Modeling binaural signal detection

Document status and date:
Published: 01/01/2001

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.
Link to publication

MODELING BINAURAL SIGNAL DETECTION

This thesis was prepared with the LaTeX 2_ϵ documentation system.
Printing: Universiteitsdrukkerij Technische Universiteit Eindhoven.

MODELING BINAURAL SIGNAL DETECTION

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
Rector Magnificus, prof.dr. M. Rem, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op woensdag 20 juni 2001 om 16.00 uur

door

DIRK JEROEN BREEBAART

geboren te Haarlem

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr. A. Kohlrausch
en
prof.dr. H. S. Colburn

Copromotor:
dr.ir. S. van de Par

Contents

I

IV

*'One of the most striking facts about our ears is that we have two of them–
and yet we hear one acoustic world; only one voice per speaker'*
E. C. Cherry and W. K. Taylor, 1954.

General Introduction

The binaural hearing system facilitates our ability to detect, localize, separate, and identify sound sources. Besides perceiving sound sources within the visual field, the perception of sounds extends to positions above, below, behind and to the left and right of the listener. The process of detecting and localizing a sound source is accurate and happens almost automatically. It is impressive that the auditory system is able to perform this task given the complexity of the information which it has to use. In the visual system, for example, there is a close relationship between the direction of a visual object and its projection on the retina. Such a place-localization map rather directly provides information for determining the absolute and relative positions of visual objects. In the peripheral auditory system, however, there is no such place-localization relation. Sound sources which exist in a 3-dimensional world give rise to a complex vibrational pattern in the surrounding air, which is only observed at two points in space, the entrances to the ear canals. Despite the complex and indirect coding of the information about the position of sound sources, the auditory system is able to reconstruct a three-dimensional aural world by clever analysis of specific properties of the waveforms arriving at both ears. The analysis of these specific properties will be discussed in the next section.

1.1 Sound source localization

In the horizontal plane, localization is mainly facilitated by two stimulus properties. For a sound source that is located to one side of the listener, the waveforms will arrive earlier at the ear oriented towards the sound source due to the finite velocity of sound travelling through air. Hence depending on the azimuth of the sound source, an interaural time delay (ITD) exists between the waveforms arriving at both ears. Furthermore, the earlier-arriving signal will generally be more intense than the opposite-ear signal due to shadowing of the head. This shadowing effect is especially strong for sounds with a wavelength that is short compared with the size of the head. Additional intensity differences can occur for small source distances, due to the longer

distance compared to the source-oriented ear. This is generally referred to as interaural intensity difference (IID). The combined effect of these cues results in the ability of human listeners to discriminate between different positions in the horizontal plane with an accuracy of 1 to 10 degrees (King and Laird, 1930; Mills, 1958; Recanzone *et al.*, 1998). Absolute localization tasks usually result in a lower accuracy between 2 and 30 degrees (Wightman and Kistler, 1989a; Makoes and Middlebrooks, 1990; Recanzone *et al.*, 1998; Brungart *et al.*, 1999). In the vertical plane, on the other hand, sound localization is facilitated by specific properties of the magnitude spectra of the waveforms arriving at the eardrum. Due to reflections in the pinna and other body parts, spectral peaks and dips are superimposed on the original sound source spectrum (cf. Wightman and Kistler, 1989b). The frequencies at which these features occur depend on the elevation of the sound source. These cues facilitate a vertical absolute localization accuracy of about 4 to 20 degrees (Wightman and Kistler, 1989a; Makoes and Middlebrooks, 1990; Perrett and Noble, 1997; Recanzone *et al.*, 1998). It has also been shown that *changes* in the localization cues, as long as the movement of the sound source is relatively slow (Perrott and Musicant, 1977), increase our ability to localize sound sources (Perrett and Noble, 1997; Wightman and Kistler, 1999).

A third dimension that the auditory system is able to cope with is the sound source distance. It is generally accepted that at least four signal properties are important for distance perception. First, the intensity of the sound source: sources further away have a lower intensity than sound sources close by. A second important distance cue available in echoic environments is the ratio between direct sound and the amount of reverberation. The intensity of the direct sound decreases with increasing distance. In most reverberant rooms, however, the amount of reverberation is approximately constant, independent of the position (Blauert, 1997). Hence, the ratio of direct and reverberant sound decreases with increasing distance. A third stimulus property is the spectral content of the sound. At greater distances, the sound-absorbing properties of air attenuate high frequencies the most. A fourth stimulus property that has been addressed recently is the interaural correlation of the waveforms arriving at both ears. It has been shown that the perceived distance decreases if the correlation of the waveforms arriving at both ears increases (Bronkhorst, 2001).

1.2 Masking

In some conditions, the auditory system fails to detect the presence of a sound source. This can be due to a very low sound level, but it may also be the result of the presence of other sound sources, i.e., the sound source is masked by other sound sources. It has been shown that the amount of masking strongly depends on the position of both sound sources. If both sounds come from

the same direction, more masking occurs than if sounds come from different directions. A well-known example of binaural properties of masking in daily life is the so-called 'cocktail party effect' (Cherry, 1953). If, in a room where several people are engaged in a conversation, a listener plugs one ear, it becomes much more difficult to understand a single conversation than with two ears.

A systematic study of the binaural phenomena of masking started with experiments that investigated the masking of signals by broadband noise as a function of the exact interaural phase relationship of signal and masker (Licklider, 1948; Hirsh, 1948b). Since that time, many of the binaural variables affecting masking have been investigated. For example, in various experiments subjects had to detect a pure tone in the presence of white noise. If the noise is presented in phase to both ears via headphones (No), and the tone is presented out-of-phase to each ear (Sπ), the masked threshold level is lower than for the case that both the noise and the signal are presented in phase (NoSo). For narrowband maskers, the difference can be as large as 25 dB (Hirsh, 1948b; Wightman, 1971; Zurek and Durlach, 1987). This release of masking is generally referred to as binaural masking level difference (BMLD). It is generally accepted that BMLDs are caused by the fact that the binaural properties (i.e., the ITD and IID) change through the addition of the signal to a masker (Jeffress *et al.*, 1962; McFadden *et al.*, 1971; Grantham and Robinson, 1977). Due to the high sensitivity to binaural cues, the auditory system is able to detect the signal at much lower intensities compared to conditions in which no binaural cues can be used in the detection task.

1.3 Towards a model

Although the phenomenon of the BMLD is known for several decades, it is still not completely understood how the auditory system processes binaural stimuli and which parameters of the stimuli are relevant. It has been shown that human listeners can detect both static ITDs and IIDs (Mills, 1960; Yost, 1972a; Yost *et al.*, 1974; Grantham and Wightman, 1978; Grantham, 1984a) or combinations of these cues (Wightman, 1969; McFadden *et al.*, 1971; Grantham and Robinson, 1977). One of the properties that has a large influence on the detectability of interaural differences is their temporal behavior. For example, the duration of a signal in an NoSπ condition has a large effect on its detectability: the threshold for a 300-ms signal may be up to 25 dB lower than for a 2-ms signal (cf. Yost, 1985; Wilson and Fowler, 1986; Wilson and Fugleberg, 1987; Bernstein and Trahiotis, 1999). Furthermore, it is well known that the rate of fluctuation in interaural differences has a large effect on the trackability. To be more specific, the binaural auditory system is known to be very sluggish (Perrott and Musicant, 1977; Grantham and Wightman, 1978, 1979; Grantham, 1984a; Holube, 1993; Holube *et al.*, 1998), especially

compared to changes in the stimulus that do not require binaural processing (Kollmeier and Gilkey, 1990; Akeroyd and Summerfield, 1999).

Another important parameter is the spectral content of the stimuli. For example, the just-detectable IID is approximately constant over frequency, while the ITD threshold strongly depends on the center frequency (Klumpp and Eady, 1956; Yost, 1972a; Grantham and Robinson, 1977). Below 1 kHz, the ITD sensitivity can very well be described by a constant interaural-phase just-noticeable difference (JND), while above 2 kHz, ITDs presented in the fine structure waveforms of pure tones cannot be detected. Also changes in the bandwidth of the stimuli have a large effect on the detectability of interaural differences (Zurek and Durlach, 1987; van de Par and Kohlrausch, 1999). The bandwidth dependence for out-of-phase pure tones presented in the background of band-limited noise agrees with the filterbank concept of Fletcher (1940). However, the apparent bandwidth of the auditory filters seems to be wider for some specific binaural conditions than for monaural conditions (Sever and Small, 1979; Hall *et al.*, 1983; Zurek and Durlach, 1987; van de Par and Kohlrausch, 1999).

A third important parameter is the similarity of the masker waveforms and signal waveforms arriving at the two ears. This similarity is usually expressed in terms of the interaural cross-correlation. An $NoS\pi$ condition as described above typically results in a BMLD for wideband noise of 15 dB (Hirsh, 1948b; Hafter and Carrier, 1970; Zurek and Durlach, 1987). For an in-phase signal combined with an out-of-phase masker (i.e., an $N\pi So$ condition), BMLDs of up to 12 dB are reported (Jeffress *et al.*, 1952, 1962). If the masker correlation ρ is varied between -1 and + 1, Robinson and Jeffress (1963) found a monotonic increase in the BMLD for an $S\pi$ signal ($N\rho S\pi$) with increasing interaural correlation. Small reductions from +1 of the interaural masker correlation in an $N\rho S\pi$ condition led to a large decrease of the BMLD, while for smaller correlations, the slope relating correlation to BMLDs to interaural correlation was shallower.

One way to gain knowledge of how various stimuli are processed and identified by the auditory system is to develop and validate a simplified version of such a system, i.e., a model. The purpose of this thesis is to present an effective signal processing model of the human binaural auditory system. This model transforms externally presented stimuli into an internal representation of these stimuli. One of the most important features of the model is to include the loss of information when sounds are processed by the various stages of the auditory system. This is obtained by including several (nonlinear) transformations that are usually based on physiological properties and psychophysical measurements of the human auditory system, and 'internal noise' as a model for inaccuracies in the internal representation.
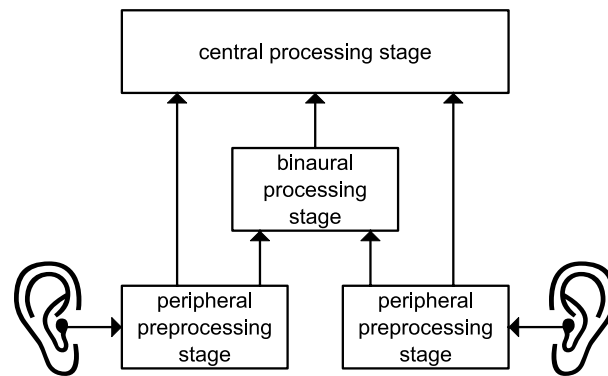
Figure 1.1: *Generic setup of binaural models.*

Over the past decades several models of binaural processing have been developed that address various aspects of binaural hearing. The general setup of the majority of these models is very similar. This bottom-up setup is shown in Fig. 1.1. The signals arriving at the eardrums are first processed by a peripheral preprocessing stage. This stage usually consists of phenomenological or physiological models of the transduction from pressure variations to spike rates in the auditory nerve. Subsequently, binaural interaction occurs in a binaural processor. In this stage, the signals from the left and right sides are compared. Basically two types of binaural interaction have been used extensively: one is based on the *similarity* of the incoming waveforms while the other is based on the *differences* of the incoming waveforms. These classes of binaural interaction are often referred to as cross-correlation based models and EC (Equalization-Cancellation) models (based on the EC theory of Durlach, 1963), respectively. A common feature of the cross-correlation models is that the binaural interaction is computed for a range of internal delays in parallel after a peripheral preprocessing stage. More sophisticated models compute the cross-correlation for several peripheral filters in parallel and supply methods of combining information across frequency bands. On the other hand, in the EC theory, only a single delay is used in the equalization step. Some variations of this theory provide the possibility to have different delays in different frequency bands (von Hövel, 1984; Kohlrausch, 1990; Culling *et al.*, 1996) . The outputs of the binaural processing stage, possibly combined with the monaural outputs of the peripheral preprocessor are fed to a central processor, which extracts certain features of the presented stimuli, such as the estimated intracranial locus of a binaural sound (Lindemann, 1985; Raatgever and Bilsen, 1986; Stern *et al.*, 1988; Shackleton *et al.*, 1992; Gaik, 1993), the presence of a signal in a binaural masking condition (Durlach, 1963; Green, 1966; Colburn, 1977; Stern and Shear, 1996; Bernstein and Trahiotis, 1996; Zerbs, 2000) or the presence of a binaural pitch (Bilsen and Goldstein, 1974; Bilsen, 1977; Raatgever and Bilsen, 1986; Raatgever and van Keulen, 1992; Culling *et al.*, 1996; Bilsen and Raatgever, 2000). For these classes of psychophysical models, Colburn and Durlach (1978) stated that the models were deficient in at least one of the following areas:

1. Providing a complete quantitative description of how the stimulus waveforms are processed and of how this processing is corrupted by internal noise.

