level, internal noise could for instance be considered between more adaptation stages. This can be seen as follows: Peripheral noise (internal noise added before any compression takes place) influences solely thresholds for a low masker level. Internal noise on the level of the optimal detector (after compression) describes the resolution of the model and changing its level shifts all threshold curves on the whole. Therefore noise added on intermediate stages should result in a decreasing influence on masked thresholds with increasing masker level, which could cause a smoother decrease of the BMLDs as a function of masker level (cf. Figure 3.9). Of course also a different kind of internal noise than additive noise might be a more appropriate description, e. g. multiplicative noise.

## 3.6 Discussion

The presented model of the “effective” binaural signal processing is able to account well for the chosen “critical” experiments. The main differences between experimental results and predicted simultaneous masked thresholds occur for very low and high frequencies ( $f = 125$  Hz and  $f \geq 2$  kHz). For the low frequency condition the predicted  $N_0S_0$ -threshold is 5 dB too low (if one considers the complete shape of the simulated threshold curve relative to experimental data). This might have its cause in the formula used to calculate the bandwidth of the peripheral low-pass filters in Equation 3.1. The critical bandwidth below about 500 Hz derived from psychophysical experiments differs between several studies. Zwicker (1961) and Zwicker and Terhardt (1980) estimate the critical bandwidth below 500 Hz at a constant value of approximately 100 Hz whereas the formula of Moore and Glasberg results in a critical bandwidth of 38 Hz for a characteristic frequency of 125 Hz. This value was adopted in this study. However, a larger critical band would result in higher thresholds since more noise falls into one filter.

The predicted BMLDs for the lowest frequency show large deviations from that of human observers. A similar experimental behavior has been observed in many binaural low-frequency conditions, for instance the dependence of BMLDs in a  $N_\phi S_0$  configuration from interaural phase for large  $\phi$ 's (Rabiner et al., 1966) or in  $N_\tau S_x$  (Jeffress et al., 1962) configuration. In these conditions the performance of threshold as a function of the interaural parameter of the noise masker is referred to as “flattening” (threshold curves show less variation and saturate earlier when the interaural parameter is increased). This behavior cannot be explained completely by the (monaural) peripheral bandwidth. Broadening the peripheral bandwidth from

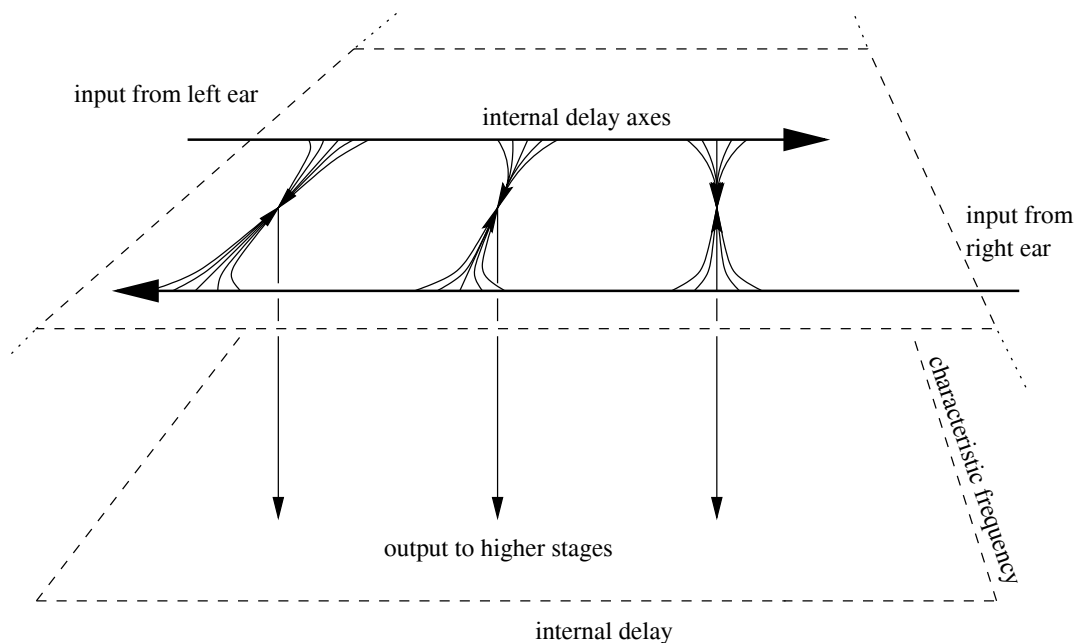
38 Hz to 100 Hz would merely result in both an increase of monaural and binaural thresholds as of about 4.2 dB. A part of the overestimated BMLDs for low frequencies in simulations may result from a frequency dependence of the compressive and adaptive mechanism that was assumed to be the same for all characteristic frequencies in the model’s signal processing. However, physiological and oto-acoustic emission studies (Cooper and Rhode, 1997), that are in contrast to the assumption stated above, show that the auditory system seems to be less compressive for very low frequencies. In combination with a higher absolute threshold such a loss in compressive properties at low frequencies could be responsible for a small BMLD. In psychophysical detection experiments (Kohlrausch, 1984) investigated whether the increased absolute (monaural) threshold for low frequencies could be responsible for reduced binaural unmasking. He found the higher absolute threshold not to be sufficient and concluded that in these configurations the binaural system itself seems to work less efficient.

In modeling the binaural mechanism it has been assumed that it is independent from frequency. This can be justified from experimental results. van de Par and Kohlrausch (1997) constructed so-called “transposed” stimuli, which were centered at 4 kHz but designed to preserve in the envelope the fine-structure of lower frequencies (125 and 250 Hz). Their results supported the hypothesis that binaural processing is very similar at low and high frequencies and that frequency-dependent differences in BMLDs probably reflects the inability of the auditory system to encode the temporal fine-structure of high-frequency stimuli.

The range of available internal time delays is in good agreement with result of a study from van der Heijden and Trahiotis (1999). They performed experiments employing both a delayed masker and signal and, additionally, a masker consisting of the sum of two independent noises having interaural delays of equal magnitude and opposite sign. They could describe their results with an EC-type model with operative internal delays up to approximately 750  $\mu$ s. Internal delays of larger magnitude up to 2 or 3 ms are probably also operative but degraded to some degree, that means that their application introduces noise. These findings and the results about the range of available delays in the present study seem to be consistent with physiological studies from Kuwada et al. (1997). They suggest a reduced density of neural elements available for the processing of larger delays.

Fassel (1989) and Kohlrausch and Fassel (1997) discussed their experiments from forward masking experiments for  $N_0S_0$  and  $N_0S_\pi$  configurations for different masker levels in the framework of cooperative effects of monaural adaptation and binaural processing and the relative order of these processing stages. The existence of a BMLD in nonsimultaneous masking is difficult to explain by present theories of binaural interaction. Kohlrausch and Fassel (1997) discussed their results in the framework of the monaural model of signal processing, which is also the basis of the present study. Binaural interaction which takes place before adaptation stages would lead to a adaptation to different effective masker levels for diotic and dichotic conditions, since the masker level is reduced by the binaural processor. Binaural

forward masking curves described by a model consisting of a binaural processor preceding adaptation would start, in contrast to their results, at a lower threshold value and decay more slowly than a diotic threshold curve, i. e. the BMLD would decrease continuously in forward masking conditions. Kohlrausch and Fassel (1997) concluded that adaptation precedes binaural interaction. For a temporal test-tone position within the masker, binaural interaction results in lower dichotic threshold. For positions after the masker the BMLD collapses within a few milliseconds, for which fine structure of the masker signal is still available in the internal representation because of the ringing of the peripheral filters. After that time, the dichotic and diotic thresholds become the same. The course of the threshold curve in the simulations is in both conditions dominated by the time constants of the adaptation loops.



**Figure 3.17:** Integration over a range of internal delays within the model in the concept of delay lines. Within the upper plane integration along continuous lines is illustrated for three different optimal delays. Delay lines receive input from right and left ear for the same characteristic frequency. Each cancellation element can be parameterized by its internal delay and its CF which is outlined by the lower plane.

The uncertainty in binaural processing reduces fine structure information in the binaural channel by smearing it over time. It is implemented similar to the “jitter error” in the Equalization and Cancellation Theory but in this description of binaural processing no actual random jitter is introduced (in a sense of equalization as a realization of a random process with internal delay as random variable). No error in performing the cancellation process is implemented. In the EC theory, the time

jitter can be interpreted as taking its effect into account by a temporal average over a complete ensemble of stimulus representations since no explicit stimulus waveform is considered. In the realization within the present model, temporal averaging takes place for each single stimulus representation. In the concept of a binaural place mechanism realized by delay lines, this could be realized by an spatial averaging over a limited range along each axis. This concept is visualized in Figure 3.17.

The uncertainty variable  $\tau_{\text{err}}$  in interaural temporal adjustment is assumed to follow a Gaussian distribution. The Gaussian distribution is chosen as a simple representation and following the EC theory. Its standard deviation  $\sigma_{\text{err}} = 105 \mu\text{s}$  has the same size as the temporal jitter in the EC theory and has proved to be a reasonable choice. Larger values of  $\sigma_{\text{err}}$  change the shape of threshold curves in  $N_0S_\pi$  and  $N_\pi S_0$ -configurations such that the BMLDs for 1 and 2 kHz become too small compared with threshold curves from human listeners.

In terminology of neural information processing the working method of the binaural processor in the present model can be described as realizing a straightforward connection of the right and left pathway in an excitatory-inhibitory (EI) manner. In this context, cross-correlation models can be considered as realizing an excitatory-excitatory (EE) connection of ipsilateral and contralateral representations of input stimuli. Both types of binaural neurons can be found in subclasses of neurons in the mammalian auditory pathway (Pickles, 1988; Caird and Klinke, 1983). Utilizing an EI-mechanism enables the model to deal with interaural level differences without that it is required to normalize the ear signals carefully with the signal energy which is necessary for cross-correlation models. Large inherent noise fluctuations can be cancelled out while leaving the test signal unchanged.

### 3.7 Summary

A model describing the effective temporal binaural signal processing combining concepts of both monaural and binaural signal processing accounts quantitatively for several aspects of binaural detection in broadband noise in simultaneous and nonsimultaneous configurations. The effects of center frequency, external interaural time delay, masker level, interaural level differences and temporal separation of masker and test signal on masked thresholds and BMLDs as well are predicted reasonably well with one set of model parameters. High frequency performance of the model can be attributed mainly to an uncertainty in binaural noise reduction. In contrast to assumptions in the literature, it can not be explained completely in term of a release from synchronization in peripheral transduction for high frequencies. Monaural compression of stationary parts of the ear signals provides a reasonable treatment of masker level and an approximate equalization of stationary interaural level differences.

## Chapter 4

# Spectro-temporal processing in binaural detection: Experiments and models

### Abstract

*The influence of spectral resolution and temporal processing in the auditory system on the performance of human subjects in binaural detection experiments is studied. The model of binaural signal processing described before is used to predict the experimental results both from the literature and from own data. The model accounts for the results of various binaural detection experiments in which the interaural parameter variations influence detection behavior in the frequency domain. The binaural model is able to integrate information across frequency by a linear combination of information across different critical bands. Temporal processing is mainly determined by the temporal properties of the monaural preprocessor. The difference between the effective “binaural” auditory filter bandwidth derived from different experiments are discussed within the framework of the present model. In addition, the model adequately describes binaural test tone integration as a spectral effect by assuming that binaural detection (of short test tones) also occurs in adjacent frequency channels. Taken together, a wide range of spectro-temporal effects in binaural detection can adequately be modeled using only a limited set of parameters and assumptions in the current model.*

## 4.1 Introduction

An important aspect in psychophysical and physiological research is the frequency resolution of the auditory system and the concept of critical bands has become an established tool to describe data from a large class of psychoacoustical and physiological experiments originally proposed by Fletcher (1940) following ideas of Helmholtz (1868). He assumed that the auditory system performs a spectral decomposition of sound signals. This concept was physiologically confirmed by findings from von Békésy (1960). It describes the peripheral auditory system by a bank of linear band-pass filters with overlapping passbands. These bands restrict spectral resolution of the auditory system, since adjacent frequency components are processed in common in one channel. The filterbank concept reveals in a great variety of different spectral masking experiments, loudness summation (Zwicker et al., 1957) or even modulation detection (Sek and Moore, 1994). All these experiments lead to similar estimates of the critical bandwidth.

One way to estimate the shape of the auditory filters is a masking experiment, where the tonal signal is centered in a spectral notch of a noise masker as described by Patterson (1976). Results from notched noise masking experiments and from auditory nerve recordings utilizing the reverse-correlation (revcor) technique (de Boer and de Jongh, 1978) can be modeled in terms of a gammatone filterbank Patterson et al. (e. g. 1987); de Boer and Kruidenier (e. g. 1990) and are in good agreement with each other. The derived filter shape also accounts for results from spectral masking configurations quite different from the original notched noise experiments: Using the monaural model of the “effective” auditory signal processing Verhey and Dau (1997) and Verhey (1998) could describe results from bandwidening and notched noise configurations as well. Derleth and Dau (1999) could describe masking patterns for sinusoids and narrow band noise for several masker-signal combinations.

In contrast, in binaural signal detection experiments there exist contradictory results concerning the “effective” critical bandwidth in different configurations. Several studies (Bourbon and Jeffress, 1965; Hall et al., 1983; Zurek and Durlach, 1987) used a bandlimited noise procedure in which the signal threshold was determined as a function of the masking noise bandwidth. All studies indicate poorer frequency resolution for dichotic presentations compared to  $N_0S_0$ . Estimates for the binaural spectral integration bandwidth are a factor 2 to 3 larger than in a comparable monaural case. Furthermore, within the bandlimited noise procedure, the estimated 3 dB bandwidth is highly dependent on the noise level. The frequency resolution improved with decreasing noise level such that for low masker level the estimated binaural critical bandwidth matches the bandwidth estimated in  $N_0S_0$  configurations (Hall et al., 1983). In addition, Hall et al. (1983) used a bandstop procedure and measured the dependence of masked thresholds as a function of the notchwidth. In contrast to the bandwidening experiment, the estimated critical bands showed no strong dependence from masker level and did not suggest large differences in critical band estimates between monaural ( $N_0S_0$ ) and binaural ( $N_0S_\pi$ ) configurations. This

coincidences with estimated binaural critical bandwidths derived from noise maskers with frequency dependent interaural parameters (e. g. Kohlrausch, 1988; Kollmeier and Holube, 1992).

This discrepancy opens the question whether the peripheral filterbank concept successfully employed for modeling monaural auditory processing still holds for binaural auditory processing utilizing the monaural inputs from both sides. The first part of this chapter therefore uses a model for the “effective” binaural auditory processing (introduced in Chapter 3) to predict both experimental results from the literature and own measurements on configurations similar to those outlined above. Wherever necessary, the detection in multiple frequency channels is modeled which appears a salient prerequisite for resolving the discrepancy described above.

In addition, the relation between temporal properties encountered in binaural detection experiments and the filterbank concept employed for binaural processing is considered in the second part of this study (apart from binaural forward masking, the present study focused upon phenomena of binaural masking observed in stationary conditions). This aspect is of considerable importance for understanding and adequately modeling the “effective” binaural auditory processing, because results from several studies concerning binaural temporal test-tone integration in broadband noise were discussed in context with the assumption of different monaural and binaural spectral resolution.

One important example is test tone integration: Blodgett et al. (1958) measured masked thresholds of a 500 Hz tonal signal in continuous broadband noise as a function of the duration of a signal in diotic and dichotic situations. They found that the increase in threshold for reducing the the duration of the signal below about 20 ms was slightly larger in monaural ( $N_0S_0$ ,  $N_mS_m$ ) configurations than for binaural ( $N_0S_\pi$ ,  $N_\pi S_0$ ) configurations. Thus, the BMLD is enlarged for short signal durations. Green (1966) obtained similar results for a signal frequency of 250 Hz and a broadband masker. In both studies very short rise/decay times of only 0.5 ms were employed for the signal. If the operative critical bands were broader in the binaural configurations, for very short test signal less signal energy would be lost as a consequence of spectral splatter. Kohlrausch (1986, 1990) employed signal frequencies and reduced the spectral splatter by imposing linear ramps of a duration of 5 ms on his tonal signals, whose shortest duration was 20 ms. Nevertheless, he replicated Blodgett et al.’s findings concerning the different behavior of thresholds for short test signals in binaural and monaural configuration. From that he concluded, that not different binaural spectral resolution is responsible for the larger BMLDs, but different temporal properties of the monaural and the binaural auditory system.

In order to evaluate systematically this assumption, multichannel simulations were performed with the binaural model. One hypothesis is that the spectral broadening of *short* test signals is advantageous for binaural detection, since signal energy becomes available also in adjacent critical bands even through signal energy is lost within the center channel by reducing duration. This is not valid for the monaural configuration, where the signal energy in adjacent filters is masked by the broadband

noise.

In these configurations, an appropriate temporal analysis within each frequency band (such as the modulation filterbank introduced by Dau et al. (1997a)) can improve the agreement between simulations and measurements for monaural temporal integration (Verhey and Dau, 1997). Hence, this extended version of the model was used to simulate temporal integration data in Section 4.4.

Although such an extended model version appears to be adequate for monaural temporal processing, the appropriate temporal behavior for binaural detection tasks still has to be considered. In monaural *and* binaural forward masking experiments considered in Chapter 3, for example, the temporal decay of masking was shown to be dominated by the monaural “channel” for large temporal distances between masker and probe tone and by the binaural “channel” for short distances.

## 4.2 Method

### 4.2.1 Procedure and subjects

Binaural masked tone thresholds were measured (when measurements were performed) and simulated using an adaptive, three-interval forced-choice (3IFC) procedure. The noise masker was presented with defined interaural parameters in three consecutive intervals separated by silence intervals of 400 ms. In one randomly chosen interval the test tone was added. The other intervals were leaved unchanged. The subject’s task was to specify the interval containing the test tone. Within a track the level of the test tone for each trial was determined according to a 1-up 2-down algorithm which results in an estimate of the test tone level necessary for 70.7% correct responses (Levitt, 1971). The step size was 8 dB at the beginning of the experiment and divided by two after every two reversals until it reached the minimum step size of 1 dB. At this value it was fixed. Thresholds were defined as the median value of the subsequent eight reversals. Trial-by-trial feedback was presented during the measurements. The procedure was repeated at least three times for each signal configuration and each subject. All figures show the median and interquartile ranges base on these measurements. All subjects had experience in psychoacoustic measurements and had normal hearing and were between 29 and 34 years old.

The results from simulations are represented by their medium value and interquartile range from at least three repetitions of every simulation.

### 4.2.2 Apparatus and stimuli

The acoustic stimuli were digitally generated at a sampling frequency of 22.05 kHz and transformed to analog signals with a two-channel 16-bit D/A converter. The stimuli were low-pass filtered and dichotically presented via headphones (Stax SR-A headphones with a Stax headphones preamplifier) in a soundproof booth. The

generation and presentation of the signals was performed and controlled by a SUN-Workstation using a signal-processing software package developed at the Drittes Physikalisches Institut in Göttingen.