2. Deriving all the predictions that follow from the assumptions of the model and comparing these predictions to all the relevant data.

3. Having a sufficiently small number of free parameters in the model to prevent the model from becoming merely a transformation of coordinates or an elaborate curve-fit.

4. Relating the assumptions and parameters of the model in a serious manner to known physiological results.

5. Taking account of general perceptual principles in modeling the higher-level, more central portions of the system for which there are no adequate systematic physiological results available.

The model described in this thesis is an attempt to satisfy these requirements as much as possible. Critical testing of the model was possible because the model was used as an 'artificial observer'. The same stimuli and the same threshold estimation procedure as in the psychophysical experiments with human observers were used to determine the detection threshold with the model. In this computational model, the detection performance is limited by two different noise sources. The first results from the limited resolution of the auditory system itself and has been termed energetic masking (cf. Lutfi, 1990). In models of binaural processing, this source of masking is included as internal noise. For example, the EC-theory summarizes the internal errors of timing and amplitude representation in the factor k, which is directly related to the BMLD. The second source of masking results from the uncertainty associated with the trial-to-trial variation of the binaural cues used to detect the signal (called informational masking by Lutfi). This stimulus uncertainty is effectively transformed into uncertainty within the internal representation of the model. Hence even an optimal detector is limited in its detection performance, if the details of the presented stimuli are not perfectly known.

1.4 Relevance

Besides the interest from a purely scientific point of view of how the human auditory system is able to detect and separate sound sources, several applications may benefit from knowledge about the auditory system, especially in the field of (digital) audio signal processing and telecommunication. In this field, speech and music signals are received, processed, transmitted and again reproduced across time and space. An example of a very popular telecommunication application is the mobile phone. Due to tightened regulations when using mobiles in traffic, hands-free usage is gaining importance. One of the resulting problems is that the signal that is picked up by the mobile

phone does not only consist of the desired speech signal, but also contains unwanted noise and reverberation. To remove these unwanted components, blind-signal separation and restauration algorithms are developed. Knowledge from the binaural auditory system may improve the perceptual quality of these separation algorithms.

One of the major concerns of sound transmission is that the amount of information that has to be transmitted should be as small as possible, without degrading its perceptual quality. A very popular example that uses this principle is defined as the MPEG-1 layer III standard, or popularly called 'mp3'. In these audio coders, the amount of information necessary to represent CD-quality audio is reduced by more than 90%. The reduction of information is facilitated by the large amount of redundancy in the original audio signal. A lot of information can be removed because its presence or absence is masked by other parts of the audio signal. To determine which information is audible and which is not, these coding algorithms heavily rely on psychoacoustic models.

With the upcoming multimedia technology, the importance of three-dimensional sound reproduction via loudspeakers or headphones is gaining interest. Several of these applications make use of knowledge about the binaural auditory system. Examples are 3D positional audio, for example in video games and teleconferencing equipment, and stereo-base widening algorithms. The availability of sophisticated auditory models enables easier development and better optimalization of such sound reproduction algorithms.

Finally, auditory models are also gaining interest from a more socially motivated view. For example people with hearing disorders may benefit from studies on the hearing system. Understanding the processing of the binaural auditory system could lead to better solutions for hearing aids, and hence improve the quality of life for people that are hearing impaired.

1.5 Outline of this thesis

Chapters 2 and 3 present psychoacoustic experiments performed with human subjects and binaural stimuli presented over headphones. These experiments were performed to gain insight in the processing of the binaural hearing system. The results of these experiments, combined with many other studies presented in literature were used to develop the binaural signal detection model presented in Chapters 4, 5 and 6, which form the core content of this thesis. In Chapter 7, a first step is made to apply the model to spatial listening conditions. A more detailed description of each chapter is given below.

In Chapter 2, experiments with human subjects are described that investigate the contribution of static and dynamically varying ITDs and IIDs to binaural detection. By using a modified version of multiplied noise as a masker and a sinusoidal out-of-phase signal, conditions with only IIDs, only ITDs or combinations of the two were realized. In addition, the experimental procedure allowed the presentation of specific combinations of static and dynamically varying interaural differences. These experiments were performed to find a single decision variable that describes the sensitivity to binaural parameters for the experiments described above.

In the experiments described in Chapter 2, subjects had to detect the *presence* of interaural differences. Chapter 3 investigates detectability of *changes* in the interaural cues if these cues are already present in the reference condition. In particular, the influence of *uncertainty* in the magnitude of these cues was investigated. This uncertainty was investigated by comparing Sπ thresholds in the background of masking noise with a certain interaural correlation for both running and frozen noise.

Chapter 4 contains a detailed description of the binaural detection model. This description includes a specification and motivation of all signal processing stages as well as the philosophy behind the model setup. Furthermore, the internal representations for a number of stimuli are demonstrated.

In Chapter 5, the model's predictive scope is tested as a function of spectral parameters of the presented stimuli. For this purpose the model is used as an 'artificial observer'. This means that the model's predictions can be obtained with exactly the same experimental procedure as with the human subjects. Hence experimentally determined thresholds can directly be compared to the predictions of the model. Both the ability of the model to *separate* as well as to *integrate* information across frequency is tested.

Analogous to the evaluation of spectral parameters in Chapter 5, Chapter 6 contains comparisons of model predictions with experimental data as a function of temporal properties of the stimuli. Both temporal integration and resolution issues are discussed.

The predictions shown in Chapters 5 and 6 were obtained for artificial stimuli, such as bandpass noises and pure tones presented over headphones. Such stimuli are not very representative for daily-life listening conditions. To test the model's predictive scope for stimuli that more closely resemble 'normal' listening conditions, tests were performed with stimuli that are filtered with head-related transfer functions (HRTFs). In particular, Chapter 7 describes the perceptual degradation due to the reduction of information present in HRTF pairs and compares the responses of subjects with model predictions.

'It is a capital mistake to theorise before one has data.
Insensibly one begins to twist facts to suit theories, instead of theories to suit facts. '
Arthur Conan Doyle.

The contribution of static and dynamically varying ITDs and IIDs to binaural detection[1]

This chapter investigates the relative contribution of various interaural cues to binaural unmasking in conditions with an interaurally-in-phase masker and an out-of-phase signal (MoSπ). By using a modified version of multiplied noise as the masker and a sinusoid as the signal, conditions with only interaural intensity differences (IIDs), only interaural time differences (ITDs) or combinations of the two were realized. In addition, the experimental procedure allowed the presentation of specific combinations of static and dynamically varying interaural differences. In these conditions with multiplied noise as masker, the interaural differences have a bimodal distribution with a minimum at zero IID or ITD. Additionally, by using the sinusoid as masker and the multiplied noise as signal, a unimodal distribution of the interaural differences was realized. Through this variation in the shape of the distributions, the close correspondence between the change in the interaural cross-correlation and the size of the interaural differences is no longer found, in contrast to the situation for a Gaussian-noise masker (Domnitz and Colburn, 1976). When analyzing the mean thresholds across subjects, the experimental results could not be predicted from parameters of the distributions of the interaural differences (the mean, the standard deviation or the root-mean-square value). A better description of the subjects' performance was given by the change in the interaural correlation, but this measure failed in conditions which produced a static interaural intensity difference. The data could best be described by using the energy of the difference signal as the decision variable, an approach similar to that of the EC model.

2.1   Introduction

Interaural time differences (ITDs) and interaural intensity differences (IIDs) are generally considered to be the primary cues underlying our ability to localize sounds in the horizontal plane. It has been shown that at low frequencies changes in either ITDs or IIDs affect the perceived locus of a sound source (Sayers, 1964; Hafter and Carrier, 1970; Yost, 1981). Besides mediating localization, it has been argued that the sensitivity to ITDs and IIDs of the auditory system is the principle basis of the occurrence of binaural masking level differences (BMLDs) (Jeffress *et al.*, 1962; McFadden *et al.*, 1971;

---

[1]This chapter is based on Breebaart, van de Par, and Kohlrausch (1999).

Grantham and Robinson, 1977). When an interaurally out-of-phase sinusoid is added to an in-phase sinusoidal masker of the same frequency, i.e., a tone-on-tone condition, *static* IIDs and/or *static* ITDs are created, depending on the phase angle between masker and signal. These interaural differences result in lower detection thresholds for the out-of-phase signal compared to an in-phase signal (Yost, 1972a). In terms of the signal-to-masker ratio, subjects tend to be more sensitive to signals producing ITDs than to those producing IIDs (Yost, 1972a; Grantham and Robinson, 1977).

Besides sensitivity to static interaural differences, the binaural auditory system is also sensitive to dynamically varying ITDs (Grantham and Wightman, 1978) and IIDs (Grantham and Robinson, 1977; Grantham, 1984a). As a consequence, BMLDs occur for stimuli with dynamically varying interaural differences. When an interaurally out-of-phase sinusoidal signal is added to an in-phase noise masker with the same (center) frequency (i.e., an $\text{MoS}\pi$ condition[2]), the detection threshold may be up to 25 dB lower than for an in-phase sinusoidal signal (Hirsh, 1948b; Zurek and Durlach, 1987; Breebaart *et al.*, 1998). For such stimuli, both dynamically varying IIDs and ITDs are present (Zurek, 1991). Experiments which allow the separation of the sensitivity to IIDs and ITDs in a detection task with noise maskers were published by van de Par and Kohlrausch (1998b). They found that for multiplied-noise maskers, the thresholds for stimuli producing only IIDs or only ITDs are very similar.

These 'classical' paradigms used in the investigation of the BMLD phenomenon with static and dynamically varying interaural differences exploited different perceptual phenomena. For the experiments that are performed with noise maskers, the average values of the IIDs and ITDs for a masker plus signal are zero, while the variances of these parameters are non-zero. The addition of an out-of-phase signal to a diotic noise masker (i.e., the production of time-varying interaural differences) is usually perceived as a widening of the sound image. For tone-on-tone masking conditions, however, a static interaural cue is introduced and detection is based on a change in the lateralization of the sound source. One notion which suggests that these situations differ from each other is that the binaural system is known to be sluggish, as has been shown by several studies (Perrott and Musicant, 1977; Grantham and Wightman, 1978, 1979; Grantham, 1984a; Kollmeier and Gilkey, 1990; Holube, 1993; Holube *et al.*, 1998). These studies show that if the rate at which interaural cues fluctuate increases, the magnitude of the interaural differences at threshold increases also. It is often assumed that this reduction in sensitivity is the result of a longer time constant for the evaluation of binaural cues compared to the constant for monaural cues (Kollmeier

---

[2]In this chapter, the notation of this condition is $\text{MoS}\pi$ instead of the regular $\text{NoS}\pi$ notation because for the stimuli described here, the masker (M) does not always consist of a noise (N).

and Gilkey, 1990; Culling and Summerfield, 1998). Another demonstration suggesting that the detection of static and dynamically varying interaural differences is different was given by Bernstein and Trahiotis (1997). They showed that roving of static IIDs and ITDs does not influence the detection of dynamically varying interaural differences, indicating that binaural detection of dynamically varying cues does not necessarily depend upon changes in laterality.

One of the proposed statistics for predicting binaural thresholds is the size of the change in the *mean* value of the interaural difference between the signal and no-signal intervals of the detection task. For example, studies of Webster (1951), Yost (1972a), Hafter (1971) and Zwicker and Henning (1985) argued that binaural masked thresholds could be described in terms of just-noticeable differences (JNDs) of the IID and ITD. For stimuli for which the mean interaural difference does not change by adding a test signal (e.g., in an MoSπ condition with band-limited Gaussian noise), it is often assumed that changes in the *width* (e.g., the standard deviation) of the distribution are used as a cue for detection (Zurek and Durlach, 1987; Zurek, 1991). The parameters of the distributions of the interaural differences are generally considered to be important properties for binaural detection. It is unknown, however, how the sensitivity for stimuli producing combinations of static and dynamically varying interaural differences can be described in terms of these parameters.

An attempt to describe the combined sensitivity to static and dynamically varying interaural differences was made by Grantham and Robinson (1977). They measured thresholds for stimuli producing static cues as well as dynamically varying cues[3]. They found that the thresholds for signals producing static cues only were very similar to thresholds for stimuli producing a fixed combination of static and dynamic cues. They discussed the data in terms of the mean interaural differences at threshold, which were very similar for the two conditions. Such an analysis does, however, ignore the contribution of dynamically varying cues for detection in those conditions where these cues are available in addition to static cues.

In the present study MoSπ stimuli will be used which contain either IIDs, ITDs or combinations of both cues for which the ratio between the static and dynamic component will be varied over a wide range. This allows one to perform a critical assessment of whether detection data can be cast within a framework based on the IIDs and ITDs. A second point of interest of this study is related to an alternative theory that has become very popular for

---

[3]The measure μ for expressing the relative amount of static and dynamically varying cues, which will be introduced in Section 2.2, was equal to 2.05 for the experiments performed by Grantham and Robinson.

describing binaural detection which relies on the cross-correlation of the signals arriving at both ears (cf. Osman, 1971; Colburn, 1977; Lindemann, 1986; Gaik, 1993; van de Par and Kohlrausch, 1995; Stern and Shear, 1996; van de Par and Kohlrausch, 1998a). In these models it is assumed that the change in the interaural correlation resulting from the addition of a signal to a masker is used as a decision variable. In fact, Domnitz and Colburn (1976) argued that for an interaurally out-of-phase tonal signal masked by a diotic Gaussian noise, a model based on the interaural correlation and a model based on the distribution of the interaural differences will yield essentially the same predictions of detection. Thus, theories based on the cross-correlation are equivalent to models based on the *width* of the probability distribution functions of the interaural differences, as long as Gaussian-noise maskers and sinusoidal signals are used. However, this equivalence is not necessarily true in general. In the discussion it will be shown that the theories discussed above do not predict similar patterns of data for the stimuli used in the present experiments. Specifically, by producing stimuli with unimodal and bimodal distributions of the interaural cues, we can make a critical comparison between theories based on the IIDs and ITDs and theories based on the interaural cross-correlation. Such a comparison is impossible for those MoSπ studies which employ Gaussian-noise maskers and sinusoidal signals.

In summary, this study has a twofold purpose. On the one hand it intends to collect more data with stimuli producing combinations of static and dynamically varying cues. On the other hand we wanted to collect data with stimuli producing different shapes of the distributions of the interaural differences. Specifically, the employed procedure enables the production of stimuli with both unimodal and bimodal distributions of the interaural differences. These data may supply considerable insight in how detection thresholds for combinations of static and dynamic cues can be described.

2.2   Multiplied noise

Because of its specific properties, multiplied noise allows control of the fine-structure phase between a noise masker and a sinusoidal signal. As already mentioned by Jeffress and McFadden (1968), control of this phase angle allows the interaural phase and intensity difference between the signals arriving at both ears in an MoSπ condition to be specified. Multiplied noise is generated by multiplying a high-frequency sinusoidal carrier by a low-pass noise. The multiplication by the low-pass noise results in a band-pass noise with a center frequency that is equal to the frequency of the carrier and which has a symmetric spectrum that is twice the bandwidth of the initial low-pass noise. For our experiments, we modified this procedure by first adding a DC value to the Gaussian low-pass noise before multiplication with the carrier. The effect of using a non-zero mean is explained in the following section.
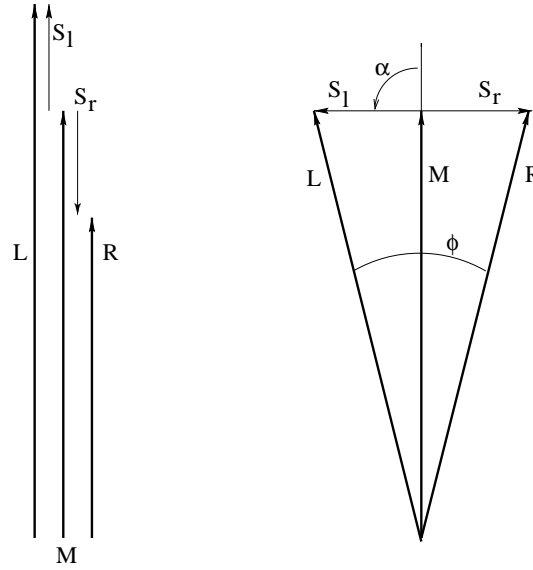
Figure 2.1: *Vector diagrams illustrating the addition of an interaurally out-of-phase signal (S_l and S_r) to an in-phase masker (M) for α=0 (left panel) and $\alpha = \pi/2$ (right panel).*

2.2.1 Multiplied noise as a masker

For the following description we assume an interaurally in-phase multiplied-noise masker and an interaurally out-of-phase sinusoidal signal (i.e., an MoSπ condition). An additional parameter is the phase angle α between the fine-structures of noise and sinusoidal signal. If the frequency and phase of the signal that is added to the left ear are equal to those of the masker (α=0), we can form a vector diagram of the stimulus as shown in the left panel of Fig. 2.1. Here, the vector **M** (the masker) rotates with a constant speed (the frequency of the carrier), while its length (i.e., the envelope of the multiplied noise) varies according to the instantaneous-value distribution of the low-pass noise. **S**$_l$ and **S**$_r$ denote the tonal signals added to the left and right ear, respectively, while **L** and **R** denote the total signals arriving at the left and right ears. Clearly, the vectors **L** and **R** differ only in length, thus only IIDs are present for this stimulus configuration.

If the fine-structure phase of the signal lags the fine-structure phase of the carrier by $\pi/2$ (i.e., $\alpha = \pi/2$), as shown in the right panel of Fig. 2.1, the resulting vectors **L** and **R** have the same length. However, **R** lags **L** by ϕ. Thus, only ITDs are produced. In a similar way, by adjusting the phase angle α to $\pi/4$ or $3\pi/4$, combinations of IIDs and ITDs can be produced.

Because the instantaneous value of the low-pass noise changes dynamically, the envelope of the multiplied noise constantly changes with a rate of fluctuation dependent on the bandwidth of the low-pass noise. The effect of the addition of a DC component to the low-pass noise before multiplication with the carrier can be visualized as follows. If no DC component is added, the instantaneous value of the low-pass noise has a Gaussian probability density

function with a zero mean and RMS=1, as shown in the left panel of Fig. 2.2 by the solid line. If the instantaneous value of the low-pass noise is positive, and an Sπ signal with $\alpha=\pi/2$ is added to the multiplied-noise masker (see the right panel in Fig. 2.1), the fine-structure phase of the right ear lags the fine-structure phase of the left ear by ϕ. If, however, the instantaneous value of the low-pass noise is negative, and the same signal is added, the fine-structure phase of the left ear lags the fine-structure phase of the right ear by ϕ. Thus, the interaural phase difference has changed its sign. Due to symmetry around zero in the instantaneous-value probability density function of the low-pass noise, the probability for a certain positive interaural difference equals the probability for a negative interaural difference of the same amount. Therefore, the distribution of the interaural difference is symmetric with a mean of zero.

The static component μ is defined as the magnitude of the DC component added to the low-pass noise with an RMS value of 1 and zero mean. For $\mu > 0$, the mean of the low-pass noise shifts to a non-zero value (dashed and dash-dotted line of Fig. 2.2, for $\mu=1$ and $\mu=2$, respectively). If the RMS value of the noise plus DC is held constant (i.e., set to 1), the width of the instantaneous-value probability density function of the low-pass noise becomes narrower with increasing μ.

The resulting envelope probability distribution of the multiplied noise is shown in the right panel of Fig. 2.2. For $\mu=0$ (solid line), the distribution function is half-Gaussian, while for increasing μ, the distribution becomes narrower; for μ approaching infinity, the envelope has a mean of one and a variance of zero.
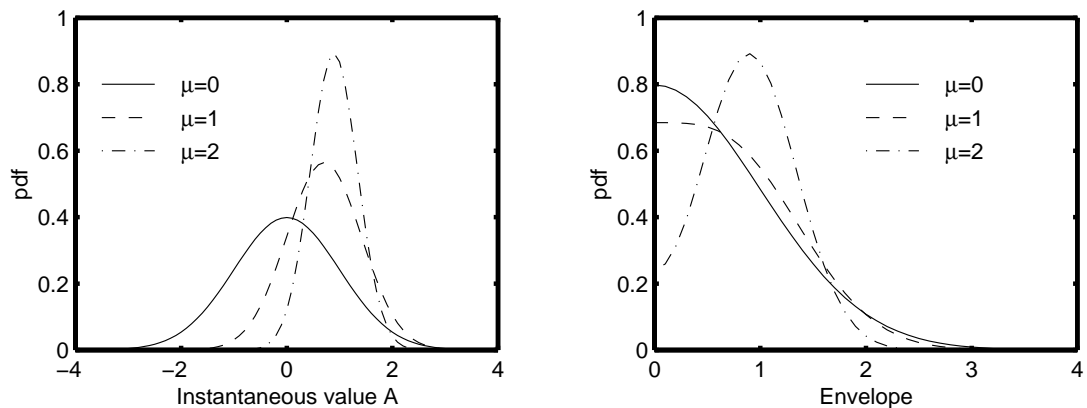
Figure 2.2: *Probability density functions of the instantaneous value of a Gaussian noise with a constant rms value of 1 (left panel) and the resulting multiplied-noise envelope (right panel). The three curves indicate different values of the static component of 0 (solid line), 1 (dashed line) and 2 (dash-dotted line).*

The decreasing variance of the envelope probability distribution with increasing static component has a strong effect on the behavior of the interaural differences that occur when an Sπ signal is added. If, at a certain time, the noise envelope is large, the phase lag in the above example is relatively small. Adding the signal to a small masker envelope, however, results in a large interaural phase lag. Thus, the width of the masker envelope probability distribution determines the range over which the interaural phase difference fluctuates. A wide distribution implies large fluctuations in the interaural difference, while a very narrow distribution implies only small fluctuations. Because an increase of the static component results in a narrower envelope probability density function, the range over which the interaural difference fluctuates becomes smaller. Consequently, the dynamically-varying part of the interaural difference decreases.

We also showed that for a zero mean of the low-pass noise, the overall probability of a positive interaural difference equals the probability of a negative interaural difference of the same magnitude. If a static component is introduced, however, the low-pass noise has a non-zero mean. Hence the probability of a positive interaural difference will be larger than the probability of a negative interaural difference. Consequently, an increase of the static component results in an increase in the mean interaural difference.

In summary, an increase of the static component of the multiplied-noise masker results for the MoSπ condition in an increase of the *mean* of the interaural difference and a decrease of the *range of fluctuations*. Thus, by controlling the value of the static component, binaural stimuli containing different combinations of static and time-varying interaural differences can be created in an MoSπ condition.