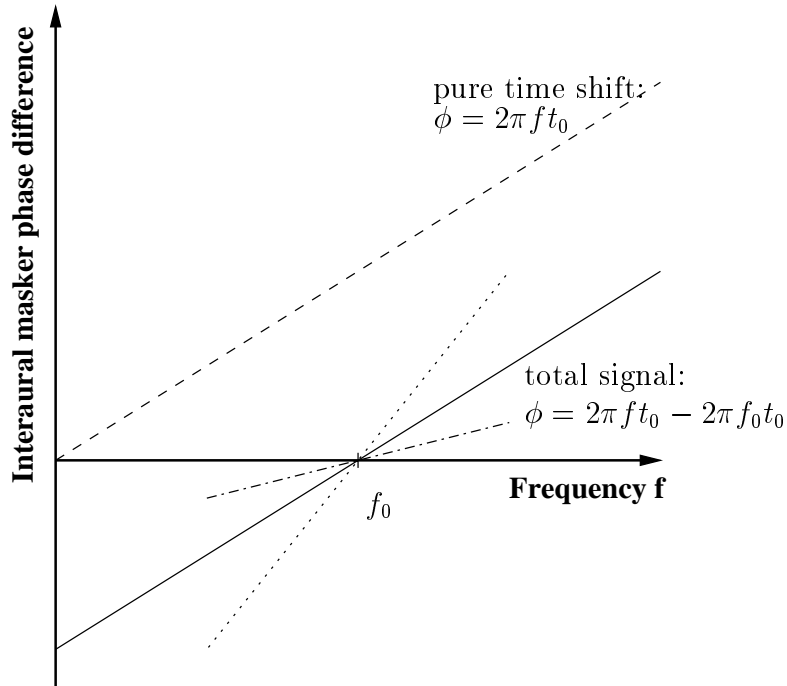
## 4.3 Results: Influence of spectral resolution on binaural detection

### 4.3.1 Noise maskers with combined interaural time and phase differences

An experiment similar to the one described by Rabiner et al. (1966) was performed and predicted with the model. An  $S_\pi$  tone was presented in a noise masker with an interaural time difference  $\tau = t_0$  and an interaural phase difference  $\phi = -2\pi f_0 t_0$  ( $f_0$  is the signal frequency), which means that for the test tone frequency the total interaural phase difference is zero. This configuration is denoted as  $N_{\tau=t_0, \phi=-2\pi f_0 t_0} S_\pi$ . The noise masker was presented at an overall level of 73 dB SPL. Figure 4.1 illustrates the interaural phase relations in terms of the interaural phase for the noise masker. A pure time shift would result in an interaural phase indicated by the dashed line. The solid line is the total interaural phase difference for the noise masker resulting from the combination of ITD and interaural phase shift. The zero internal delay which results in an optimal cancellation of the masker is best for the components of the test tone frequency  $f_0$ , but is suboptimal for frequencies different from  $f_0$ . Since all signal components are processed in the same channel where the same internal time delay is applied, the components different from  $f_0$  are not completely canceled. Thus, a large external time delay reduces the spectral range around  $f_0$ , where the binaural mechanism works efficiently. The BMLD decreases with increasing delay. An example for a larger external delay and the resulting interaural phase difference is given by the dotted line in Figure 4.1. The dashed-dotted line shows interaural relations for a time delay closer to zero.

Figure 4.2 shows the binaural masked thresholds as a function of the external time delay for three subjects and the corresponding simulations. Thresholds have been measured for ITDs from 0 to 7 ms in steps of 1 ms. At the signal frequency 500 Hz this means an interaural phase shift of 0 to  $-7\pi$  in steps of  $\pi$ . The reference  $N_0 S_0$ -thresholds for all subjects and from the simulations are also included in the Figure 4.2. The results are represented by their medium value and interquartile range from at least three repetitions of each measurement and five repetitions of each simulation.

The difference between the threshold functions for the three subjects lies between 3 and 5 dB. The interquartile range for all subjects is smaller than 3 dB. Thresholds from human listeners decrease with increasing ITD and tend to take an asymptotic course against a constant BMLD for large interaural delays. The observed threshold functions agree quite well with the results from Rabiner et al. (1966) (who used a

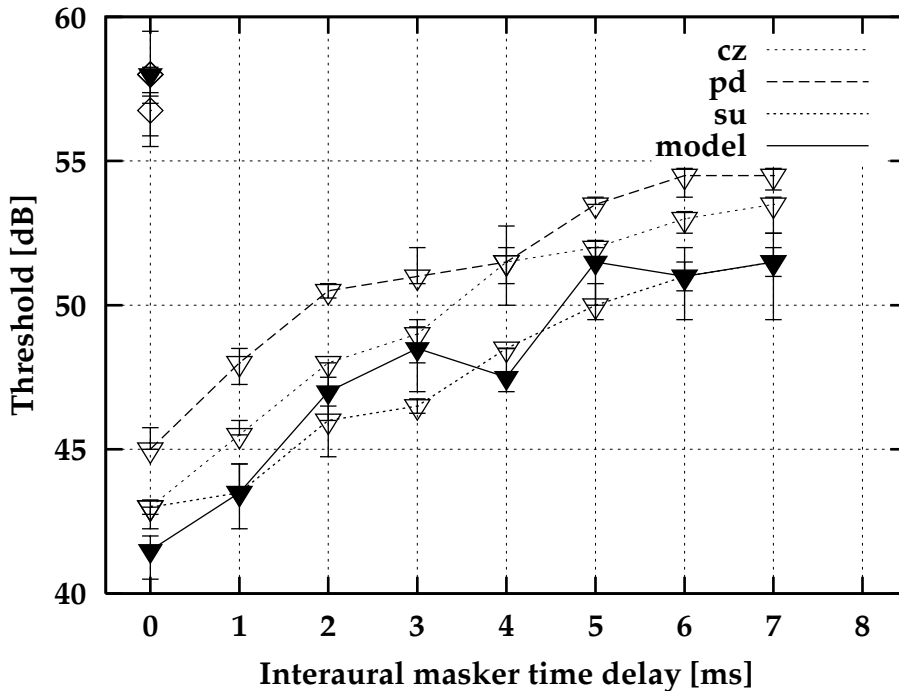


**Figure 4.1:** Illustration of the interaural masker phase difference for the configuration  $N_{\tau=t_0, \phi=-2\pi f_0 t_0} S_\pi$ . The dashed line shows the phase difference for a pure ITD  $\tau = t_0$ . An additional phase shift results in a zero phase difference for the component equal to the signal frequency (continuous line). The dotted line shows the phase difference for a larger ITD, the dashed-dotted line from a smaller ITD than  $t_0$ .

different threshold estimation procedure in continuous noise at a different level as in the present study) and the assumption of a reduced efficiency of the binaural noise reduction with increasing external delay. The simulated thresholds show a good correspondence with the experimental data except for the outlier at an ITD of 4 ms. It matches appropriately the threshold of the subject with the lowest thresholds. Compared to the mean threshold values of all subjects, the simulated thresholds are slightly too low, i. e. the BMLD is slightly overestimated.

### 4.3.2 Noise maskers with frequency dependent interaural phase differences

Kohlrausch (1988) estimated the auditory filter shape from binaural detection experiments with noise maskers exhibiting an interaural phase difference of 0 below and of  $\pi$  above a transition frequency of  $f_{tr} = 500$  Hz (and vice versa). The test signal frequency was varied between 200 and 800 Hz and was presented binaurally in antiphase,  $S_\pi$ . These configurations are noted as  $N_{0\pi} S_\pi$  and  $N_{\pi 0} S_\pi$  respectively. Results from these configuration are compared with results from  $N_0 S_0$  and  $N_0 S_\pi$

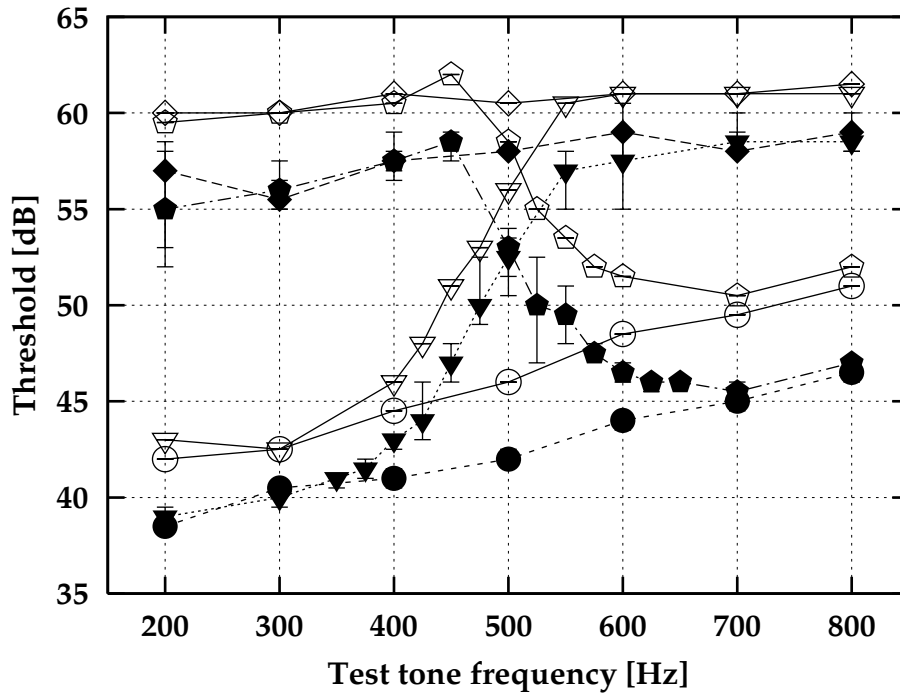


**Figure 4.2:** Thresholds for the configuration  $N_{\tau=t_0}, \phi=-2\pi f_0 t_0 S_\pi$  (downward triangles) for three subjects (open symbols) and predictions from simulations (filled symbols). Reference thresholds ( $N_0 S_0$ ) are indicated by diamonds. Values for the three subjects are: su 56.75 dB, pd and cz 58 dB.

configurations using the same probe signals and monaural masker spectra.

Figure 4.3 shows the experimental data from Kohlrausch (1988) (open symbols) compared with results from simulations (filled symbols). For the medium value and the interquartile range within simulations, nine repetitions of every simulation have been considered.

Experimental threshold curves for the configurations  $N_{\pi_0} S_\pi$  and  $N_{0\pi} S_\pi$  show an asymptotic behavior for signal frequencies far from the transition frequency 500 Hz:  $N_{\pi_0} S_\pi$  thresholds approach  $N_0 S_0$  curves for small frequencies and  $N_0 S_0$  curves for large frequencies. For  $N_{0\pi} S_\pi$  the relations are just the other way round. The transition between both curves takes place roughly between 400 and 600 Hz. Apart from a general shift of the threshold curves (predicted thresholds are between three and five dB lower than thresholds of human subjects) and a slightly enlarged BMLD in the simulations for the higher frequencies, the correspondence between experimental data and predictions is very good: the shape of the simulated threshold curves agrees with the experiments. Especially the slope and the range in which the transition of thresholds takes place for the  $N_{\pi_0} S_\pi$  and  $N_{0\pi} S_\pi$  curves is predicted quantitatively by the present model.

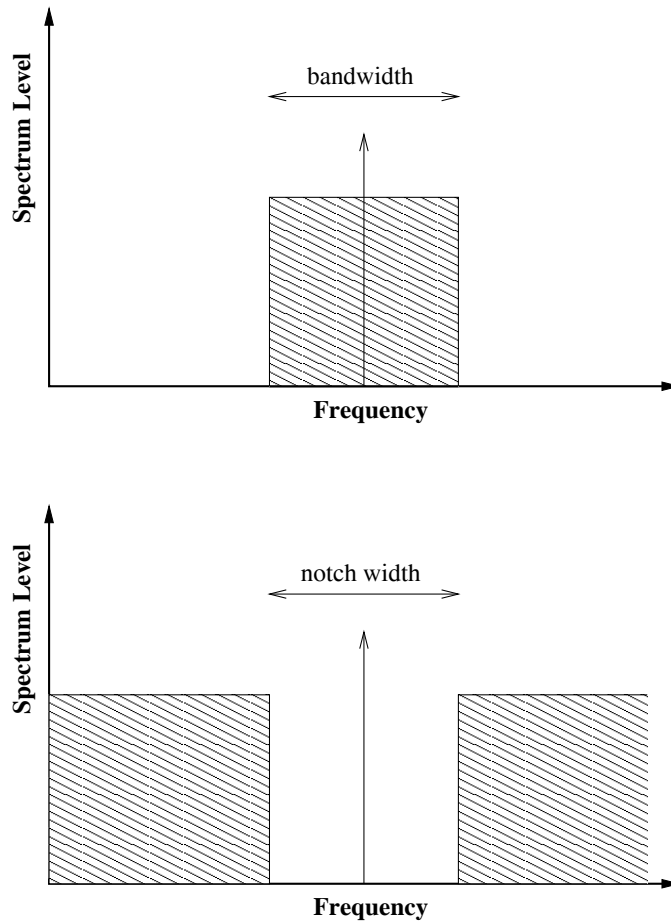


**Figure 4.3:** Binaurally masked thresholds for the configurations  $N_0S_0$  (diamonds),  $N_0S_\pi$  (circles),  $N_{\pi 0}S_\pi$  (pentagons) and  $N_{0\pi}S_\pi$  (downward triangles). Experimental results (open symbols) from Kohlrausch (1988) and results from simulations (filled symbols).

### 4.3.3 Spectral integration in binaural detection experiments

Hall et al. (1983) measured  $N_0S_0$  and  $N_0S_\pi$  thresholds in notched noise and bandpass noise configurations that are schematically plotted in Figure 4.4. In order to test the role of information processing across frequency, all simulations of these experiments have been performed considering the output of multiple off-frequency auditory filters that are adjacent to the one with the center frequency equal to the signal frequency and are regularly arranged on an ERB-scale.

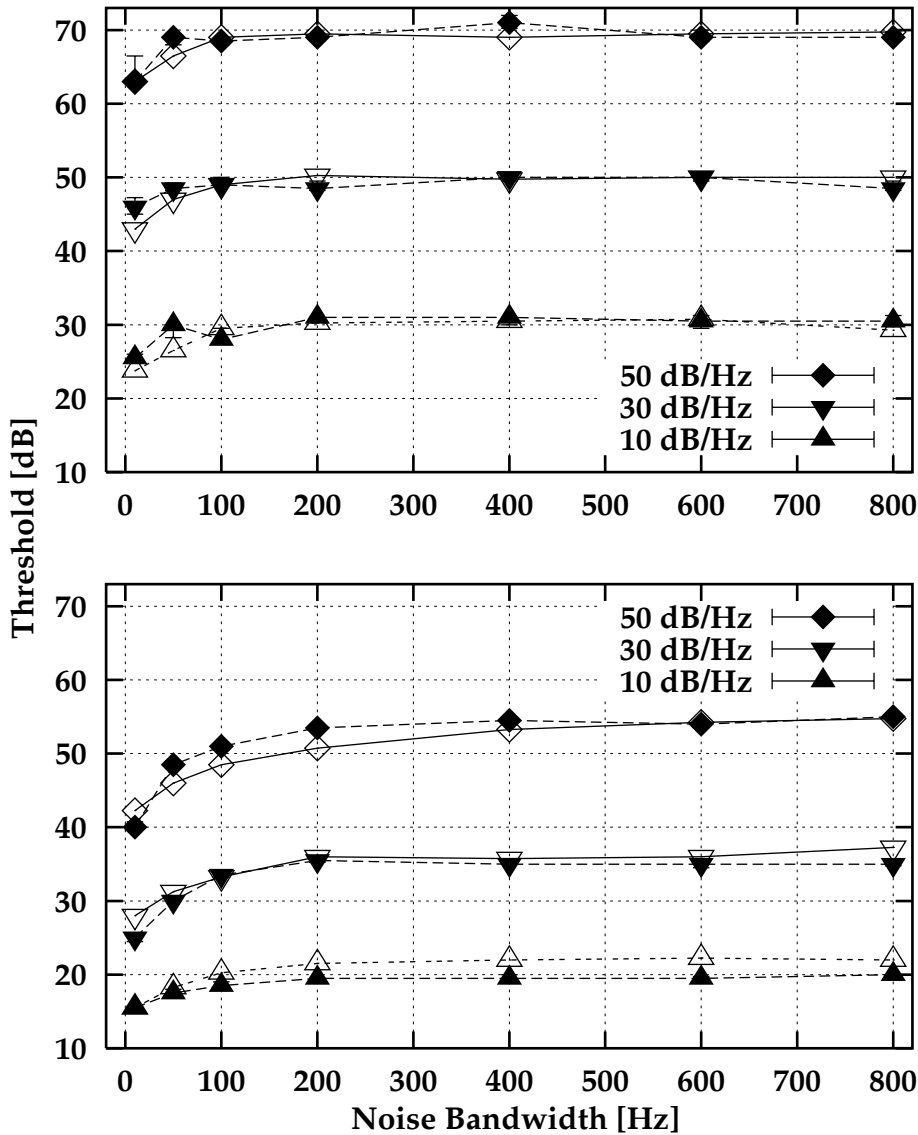
Figure 4.5 shows the  $N_0S_0$  and  $N_0S_\pi$  thresholds from the study of Hall et al. (1983) and the corresponding predicted thresholds in the bandpass noise configurations. Thresholds were measured for three different spectrum levels of the masker (at 10, 30, and 50 dB/Hz). The figure shows thresholds of the sinusoidal test signal in the presence of a bandpass-noise spectrally centered at the signal frequency (cf. Figure 4.4, lower panel) as a function of the masker bandwidth. The upper panel shows thresholds from the diotic condition ( $N_0S_0$ ), the lower panel for binaural conditions ( $N_0S_\pi$ ). Different symbols indicate different masker spectrum levels: diamonds 50 dB/Hz, downward triangles 30 dB/Hz and upward triangles 10 dB/Hz. In the monaural condition, thresholds increase with increasing masker bandwidth



*Figure 4.4: Spectral arrangement of the noise masker and the test tone in the bandwidthing experiment (upper panel) and in the notched noise condition (lower panel). Parameters are the bandwidth and the notch width, respectively.*

up to the critical bandwidth. A further increase does not change thresholds. In the investigated masker level range, the critical bandwidth is level independent.

The three different masker spectrum levels are sufficiently above the absolute threshold so that the masked threshold increases in a linear way with increasing masker level. In the binaural condition, the general shape of the threshold curve agrees with those in the monaural condition. However, in contrast to the monaural condition, the derived critical bandwidth decreases with decreasing masker spectrum level. For the masker spectrum level of 50 dB/Hz, the slope of the threshold curve is steep at the beginning and decreases with increasing bandwidth, but the increase takes place for the whole range of presented masker bandwidths. With decreasing masker level the initial slope becomes smaller and the threshold curves reach a stationary value earlier (at 200 Hz bandwidth for a spectrum level of 30 and 10 dB/Hz). For the high levels and narrowband maskers the BMLD from human observers amounts



**Figure 4.5:** Experimental data from Hall et al. (1983) (open symbols) and results from corresponding simulations (filled symbols) for the experiment utilizing bandpass noise in an  $N_0S_0$  (upper panel) and an  $N_0S_\pi$  configuration (lower panel). Masked thresholds are plotted as function of the masker bandwidth. The parameter for the different threshold curves in each plot is the spectrum level of the Gaussian noise masker.

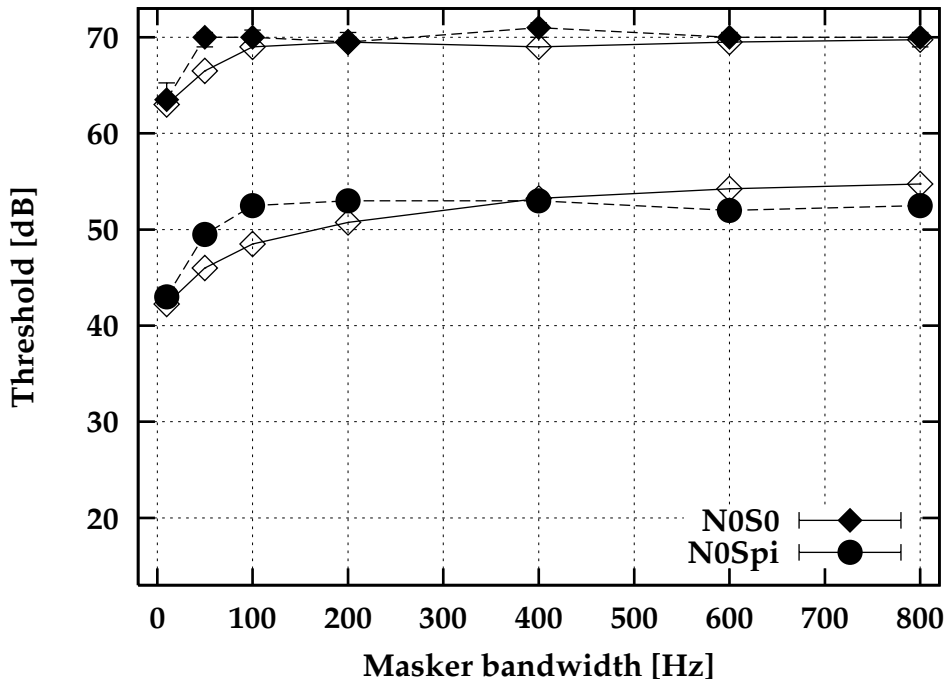
about 20 dB.

The model predictions agree quite well with the observed thresholds for the  $N_0S_0$  configuration, a slight deviation of experimental and simulated thresholds ( $\leq 3.5$  dB) can be observed at small bandwidths below 100 Hz.

For the simulations of the  $N_0S_\pi$  configuration, the general correspondence be-

tween simulations and experiments is satisfying. Measured and simulated thresholds are in good agreement for the intermediate masker spectrum level. However, threshold curves from simulations do not show the increase of the masker thresholds over the whole range of bandwidths for the highest spectrum level. The initial slope of that simulated threshold curve is steeper than for human observers and it reaches the stationary threshold value already for a bandwidth of 200 to 400 Hz. For the lowest masker spectrum level, the simulated thresholds are slightly too low for large masker bandwidths.

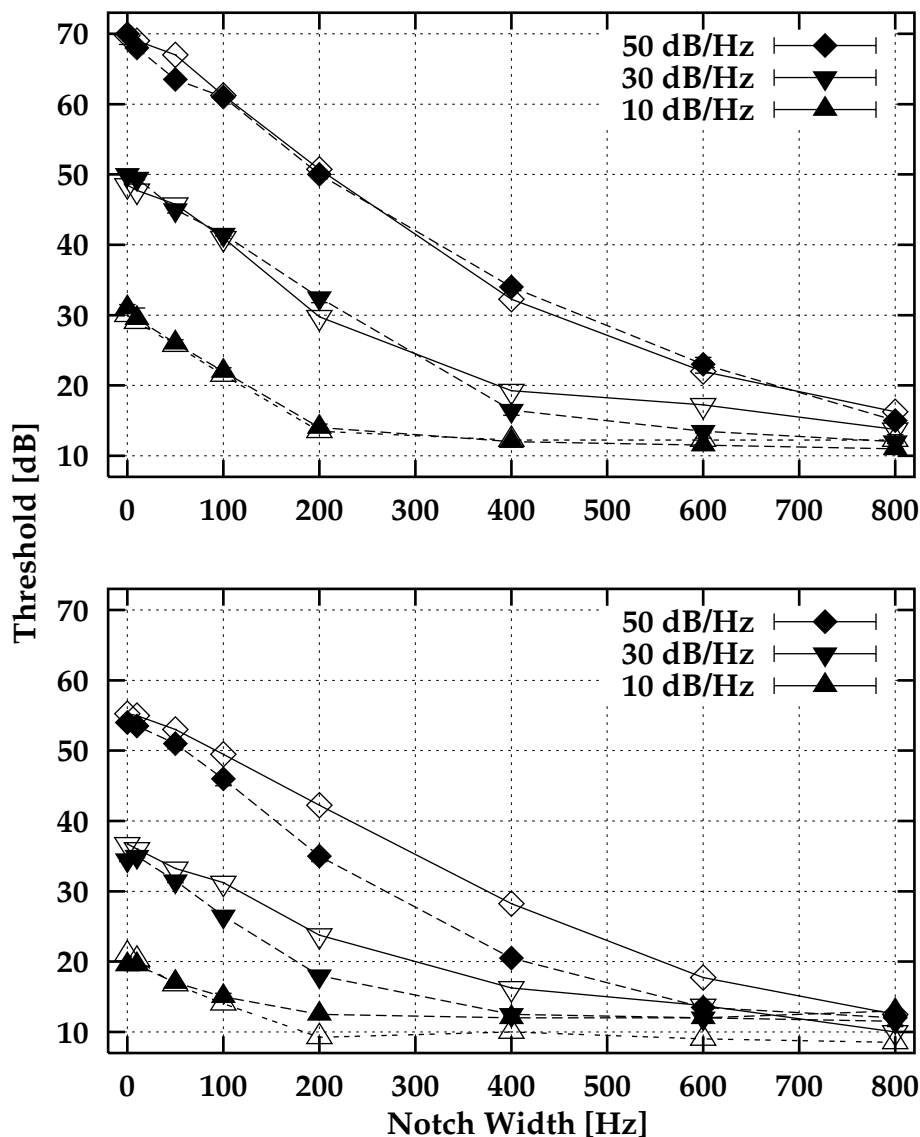
It should be remarked that the  $N_0S_0$ -simulations are repeated for the on-frequency auditory filter alone give the same results as the multichannel simulations shown above, which is not valid for the  $N_0S_\pi$ -configurations. To illustrate the role of off-frequency detection in the experiments considered here, Figure 4.6 compares results from measurements and simulations for the bandwidening paradigm in the  $N_0S_0$  and  $N_0S_\pi$  configuration at a masker spectrum level of 50 dB/Hz for the single on-frequency channel version.



**Figure 4.6:** Experimental data from Hall et al. (1983) (open symbols) and results from corresponding simulations (filled symbols) for the experiment utilizing band-pass noise in an  $N_0S_0$  and  $N_0S_\pi$  configuration at a masker spectrum level of 50 dB/Hz. Simulations were performed only in the on-frequency channel.

For the  $N_0S_0$  configuration, results from simulations are unchanged compared with the multichannel model. However, simulated thresholds for the  $N_0S_\pi$  configuration increase much faster than experimental thresholds and reach their stationary value

already at bandwidths of 100 Hz. Note that this bandwidth is much smaller than in the simulations performed in multiple channels.



**Figure 4.7:** Experimental data from Hall et al. (1983) (open symbols) and results from corresponding simulations (filled symbols) for the experiment utilizing notched noise in an  $N_0S_0$  (upper panel) and an  $N_0S_\pi$  configuration (lower panel). Masked thresholds are plotted as function of the masker notchwidth. The parameter for the different threshold curves in each plot is the spectrum level of the Gaussian noise masker.

Figure 4.7 shows the  $N_0S_0$  (upper panel) and  $N_0S_\pi$  (lower panel) thresholds for the notched noise configuration (cf. Figure 4.4, lower panel) with results from the

corresponding simulations. The same symbols as in Figure 4.5 indicate the different masker spectrum levels (again 50, 30 and 10 dB/Hz). For all masker levels thresholds decrease with increasing notch width in monaural ( $N_0S_0$ ) and binaural conditions ( $N_0S_\pi$ ). The slope of the threshold curves is nearly independent of the spectrum level. However, in the binaural condition a shallower slope as in the monaural condition is observed. Thus the BMLD decreases with increasing notch width from 0 to 400 Hz.

The correspondence between measured and experimental masked thresholds is very good for the  $N_0S_0$  configurations. The behavior of the simulated thresholds in the  $N_0S_\pi$  configurations deviates from the measured thresholds. In the binaural conditions, simulated threshold curves agree with the experimental thresholds only for no notch (0 Hz), which corresponds to a broadband noise masker. For large notch-widths simulated and experimental thresholds as well converge against the absolute threshold. However, in the range of notchwidths from 50 to 600 Hz, simulated threshold decrease faster than in the experimental data. Compared with the  $N_0S_0$  thresholds, the simulations show an approximately constant BMLD up to a notch-width of 400 Hz, whereas the BMLD for human observers decreases continuously with increasing notchwidth.

## 4.4 Influence of temporal processing on binaural detection

### Test-tone integration in binaural and monaural configurations

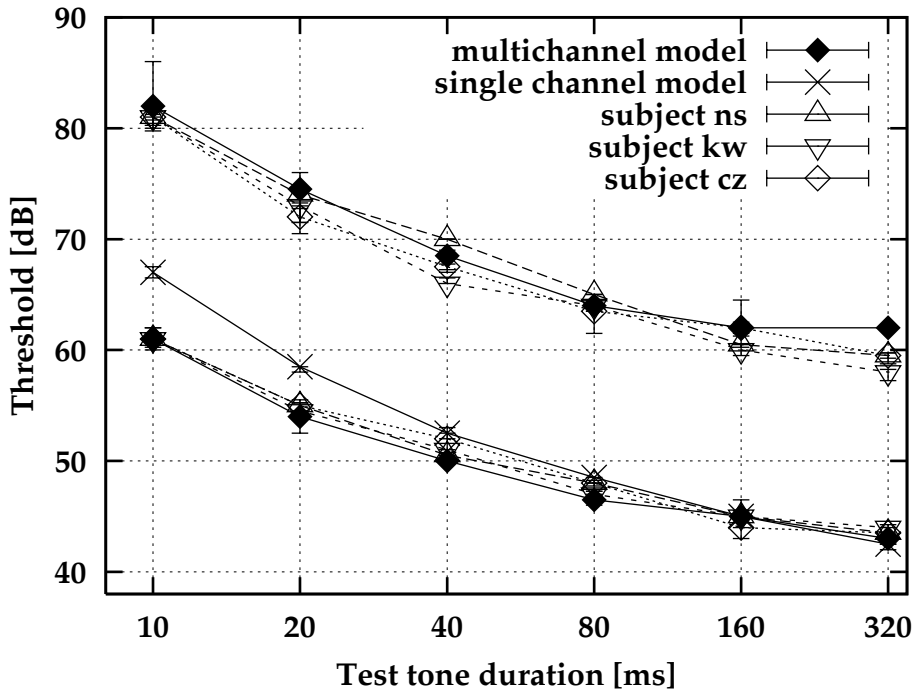
Masked thresholds of a 500 Hz sinusoidal test signal masked by diotic ( $N_0$ ) Gaussian noise were measured as a function of the signal duration (10, 20, 40, 80, 160, and 320 ms). The interaural phase of the test signal was 0 or  $\pi$ . The onset for all signals occurred 100 ms after the masker onset. The signals were windowed with 5 ms Hanning ramps to present the same on- and offset for all durations. The running noise masker had a duration of 500 ms including 10 ms Hanning ramps. It was digitally generated with a center frequency of 2.5 kHz and a bandwidth of 5 kHz at an overall level of 77 dB SPL. In the simulations the peripheral frequency on-frequency channel was used with four adjacent channels (two with a lower and two with a higher center frequency, all regularly arranged on an ERB-scale).

Figure 4.8 shows the level of the test signal at threshold as a function of the duration of the test signal on a logarithmic time scale. Results of model simulations are plotted with filled symbols, results from three subjects with open symbols. For the measured thresholds the inter-individual variability is smaller than 4 dB and the intra-individual variability is below 2 dB. Probably due to the statistics of the external noise variability within the simulations is somewhat larger, up to 6 dB. For the short signal durations the decrease in threshold is slightly steeper than the 3 dB per doubling the signal duration observed by Dau (1996). This is a result from

the different rise/decay ramps utilized in both studies since Dau imposed Hanning ramps over the whole duration of the signal, whereas in the present study short Hanning ramps with a constant duration were applied for all signal durations. The test signals longer than 10 ms employed in the present study contain more energy than those in Dau's study. Furthermore, they exhibit a stronger overshoot in their internal representation.

In both the  $N_0S_0$  and  $N_0S_\pi$  configuration thresholds decrease with increasing signal duration. The slope becomes shallower for long durations. This is in agreement with the data from the literature (see above). The difference between the unmodulated and measured data is smaller than 3 dB.

The model can replicate the test-tone integration over the whole range of signal durations for both the  $N_0S_0$  and  $N_0S_\pi$  configuration quite well. One deviation occurs for the longest durations in the  $N_0S_0$  configuration, where simulations have already reached a stationary level. The measured thresholds still show a decrease of 2 dB.



**Figure 4.8:** Masked thresholds of a signal presented in running noise as a function of the duration of the test tone. Experimental results for three subjects are plotted different with open symbols, results for simulations performed with the multichannel model are plotted with filled symbols. Simulations performed only within the on-frequency channel are plotted with the  $\times$ -symbol.

When the simulations are repeated with the single on-frequency channel model, predicted thresholds for the  $N_0S_0$  configuration remain essentially unchanged. In the

$N_0S_\pi$  configuration (represented by the  $\times$ -symbol in Figure 4.8), predicted thresholds are about 6 dB higher for the smallest signal duration and about 4 dB for the 20 ms signal. This shows that in these adjacent frequency channels information about short signals is present after binaural processing. If the same signal-to-noise ratio had been present in all five channels considered in the simulations, this would have resulted in a threshold of about 7 dB below the corresponding threshold for a single-channel model. This difference approximately holds for the smallest signal duration whereas the smaller difference for larger durations are due to smaller signal-to-noise ratios in the off-frequency channels in comparison with the on-frequency channel.

## 4.5 Discussion

### 4.5.1 Information processing within one critical band

The binaural model of the present study is able to account for effects in binaural masking, where the spectral resolution or a combination of spectral resolution and spectral integration of the auditory system is the limiting factor for the performance in the detection task. The experiments described in Sections 4.3.1 and 4.3.2 and corresponding simulations concerning spectral resolution in binaural detection tasks were performed in broadband noise where off-frequency detection provides no advantage over detection in the on-frequency channel, i. e. the peripheral filter centered on the respective signal frequency. Hence, these experiments can be described by only considering one peripheral auditory filter in the model.

The “classical” way of modeling these experiments is to assume an “effective” peripheral filter shape that operates only for the binaural conditions and to consider the “internal” power spectrum of the masker at the output of the EC-mechanism. Since a frequency-varying interaural correlation produces an appropriate spectral shape at the output of the EC-mechanism, the binaural masked threshold can be predicted by this approach by assuming a certain signal-to-noise ratio necessary for detection.

In the framework of the EC-theory Durlach (1972) estimated from the results of the experiment performed by Rabiner et al. (1966) which is described in Section 4.3.1 a critical bandwidth, which is in reasonable agreement with monaural critical bandwidths for the frequencies employed. In the present study the approach was to assume a peripheral filterform with a bandwidth adjusted to monaural masking conditions and to test whether an appropriate model incorporating such a filter is able to predict the results in this binaural experiment. The simulated threshold function reaches its stationary part slightly earlier than the measured functions, indicating that critical bandwidths in the model description are slightly overestimated. This is due to the fact that for broader auditory filters already a relatively small combined interaural time and phase shift within the channel would result in large interau-

ral phase differences for off-frequency components that can not be canceled by the binaural processor.

The same approach described above was used for the second experiment described in Section 4.3.2 by Kohlrausch (1988) to estimate an auditory filter shape (assuming that masked thresholds are only dependent on the interaural cross correlation of the masker within the filter band at the signal frequency). He estimated an equivalent rectangular bandwidth of 80 Hz for an auditory filter in the spectral region around the transition frequency of 500 Hz (assuming that its bandwidth and shape does not depend strongly on the center frequency), which is close to estimates of monaural critical bands at that frequency (e. g. Moore and Glasberg, 1987).

Within the present model of binaural signal processing the argumentation is similar, only that as binaural cue the interaural correlation of the output of the auditory filter centered at the test tone frequency has to be replaced by the output signal of the binaural processor, which depends upon the effective interaural correlation in the corresponding auditory filter.

#### 4.5.2 External and internal noise statistics in multichannel processing

In some experiments large differences between monaural and binaural frequency resolution are observed. In the present study this difference was shown for the bandwidening experiment, whereas in the other experiments the same bandwidth was observed for both monaural and binaural conditions (see above). The predictions are expected to be very close to the “classical” model prediction if only stationary conditions with broadband maskers and narrowband target signals are considered. Different explanations have to be considered as soon as maskers are employed with a strong frequency-dependent spectral shape (such as those conditions condition employed in the experiments described in Section 4.3.3).

To account for the wider binaural critical bandwidths in the bandwidening configuration, Hall et al. (1983) suggest that notched noise and bandpass noise data reflect different aspects of binaural frequency resolution. In wideband noise they assume that the binaural system is not only influenced by the central auditory filter but also by adjacent filters, i. e. a combination of information across peripheral filters. Information from these adjacent filters include no interaural differences and is detrimental for detecting the  $S_\pi$  signal. Their influence and the number of activated filters involved in this process are reduced by narrowing the noise bandwidth. So a larger proportion of activated auditory filters will indicate the presence of interaural cues which improves the detectability of the signal.

This explanation cannot account for both bandwidening experiment and notched-noise experiment since in the latter condition the estimated critical bandwidth matches approximately the monaural critical bandwidth. Hall et al. (1983) argued, that the shallower slope in notched noise is simply a consequence of a reduced BMLD for low masker levels. However, assuming a different mechanism op-

erating only in notched-noise experiments seems to be in contradiction to studies that estimated binaural critical bandwidths using noise maskers with frequency dependent interaural parameters already mentioned above, in which a relative good correspondence between binaural and monaural critical bandwidths turns out.

Kohlrausch et al. (1998b,a) considered the role of external vs. internal noise statistics employed in multichannel detection. They explained the difference in the estimated critical bandwidth in dependence from the experimental paradigm with the different cues available in monaural (or diotic) and binaural detection tasks due to the different influence of the masker variability in both configurations. According to Bos and de Boer (1966), *monaural* signal detection is based on changes in the overall energy of the masker and it is limited by the inherent fluctuations. Assuming the auditory system performing an averaging mechanism, increasing noise bandwidth leads to a better estimate of the masker energy. Assuming that on the other hand binaural detection is based on a change in interaural correlation, the statistics of the external masker should not influence the detection cue in these situations. Following these arguments, binaural detection can be considered as only limited by an amount of internal noise.

As a second important difference between monaural and binaural detection Kohlrausch et al. (1998b,a) propose the combination of information from different auditory filters if a binaural narrowband masker is used at a sufficiently high masker level. Due to spread of excitation in narrowband masker conditions in adjacent auditory filters the change of masker correlation is available as a cue. Assuming that the internal noise limiting binaural detection is independent for the various auditory filters, combining information from several auditory filters is advantageous for detection since it reduces the influence of the internal noise. These explanations do not hold for a binaural broadband masker since the change of correlation in a remote auditory filter becomes much smaller than for a narrowband masker. Thus, the decreasing BMLD with increasing bandwidth is a consequence of the decreasing change of correlation in adjacent filters. In the monaural condition, signal information from adjacent auditory filters is not advantageous for detection for both narrowband and broadband maskers, since the detection-limiting external noise fluctuations are highly correlated across auditory filters. Thus, a combination of information across these filters would not result in a statistical advantage.

These arguments can be transferred into the framework of the present model, keeping in mind that binaural cues in the model are not expressed in terms of interaural correlation but in the size of the output of the binaural processor which is itself strongly depending on the interaural correlation. In the course of binaural signal processing of the present model, the external masker is canceled or at least strongly reduced. This eliminates or reduces also the external statistics as limiting factor for binaural detection in narrowband conditions, so that the argumentation above remains also valid within the framework of the model.