2.2.2 Multiplied noise as a signal

We now consider the situation where the roles of the multiplied noise and the sinusoid are reversed. The masker consists of an in-phase sinusoid, and the signal consists of an interaurally out-of-phase multiplied noise with a carrier having the same frequency as the sinusoidal masker. If the phase lag between the left-ear carrier and masker is zero ($\alpha = 0$), this stimulus produces only IIDs. For $\alpha = \pi/2$, only ITDs are present. A phase lag of $\alpha = \pi/4$ results in IIDs and ITDs favoring the same ear, while a phase lag of $\alpha = 3\pi/4$ results in IIDs and ITDs pointing in opposite directions. Again, by adding a static component to the low-pass noise, a mixture of static and dynamically varying interaural differences is achieved.

Two important differences exist between the stimulus described here (with a multiplied-noise *signal*) and the stimulus described in Section 2.2.1 (with

a multiplied-noise *masker*): (1) in the present condition, the envelope of the masker is flat, and besides interaural differences, the signal also produces fluctuations in the envelope of the waveforms arriving at both ears, and (2) an *increase* in the multiplied-noise signal envelope results in an *increase* in the interaural difference, while the opposite is true for the case with a multiplied-noise masker envelope. This reversed relation between multiplied noise envelope and interaural difference has a strong effect on the probability density functions of the interaural differences that occur. This aspect will be discussed in the next section.

2.2.3   Probability density functions of the interaural cues

For a given phase angle α between sinusoid and masker carrier, a certain static component μ and a fixed signal-to-masker ratio S/M, the probability densities of the resulting IIDs and ITDs can be calculated as shown in Appendix 2.A. In Fig. 2.3, the probability density functions for the interaural intensity difference are given for three values of μ and S/M for the two conditions that the masker consists of multiplied noise (left panels) and that the signal consists of multiplied noise (right panels). The upper panels show the IID probability density function for $\alpha=0$ (i.e., only IIDs), the lower panels show the ITD probability density function for $\alpha=\pi/2$ (i.e., only ITDs). The solid line represents no static component ($\mu=0$) and a signal-to-masker ratio of -15 dB, the dashed line represents $\mu=0$ and S/M=-30 dB, while the dotted line represents S/M=-30 dB but with a static component of $\mu=1$. Clearly, for $\mu=0$, the probability density functions are symmetric around zero. Furthermore, a smaller S/M ratio results in narrower distributions. Finally, we see that if the *masker* consists of multiplied noise and $\mu = 0$, the probability density function has a *minimum* at zero (the distribution is bimodal), while for a multiplied noise *signal*, the probability density function shows a *maximum* at zero (the distribution is unimodal).

2.3   Method

2.3.1   Procedure

A 3-interval 3-alternative forced-choice procedure with adaptive signal-level adjustment was used to determine masked thresholds. Three masker intervals of 400-ms duration were separated by pauses of 300 ms. The subject's task was to indicate which of the three intervals contained the 300-ms interaurally out-of-phase signal. This signal was temporally centered in the masker. Feedback was provided to the subject after each trial. In some experiments, the reference intervals contained an Mo masker alone, while in other experiments, an MoSo stimulus (i.e., both masker and signal interaurally in phase) was used. The rationale for these different procedures is explained in the next section.
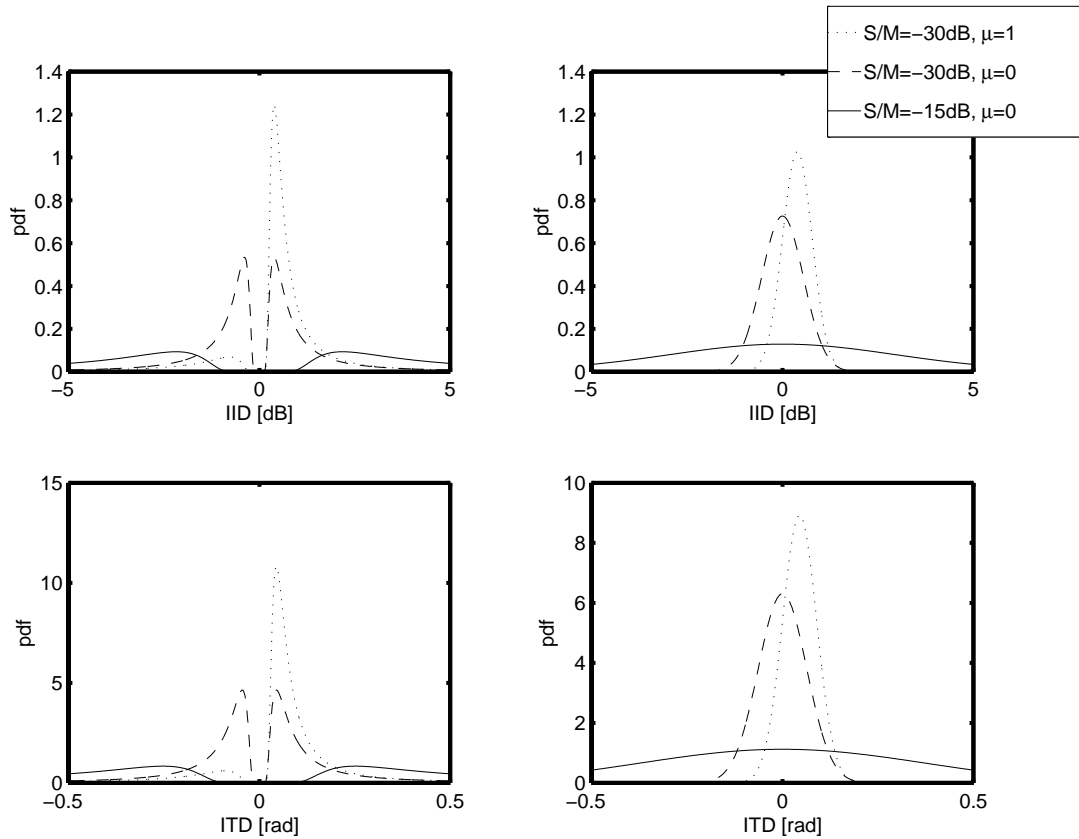
Figure 2.3: *Probability density functions for the interaural intensity difference for $\alpha=0$ (upper panels) and for the interaural phase difference for $\alpha = \pi/2$ (lower panels). Left panels: multiplied-noise masker, sinusoidal signal. right panels: sinusoidal masker, multiplied-noise signal. Solid line: S/M=-15 dB, $\mu=0$. Dashed line: S/M=-30 dB, $\mu=0$. Dotted line: S/M=-30 dB, $\mu=1$.*

The signal level was adjusted according to a two-down one-up rule (Levitt, 1971). The initial step size for adjusting the level was 8 dB. After each second reversal of the level track, the step size was halved until it reached 1 dB. The run was then continued for another 8 reversals. From the level of these 8 reversals, the median was calculated and used as a threshold value. At least four threshold values were obtained and averaged for each parameter setting and subject.

2.3.2 Stimuli

All stimuli were generated digitally and converted to analog signals with a two-channel, 16-bit D/A converter at a sampling rate of 32 kHz with no external filtering other than by the headphones. The maskers were presented to the subjects over Beyer Dynamic DT990 headphones at a sound pressure level of 65 dB. The multiplied-noise samples were obtained by a random selection of a segment from a 2000-ms low-pass noise buffer with an appropriate DC component and a multiplication with a sinusoidal carrier. The low-pass noise buffer was created in the frequency domain by selecting the frequency range

from a 2000-ms white-noise buffer after a Fourier transform. After an inverse Fourier transform, the addition of a DC component and rescaling the signal to the desired RMS value, the noise buffer was obtained. All thresholds were determined at 500-Hz center frequency. In order to avoid spectral splatter, the signals and maskers were gated with 50-ms raised-cosine ramps. Thresholds are expressed as the signal-to-masker power ratio in decibels.

Thresholds were obtained by measuring the detectability of an interaurally out-of-phase signal in an in-phase masker (MoSπ) in the following four experiments:

1. The masker consisted of in-phase multiplied noise, while the signal consisted of an interaurally out-of-phase sinusoid. In this experiment, the reference intervals contained only an Mo masker. Thresholds were obtained as a function of the static component (μ=0, 0.5, 1, 1.5 and 2) for $\alpha = 0$ and $\alpha = \pi/2$ and masker bandwidths of 10 and 80 Hz. The rationale for this experiment was to investigate binaural masked thresholds for combinations of static and dynamic interaural differences, for bimodal distributions of the interaural cues. Two bandwidths were applied; a narrow one in order to produce slowly varying interaural differences and a bandwidth corresponding to the equivalent rectangular bandwidth at 500 Hz (Glasberg and Moore, 1990), producing interaural cues which fluctuate faster. In this way the influence of the rate of fluctuations is investigated.

2. The masker consisted of an in-phase sinusoid, while the signal consisted of interaurally out-of-phase multiplied noise. Thresholds were obtained for the same parameter settings as in experiment 1. This experiment served to study thresholds for unimodal distributions of the interaural differences. The reference intervals consisted of in-phase sinusoids combined with in-phase multiplied noise (i.e., MoSo). Thus, the task was to discriminate between MoSo and MoSπ, so that the subjects could not use the fluctuations in the envelope produced by the signal as a cue for detection.

3. The masker consisted of an in-phase sinusoid, while the signal consisted of interaurally out-of-phase multiplied noise. For similar reasons as in experiment 2, the reference intervals consisted of in-phase sinusoids combined with in-phase multiplied noise. Thresholds were obtained as a function of the bandwidth (10,20,80,160,320 and 640 Hz) of the noise for μ=0 and $\alpha = 0$ and $\alpha = \pi/2$. This experiment served to check for possible effects of off-frequency listening in experiment 2. Because the noise bandwidth is larger than the bandwidth of the masker (the sinusoid), an auditory filter that is tuned to a frequency just above or below the masker frequency receives relatively more noise (signal) intensity than masker intensity. Furthermore, this difference increases with increasing

signal bandwidth. It is therefore expected that for signal bandwidths beyond the critical band, off-frequency listening will result in lower thresholds compared with the case of a signal of subcritical bandwidth. If off-frequency listening influences the results in experiment 2, the parameters of the distributions of the interaural differences cannot be compared between experiments 1 and 2, since peripheral filtering would alter these parameters significantly. To investigate at which signal bandwidth this effect starts to play a role, we determined the bandwidth dependence of the thresholds for this stimulus configuration.

4. Similar to experiment 1, the masker consisted of an in-phase multiplied noise, while the signal consisted of an interaurally out-of-phase sinusoid. The reference intervals contained an Mo masker alone. In this experiment, thresholds were obtained as a function of the fine-structure phase angle between masker and signal for α=0 (only IIDs), $\pi/4$ (IIDs and ITDs which favor the same ear), $\pi/2$ (only ITDs) and $3\pi/4$ (IIDs and ITDs pointing in opposite directions). No static component was present (μ=0). The masker had a bandwidth of 10 or 80 Hz. In addition, an in-phase sinusoid was used as a masker. This experiment served to investigate the effect of the phase angle α, for both dynamically varying and static interaural differences.

Table 2.1 shows a summary of the experimental conditions that were used.

Exp. Number	Masker type	Signal type	Noise band-width [Hz]	μ	α	Reference intervals
1	Multiplied noise	Sinusoid	10, 80	0, 0.5, 1, 1.5, 2	0, $\pi/2$	Mo
2	Sinusoid	Multiplied noise	10, 80	0, 0.5, 1, 1.5, 2	0, $\pi/2$	MoSo
3	Sinusoid	Multiplied noise	10, 20, 40, 80, 160, 320, 640	0	0, $\pi/2$	MoSo
4	Multiplied noise	Sinusoid	10, 80	0, infinity	0, $\pi/4$, $\pi/2$, $3\pi/4$	Mo

Table 2.1. Table showing the experimental variables of experiments 1 to 4.