In the framework of the present study binaural test-tone integration is investigated in the same context of integration of temporal information across *spectral*

channels. We assume, like it was discussed above for detection in narrow bands of noise, that internal rather than external noise restricts the performance in binaural detection experiments.

To account for the different properties of temporal processing in monaural and binaural configurations, a modulation filterbank was introduced in the monaural channel. Envelope fluctuations of the stimuli are considered by analyzing them after monaural preprocessing in a modulation filterbank at the place of the integration low-pass filter. Dau et al. (1997a) showed that such an expanded version of the monaural model of auditory signal processing improves the agreement between simulations and measurements of thresholds for monaural test-tone integration<sup>1</sup>. The binaural processor was left unchanged, i. e. temporal integration was described by the same low-pass filter with a cut-off frequency of 8 Hz as in the simulations before. This corresponds to the minimal assumption that interaural stimulus fluctuations are only evaluated in the lowest modulation filter<sup>2</sup>.

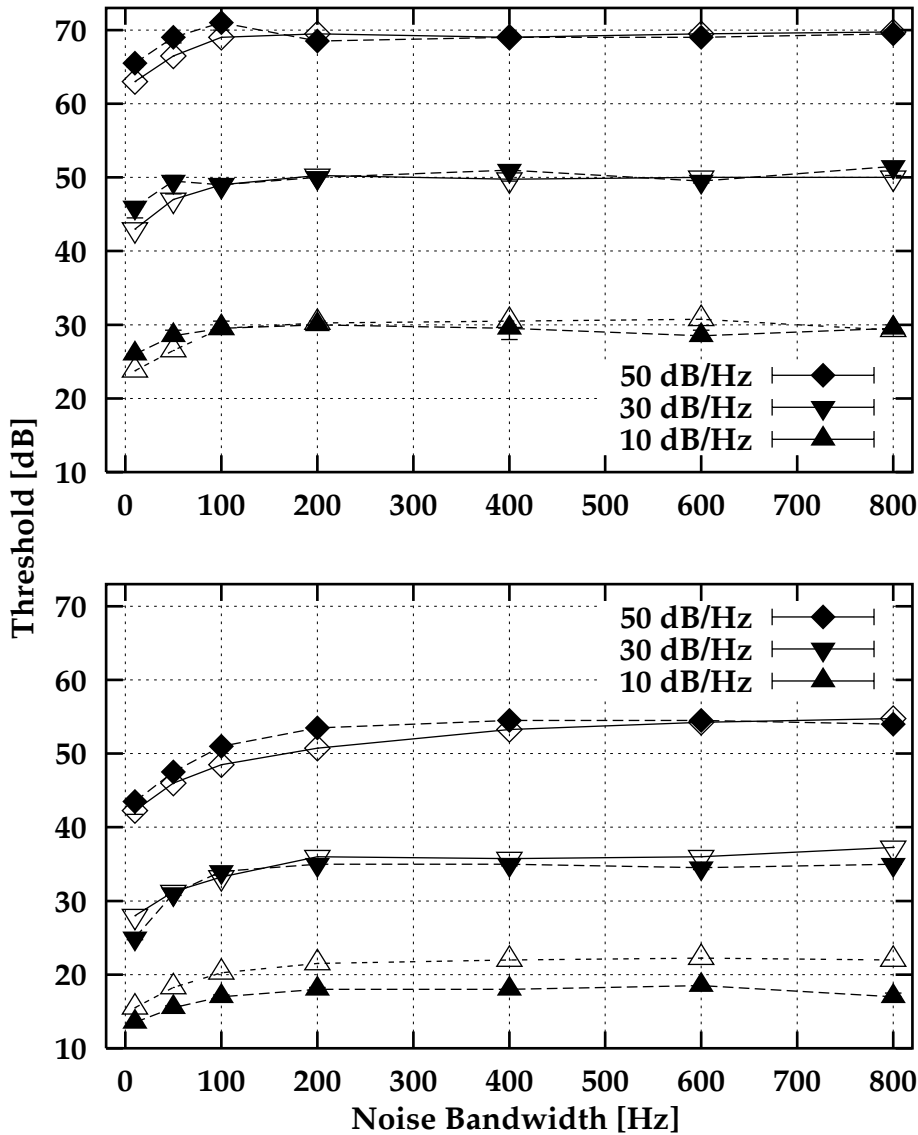
Different reasons may account for the discrepancies between predicted and measured  $N_0S_\pi$  thresholds from the detection experiment using the notched noise paradigm and, to a smaller degree, using the bandpass noise.

One possible reason for the discrepancy may be the shape of the auditory filters chosen in this study. Auditory filter shapes estimated using the notched-noise method have been characterized as having a sharply tuned “tips” and broader “tails” (Patterson et al., 1982; Glasberg et al., 1999). Patterson et al. (1982) used the sum of two rounded-exponential functions to describe both sides of the auditory filter. These wider tails are not considered by the description of the auditory filters within the model. To investigate the influence of wider tail on the predicted thresholds for the bandwidening and the notched-noise configurations, both experiments have been repeated using a filterbank of 3<sup>rd</sup> gammatone filters instead of 4<sup>th</sup> order filters. The equivalent rectangular bandwidth according to Equation 3.1 was kept constant. These filters are chosen since they represent the required properties of wider tails and a sharper tip. The results of the simulations are plotted together with the experimental data in Figures 4.9 (bandpass noise) and 4.10 (notched noise). For the simulations of the bandpass noise experiment the agreement with experimental thresholds is slightly better than in the simulation performed using the 4<sup>th</sup> order filters (Figure 4.5). For small masker bandwidths the slope of the simulated threshold curve is less steep and the increase in threshold is predicted over a larger range of bandwidths than for the simulations performed with the 4<sup>th</sup> order gammatone filters, indicating the shallower slope of the 3<sup>rd</sup> order filters.

---

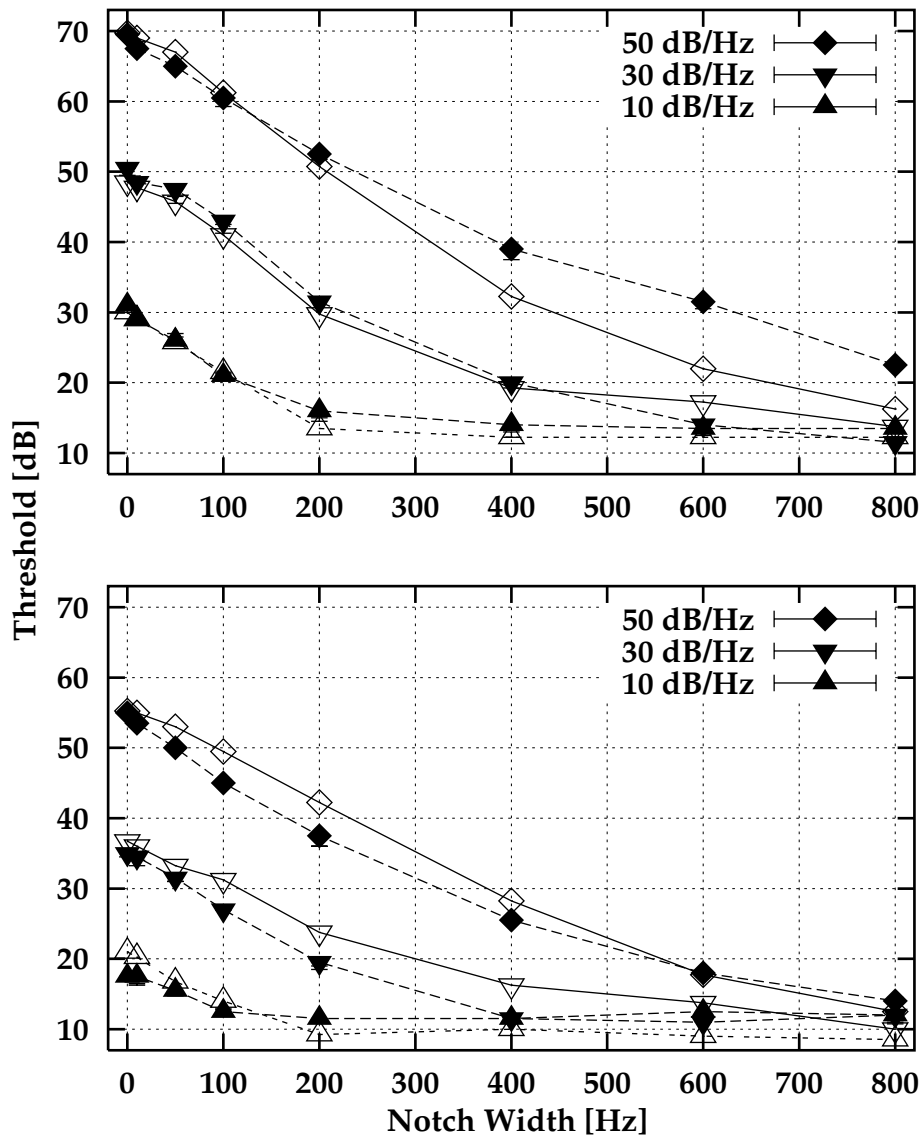
<sup>1</sup>The modulation filterbank was originally developed to account for modulation detection and masking data

<sup>2</sup>Studies about binaural modulation are relatively scarce. Grantham and Bacon (1991) investigated binaural modulation masking and concluded that at higher modulation rates the task is performed by listening for interaural decorrelation rather than changing IID cues. Since interaural correlation has to be calculated over a certain temporal window, taking into account only the longest time constants may be sufficient to describe binaural detection.



*Figure 4.9: Simulations of the same experimental bandwidening configuration as in Figure 4.5. Simulations were performed with a gammatone filterbank of 3<sup>rd</sup> order filters.*

Also the correspondence between simulations and experimental data for the  $N_0S_\tau$  configuration using gammatone filters of 3<sup>rd</sup> order in the notched noise configuration is better than for the original form of the auditory filters. Results from simulations are plotted in Figure 4.10. In contrast to this reconciliation of both threshold curves the simulated and measured  $N_0S_0$  thresholds differ significantly for the highest masker spectrum level and large notch widths (above 200 Hz). Again, in these simulations the BMLD remains constant over a large range of notch widths and



*Figure 4.10: Results of simulations and experimental data of the same notched noise experiment as described in Figure 4.7. Simulations were performed with a gammatone filterbank of 3<sup>rd</sup> order filters.*

masker levels. Since changing the filter shape results always in a shift of both  $N_0S_0$  and  $N_0S_\pi$  thresholds, also simulations employing alternative filter shapes, such as a linear combination of filters with different slope, e. g. 4<sup>th</sup> and 2<sup>nd</sup> order with the same center frequency are not able to predict (the results for both conditions at the same time. the simulations are not shown here). The deviation of simulated and experimental threshold curves is transferred from the binaural to the monaural curve. We conclude that the peripheral filter form alone is not responsible for the deviation.

Further mechanisms responsible for the decreasing BMLD with increasing notch width could be a contralateral interaction (based e. g. on interaural correlation), which causes an effective change in the auditory filter shape in conditions when interaural differences are occurring.

Zurek and Durlach (1987) interpreted the wider “operational” critical bands found in these experiment within the “classical” model in terms of spectral interference or dominance effects. That means that there exist dominant spectral regions when conflicting interaural information is presented to the auditory system. Furthermore they argued that interaural parameters fluctuate at a faster rate with larger masker bandwidths. So a decreasing BMLD with increasing noise bandwidth can be interpreted as evidence of the binaural system’s smoothing over rapid fluctuations in interaural cues (binaural “sluggishness”, where this notion has to be understood as a sluggishness of signal processing in the binaural canal).

It is also possible, that the mechanism of binaural noise reduction works less effective with decreasing masker energy within a critical band. In the description of the signal processing of the present model, the exact internal delay is calculated from the position of the maximum of an *averaged* version interaural cross-correlation function of the internal representations of the masker signals. That means, that the estimate for position of the maximum has a small error, decreasing with the number of representations over which the pattern of the ICF is averaged. Therefore, assuming that the calculation of the value of the optimal internal time delay happens separately for each representation of the masker (e. g. from a running cross-correlation function) it becomes less accurately defined. A second mechanism which enlarges the error in performing the internal delay, is the peripheral noise. With decreasing masker level, the influence of its internal decorrelation increases. Altogether, the maximum of the ICF and so of the value of the internal time delay becomes a stochastic variable, whose standard deviation increases with decreasing masker level.

## 4.6 Summary

Results from binaural masking experiments (from new measurements and from the literature) were compared with model simulations of the same experiments. The model of binaural signal processing described in Chapter 3 that combines both concepts of monaural and binaural signal processing in the auditory system is able to account for experimental results from binaural unmasking experiments where the masker is stationary broadband noise. In this configuration detection can be restricted to the on-frequency channel. For bandlimited noise, agreement between experiments and simulations becomes better when additional adjacent off-frequency channels are considered in the simulations. In the same way binaural test-tone integration can be described as spectral integration over adjacent off-frequency channels which is possible due to spectral splatter of the short test signals. Simulated binaural thresholds for tones in notched noise are too low for the range of medium notch widths.

# Chapter 5

## Binaural sluggishness

### Abstract

*The model of functional binaural signal processing described in the previous chapter in its present form does not account for experimental configurations with time varying interaural parameters that exhibit a “binaural sluggishness”. Therefore, a modification of the binaural processor is proposed that includes a sluggishness in the strategy employed for binaural noise reduction. This expanded model shows the temporal behavior of experimental threshold curves. Additionally, the effect of a sluggishness in the evaluation of fast fluctuating interaural differences is investigated in the simulation of notched noise and bandwidening experiments.*

## 5.1 Introduction

To investigate the transient properties of the binaural system compared with the monaural system to one single, nonperiodic transition in the interaural phase of the noise masker, Kollmeier (1986) and Kollmeier and Gilkey (1990) performed various binaural detection experiments analogous to monaural forward and backward masking situations. In a typical configuration, they observed the detectability of a short tonal test signal in dependence to the relative position to an instantaneous interaural phase change of the masker from  $N_\pi$  to  $N_0$ . The signal was a short  $S_\pi$  tone. Thresholds from this  $N_\pi N_0 S_\pi$  configuration were compared with  $N_\pi S_\pi$  thresholds, where the masker phase was held constant, but the masker level was lowered at a fixed time by 15 dB (noted as  $N_\pi N_\pi(-15\text{ dB})S_\pi$ ). The response to a change in overall level is assumed to be governed by the properties of the temporal processing of the monaural system. Since the level is not changed in the  $N_\pi N_0 S_\pi$  configuration, Kollmeier and Gilkey assumed that performance is governed by the binaural system. Kollmeier and Gilkey (1990) investigated their results in the framework of the EC-theory in terms of effective interaural correlation within a temporal window which integrates binaural information. They found that time constants attained from fitting the experimental data are a factor 2–4 higher than time constants from the corresponding  $N_\pi S_\pi$  experiment.

They termed the extended temporal range of the binaural response as “binaural sluggishness”. This is in accordance with earlier studies of Grantham and Wightman (1979) (see review by Holube et al. (1998)). They investigated the detectability of a short phase-inverted ( $S_\pi$ ) test signal in noise with a non-stationary interaural correlation. The noise maskers include no monaural changes. The interaural correlation was sinusoidally varied between -1 and 1. At low interaural variation rates the BMLD follows the “instantaneous” interaural correlation within the range of the temporal position of the test signal (the BMLD was large for interaural correlations close to 1 and 0 for correlations near -1). With increasing rate the difference between BMLDs for positive and negative interaural correlations decreases and disappears for interaural variation rates above 4 Hz, which is a much lower cut-off frequency as obtained in monaural experiments with amplitude-modulated stimuli. Grantham and Wightman (1979) estimated from these results a “minimum binaural integration time” of 44–243 ms.

This chapter investigates how this dynamic properties of the binaural system can adequately be described by the model of the “effective” binaural signal processing introduced in the previous chapters. The binaural model accounts also for some temporal effects occurring in binaural detection experiments. As it has been demonstrated in Chapter 3, it can replicate thresholds in binaural forward masking experiments. If the output for multiple frequency channels is considered, it predicts very well thresholds and BMLDs in binaural test-tone integration.

In the literature sometimes the terms of a binaural “strategy sluggishness” (Holube et al., 1998) or “analyzer sluggishness” (Kollmeier and Gilkey, 1990) are used.

They describe the detection strategy in more complex binaural configurations, e. g. with dynamic localization or in detection experiments with a masker with periodic interaural correlation. A binaural processor has to monitor the output of several channels according to different internal delays. Establishing and comparing the distribution of the output of multiple channels might result in a sluggish performance of the total binaural system.

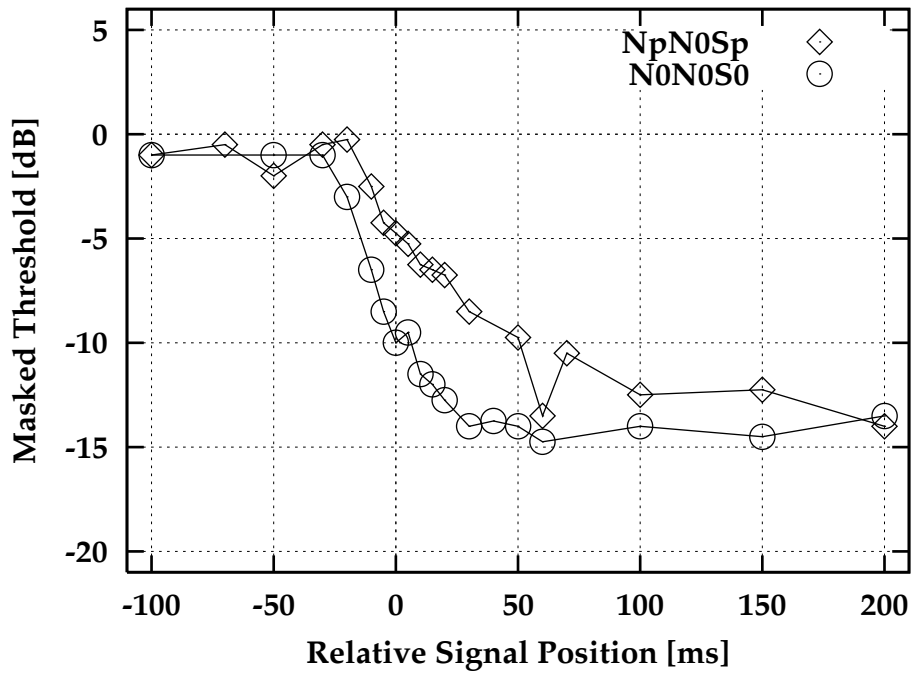
Since in contrast to many other binaural model the model employed in the present study is a model of temporal signal processing, it allows to test the hypothesis of Zurek and Durlach (1987) who assumed that the binaural system is sluggish in the evaluation of fast fluctuating interaural differences to account for results from binaural masking experiments in bandwidening configurations.