2.4 Results

2.4.1 Experiment 1: Multiplied noise as masker

In Fig. 2.4, the four lower panels show the detection thresholds for 4 subjects as a function of the static component for experiment 1. The upper panel shows the mean thresholds. The filled symbols denote the IID conditions (α=0), the open symbols denote the ITD conditions ($\alpha = \pi/2$). The upward triangles correspond to a masker bandwidth of 80 Hz, the downward triangles to 10 Hz. Most of the thresholds are in the range of -30 to -20 dB. Generally we see that the mean thresholds (upper panel) show only small differences across bandwidth or physical nature of the cue (i.e., IIDs vs ITDs). Within subjects, however, some systematic differences are present. Subjects MV and JB show higher

thresholds for the 80-Hz conditions than for the 10-Hz conditions, while for subject MD, the 80-Hz IID thresholds are lower than the 10-Hz IID data. Although within and across subjects thresholds vary by about 10 dB, the mean data do not show effects of that magnitude.
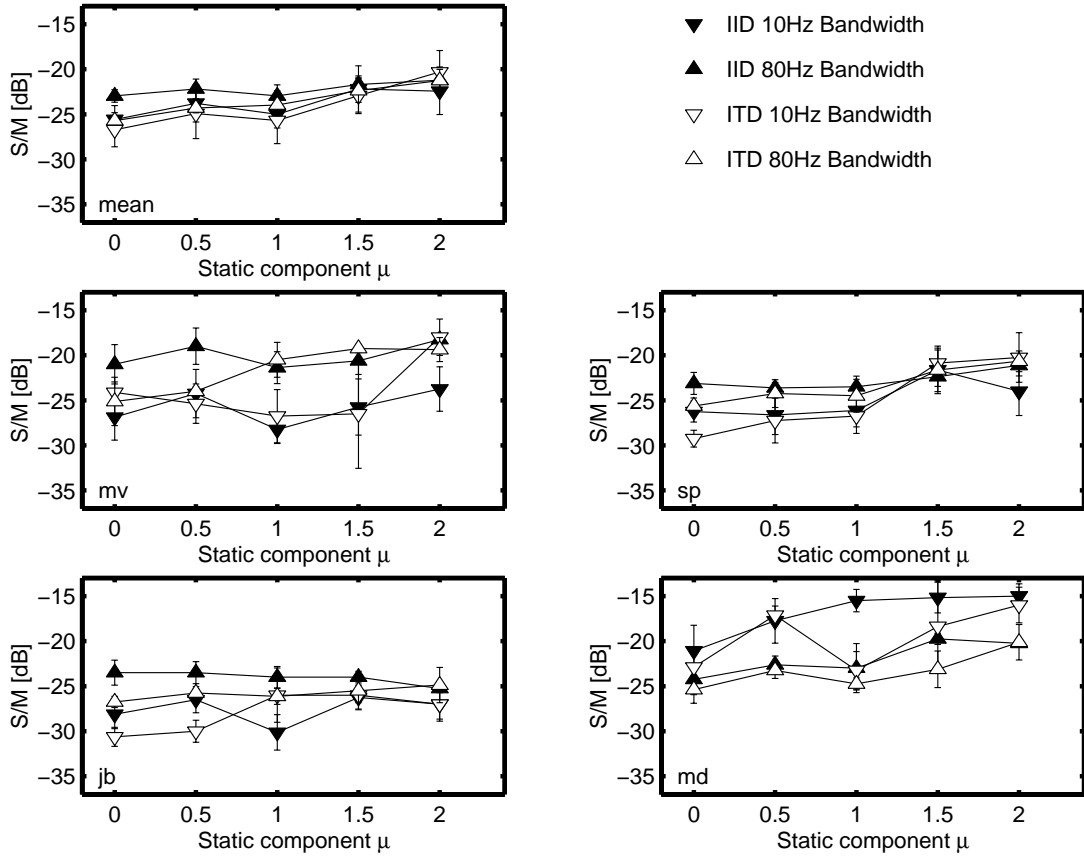
Figure 2.4: *Detection thresholds for an out-of-phase sinusoidal signal added to an in-phase multiplied-noise masker for $\alpha=0$, 10-Hz bandwidth (filled downward triangles), $\alpha=0$, 80-Hz bandwidth (filled upward triangles), $\alpha=\pi/2$, 10-Hz bandwidth (open downward triangles) and $\alpha=\pi/2$, 80-Hz bandwidth (open upward triangles). The four lower panels show thresholds for individual subjects, the upper panel represents the mean across four subjects. Errorbars for the individual plots denote the standard error of the mean based on 4 trials of the same condition. The errorbars in the upper panel denote the standard error of the mean across the mean data from the four subjects.*

2.4.2 Experiment 2: Multiplied noise as signal

In Fig. 2.5, the detection thresholds for experiment 2 are shown as a function of the static component. The format is the same as in Fig. 2.4. The 80-Hz ITD data (open upward triangles) are systematically 4 to 5 dB lower than thresholds for the other stimulus configurations (especially for subjects MD and MV). For these subjects, the thresholds show an increase of up to 6 dB with increasing static component for the 80-Hz ITD condition. The other conditions show approximately constant thresholds for the mean data, independent of bandwidth and physical nature of the interaural cue.
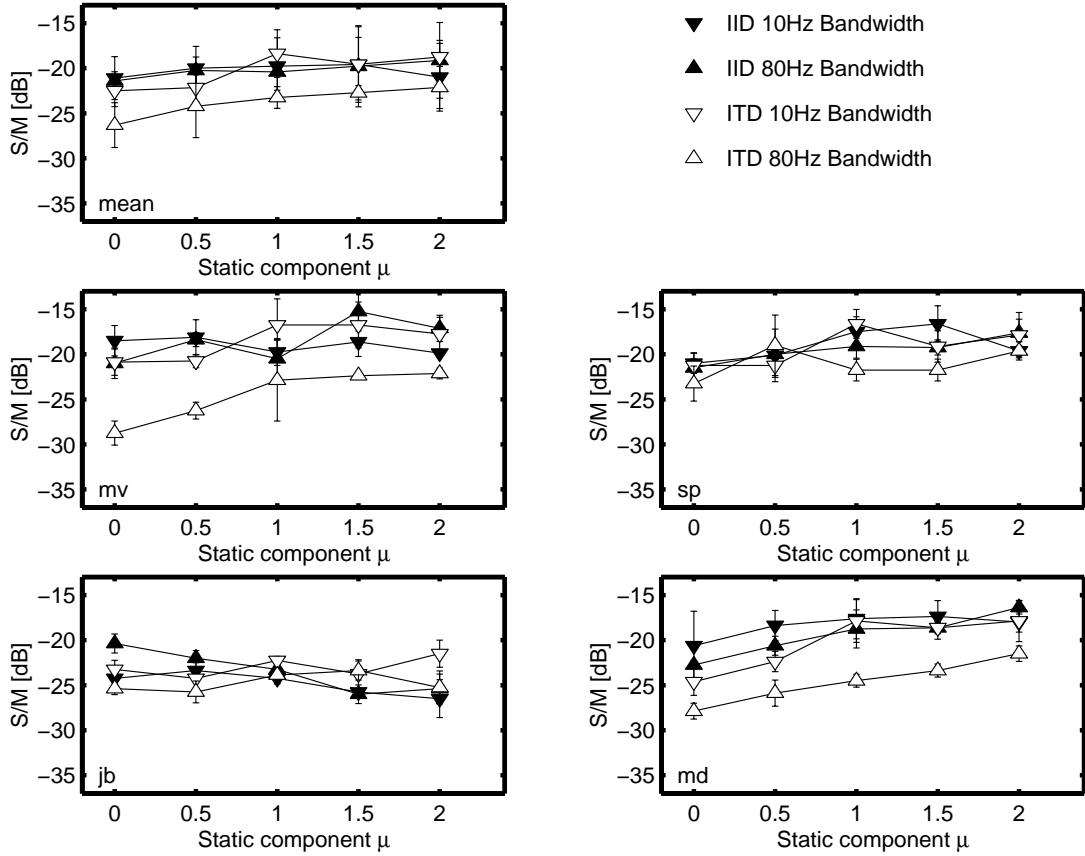
Figure 2.5: *Detection thresholds for an out-of-phase multiplied-noise signal added to a diotic sinusoidal masker as a function of the DC component μ. Same format as Fig. 2.4.*

Because of the small differences that were found in these two experiments, a multifactor analysis of variance (MANOVA) was performed for the results shown in Figs. 2.4 and 2.5 to determine the significance of the different experimental variables used in the experiments. The factors that were taken into account were: (1) the multiplied-noise bandwidth, (2) the masker-signal phase angle α, (3) the static component μ, (4) the masker type (multiplied noise as masker or signal). The p-values for the effects that were significant at a 5% level are shown in Table 2.2.

Thus, significant factors are

1. the masker-signal phase angle α: a change from α=0 to $\pi/2$ results in a mean decrease in thresholds of 1.4 dB,

2. the static component μ: An increase from μ=0 to 2 results in an increase of the thresholds by 3 dB,

3. the masker type: on average, conditions with a multiplied-noise masker have 2.2 dB lower thresholds than conditions with a multiplied-noise signal.

Significant interactions are

1. the multiplied-noise bandwidth combined with α: an increase from 10 to 80-Hz bandwidth results in a decrease of the thresholds by 5 dB for the ITD-only conditions, while the IID-only conditions are similar,

2. the multiplied-noise bandwidth combined with the masker type: the above interaction is only seen for a multiplied-noise signal. For a multiplied-noise masker, the thresholds for $\alpha=0$ and $\alpha=\pi/2$ remain similar with changes in the masker bandwidth.

Effect	p-value
phase angle α	0.01120
static component	0.00073
masker type	0.00001
noise bandwidth and α	0.02424
noise bandwidth and masker type	0.00642

Table 2.2. Factors and their significance levels according to a multifactor analysis of variance of the data shown in Figs. 2.4 and 2.5. Only those factors (upper three) and interactions (lower two) which are significant at a 5% level are given.

2.4.3 Experiment 3: Bandwidth dependence of a multiplied-noise signal

In this experiment, thresholds were determined as a function of the bandwidth of a multiplied-noise test signal added to a sinusoidal masker. Figure 2.6 shows the detection thresholds as a function of the bandwidth of the multiplied noise for $\alpha=0$ (IIDs, filled triangles) and $\alpha=\pi/2$ (ITDs, open triangles). Both for the ITD and IID conditions, the thresholds remain approximately constant for bandwidths up to a bandwidth of 80 to 160 Hz, while for wider bandwidths, the thresholds decrease with a slope of 7 dB/oct of signal bandwidth. The measure of 80 Hz of the auditory filter bandwidth agrees with the monaural equivalent rectangular bandwidth estimates of 79 Hz at 500 Hz center frequency from Glasberg and Moore (1990). Furthermore, we see that, on average, the ITD thresholds are approximately 5 dB lower than the IID thresholds for intermediate bandwidths (i.e., 40 and 80 Hz), which is consistent with the data from experiment 2.