## 5.2 Results

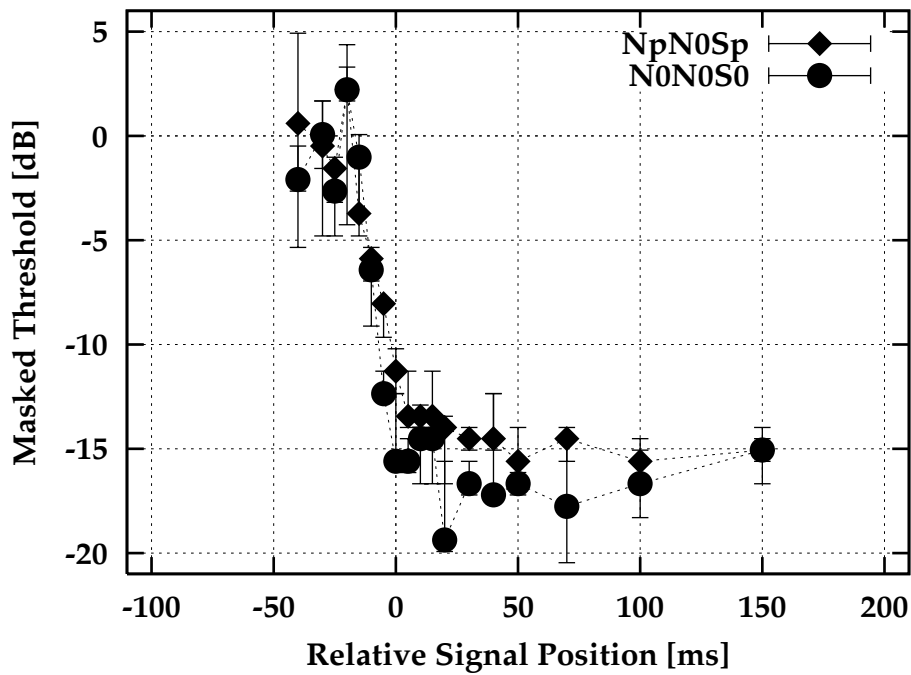
Kollmeier (1986) investigated the temporal resolution of the binaural auditory system in detection experiments for the transient configurations  $N_\pi N_0 S_\pi$  and  $N_0 N_0 (-15 \text{ dB}) S_0$ . The running noise masker was presented with a duration of 750 ms at a reference spectrum level of 40 dB SPL/Hz. The signal was a 500 Hz sinusoid with a duration of 20 ms including 5 ms raised-cosine onset and offset ramps. Thresholds were measured in a two-interval forced-choice (2IFC) procedure employing a one up/three down algorithm. In the  $N_\pi N_0 S_\pi$  condition the interaural phase of the noise was inverted (from  $\pi$  to 0) 375 ms after its onset. In the  $N_0 N_0 (-15 \text{ dB}) S_0$  configuration the masker level was attenuated by 15 dB. Figure 5.1 shows a typical result for one subject for both configurations. The abscissa denotes the relative time between the onset of the test tone and the masker transition, the test tone level at threshold is plotted on the ordinate. The 0-dB point corresponds to the subject's nontransient  $N_0 S_0$  thresholds.

In the  $N_\pi N_0 S_\pi$  conditions a relatively high threshold level (about -2 dB) is obtained for delay times less than 30 ms relative to the masker transition. Thresholds are at a low level (-12 to -15 dB) for delay times larger than about 150 to 200 ms. Between both delay times a continuous transition from high to low threshold levels occurs. The corresponding decay of thresholds in the purely monaural configuration  $N_0 N_0 (-15 \text{ dB}) S_0$  happens much faster than in the binaural configuration; the threshold curve reaches a stationary level for a relative delay of about 50 ms.

Masked thresholds resulting from the corresponding simulations using the model from Chapter 3 are shown in Figure 5.2. The most striking difference to the experimental results in Figure 5.1 is that  $N_\pi N_0 S_\pi$  and  $N_0 N_0 S_0$  threshold curves decay on the same time scale. Additionally, the transition from high to low threshold takes place faster than in the experiment: the binaural thresholds are at the lower level after a delay of about 30 ms. Besides that difference, the monaural threshold curve shows an "undershoot" effect.



*Figure 5.1: Experimental results from Kollmeier (1986) for the configurations  $N_{\pi}N_0S_{\pi}$  and  $N_0N_0(-15\text{ dB})S_0$  for one subject.*



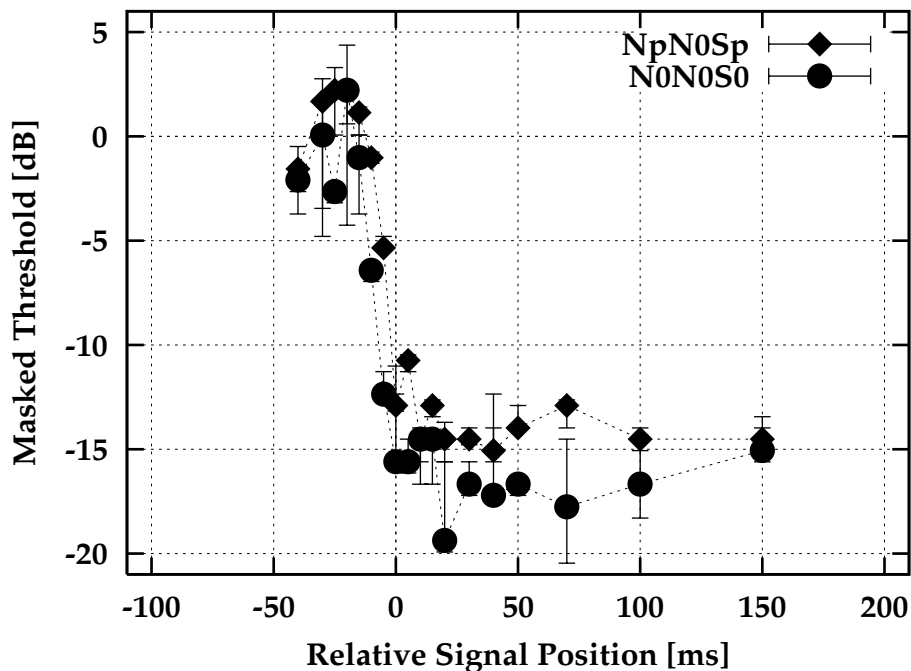
*Figure 5.2: Masked thresholds from simulations for the same configurations as in Figure 5.1 using the model described in Chapter 3.*

### 5.3 Strategy sluggishness

Several reasons may be responsible for the inability of the model from Chapter 3 to predict the “binaural sluggishness” encountered in the experimental conditions considered here. In the following, they are considered as hypotheses (a)-(d).

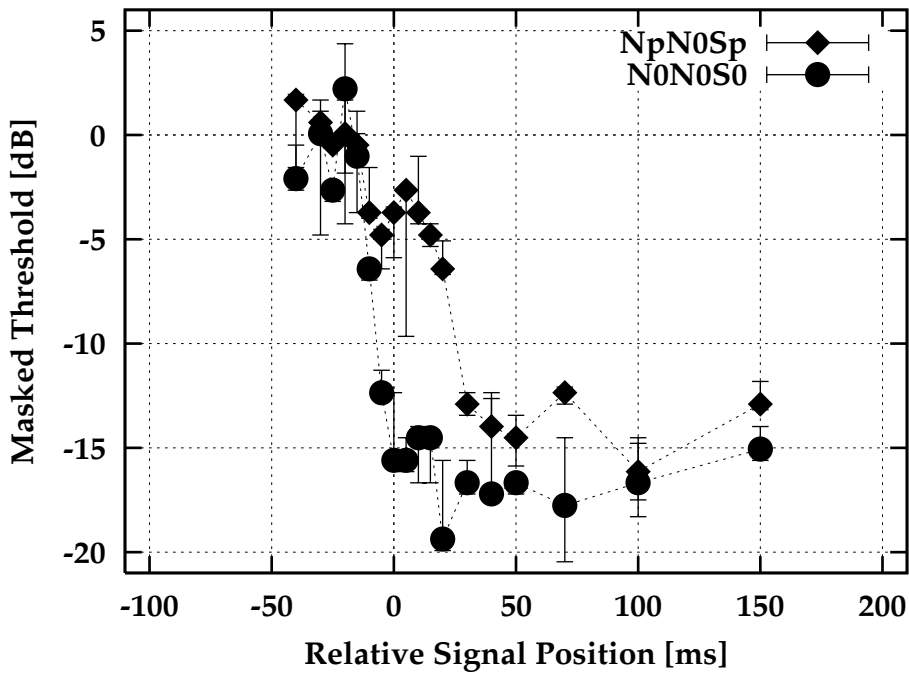
(a) A special “binaural” processing channel exhibiting longer adaptation time constants than the respective monaural channels (“channel sluggishness”). In such a channel the “effective masking level” would decay more slowly than in the monaural channels and the observed sluggishness could be interpreted as forward masking in this binaural channel. However, that hypothesis would be in conflict with our model assumptions that binaural processing follows monaural processing (with the same time constants for the monaural *and* binaural channel). Therefore, we discard hypothesis (a).

(b) A longer time over which binaural information is integrated in the binaural system compared to the monaural system. We denote this as “integration sluggishness”. In the signal processing of the binaural model, hypothesis (b) can be modeled by changing the time constant of the integration low-pass filter following the binaural processor. This would result in an interaction of earlier parts of the masker with the test tone. Figure 5.3 shows the masked thresholds from simulations per-



**Figure 5.3:** Masked thresholds from simulations for the same configurations as in Figure 5.1. Simulation were performed utilizing a integration low-pass filter with a time constant of 200 ms in the binaural channel.

formed with a low-pass filter with a time constant of 200 ms (corresponding to a cut-off frequency of 0.8 Hz) following the binaural processor. There are only small differences between the  $N_\pi N_0 S_\pi$  in this simulation and the corresponding simulations considered in Figure 5.2 performed with the original model. The predicted binaural thresholds change on the same time scale as the binaural thresholds, i. e. the temporal behavior in purely binaural conditions follows the monaural temporal performance of the model, which are mainly imposed by the time constants of the adaptation loops. Hence, the temporal behavior of the binaural system can not be described by only using a prolonged binaural integration time.



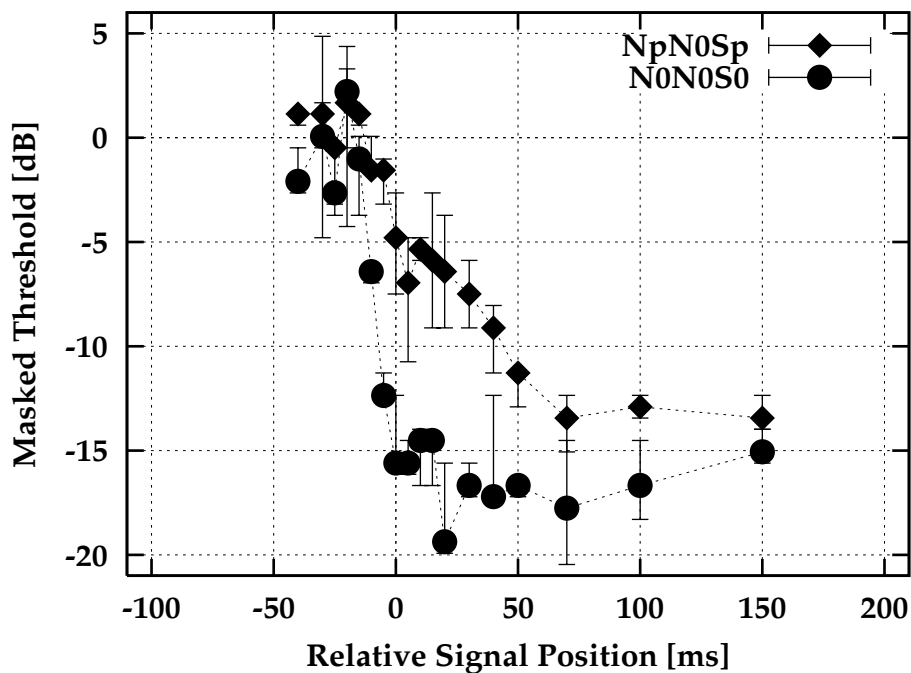
**Figure 5.4:** Simulated thresholds for the same configurations as in Figure 5.2. The internal delay has been calculated from a delayed version of ICF of the noise masker (see text).

(c) A simple modification of the algorithm determining the size of the internal time delay is to include only information from the past relative to the temporal occurrence of the test tone to calculate the internal time delay. This can be described by calculating the interaural cross correlation function not from the temporal position of the test to (as described in Chapter 3) but from a range before the occurrence of the test signal.

Figure 5.4 shows results from such a simulation for the same configurations as in Figure 5.2. The internal delay of the binaural processor has been calculated from the ICF of the noise maskers in the range of 60 ms *before* the beginning of the test tone. The algorithm results essentially in a *shift* of the position of the transition

from high towards low thresholds for the predicted  $N_\pi N_0 S_\pi$  thresholds. The size of the shift is about 30 ms corresponding to the temporal center of the window employed for the calculation of the ICF. For that reason, we denote hypothesis (c) as “delay sluggishness”. This performance can be explained as a consequence of the position of the maximum of the ICF, when the temporal window in which the correlation function is calculated, is shifted across the position of the transient in interaural phase. The position of the maximum, and consequently the performed internal delay, remains constant at the value of 1 ms for the first 30 ms and switches abruptly to the value of 0 ms, the optimal delay to cancel the second half of the noise masker.

(d) The results from simulations following hypothesis (c) propose that, in the framework of the present binaural model, the mechanism of selecting or steering the internal time delay itself has to be phenomenologically described as sluggish. This might be modeled by the low-pass filtered output of a running estimate of the optimal delay, which may itself be determined from the position of the maximum of a running cross-correlation. To test such a sluggishness of the equalization process,



**Figure 5.5:** Simulated thresholds for the same configurations as in Figure 5.2. The internal delay has been chosen from a low pass-filtered version of the optimal delay for the position of the beginning of the test tone (see text).

simulations have been performed in which a low-pass filtered version of the optimal delay for the noise masker is used. As described above, the optimal internal delay for an  $N_\pi N_0 S_\pi$  configuration is 1 ms for the first half of the masker and 0 ms for the

second half. No assumptions were made how the binaural processor estimates the delay. The results of the simulations are shown in Figure 5.5. In these simulations threshold curves show a behavior similar to experimental threshold curves.

A sluggish mechanism steering as described above should comply with the demand that it conserves the hitherto existing model predictions. Apart from forward masking configurations, all previously considered binaural masker configurations consisted stationary interaural relations, for which the sluggish mechanism of selecting an interaural delay results in the same estimates as the old mechanism. This is true at least for the temporal average of the estimate. For the binaural forward masking configurations the performance of the model is determined by monaural detection. The employed mechanism of steering the binaural noise reduction should play no significant role.

Nevertheless, similar results as in the  $N_\pi N_0 S_\pi$  described above occur also for transient binaural configurations that are constructed analogous to a monaural forward masking experiment, e. g.  $N_0 N_\pi S_\pi$ . In such a configuration an increase in threshold is already observed for test signal positions about 60 ms before a position in which in a corresponding diotic condition thresholds are influenced by forward masking. Therefore in the description of the process used to evaluate the internal time delay a temporal range *around* the test signal should be used. Holube et al. (1998) described results from binaural detection experiment employing noise maskers with different shapes of temporal binaural transitions of interaural correlation (step or sinusoidal, periodic or nonperiodic) in terms of temporal windows and their respective time constants (the sum of upper and lower time constant). They found significantly different time constants for the stepwise (43 Hz) and the sinusoidal (109 ms) masker correlation variation. Following these results, it is not clear whether the results from simulations performed with the binaural model of the present study and utilizing hypothesis (d), can be explained with with one single temporal window, in which the ICF is calculated and from which the internal delay is estimated considering an additional binaural strategy inertia. It should be remarked that the proceeding in the present study is different from that in the study of Holube et al. (1998) where the binaural masked threshold is assumed to be a function of the effective masker correlation within the assumed temporal window (Kollmeier and Holube, 1992). Further investigations with the present model in configurations with transient binaural interaural parameters are required.

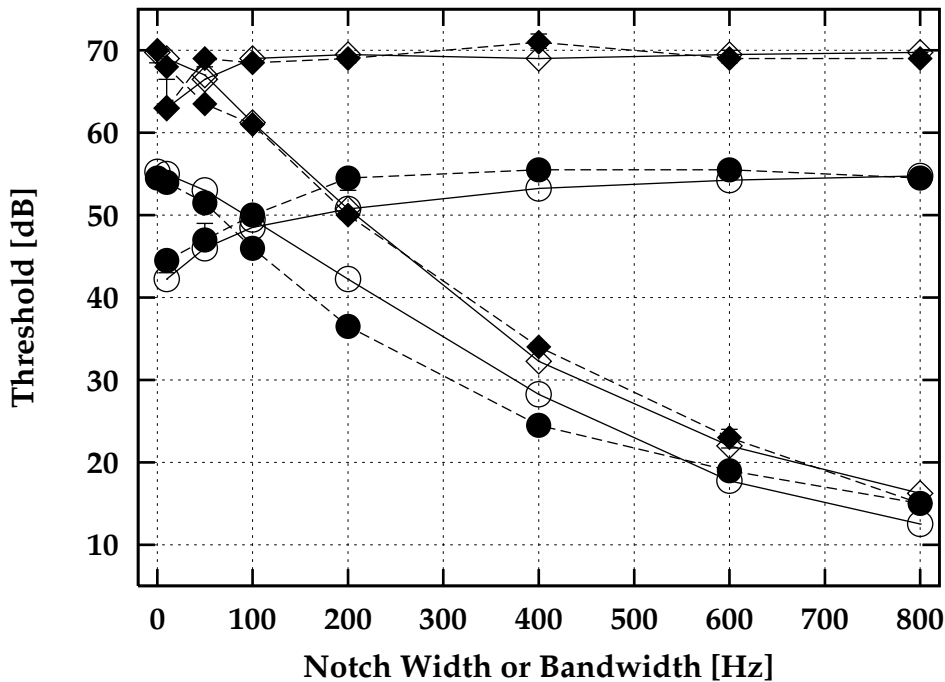
## 5.4 Binaural sluggishness in evaluating fast fluctuating interaural differences

Zurek and Durlach (1987) interpreted results from binaural masking experiments utilizing bandpass noise maskers in context with the faster fluctuation of interaural parameters for larger masker bandwidths. So a decreasing BMLD with increasing noise bandwidth might be interpreted as evidence of the binaural system's smoothing

over rapid fluctuations in interaural cues.

In the binaural processor interaural differences are coded in the difference signal of the left and right monaurally processed ear signals after adaptation. A sluggish binaural canal can be described by an low-pass filter following the cancellation stage and preceding the full-wave rectification stage (in contrast to the modified integration low-pass filter which was employed in the simulations for Figure 5.3).

Figure 5.6 shows experimental (open symbols) and predicted results (filled symbols) from such a modified model. They are for the binaural detection experiment in bandpass and notched noise respectively performed by Hall et al. (1983) which was already described in Chapter 4. Only the configurations with a noise masker of presented at spectrum level of 50 dB/Hz is investigated in  $N_0S_0$  (diamonds) and  $N_0S_\pi$ -configurations (circles). In the simulation a “sluggishness low-pass filter” of first order with a cut-off frequency of 80 Hz (i. e. a time-constant of 2 ms) was employed. As for the “standard” model, also multiple off-frequency channels are considered in the simulations. The predicted results for the bandwidening experi-



**Figure 5.6:** Experimental and simulated threshold curves for the Hall experiment in the notched noise and the bandwidening configurations at a masker spectrum level of 50 dB/Hz. Simulations have been performed employing a 80 Hz low-pass filter behind the binaural cancellation mechanism.

ments are scarcely different from the corresponding results presented in Section 4.3.3. For the notched noise configuration, the large deviation between experimental data and model predictions is reduced. Also in the simulations a continuously de-

ing BMLD can be observed with increasing notchwidth. Nevertheless, compared with the experimental data, the BMLD is slightly too large for medium notchwidths (100-400 Hz) and disappears for large notchwidths.

The cut-off frequency of 80 Hz employed for the sluggishness low-pass filter yields reasonable results for both the notched noise and the bandpass noise configuration. A smaller cut-off frequency results in a smaller notch width for which the BMLD disappears. Also the  $N_0 S_\pi$  thresholds become affected, since the effect of spectral integration across off-frequency channels is reduced. A larger cut-off frequency results in predictions similar to those in Section 4.3.3. However, the time constant of the investigated filter is much too low to explain results from experiments concerning binaural canal sluggishness described in the previous section.