2.4.4 Experiment 4: Dependence on α

Figure 2.7 shows thresholds for experiment 4 as a function of the phase angle between masker carrier and signal. The lower four panels show thresholds of 4 individual subjects, the upper panel shows the mean thresholds. The downward triangles refer to a masker bandwidth of 10 Hz, the upward triangles refer to a masker bandwidth of 80 Hz and the squares to the tone-on-tone condition. The latter has almost always the highest thresholds
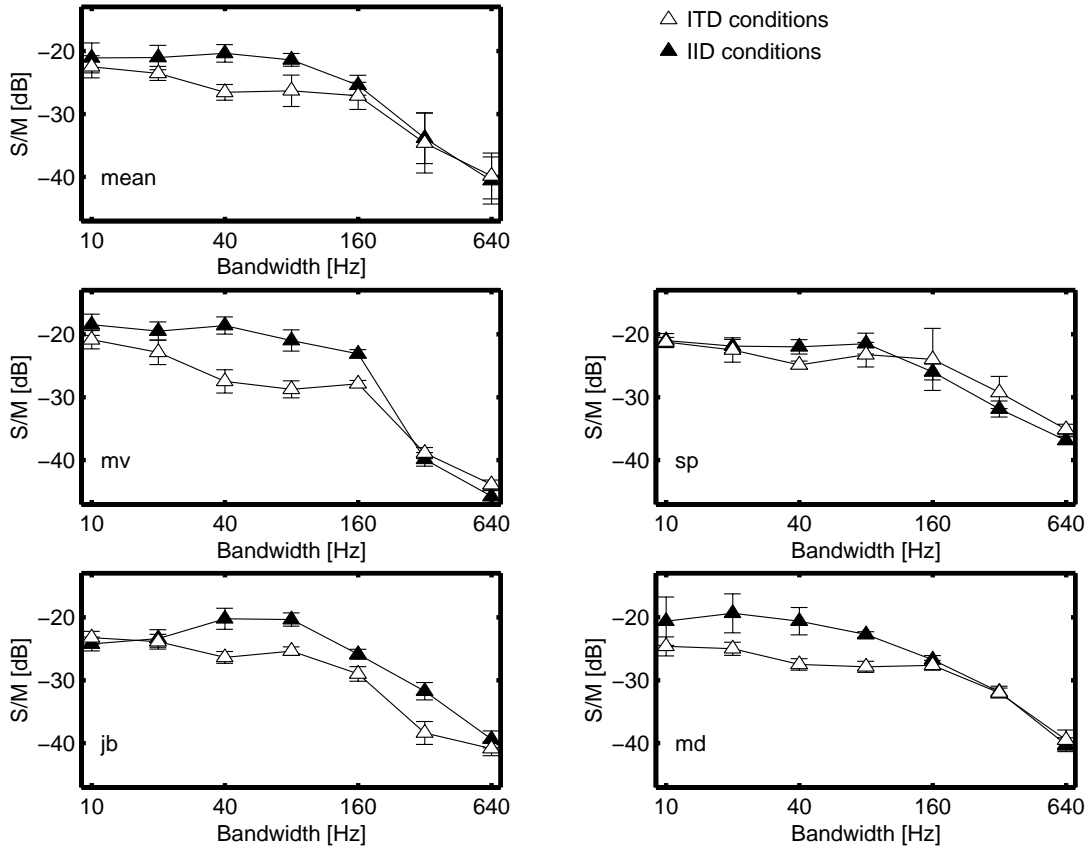
Figure 2.6: *Detection thresholds for an interaurally out-of-phase multiplied noise signal added to an in-phase sinusoidal masker as a function of the bandwidth of the noise for α=0 (filled triangles) and α=π/2 (open triangles). The upper panel shows the mean thresholds.*

being 3 to 7 dB higher than thresholds for the noise maskers. Furthermore, a small decrease in thresholds is observed if α is increased from 0 to π/2 for the 80-Hz-wide condition. For the 10-Hz-wide and the tone-on-tone conditions, the thresholds are independent of α.

2.5 Discussion

2.5.1 Effect of α

If the overall means of the data presented in Figs. 2.4 and 2.5 are considered, the IID thresholds are on average 1.4 dB higher than the ITD thresholds. This value is roughly in line with the observed 3 dB found by van de Par and Kohlrausch (1998b). Furthermore, the data shown in Fig. 2.7 show a minor influence of the masker-signal phase α, for both static and dynamically varying interaural differences. Many studies have been published which present differences between ITD-only and IID-only conditions varying between -8 and +6.5 dB (Jeffress *et al.*, 1956; Hafter *et al.*, 1969; Wightman, 1969; Jeffress and McFadden, 1971; McFadden *et al.*, 1971; Wightman, 1971; Yost, 1972b; Yost *et al.*, 1974; Robinson *et al.*, 1974). Only one study reports differences that deviate from these data with differences of up to 16 dB (Grantham and
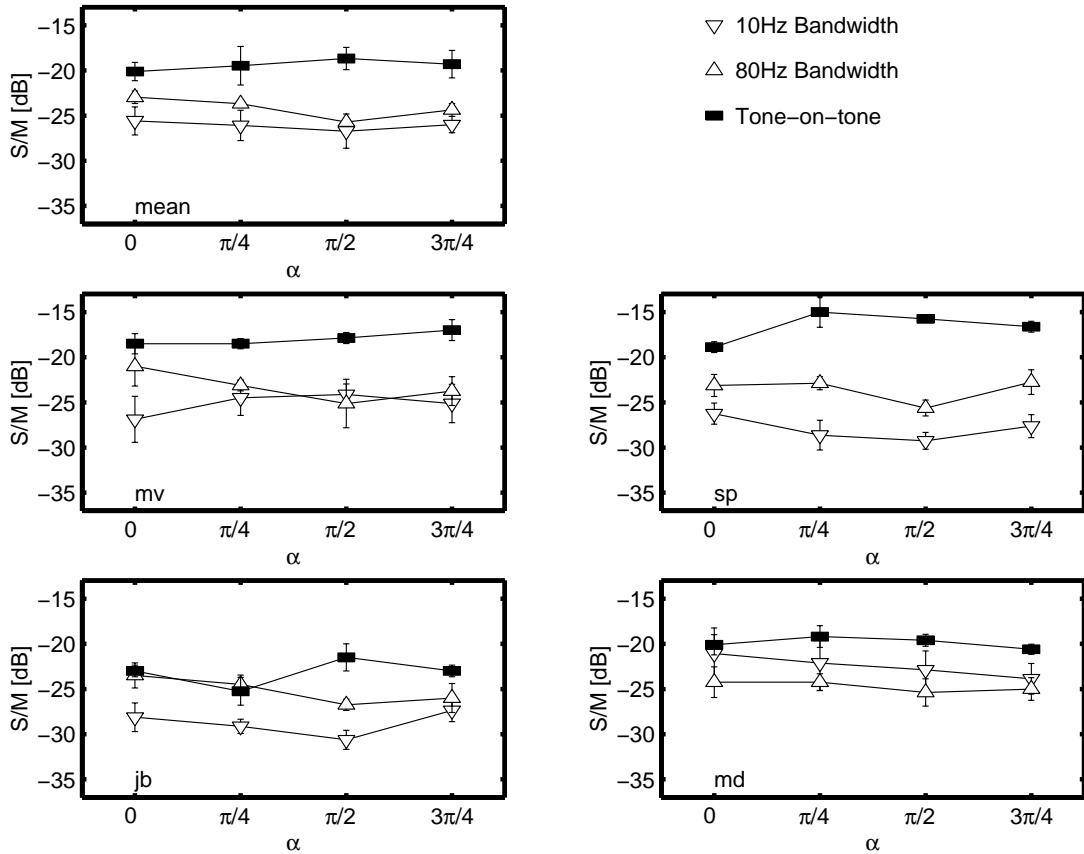
Figure 2.7: *Detection thresholds for an interaurally out-of-phase sinusoid added to an in-phase sinusoid (squares), a 10-Hz-wide multiplied noise (downward triangles) and a 80-Hz-wide multiplied noise (upward triangles) as a function of the fine-structure phase angle between signal and masker carrier. The lower 4 panels show thresholds for 4 subjects, the upper panel shows the mean thresholds.*

Robinson, 1977). We therefore conclude that our results are well within the range of other data, although there does not exist much consistency about the influence of α on detection thresholds.

If the thresholds for $\alpha = \pi/4$ and $\alpha = 3\pi/4$ are compared, only small threshold differences of less than 3 dB are found. Grantham and Robinson (1977) reported differences varying between -5 and +8 dB across different subjects. Also studies of Robinson *et al.* (1974) and Hafter *et al.* (1969) report differences within that range.

Corresponding to results from other studies (cf. McFadden *et al.*, 1971; Jeffress and McFadden, 1971; Grantham and Robinson, 1977), large differences exist across subjects when the effect of α is concerned. Some subjects seem to be more sensitive to signals producing ITDs, and some to IIDs. Thus, one model with a fixed set of parameters can never account for these interindividual differences. But since we are comparing theories and trying to model the general trend, we focus on the mean data knowing that individual differences are not taken into account.

If binaural detection were based on changes in laterality resulting from a combined time-intensity image, different thresholds would be expected for $\alpha=\pi/4$ and $\alpha=3\pi/4$. For $\alpha=\pi/4$, the interaural differences in time and intensity point in the same direction and the combined image would be lateralized more than for each cue separately, while for $\alpha=3\pi/4$, the ITDs and IIDs would (at least) partially cancel each other. The very similar threshold values suggest that detection is not based on changes in laterality resulting from a combined time-intensity image.

2.5.2 Binaural sluggishness

Several studies have provided evidence that the binaural auditory system is sluggish. We can classify these studies into two categories. The first category comprises experiments that determine the ability of human observers to detect interaural differences against a reference signal that contains no interaural differences. For example, if observers have to discriminate a binaural amplitude modulated noise in which the modulating sinusoid is interaurally in-phase, from the same amplitude modulated noise in which the modulator is interaurally out-of-phase, a substantial increase in the modulation depth at threshold is observed if the modulation frequency is increased from 0 to 50 Hz (Grantham, 1984a). Similar results were found for dynamically varying ITDs (Grantham and Wightman, 1978). However, the time constant of processing dynamically varying ITDs seems to be longer than for IIDs. Estimates for these constants are approximately 200 ms and 50 ms, respectively (Grantham, 1984a). Also many binaural masking conditions like MoSπ fall into this category of detection against a monaural reference signal (Zurek and Durlach, 1987). The second category comprises binaural detection experiments in which the masker has a time-varying correlation (cf. Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998). These studies show that modulation rates of interaural correlation as low as 4 Hz result in large increases in detection thresholds.

The experiments performed in our study clearly belong to the first group, because the correlation of the masker is always one. The fact that our results do not show any difference between the 10 and 80-Hz-wide conditions runs counter to an expectation based on binaural sluggishness. If one tries to characterize the rate at which the interaural differences change from leading to lagging in each ear, one could take the expected number of zero-crossings of the low-pass noise used in generating the multiplied noise. Roughly, if the low-pass noise changes its sign, the resulting interaural difference in an MoSπ condition also changes its sign. Thus, the number of zero-crossings represents the number of changes per second in lateralization pointing to the left or right ear. For a 10-Hz-wide noise, the expected number of lateralization changes amounts to 5.8 per second, while for the conditions at 80 Hz bandwidth, the

expected number is 46.2 (Rice, 1959). On the basis of the expected number of zero-crossings, assuming that the binaural system is sluggish in its processing of binaural cues, a difference in detection thresholds is expected between conditions at 10 and 80-Hz bandwidth. Furthermore, assuming that the time constant for processing ITDs is longer than for IIDs (Grantham, 1984a), the ITD-only thresholds should be higher than the IID-only thresholds for the 80-Hz-wide condition. The MANOVA analysis shows that the bandwidth of the multiplied noise is not a significant factor, indicating that the thresholds between the 10-Hz-wide and the 80-Hz-wide conditions are similar. Further-more, the data do not show the expected difference between the IID-only and the ITD-only conditions for a bandwidth of 80 Hz. Thus, effects of sluggishness, although expected, were not found in this study.

2.5.3 Off-frequency listening

For bandwidths beyond 80 Hz using a multiplied-noise signal, the thresholds decrease with increasing bandwidth (see Fig. 2.6). This is probably caused by the fact that the signal bandwidth exceeds the equivalent rectangular bandwidth of the auditory filters. Thus, the signal-to-masker ratio within an auditory filter tuned to a frequency just below or just above the masker frequency will be larger than for an on-frequency filter, resulting in lower detection thresholds if off-frequency filters can be used for detection. These off-frequency effects start to play a role for a signal bandwidth of 160 Hz. This indicates that for the results of experiment 2, where the maximum employed bandwidth was 80 Hz, off-frequency listening is not likely to influence detection thresholds. Hence the externally presented interaural differences are very similar to the differences after peripheral filtering for all experiments. Therefore we can validly compare the parameters of the distributions of the interaural differences at threshold across experiments with multiplied noise as masker and as signal.