## Chapter 6

# Summary and General Conclusions

In this thesis a new functional model of the temporal auditory signal processing is introduced and evaluated. The model is based on a combination of a well evaluated model of the monaural auditory signal processing which is employed as monaural preprocessor and a new binaural signal processing module.

The monaural model of auditory signal processing was employed in numerous previous studies (Dau et al., 1996a,b) to model detection thresholds in a variety of simultaneous and nonsimultaneous psychacoustical detection experiments. For the present study, the same set of parameters as in the original model was employed.

The binaural processor is constructed following ideas of the Equalization and Cancellation (EC) theory (Durlach, 1972). It follows monaural processing and performs a relative temporal delay of both ear signals and subsequently calculates their difference. The subsequent decision device which works on the monaural and binaural “internal representations” of the ear signals is implemented as an optimal detector.

Results from simulations of several critical experiments were presented in Chapter 3. These simulations were used to fix the value of parameters of the binaural processor (the range of available internal delays and the size of the uncertainty of the internal equalization of interaural time differences, the “jitter error”) and to verify the effects of the different stages of monaural and binaural signal processing on the performance of the complete model. The reduced BMLD for high frequencies is a cooperative effect of the loss of phase locking in the description of the transduction process for frequencies above 1 kHz and of the jitter error. In contrast to descriptions in the literature where the BMLD is described in terms of the interaural correlation of the half-wave rectified and low-pass filter noise masker, in the present model the uncertainty in performing an interaural delay is necessary to account for experimental results at high frequencies. The binaural processing stage profits from the compression of stationary parts of the ear signals in the monaural preprocessing stage. It could be shown that in the present model the interaural amplitude adjustment from the EC-theory can be omitted since the peripheral compression already performs this task in a satisfactory way. Deviations of some dB of predicted

and measured absolute thresholds are mainly determined by the description of the middle ear and the peripheral noise added after the gammatone filterbank. Since the internal decorrelation is necessary to describe the BMLD at low input levels, the description of the middle ear transfer function may be chosen too strict in the framework of the monaural model. Since the binaural processor follows the monaural adaptation stages, the shape of threshold curves in monaural and binaural forward masking experiments as well is determined mainly by the temporal properties of the *monaural* model, which is in line with the experimental results.

In Chapter 4 the performance of the binaural model in more complex spectro-temporal masking configurations was investigated. Whereas one gammatone on-frequency filter provides appropriate results in the model description of monaural and binaural detection configurations when broadband noise maskers are employed in stationary conditions, more off-frequency channels have to be taken into account when a narrowband noise masker or short test signals are utilized. In notched noise configurations, the binaural model does not give correct predictions of binaural masked thresholds. Hence, the binaural model applied in multiple channels appears to be adequate to predict the data.

In Chapter 5 different hypotheses were investigated in order to find a description of the “sluggish” performance in binaural masking experiments with time varying interaural parameters where larger “effective” time constants are found than in corresponding monaural experiments. Within the framework of the model of binaural signal processing, this performance could be described with the assumption that the mechanism which steers the internal time delay is sluggish. A sluggishness in the evaluation of fast fluctuating interaural differences describes by a low-pass filter following the cancellation unit of the binaural processor could be used to describe thresholds in binaural notched noise experiments with smaller deviations from the experimental data than in the original model. This approach is not suitable to describe the binaural inertia described above.

When detection in the binaural model of auditory signal processing took place in multiple frequency (on-frequency and off-frequency) channels, in the present study the same strategy of binaural unmasking, i. e. the same internal delay has been utilized in all frequency channels. This is a consequence of the noise masker - test tone configurations employed in the present study. As an expansion in further experiments and simulations, binaural masking configurations should be investigated in which different internal interaural delays in different frequency channels would lead to an optimal strategy of binaural unmasking. Such a configuration can for example be generated in binaural masking experiments in broadband noise where the signal in a (harmonic) tone complex consisting of components with different ITDs or different interaural phase.

Further interesting possible applications of the binaural model are automatic predictions of the performance of human listeners in realistic binaural situations with spatially separated target and noise signals. This implies that realistic head related transfer functions (HRTFs) preceding monaural signal processing have to be

taken into account. Especially the prediction of speech intelligibility (cocktail-party effect) is of interest. The binaural model of auditory signal processing might be employed as a preprocessor to produce a binaural auditory-based representation for automatic speech recognition based on two input (ear) signals. Furthermore, with an adequate description of the impaired auditory system, the effect of a (monaural) hearing impairment on binaural performance could be investigated.

# Appendix A

## Internal adjustment of external parameters

### A.1 Factor $q$ for internal amplitude adjustment

Let  $\bar{R}$  and  $\bar{L}$  be the RMS values of the monaurally preprocessed masker-alone signal of the right and left ear,  $R$  and  $L$ :

$$\bar{R} = \text{rms}(R) = \sqrt{\frac{1}{T} \int_0^T |R(t)|^2 dt} \quad (\text{A.1})$$

$$\bar{L} = \text{rms}(L) = \sqrt{\frac{1}{T} \int_0^T |L(t)|^2 dt}. \quad (\text{A.2})$$

The factor  $q$  to adjust the signal energy of both masker signals is chosen such that the *overall* masker energy for right and left ear is equalized. The transformation  $R \rightarrow r_q R$  and  $L \rightarrow l_q L$  yields the first condition:

$$r_q \bar{R} = l_q \bar{L} \quad (\text{A.3})$$

since  $\text{rms}(af) = |a| \text{rms}(f)$  for every real  $a$  and  $r_q, l_q > 0$ . The constancy of the arithmetic mean for the scaled and unscaled RMS-values results in the second condition:

$$r_q \bar{R} \cdot l_q \bar{L} = \bar{R} \bar{L} \quad (\text{A.4})$$

i. e.

$$r_q \cdot l_q = 1.$$

Equation A.3 and A.4 result in Equation 3.2 for the factors  $r_q$ ,  $l_q$ , and  $q$  respectively.

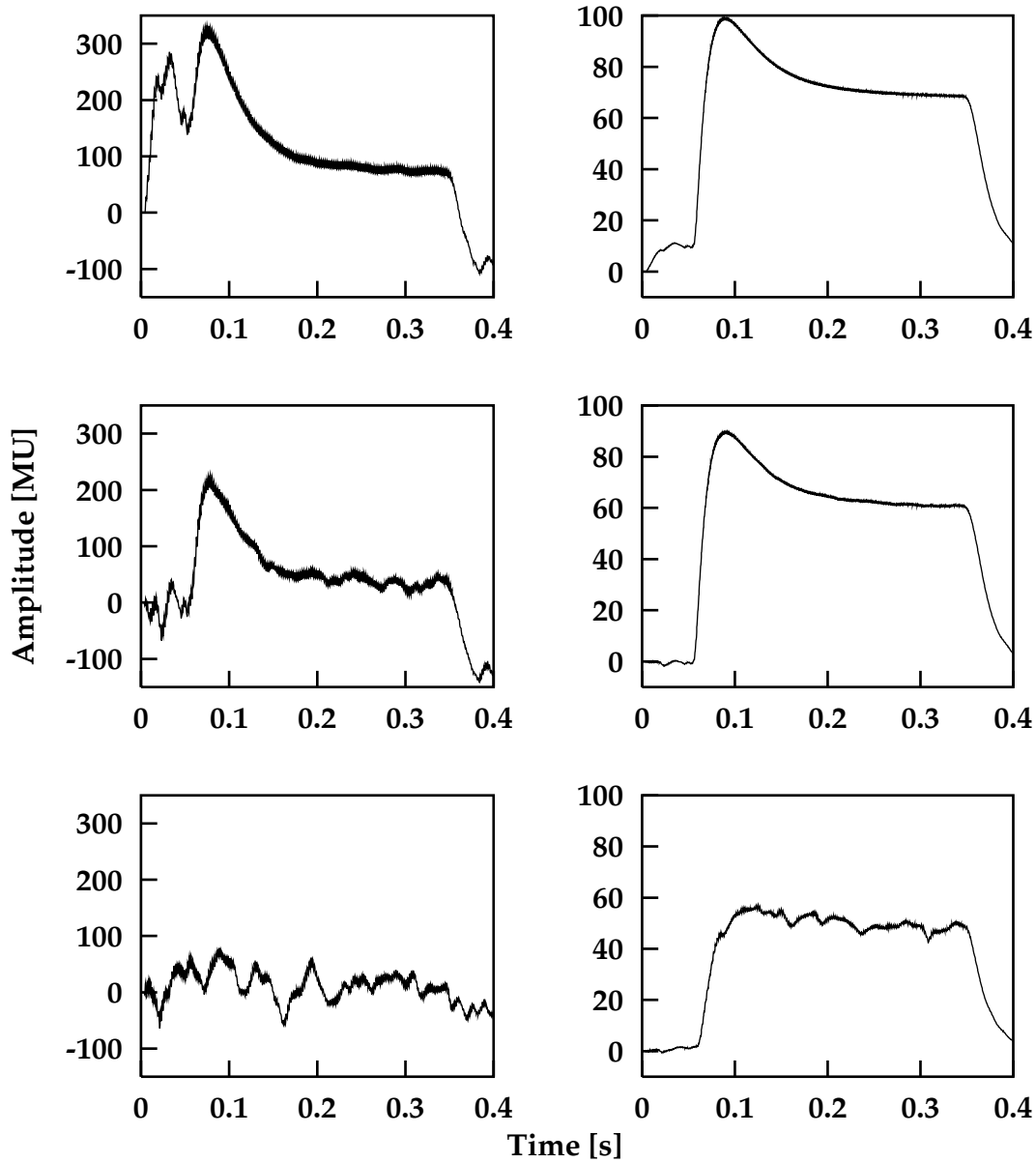
## Appendix B

# Examples of the binaural auditory signal processing in the model

As an illustration of the outcome of binaural auditory processing within the model some examples of typically occurring internal representations of signals from simulation of the first experiments for an  $N_0S_\pi$  situation (Section 3.4.1) are shown in Figure B.1. Only the output of the filter tuned to the signal frequency of 500 Hz was considered. All plots show internal representations without peripheral or internal noise.

The left panels show the result from monaural processing (right ear). They have the same temporal behavior as the internal representations described by Dau et al. (1996a) for a higher characteristic frequency. Additionally after the integration low-pass filter some signal fine structure with the frequency of the test signal (and the corresponding gamma tone filter) can be observed. This fine structure is somewhat reduced in the binaural signals which are shown in the right panels. Because of the limited graphical resolution, the 500 Hz fine structure appears as widening of the line. The top panels show the internal representations of a suprathreshold test-tone with a level of about 20 dB above monaural threshold in a noise masker. The shape of the monaurally processed signals is the result of peripheral processing, adaptation and the final 8 Hz low-pass filter. When the masker is turned on a pronounced overshoot occurs and causes the first peak in the signal. The second peak results from the overshoot when the test signal is turned on after 50 ms. Since the signal is presented at a high level, it dominates the total signal. After the signal is switched off an undershoot is observed.

The binaural processing unit reduces the masker significantly. Nevertheless, the masker is not eliminated completely even in the part of the signal where only the noise alone is present (the first 50 ms of the total signal). The remaining masker results from incomplete noise reduction because of the uncertainty error. The binaural processor operates *before* the integration stage and has more fine structure available than it is obvious from the plots of signals after the integration low-pass filter (compare Figure B.2).



**Figure B.1:** Internal representations of acoustical waveforms for the peripheral channel centered at 500 Hz. Left panels show monaurally, right panels binaurally processed signals for an  $N_0S_\pi$  configuration. Top panels show internal representations of the masker plus suprathreshold test signal. The test signal was a 200 ms sinusoid windowed with 20 ms Hanning ramps. The middle panels show the difference between the internal representations of masker plus test signal and of the masker alone. The bottom panels show the same difference for a level of the test signal below monaural masked threshold. The amplitude of all signals is given in model units.

In addition the overshoot caused by the test tone is reduced by binaural processing. As discussed in the description of the temporal behavior of the adaptation loops by Dau (1996) the overshoot is mainly due to the stationary and slowly varying parts of the input of the adaptation loops. These parts are canceled out nearly completely by calculating the difference in binaural processing (since the level of the right and left noise masker is equal). This reduces also the amplitude of the binaural representations compared to the monaural representations. The binaural signal becomes again a positive signal after the full-wave rectification following the cancellation process.

The middle panels show the difference between the internal representations of the masker plus signal and the representation of the masker alone (differences are calculated for monaural and binaural channel separately). Since in the binaural representation the noise is already almost extinct there occurs no large difference between the upper and middle binaural signal.

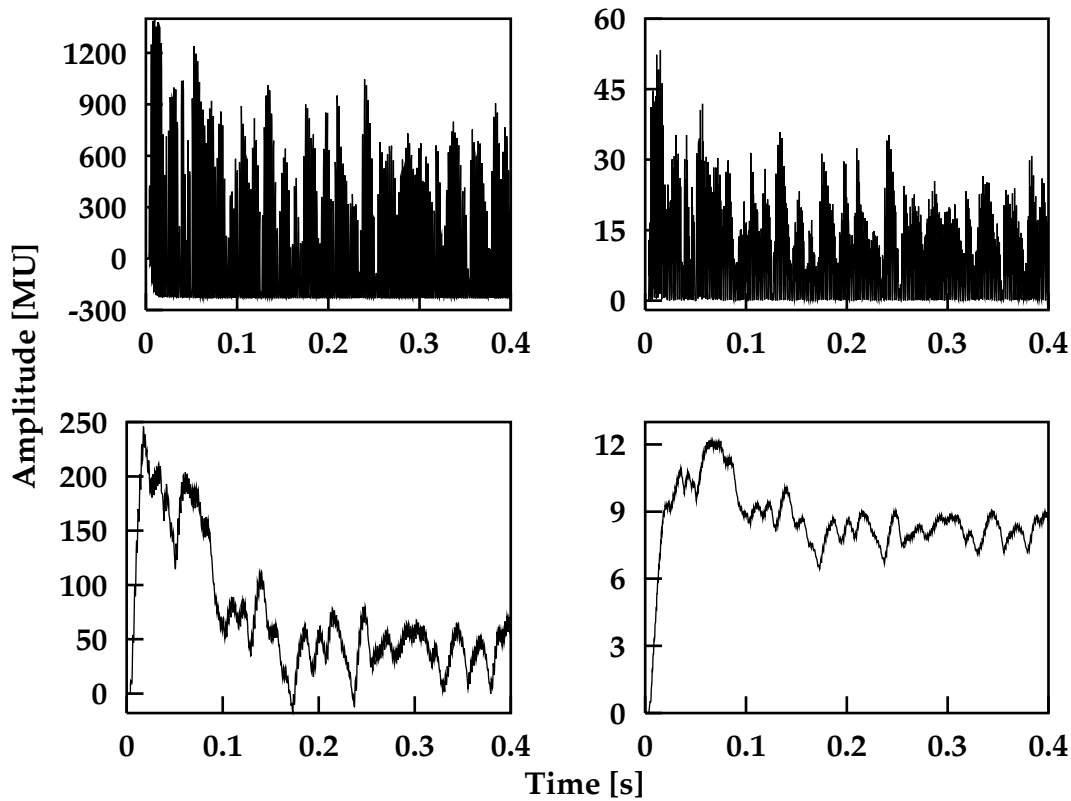
The lower panels show again the difference between internal representations of masker plus signal and masker alone but for a signal level slightly above monaural threshold. Whereas in the graph of the monaural representation the test signal is barely visible and the total representation is dominated by noise the signal is still quite large in the binaural representation although it has become more noisy. The noise fluctuations dominating the monaural representation in the left panel can also be recognized in a reduced (smoothed) form in the binaural representation in the right panel.

The conditions and signals on which the binaural processor is working are illustrated in the upper panels of Figure B.2. Both panels show the output pattern which results from an  $N_0$ -masker from the same simulations like presented in Figure B.1. The upper left panel shows the noise masker after monaural preprocessing, the same noise masker after binaural processing (with the appropriate internal time delay  $\tau = 0$  ms) is shown in the upper right panel. Finally, the lower panels show both signals 8 Hz low-pass filtered.

Figure B.3 shows internal representations for the same interaural relations for a signal and characteristic frequency of 4 kHz. The plots are arranged in the same way like in Figure B.1. To emphasize the distinction between high and low frequencies in the simulations a 4<sup>th</sup> order low-pass filter after the half-wave rectification stage has been utilized (like for the simulated thresholds in Figure 3.14).

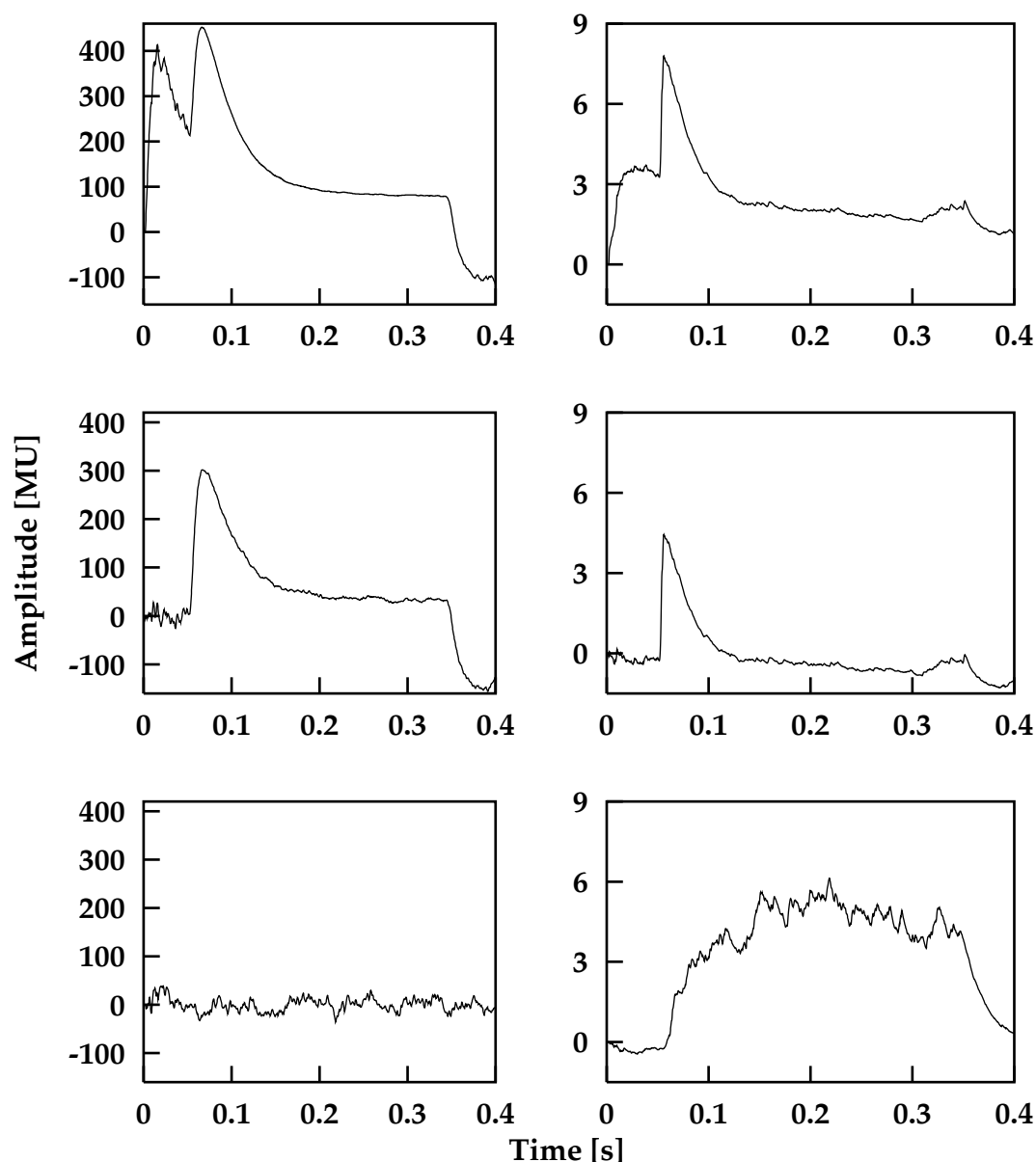
Some differences between the graphs of high and low frequency representations can be observed:

1. Due to the modeled loss of phase locking for high frequencies the fine structure fluctuating with the signal frequency is no longer visible, even in the monaural internal representations.
2. The monaural signal amplitude takes on higher values. This is a result of the stationary part of the signal which is larger for higher frequencies when less fast fluctuations are transmitted.