One noteworthy effect seen in the data which is a significant factor according to our statistical assessment is that the ITD-only thresholds for a multiplied-noise signal decrease by 5 dB when the bandwidth is increased from 10 to 80 Hz, while for a multiplied-noise masker, this decrease does not occur. It is not clear what causes this effect.

2.5.4 Models based on the evaluation of IIDs and ITDs

In this section we analyze the contribution of static and dynamic cues to binaural detection. For this purpose we consider the mean, the standard deviation and the RMS of the probability density functions for IIDs and ITDs at threshold for the mean data shown in Figs. 2.4 and 2.5. The left panel of Fig. 2.8 shows the standard deviations of the probability density functions for IID only conditions as a function of the mean IID, while the right panel

shows the standard deviations of the ITD functions as a function of the mean ITD for the ITD only conditions. The open symbols represent thresholds from experiment 1, the filled symbols represent thresholds from experiment 2. The squares represent the 10-Hz-wide noise, the upward triangles the 80-Hz-wide noise and the downward triangles the tone-on-tone conditions. The data at the left side in each panel represent μ=0, while from left to right, the static component increases. With increasing static component, the mean of the interaural difference at threshold level increases also, while the standard deviation shows a minor decrease.
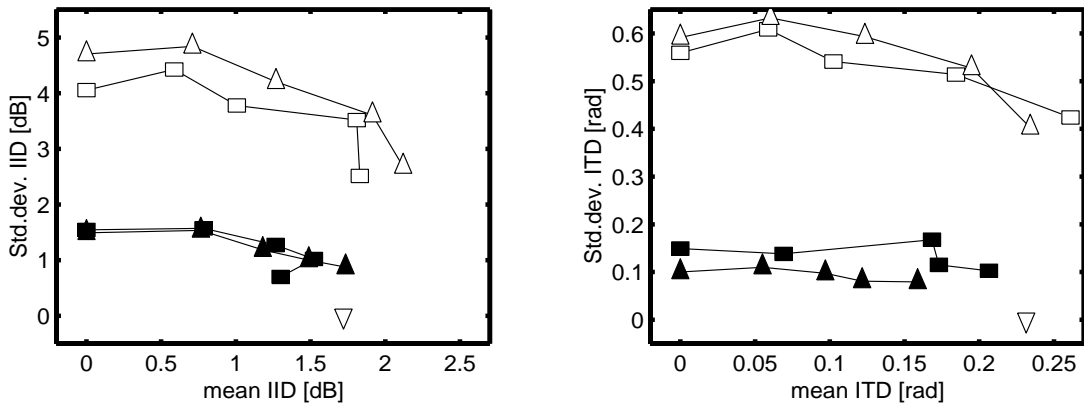
Figure 2.8: *Mean and standard deviation of the interaural cues at threshold level at 10-Hz bandwidth (squares) and 80-Hz bandwidth (upward triangles). The open symbols represent data for a diotic multiplied-noise masker (experiment 1), the filled symbols represent data for an interaurally out-of-phase multiplied noise signal (experiment 2). The downward triangles represent the data for the tone-on-tone conditions (experiment 4). The left panels shows data for the IID-only conditions, the right panel shows data for the ITD-only conditions.*

Clearly, the mean interaural differences at μ=2 for the conditions with multiplied noise (upward triangles and squares at the right side of each panel) are very similar to the tone-on-tone conditions (downward triangles). Furthermore, we see that points for 10-Hz bandwidth lie very close to points for 80-Hz bandwidth. This is expected, because these conditions have very similar detection thresholds. Because the mean and standard deviation of the interaural difference are independent of the bandwidth of the signals, similar thresholds result in similar statistics of the interaural differences.

The standard deviations for experiment 2 (filled symbols) are approximately 4 times smaller than for experiment 1 (open symbols). To end up with a similar standard deviation for IIDs at threshold for experiment 2 as in experiment 1, the signal-to-masker ratio must amount to -11 dB for μ=0. However, the data show a threshold of -21 dB. Thus, the standard deviations of the interaural differences cannot be used to correctly predict binaural masker thresholds for both experiments[4].

---

[4]In addition, other moments of the PDFs of the interaural differences were evaluated at

The only data available in the literature using comparable stimuli are those published by Grantham and Robinson (1977). Similar to our procedure, they used a noise stimulus with a certain DC offset and multiplied this noise with a sinusoidal carrier. The resulting bandpass noise (120-Hz wide) was used as a masker in an MoSπ condition. They did not vary the DC offset, however, but fixed it to a value that corresponds to μ=2.05 in our framework. In addition, they included the tone-on-tone conditions in their experiment. Because of the fact that their IID data are relatively high and their ITD data are relatively low compared to our results, we focus on the relative difference between the conditions with noise maskers and tonal maskers. For IIDs only, Grantham and Robinson (1977) found a mean IID at threshold of 3.2 dB and 3.1 dB for tonal and noise maskers, respectively. For ITD-only conditions, these values amounted to 0.080 and 0.106 rad, respectively. Thus, in correspondence with our data, the mean interaural differences for the tonal masker are slightly lower than for the noise masker for α=$\pi/2$. From our analysis based on PDFs, it is clear that the close correspondence between the mean values of the interaural cues found by Grantham and Robinson relies on their specific choice of μ. Had they chosen a lower value, then they probably would have found larger discrepancies: for μ=0, the mean interaural cue is equal to zero at threshold, while for large values of μ, mean interaural differences of up to 4 dB or 0.1 rad may be found at threshold. Thus, the mean interaural difference cannot account for the complete set of data either.

A straightforward way to combine the sensitivity for static and dynamically varying interaural differences is to consider the RMS value of the interaural differences. Fig. 2.9 shows the RMS values of the interaural cues of the mean data of experiment 1 and 2 as a function of the mean interaural cues. The format is the same as in Fig. 2.8. Within one experiment, the RMS value remains fairly constant, although there is a tendency for the RMS to decrease with increasing mean for the data with a multiplied-noise masker and to increase for the multiplied-noise signal. However, we can reject the RMS as a valuable detection variable because for this measure too, the values of the two experiments differ by a factor 2 to 4.

2.5.5 Models based on the interaural correlation
Another detection statistic that is often proposed to account for binaural masked thresholds is the interaural correlation. Domnitz and Colburn (1976) argued that for an out-of-phase sinusoidal signal combined with a diotic Gaussian-noise masker, models based on the PDFs of the interaural differences and models based on the interaural correlation are equivalent. We will now explore whether this statement also holds for our stimuli. For

---

threshold level. These properties also resulted in significantly larger differences between the thresholds of experiments 1 and 2 than the observed difference of 2.2 dB.
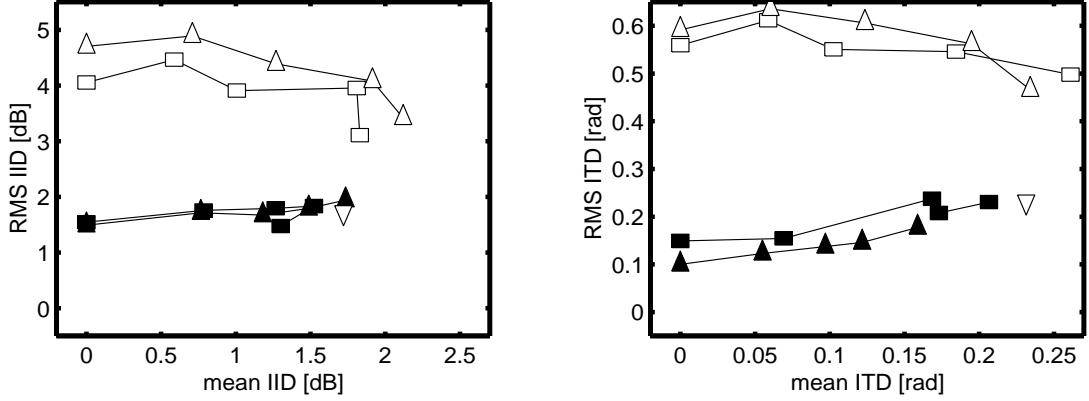
Figure 2.9: *Mean and RMS value of the interaural cues at threshold level in the same format as Fig. 2.8.*

a diotic masker alone, the interaural correlation equals +1. The interaural correlation for the masker plus signal is given by (Durlach *et al.*, 1986):

$$\rho = \frac{1 - \frac{<S^2>}{<M^2>}}{1 + \frac{<S^2>}{<M^2>}}, \tag{2.1}$$

where $\frac{<S^2>}{<M^2>}$ denotes the signal-to-masker power ratio. This equation holds provided that masker and signal are statistically independent. For our stimuli, this is true provided that $\mu=0$ or $\alpha=\pi/2$. Thus, the correlation is only dependent on the signal-to-masker ratio and does not depend on the physical nature of the interaural difference (i.e., IIDs or ITDs). Furthermore, the correlation is not dependent on the shape of the PDFs of the interaural differences. Therefore, contrary to models based on the PDF of the interaural cues, a model based on the cross-correlation will yield *equal* thresholds for experiment 1 and 2, on the condition that $\mu=0$ or $\alpha=\pi/2$ (i.e., masker and signal uncorrelated). This implies that the statement from Domnitz and Colburn (1976) cannot be generalized to our stimuli and that with our stimuli a valuable way to distinguish between cross-correlation models and binaural-cue-based models is available.

Equation 2.1 is, however, not applicable under conditions where $\alpha \neq \pi/2$ and $\mu > 0$. For $\alpha=0$, the interaural correlation can be written as (see Appendix 2.B)

$$\rho = \frac{<M^2>-<S^2>}{\sqrt{\left(<M^2>+<S^2>+\frac{<2SM\mu>}{\sqrt{1+\mu^2}}\right)\left(<M^2>+<S^2>-\frac{<2SM\mu>}{\sqrt{1+\mu^2}}\right)}}. \tag{2.2}$$

From the above equation, we see that the static component has a strong influence on the interaural correlation of an MoSπ stimulus. We therefore computed the predictions according to a simple interaural correlation model for an MoSπ condition with a multiplied-noise masker as a function of the

static component μ. The overall mean signal-to-masker ratio at threshold for α=0 and μ=0 in our experiments was -24.3 dB resulting in an interaural correlation of 0.9926. We therefore used a *decorrelation* of 0.0074 as a just noticeable difference (JND) in the interaural correlation. From this correlation JND, we computed the signal-to-masker ratios that produce the same amount of decorrelation as a function of the static component μ. The computed thresholds are shown by the filled squares in Fig. 2.10. The open symbols represent the mean experimental data from experiment 1 for α=0. Clearly, the predicted values based on the change in the interaural correlation show a large increase in threshold with increasing static component. This results from the insensitivity of the interaural correlation to static interaural intensity differences. This strong increase was not found in our experimental data, implying that a simple cross-correlation model without extensions for static interaural intensity differences cannot account for the data.
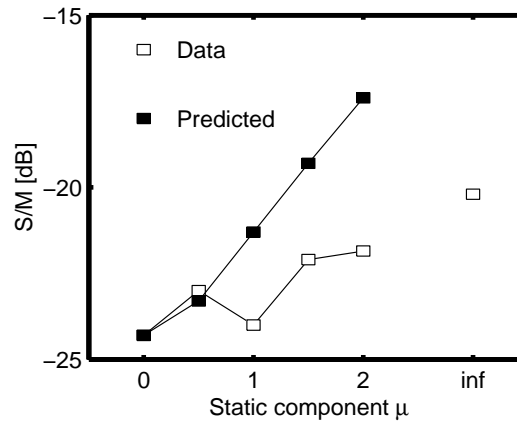
Figure 2.10: *Predicted values according to a simple interaural correlation model (filled symbols) and experimental data (open symbols) for an MoSπ condition with a multiplied-noise masker and α=0 as a function of the static component.*

2.5.6 A new model

The question arises which other detection statistic can be used to characterize our data. Because the experimental data show approximately similar thresholds across all the experimental conditions, we propose that a model based on the *difference intensity* of the signals arriving at both ears could be a valid detection statistic. We define difference intensity as the intensity in the stimulus obtained when the waveforms to the two ears are equalized and differenced. This approach is related to Durlach's EC-theory (Durlach, 1963), but the two are not equivalent: the EC-theory predicts BMLDs, while this approach describes binaural thresholds directly. Such an approach also differs from a cross-correlation model for stimuli containing static IIDs. For tone-on-tone conditions with α=0 (i.e., only static IIDs are present), a cross-correlation model fails to detect the static IID, while the present approach is sensitive to this cue, as will be explained below.