*Figure B.2:* Upper left panel: Monaurally processed  $N_0$  noise masker after the fifth adaptation loop. The signal is calculated as output from the 500 Hz gammatone filter. Upper right panel: output of the binaural processor for the same noise signal. Lower panels: the same signals after the 20 ms integration low-pass filter.

3. Also of a consequence of phase locking there are merely small interaural differences between the internal representations of the right and left ear signals. Therefore the amplitude of the binaural representations is much smaller (more than one order) than in the low-frequency configuration. For all plots (and all test signal levels) it is even smaller than the standard deviation of the internal noise which restricts the internal amplitude resolution within the model. So detection always takes place in consequence of monaural processing.
4. For a suprathreshold level of the high-frequency signal binaural processing only emphasizes its onset. Increasing the cut-off frequency of the low-pass filter limiting phase-locking or decreasing its order makes the internal representations becoming more similar to those in a low-frequency configuration. That means also the part where the adaptation loops provide stationary output for the test tone it becomes recognizable in the binaural signal.
5. At a level of the test signal at or slightly below monaural threshold this behav-



*Figure B.3: Internal representations of acoustical waveforms for the peripheral channel centered at 4kHz. The arrangement of the panels is in the same order like in Figure B.1.*

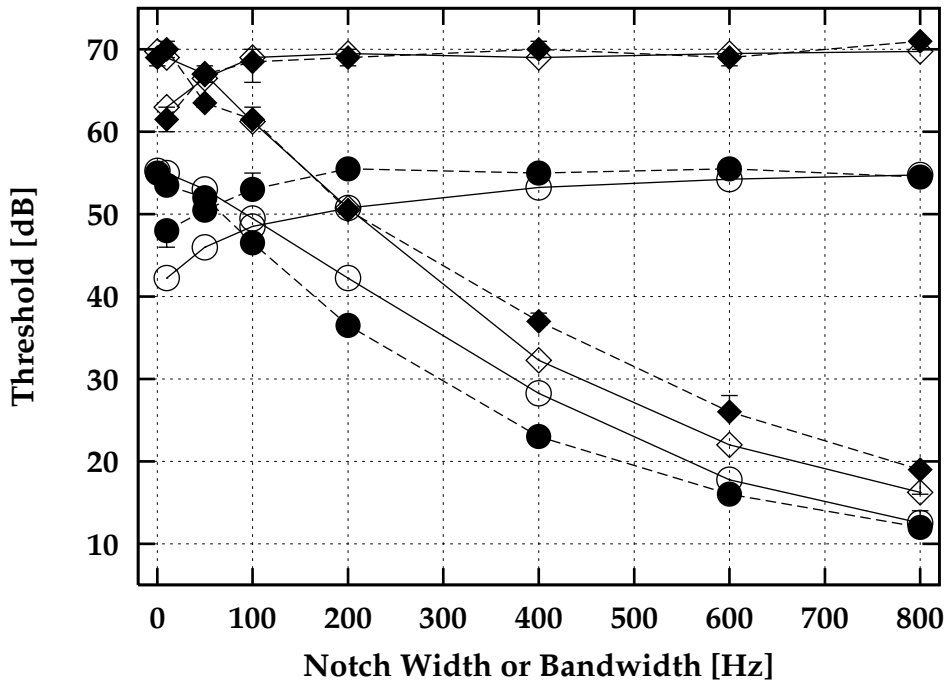
ior changes. The binaural representation now shows excitation over the whole temporal range in which the test signal is presented. This seems to be a consequence of the addition of masker and signal with a similar amplitude before nonlinear processing takes place. The superposition of two sinusoids results in an amplitude modulated sinusoid with a modulation frequency equal to the difference of both sinusoid frequencies. The amplitude modulation resulting

from a superposition of the filtered noise masker and the test signal is stochastic and relatively low-frequent. It is transferred by the hair cell model and the adaptation loops. Since the interaural phase of the test signal is non-zero the fluctuations are different for both ear signals and result also in a binaural output. Nevertheless the amplitude in the binaural channel is too small to result in a detection due to binaural processing.

# Appendix C

## Level-dependent peripheral noise

Several mechanism not described correctly or not included into the description of the signal processing of the model of auditory signal processing may be responsible for the inconveniences described above. The behavior of the binaural thresholds is similar to the different course of predicted and experimental BMLDs as a function of the masker level, where the transition between a minimum BMLD at low masker levels and a maximum BMLD at high masker levels in the simulations takes place less gradually than in the experiments. The mechanism of binaural noise reduction working on the inputs from the same critical band of the two ears seems to be less effective when less masker energy falls in the auditory filter, which may result from the fact the internal decorrelation at low and moderate masker energies is too low in the model's signal processing. To test this hypothesis, simulations for the notched noise and the bandwidening condition have been performed utilizing level dependent peripheral noise. This is motivated in the description of the neural transduction process and the resulting firing patterns on individual auditory nerve fibers as the output of an inhomogeneous Poisson process (Siebert, 1968). Since the present model does not work on firing patterns, the peripheral noise is constructed mimicking a Poisson process. Since its standard deviation of the output is proportional to the square root of the input signal, the internal noise is constructed as the multiplication of a noise pattern with the square root of the Hilbert envelope of the half-wave rectified and low-pass filtered waveform in the frequency channel. Results from the corresponding simulations are shown for a masker spectrum level of 50 dB/Hz in Figure C.1. The plot shows threshold curves for the notched noise and the bandwidening configuration. For the notched noise condition, the peripheral noise influences both simulated  $N_0S_0$  and  $N_0S_\pi$  thresholds for large notchwidths, i. e. low masker energy in the on-frequency channel.  $N_0S_0$  thresholds are too high for notchwidths above 200 Hz, whereas  $N_0S_\pi$  thresholds are still too low for the medium notchwidths. In the bandwidening condition,  $N_0S_\pi$  thresholds are too high for narrow and medium bandwidths ( $\leq 400$  Hz). Altogether, describing the peripheral noise as level dependent does not remove the deviations between observed and simulated binaural thresholds.

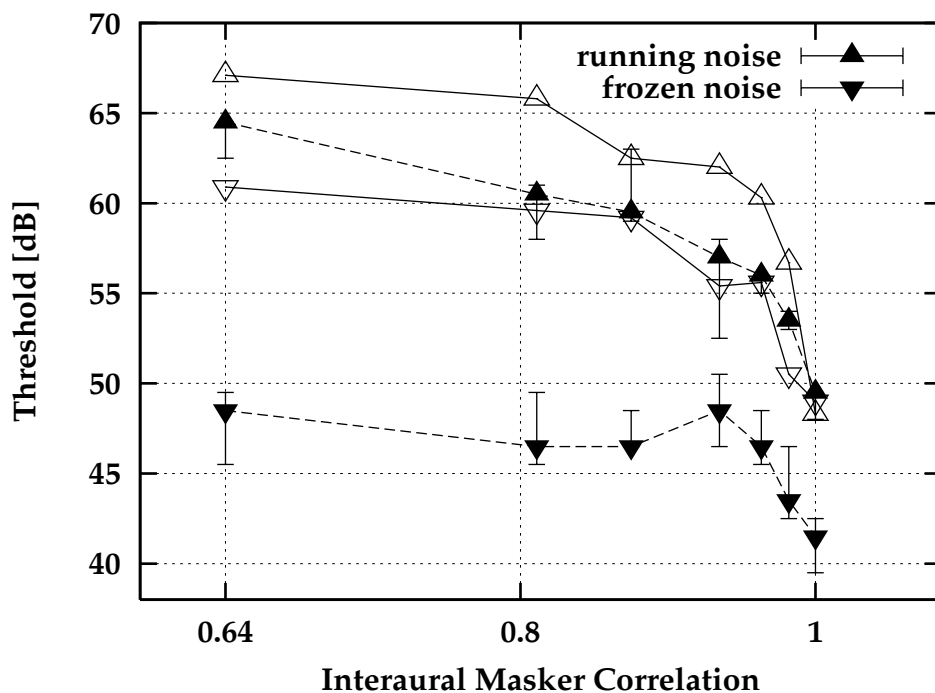


*Figure C.1: Experimental and simulated threshold curves for the Hall experiment in the notched noise and the bandwidening configurations at a masker spectrum level of 50 dB/Hz. Simulations have been performed utilizing level dependent peripheral noise (see text).*

The concept of introducing level-dependent internal noise has been motivated by experimental results from Kohlrausch et al. (1999) who investigated the influence of the masker correlation uncertainty in binaural detection experiments. They measured the masked thresholds of a 500 Hz sinusoid for both running and frozen noise masker as a function of the masker correlation. Experiments for maskers of a bandwidth of 10 Hz were repeated in simulations with the original model of binaural signal processing described in Chapter 3. The noise masker was generated with *exact* correlation. For the frozen noise configuration, a realization was generated for each track. Results from experiments (open symbols) and simulations (filled symbols) are plotted in Figure C.2.

Experimental frozen noise thresholds are lower than those in running noise but converge for correlations close to unity. Threshold curves for running noise show a steeper increase with decreasing correlation than those in frozen noise. All predicted thresholds are below experimental thresholds, apart from those in running noise at unity correlation. The deviation is much larger for frozen noise thresholds where it reaches 13 dB. Also the increase with decreasing correlation is weaker for the predicted thresholds.

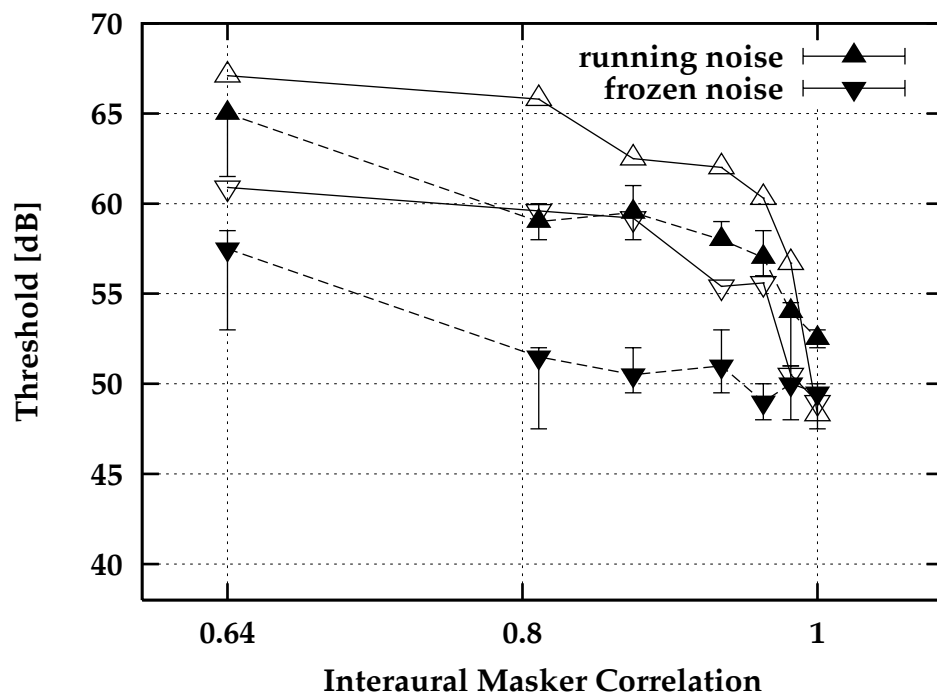
These results suggests that internal decorrelation is not strong enough in the



**Figure C.2:** Binaural masked thresholds for 500 Hz test signal in 10 Hz wide running (upward triangles) and frozen noise (downward triangles). Open symbols show experimental results from Kohlrausch et al. (1999), filled symbols results from simulations performed with the model of binaural signal processing.

description of the signal processing in the model. At the other hand it has been shown that the peripheral noise provides a sufficient description of absolute monaural and binaural thresholds. So a level-dependent internal noise (e. g. mediated by the transduction process) could be an expansion which conserves the former results at the absolute threshold.

Figure C.3 shows the masked thresholds from simulations (filled symbols) performed with the level-dependent peripheral noise described above together with the experimental results (open symbols). Predicted thresholds are plotted as medium value and interquartile range of nine repetitions for each value of the masker correlation. Predicted thresholds in running noise are only slightly changed for large interaural masker correlations. Predictions for frozen noise are up to 8 dB higher than in simulations with peripheral noise of a constant level, but the increase in threshold with decreasing correlation is even more reduced.



*Figure C.3: Binaural masked thresholds for 500 Hz test signal in 10 Hz wide noise. Open symbols show experimental results from Kohlrausch et al. (1999), filled symbols results from simulations performed with the model of binaural signal processing including level dependent peripheral noise (see text).*

# Bibliography

- Berger, E. H., (1983). "Influence of physiological noise and the occlusion effect on the measurement of real-ear attenuation at threshold", *J. Acoust. Soc. Am.* **74**, 81–94.
- Bernstein, L. R. and Trahiotis, C., (1996). "The normalized correlation: Accounting for binaural detection across center frequency", *J. Acoust. Soc. Am.* **100**, 3774–3784.
- Bernstein, L. R., (1997). "Detection and discrimination of interaural disparities: modern earphone-based studies". In Gilkey, R. H. and Anderson, T. R., editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, chapter 6, 117–138. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Blauert, J. and Cobben, W., (1978). "Some consideration of binaural cross correlation analysis", *Acustica* **39**, 96–103.
- Blauert, J., (1997). *Spatial Hearing*. MIT Press, Cambridge, Massachusetts, revised edition.
- Blodgett, H. C., Jeffress, L. A. and Taylor, R. W., (1958). "Relation of masked threshold to signal-duration for various interaural phase-conditions", *Am. J. Psychol.* **71**, 283–290.
- Bodden, M., (1993). "Modeling human sound-source localization and the cocktail-party effect", *acta acustica* **1**, 43–55.
- Bos, C. E. and de Boer, E., (1966). "Masking and discrimination", *J. Acoust. Soc. Am.* **39**, 708–715.
- Bourbon, W. T. and Jeffress, L. A., (1965). "Effect of bandwidth of masking noise on detection of homophasic and antiphase tonal signals", *J. Acoust. Soc. Am.* **37**, 1180.
- Burgdorf, W. and Wagener, B., (1968). "Verdeckung durch subjektiv diffuse Schallfelder", *Acustica* **19**, 72–79.

- Burgdorf, W., (1963). "Zur subjektiven Wirkung von Schallfeldern in Räumen (Rückverdeckung, Phantomschallquellen)", *Acustica* **13**, 86–91.
- Caird, D. and Klinke, R., (1983). "Processing of binaural stimuli by cat superior olivary complex neurons", *Exp. Brain Res.* **52**, 379–392.
- Cherry, E. C., (1953). "Some experiments on the recognition of speech, with one and with two ears", *J. Acoust. Soc. Am.* **25**, 975–979.
- Chouard, N., (1997). *Loudness and unpleasantness perception in dichotic conditions*. PhD thesis, Université du Maine, Le Mans.
- Colburn, H. S. and Durlach, N. I., (1978). "Models of binaural Interaction". In Carterette, E. C. and Friedman, M. P., editors, *Handbook of Perception*, volume IV, chapter 11, 467–518. Academic Press, New York. follows durlach:78:BPH.
- Colburn, H. S. and Latimer, J. S., (1978). "Theory of binaural interaction based on auditory-nerve data. III. Joint dependence on interaural time and amplitude differences", *J. Acoust. Soc. Am.* **64**, 95–106.
- Colburn, H. S., (1973). "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination", *J. Acoust. Soc. Am.* **54**, 1458–1470.
- Colburn, H. S., (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. Supplementary Material". Am. Inst. Physics, Physics Auxiliary Publication Service document No. PAPS, JASMA-61-525-98.
- Colburn, H. S., (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise", *J. Acoust. Soc. Am.* **61**, 525–533.
- Colburn, H. S., (1996). "Binaural Psychoacoustics and Models". In Kollmeier, B., editor, *Psychoacoustics, Speech and Hearing Aids*, 211–220, Singapore. World Scientific.
- Colburn, H. S., (1996). "Computational Models of Binaural Processing". In Hawkins, H. L., McMullen, T. A., Popper, A. N. and Fay, R. R., editors, *Auditory Computation*, Springer Handbook of Auditory Research, chapter 8, 332–400. Springer, New York.
- Cooper, N. P. and Rhode, W. S., (1997). "Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea", *J. Neurophysiol.* **78**, 261–270.
- Dau, T. and Püschel, D., (1992). "Ein Computermodell zur Detektion komplexer Schallreize". In *Fortschritte der Akustik - DAGA '92*, 877–880, Bad Honnef. DPG-Kongreß Verlag.

- Dau, T., Püschel, D. and Kohlrausch, A., (1996). "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure", *J. Acoust. Soc. Am.* **99**, 3615–3622.
- Dau, T., Püschel, D. and Kohlrausch, A., (1996). "A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements", *J. Acoust. Soc. Am.* **99**, 3623–3631.
- Dau, T., Kollmeier, B. and Kohlrausch, A., (1997). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers", *J. Acoust. Soc. Am.* **102**, 2893–2905.
- Dau, T., Kollmeier, B. and Kohlrausch, A., (1997). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration", *J. Acoust. Soc. Am.* **102**, 2906.
- Dau, T., (1996). *Modeling auditory processing of amplitude modulation*. PhD thesis, Universität Oldenburg, BIS-Verlag Oldenburg.
- de Boer, E. and de Jongh, H. R., (1978). "On cochlear encoding: Potentialities and limitations of the reverse-correlation technique", *J. Acoust. Soc. Am.* **63**, 115–135.
- de Boer, E. and Kruidenier, C., (1990). "On ringing limits of the auditory periphery", *Biol. Cybernet.* **63**, 433–442.
- de Boer, E., (1966). "Intensity discrimination of fluctuating signals", *J. Acoust. Soc. Am.* **40**, 552–560.
- Derleth, R. P. and Dau, T., (1999). "On the role of envelope fluctuation processing in spectral masking", submitted to *J. Acoust. Soc. Am.* .
- Derleth, R. P., Verhey, J. and Kollmeier, B., (1996). "Modell zur Modulationswahrnehmung bei Innenohrschwerhörigen". In *Fortschritte der Akustik - DAGA '96*, 148–149, Oldenburg. DEGA e. V.
- Derleth, R. P., (1999). *Temporal and compressive properties of the normal and impaired auditory system*. PhD thesis, Universität Oldenburg, Oldenburg.
- Dierks, K. J. and Jeffress, L. A., (1962). "Interaural phase and the absolute threshold of tones", *J. Acoust. Soc. Am.* **34**, 981–984.
- Dolan, T. R., (1968). "Effect of masker spectrum level on masking-level differences at low signal frequencies", *J. Acoust. Soc. Am.* **49**, 1507–1512.
- Durlach, N. I. and Colburn, H. S., (1978). "Binaural Phenomena". In Carterette, E. C. and Friedman, M. P., editors, *Handbook of Perception*, volume IV., chapter 10, 365–466. Academic Press, New York.

- Durlach, N. I., (1963). "Equalization and cancellation theory of binaural masking-level differences", *J. Acoust. Soc. Am.* **35**, 1206–1218.
- Durlach, N. I., (1972). "Binaural signal detection: Equalization and cancellation theory". In Tobias, J. V., editor, *Foundations of modern auditory theory*, volume II, chapter 10, 369–462. Academic Press, New York.
- Ebata, M., Sone, T. and Nimura, T., (1968). "Improvement of hearing ability by directional information", *J. Acoust. Soc. Am.* **44**, 542–547.
- Fassel, R., (1989). "Vergleich monauraler und binauraler Nachverdeckungskurven". Diploma thesis, Universität Göttingen.
- Fletcher, H., (1940). "Auditory patterns", *Reviews of Modern Physics* **12**, 47–65.
- Gaik, W., (1990). *Untersuchungen zur binauralen Verarbeitung kopfbezogener Signale*, volume 63 of *17: Biotechnics*. VDI-Verlag, Düsseldorf.
- Gaik, W., (1993). "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling", *J. Acoust. Soc. Am.* **94**, 98–110.
- Gilkey, R. H. and Anderson, T. R., editors, (1997). *Binaural and Spatial Hearing in Real and Virtual Environments*. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Glasberg, B. R. and Moore, B. C. J., (1990). "Derivation of auditory filter shapes from notched-noise data", *Hear. Res.* **47**, 103–138.
- Glasberg, B. R., Moore, B. C. J. and Stone, M. A., (1999). "Modelling changes in frequency selectivity with level". In Dau, T., Hohmann, V. and Kollmeier, B., editors, *Psychophysics, Physiology and Models of Hearing*, 143–154, Singapore. World Scientific.
- Grantham, D. W. and Bacon, S. P., (1991). "Binaural modulation masking", *J. Acoust. Soc. Am.* **89**, 1340–1349.
- Grantham, D. W. and Wightman, F. L., (1979). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation", *J. Acoust. Soc. Am.* **65**, 1509–1517.
- Green, D. M. and Swets, J. A., (1988). *Signal Detection Theory and Psychophysics*. Peninsula Publishing, Los Altos, California, reprint edition.
- Green, D. M., (1966). "Interaural phase effects in the masking of signals in different durations", *J. Acoust. Soc. Am.* **39**, 720–724.
- Gulick, W. L., Gescheider, G. A. and D., F. R., (1989). *Hearing — Physiological Acoustics, Neural Coding, and Psychoacoustics*. Oxford University Press, Oxford.

- Hafter, E. R., (1971). "Quantitative Evaluation of a Lateralization Model of Masking-Level Differences", *J. Acoust. Soc. Am.* **50**, 1116–1122.
- Hall, J. W., Tyler, R. S. and Fernandes, M. A., (1983). "Monaural and binaural auditory frequency resolution measured using bandlimited noise and notched-noise masking", *J. Acoust. Soc. Am.* **73**, 894–898.
- Hansen, M., (1998). *Assessment and prediction of speech transmission quality with an auditory processing model*. PhD thesis, Universität Oldenburg, Oldenburg.
- Hartung, K., (1995). "Messung, Verifikation und Analyse von Aussenohrübertragungsfunktionen". In Arnold, W. and Hirsekorn, S., editors, *Fortschritte der Akustik - DAGA '95*, 755–758, Oldenburg. DEGA e. V.
- Hawkins, H. L., A., M. T., Popper, A. N. and Fay, R. R., editors, (1996). *Auditory Computation*. Springer Handbook of Auditory Research. Springer, New York.
- Helmholtz, H. L. F., (1868). *Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. F. Vieweg, Braunschweig.
- Hirsh, I. J. and Burgeat, M., (1958). "Binaural effects in remote masking", *J. Acoust. Soc. Am.* **30**, 827–832.
- Hirsh, I. J., (1948). "The influence of interaural phase on interaural summation and inhibition", *J. Acoust. Soc. Am.* **20**, 536–544.
- Holube, I., Colburn, H. S., van de Par, S. and Kohlrausch, A., (1995). "Simulationen der Mithörschwellen in dichotischen Rauschmaskierern". In Arnold, W. and Hirsekorn, S., editors, *Fortschritte der Akustik - DAGA '95*, 783–786, Oldenburg. DEGA e. V.
- Holube, I., Kinkel, M. and Kollmeier, B., (1998). "Binaural and monaural filter bandwidths and time constants in probe tone detection experiments", *J. Acoust. Soc. Am.* **104**, 2412–2425.
- Hudde, H. and Engel, A., (1998). "Measuring and modeling basic properties of the human middle ear and ear canal. Part III: Eardrum impedances, transfer functions and model calculations", *acta acustica* **84**, 1091–1109.
- Hudde, H., (1998). "Grundeigenschaften des Mittelohres auf der Basis modellbezogener Messungen". In Sill, A., editor, *Fortschritte der Akustik - DAGA '98*, 320–321, Oldenburg. DEGA e. V.
- Jeffress, L. A., Blodgett, H. C. and Deatherage, B. H., (1952). "The masking of tone by white noise as a function of interaural phases on both components. I. 500 cycles", *J. Acoust. Soc. Am.* **24**, 523–527.

- Jeffress, L. A., Blodgett, H. C., Sandel, T. T. and Wood, C. L., (1956). "Masking of tonal signals", J. Acoust. Soc. Am. **28**, 416–426.
- Jeffress, L. A., Blodgett, H. C. and Deatherage, B. H., (1962). "Masking and interaural phase. II. 167 cycles", J. Acoust. Soc. Am. **34**, 1124–1126.
- Jeffress, L. A., (1948). "A place theory of sound localization", Journal of Comparative and Physiological Psychology **41**, 35–39.
- Jeffress, L. A., (1958). "Medial geniculate body – a disavowal", J. Acoust. Soc. Am. **30**, 802–803.
- Jeffress, L. A., (1972). "Binaural signal detection: Vector theory". In Carterette, E. C. and Friedman, M. P., editors, *Foundations of modern auditory theory*, volume II, chapter 9, 349–368. Academic Press, New York.
- Koch, E. S., editor, (1959). *Psychology: A study of a Science. Study 1*, volume 1. McGraw-Hill, New York.
- Kohlrausch, A. and Fassel, R., (1997). "Binaural masking level differences in nonsimultaneous masking". In Gilkey, R. H. and Anderson, T. R., editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, 169–190. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Kohlrausch, A., van de Par, S. and Breebaart, J., (1998). "Warum sind "binaurale Frequenzgruppen" (manchmal) breiter als monaurale?". In *Fortschritte der Akustik - DAGA '98*, 492–493. Deutsche Gesellschaft für Akustik e. V.
- Kohlrausch, A., van de Par, S. and Breebart, J., (1998). "The influence of masker variability on estimates of monaural and binaural and binaural critical bandwidths". In *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America*, 2895–2896, Seattle.
- Kohlrausch, A., van de Par, S. and Breebart, J., (1999). "Measuring the role of masker-correlation uncertainty in binaural masking experiments", *acta acustica* **85**, 465.
- Kohlrausch, A., (1984). *Psychoakustische Untersuchungen spektraler Aspekte beim binauralen Hören*. PhD thesis, Universität Göttingen.
- Kohlrausch, A., (1986). "The influence of signal duration, signal frequency and masker duration on binaural masking level differences", *Hear. Res.* **23**, 267–273.
- Kohlrausch, A., (1988). "Auditory filter shape derived from binaural masking experiments", J. Acoust. Soc. Am. **84**, 573–583.

- Kohlrausch, A., (1990). "Binaural masking experiments using noise maskers with frequency-dependent interaural phase differences. I: Influence of signal and masker duration", J. Acoust. Soc. Am. **88**, 1737–1748.
- Kollmeier, B. and Gilkey, R. H., (1990). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection", J. Acoust. Soc. Am. **87**, 1709–1719.
- Kollmeier, B. and Holube, I., (1992). "Auditory filter bandwidths in binaural and monaural listening configurations", J. Acoust. Soc. Am. **92**, 1889–1901.
- Kollmeier, B., (1986). *Entwicklung zeitoptimierter psychoakustischer Meßverfahren und ihre Anwendung auf binaurale Zeiteffekte*. PhD thesis, Universität Göttingen, Göttingen.
- Kollmeier, B., (1990). *Meßmethodik, Modellierung und Verbesserung der Verständlichkeit von Sprache*. Habilitationsschrift, Göttingen.
- Kuwada, S., Batra, R. and Fitzpatrick, D. C., (1997). "Neural processing of binaural temporal cues". In Gilkey, R. H. and Anderson, T. R., editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, 399–425. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Langford, T. L. and Jeffress, L. A., (1964). "Effect of noise crosscorrelation on binaural signal detection", J. Acoust. Soc. Am. **36**, 1455–1458.
- Langhans, A. and Kohlrausch, A., (1992). "Differences in auditory performance between monaural and diotic conditions. I: Masked thresholds in frozen noise", J. Acoust. Soc. Am. **91**, 3456–3470.
- Levitt, H., (1971). "Transformed up-down methods in psychoacoustics", J. Acoust. Soc. Am. **49**, 467–477.
- Licklider, J. C. R., (1948). "The influence of interaural phase relations upon the masking of speech by white noise", J. Acoust. Soc. Am. **28**, 150–159.
- Licklider, J. C. R., (1959). "Three auditory theories". In Koch, E. S., editor, *Psychology: A Study of a Science. Study 1*, volume 1, 41–144. McGraw-Hill, New York.
- Lindemann, W., (1986). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for nonstationary signals", J. Acoust. Soc. Am. **80**, 1608–1622.
- Lindemann, W., (1986). "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.

- McFadden, D., (1968). "Masking-level differences determined with and without interaural disparities in masker intensity", *J. Acoust. Soc. Am.* **44**, 212–223.
- Mehrgardt, S. and Mellert, V., (1977). "Transformation characteristics of the external human ear", *J. Acoust. Soc. Am.* **61**, 1567–1576.
- Metz, P. J., von Bismark, G. and Durlach, N. I., (1968). "Further Results on binaural unmasking and the EC model. II. Noise bandwidth and interaural phase", *J. Acoust. Soc. Am.* **43**, 1085–1091.
- Moore, B. C. J. and Glasberg, B. R., (1987). "Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns", *Hear. Res.* **28**, 209–225.
- Moore, B. C. J., editor, (1986). *Frequency Selectivity in Hearing*. Academic, London.
- Narins, P. M., Benedix, J. H. and Moss, F., (1997). "Does stochastic resonance play a role in hearing?". In Lewis, E. R., Long, G. R., Lyon, R. F., Narins, P. M., Steele, C. R. and Hecht-Poinar, E., editors, *Diversity in Auditory mechanics*, 83–90, Singapore. World Scientific.
- Osman, E., (1971). "A correlation model of binaural masking level differences", *J. Acoust. Soc. Am.* **50**, 1494–1511.
- Patterson, R. D. and Moore, B. C. J., (1986). "Auditory filters and excitation patterns as representations of frequency resolution". In Moore, B. C. J., editor, *Frequency Selectivity in Hearing*, 123–178. Academic, London.
- Patterson, R. D., Nimmo-Smith, I., Weber, D. L. and Milroy, R., (1982). "The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold", *J. Acoust. Soc. Am.* **72**, 1788–1803.
- Patterson, R. D., Nimmo-Smith, J., Holdsworth, J. and Rice, P., (1987). "An efficient auditory filterbank based on the gammatone function". Paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE.
- Patterson, R. D., (1976). "Auditory filter shape derived with noise stimuli", *J. Acoust. Soc. Am.* **59**, 640–654.
- Pickles, J. O., (1988). *An Introduction to the Physiology of Hearing*. Academic Press, London, San Diego, New York, 2nd edition.
- Plomp, R. and Bouman, M. A., (1959). "Relation between hearing threshold and duration for tone pulses", *J. Acoust. Soc. Am.* **31**, 749–758.
- Püschel, D., (1988). *Prinzipien der zeitlichen Analyse beim Hören*. PhD thesis, Universität Göttingen.

- Rabiner, L. R., Laurence, C. L. and Durlach, N. I., (1966). "Further results on binaural unmasking and the EC model", *J. Acoust. Soc. Am.* **40**, 62–70.
- Rayleigh, L. S. J. W., (1907). "On our perception of sound direction", *Philos. Mag.* **13**, 214–232.
- Robinson, D. E. and Jeffress, L. A., (1963). "Effect of varying the interaural correlation on the detectability of tonal signals", *J. Acoust. Soc. Am.* **65**, 1947–1952.
- Ruggero, M. A., (1992). "Physiology and Coding of Sound in the Auditory Nerve". In Popper, A. N. and Fay, R. R., editors, *The Mammalian Auditory Pathway: Neurophysiology*, Springer Handbook of Auditory Research, chapter 2, 34–93. Springer, New York.
- Sayers, B. M. and Cherry, E. C., (1957). "Mechanisms of binaural fusion in the hearing of speech", *J. Acoust. Soc. Am.* **29**, 973–987.
- Schenkel, K. D., (1964). "Über die Abhängigkeit der Mithörschwellen von der interauralen Phasenlage des Testschalls", *Acoustica* **14**, 337–346.
- Sek, A. and Moore, B. C. J., (1994). "The critical modulation frequency and its relationship to auditory filtering at low frequencies", *J. Acoust. Soc. Am.* **95**, 2606–2615.
- Shaw, E. A. G. and Piercy, J. E., (1962). "Physiological noise in relation to audiometry", *J. Acoust. Soc. Am. Suppl. 1* **34**, S45 (abs.).
- Siebert, W. M., (1968). "Stimulus transformation in the peripheral auditory system". In Kolers, P. and Eden, M., editors, *Recognizing Patterns*. MIT Press, Cambridge, Mass.
- Soderquist, D. R. and Lindsey, J. W., (1972). "Physiological noise as a masker of low frequencies: the cardiac cycle", *J. Acoust. Soc. Am.* **52**, 1216–1220.
- Stern, R. S. and Colburn, H. S., (1978). "Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position", *J. Acoust. Soc. Am.* **64**, 127–140.
- Stumpf, C., (1905). "Differenztöne und Konsonanz", *Z. Psychol.* **39**, 269–283.
- Tchorz, J. and Kollmeier, B., (1999). "A model of auditory perception as front end for automatic speech recognition", *J. Acoust. Soc. Am.* **106**, 2040–2050.
- van de Par, S. and Kohlrausch, A., (1997). "A new approach to comparing binaural masking level differences at low and high frequencies", *J. Acoust. Soc. Am.* **101**, 1671–1680.

- van der Heijden, M. and Trahiotis, C., (1999). "Masking with interaurally delayed stimuli: The use of internal delays in binaural detection", *J. Acoust. Soc. Am.* **105**, 388–399.
- Verhey, J. and Dau, T., (1997). "Modeling comodulation masking release". In Schick, A. and Klatte, M., editors, *Contributions to Psychological Acoustics*, 389–396, Oldenburg. Seventh Oldenburg Symposium on Psychological Acoustics, Bibliotheks- und Informationssystem der Carl von Ossietzky Universität Oldenburg.
- Verhey, J. L., (1998). *Psychoacoustics of spectro-temporal effects in masking and loudness patterns*. PhD thesis, Universität Oldenburg, BIS-Verlag Oldenburg.
- von Békésy, G., (1960). *Experiments in Hearing*. McGraw-Hill, New York.
- von Hövel, H., (1984). *Zur Bedeutung der Übertragungseigenschaften des Außenohres sowie des binauralen Hörsystems bei gestörter Sprachübertragung*. PhD thesis, RWTH Aachen.
- Webster, F. A., (1951). "The influence of interaural phase on masked thresholds: I. The role of interaural time-deviation", *J. Acoust. Soc. Am.* **23**, 452–462.
- Weiss, T. F. and Rose, C., (1988). "A comparison of synchronization filters in different auditory receptor organs", *Hear. Res.* **33**, 175–180.
- Wiesenfeld, K. and Moss, F., (1995). "Stochastic resonance and the benefits of noise: From ice ages to crayfish and SQUIDS", *Nature* **373**, 33–36.
- Yost, W. A., (1988). "The masking-level difference and overall masker level: Re-stating the internal noise hypothesis", *J. Acoust. Soc. Am.* **83**, 1517–1521.
- Zurek, P. M. and Durlach, N. I., (1987). "Masker-bandwidth dependence in homophasic and antiphase tone detection", *J. Acoust. Soc. Am.* **81**, 459–464.
- Zurek, P. M., (1981). "Probability distributions of interaural phase and level differences in binaural detection stimuli", *J. Acoust. Soc. Am.* **90**, 1927–1932.
- Zwicker, E. and Terhardt, E., (1980). "Analytical expression for critical band rate and critical bandwidth as a function of frequency", *J. Acoust. Soc. Am.* **68**, 1523–1525.
- Zwicker, E., Flottorp, G. and Stevens, S. S., (1957). "Critical band width in loudness summation", *J. Acoust. Soc. Am.* **29**, 548–557.
- Zwicker, E., (1961). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)", *J. Acoust. Soc. Am.* **33**, 248.

# Danksagung

An dieser Stelle möchte ich mich bei allen Menschen bedanken, die auf ihre Weise zu dieser Arbeit beigetragen haben.

Herrn Prof. Dr. Dr. Birger Kollmeier danke dafür, daß er mir spontan und unbürokratisch diese Arbeit und den “Quereinstieg” in eine vielseitige Arbeitsgruppe mit ausgezeichneten Arbeitsbedingungen ermöglicht hat und für sein stetes Interesse an der Entwicklung der vorliegenden Arbeit.

Mein besonderer Dank gilt Dr. Jesko Verhey für seine freundschaftliche und fachliche Unterstützung in Diskussionen und während der Korrekturphase. Dr. Torsten Dau danke ich für seine hilfreiche Beratung in allen Bereichen des optimalen Detektors und des Perzeptionsmodells.

Ganz besonders möchte ich mich bei Prof. Dr. Steve Colburn für zahlreiche sehr interessante und fruchtbare Diskussionen, insbesondere während seines Aufenthaltes in Oldenburg, bedanken. Mein Dank gilt auch Dr. Steven van der Par, Prof. Dr. Armin Kohlrausch und Jeroen Breebaart für ihr Interesse in binauralen Fragestellungen und meiner Arbeit und für viele interessante Diskussionen und Anregungen.

Bei allen Mitgliedern der Arbeitsgruppe “Medizinische Physik” und des Graduiertenkollegs “Psychoakustik”, sowie Mitgliedern der “Arbeitsgruppe Akustik” möchte ich mich für die angenehme Arbeitsatmosphäre (und die vielen freundschaftlichen Kontakte die sich daraus ergaben) bedanken. Oliver Wegner danke ich für die prompte Unterstützung bei Software- und Netzproblemen. Für die Teilnahme an den Messungen sei allen Versuchspersonen gedankt.

Meinen langjährigen “Zimmergenossen” (besonders Dr. Leo Ensel, Christoph Eichenlaub und Dr. Nicolas Chouard) danke ich besonders für das freundschaftliche Klima und den Spaß am alltäglichen Wahnsinn.

Nicht zuletzt möchte ich Edith Bakenhus von der Geschäftsstelle Physik und den Mitarbeiterinnen des Sekretariats der Arbeitsgruppe für ihre Hilfestellung nicht nur bei verwaltungstechnischen Fragen danken.

Diese Arbeit wurde im Rahmen des Graduiertenkollegs “Psychoakustik” von der Deutschen Forschungsgemeinschaft (DFG) finanziell unterstützt.