In our approach, an internal interaural delay and an internal interaural intensity difference are determined which tend to equalize the masking signal arriving at the two ears. These parameters can be obtained from the masker-alone intervals. For the signal interval, the *masker* is equalized and subsequently eliminated by a cancellation process. For signals producing interaural intensity differences, the amount of signal remaining after the described equalization process increases with an increasing interaural intensity difference; hence, static IIDs can be detected. For MoSπ stimuli, the masker-elimination process is simply performed by subtracting the waveforms at both ears, computing the power of the remaining signal and using this as a decision variable. If the signals arriving at the left and the right ears are denoted by L(t) and R(t), respectively, the difference intensity D is defined as:

$$D = \int_0^T \left(L(t) - R(t) \right)^2 dt. \tag{2.3}$$

Here, T denotes the interval length. In fact, for an MoSπ condition, D is exactly equal to four times the energy of the out-of-phase signal. Hence a model based on this processing scheme would give equal thresholds for all subcritical conditions (i.e., with a noise bandwidth of 80 Hz or less), because the difference intensity D is directly proportional to the signal intensity. For all MoSπ conditions as presented in this study, the only limiting factor in the detection process will be the internal errors, since the masker can be cancelled completely. The magnitude of this internal error can in principle be set to any (fixed) value. We can therefore simply derive predictions for the experiments 1, 2 and 4. We adjusted this value to result in a signal-to-masker threshold value of -22 dB. The predictions according to this model are shown in Figs. 2.11 and 2.12. Figure 2.11 shows the thresholds for experiments 1 and 2 (left and right panels, respectively) in the same format as Fig. 2.4. The solid line represents the model predictions. Figure 2.12 shows the mean data of Fig. 2.7 (experiment 4) combined with the model predictions (solid line). We did not simulate the data for experiment 3, because modeling off-frequency effects needs a much more complicated model. We are in the process of describing such a model but it is far beyond the scope of this chapter to include it here.

The predictions match the experimental data quite well. There are, however, some differences between data and model predictions, which can be summarized as follows:
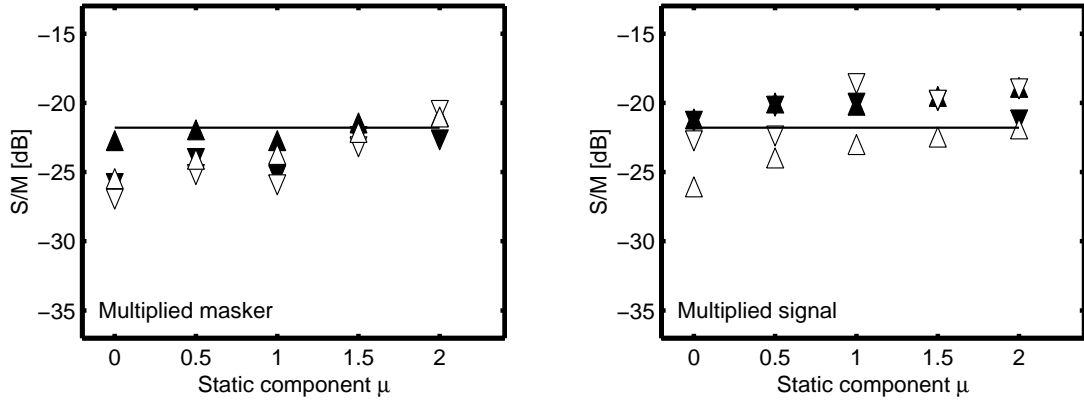
Figure 2.11: *Predicted values of experiments 1 and 2 according to an EC-like model (solid line) and experimental data (symbols, in the same format as in Fig. 2.4) for an MoSπ condition with a multiplied-noise masker (left panel) and a multiplied-noise signal (right panel) as a function of the static component.*

- The slight increase of thresholds with increasing static component (see Fig. 2.11) is not present in the model simulations

- the fact that the thresholds for a multiplied-noise masker are on average 2.2 dB lower than the data for a multiplied-noise signal is not represented in the model predictions

- the fact that the tone-on-tone conditions (filled symbols in Fig. 2.12) give higher thresholds than the multiplied-noise maskers (open symbols in Fig. 2.12) cannot be understood by this simple model.

- the model's performance is independent of the masker-signal phase difference α, while the experimental data show an overall difference of 1.4 dB if α is changed from 0 to $\pi/2$.

However, some of these effects can be understood by considering the presence of nonlinearities in the peripheral auditory system. For example, we can qualitatively account for the fact that the ITD thresholds are 1.4 dB lower than the IID thresholds by assuming that peripheral compression at the level of the basilar membrane has an effect on binaural masked thresholds. This issue was already discussed by van de Par and Kohlrausch (1998c, 2001). Following their hypothesis, basilar membrane compression results in a decrease of the IIDs in the internal representation and has no effect on ITDs, resulting in higher IID thresholds if the difference intensity is used as a decision variable. Peripheral compression also has a strong effect on binaural masked thresholds with different masker-envelope statistics (van de Par and Kohlrausch, 1998b). Compression reduces interaural intensity differences most strongly for high envelope values. Because a sinusoidal masker has no valleys in the envelope and multiplied noise has many valleys in its envelope, it is expected that IID thresholds for a sinusoidal masker (see Fig. 2.5) are higher than for a multiplied noise masker (see Fig. 2.4), in line with the observed overall
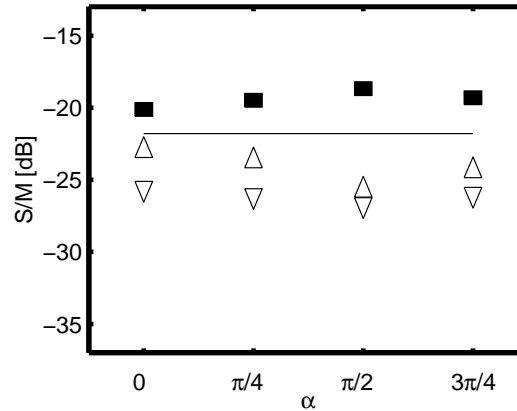
Figure 2.12: *Predicted values of experiment 4 according to an EC-like model (solid line) and experimental data (symbols, same format as in Fig. 2.7) for an MoSπ condition as a function of the masker-signal phase difference α.*

difference of 2.2 dB. This rationale also holds for the increase in thresholds with an increase of the static component. As described in Section 2.2, the static component has a strong influence on the envelope statistics of the stimuli, resulting in fewer valleys in the envelope if the static component is increased.

However, an EC-like model fails to account for the tone-on-tone data shown in Fig. 2.12. Here, the tone-on-tone conditions show distinctively higher thresholds than the conditions with a multiplied-noise masker, while an EC-like model without peripheral compression predicts equal thresholds. The inclusion of peripheral compression might account for the fact that the IID tone-on-tone thresholds are higher than the conditions with noise maskers. It is more difficult, however, to see how compression can explain the difference in threshold between tonal and noise maskers with ITDs only.

In summary, the binaural masked thresholds for MoSπ stimuli in the present study seem to be best described by a peripheral preprocessing stage followed by a differencing device that calculates the difference intensity of the signals from the left and right ears and uses this output as a detection variable. Although such an approach cannot account for all data presented here, it provides better predictions than a model based on the probability distributions of the interaural differences or a cross-correlation model per se.

2.A Appendix: Distributions of interaural differences

2.A.1 ITD probability density for a multiplied-noise masker

Figure 2.1 shows a vector diagram illustrating an interaurally out-of-phase signal $\mathbf{S}_l$ and $\mathbf{S}_r$, a noise masker $\mathbf{M}$ and the resulting signals $\mathbf{L}$ and $\mathbf{R}$ presented at the two ears. The phase angle between masker and signal fine structures is denoted by α, the phase angle between $\mathbf{L}$ and $\mathbf{R}$ by ϕ (with $-\pi < \phi \le \pi$). For the interaural phase ϕ, a convenient expression relating the variables was given by Zurek (1991):

$$\phi = \pi/2 - \arctan \frac{A\cos\alpha - S}{A\sin\alpha} - \arctan \frac{A\sin\alpha}{S + A\cos\alpha}. \tag{2.A.1}$$

Here, A represents the instantaneous value of the low-pass noise. Note that A is defined as the sum of a DC component μ and a Gaussian noise with mean zero and RMS=1, which is rescaled to have unit power. This results in a probability density function for A given by:

$$p(A) = \sqrt{\frac{1 + \mu^2}{2\pi}} \exp\left(-1/2(-\mu + A\sqrt{1 + \mu^2})^2\right). \tag{2.A.2}$$

The phase probability density p(ϕ) can be written as the product of the probability density for A(ϕ), multiplied by the absolute derivative of A to ϕ:

$$p(\phi) = p(A(\phi)) \left| \frac{dA(\phi)}{d\phi} \right|. \tag{2.A.3}$$

Equation 2.A.1 gives an expression for ϕ(A). However, to derive an expression for p(ϕ), an expression for A(ϕ) is needed. We have to study two distinct cases. We see from Fig. 2.1 that for A$\ge$0, $\phi \ge 0$ and that for A<0, $\phi < 0$. A(ϕ) can be derived by inverting Eq. 2.A.1, resulting in a second order polynomial equation which normally has two roots. However, according to the above restriction, the solution for A(ϕ) results in:

$$A(\phi) = -S\sin\alpha\tan(\phi - \pi/2) + \beta S\sqrt{\sin^2\alpha\tan^2(\phi - \pi/2) + 1}, \tag{2.A.4}$$

where β equals 1 for $\phi \ge 0$ and -1 otherwise. The derivative of A to ϕ becomes

$$\frac{dA}{d\phi} = \frac{-S\sin\alpha}{\cos^2(\phi - \pi/2)} + \frac{\beta S\sin^2\alpha\tan(\phi - \pi/2)}{\cos^2(\phi - \pi/2)\sqrt{\sin^2\alpha\tan^2(\phi - \pi/2) + 1}}. \tag{2.A.5}$$

Now, all parameters for Eq. 2.A.3 are known and p(ϕ) can be calculated.

2.A.2 IID probability density for a multiplied-noise masker

The probability density function for the IID can be derived in a very similar way as was done for the ITD. The interaural intensity difference λ is defined as:

$$\lambda = 20 \log \frac{|R|}{|L|} = 10 \log \frac{A^2 + S^2 + 2AS \cos \alpha}{A^2 + S^2 - 2AS \cos \alpha}. \tag{2.A.6}$$

Inverting the above equation results in:

$$A(\lambda) = -S \cos \alpha \frac{1 + 10^{\lambda/10}}{1 - 10^{\lambda/10}} + \beta S \sqrt{-1 + \left(\frac{1 + 10^{\lambda/10}}{1 - 10^{\lambda/10}} \right)^2 \cos^2 \alpha}, \tag{2.A.7}$$

for $\lambda \geq 0$, A$\geq$**0** and λ <**0**, A<**0**. Therefore, $\beta = 1$ for $\lambda \geq 0$ and -1 otherwise. The probability density function for λ is given by:

$$p(\lambda) = p(A(\lambda)) \left| \frac{dA(\lambda)}{d\lambda} \right|. \tag{2.A.8}$$

Equations 2.A.2, 2.A.7 and 2.A.8 give all necessary parameters to calculate p(λ).

2.A.3 ITD and IID probability density for a multiplied-noise signal

When exchanging the role of multiplied noise and sinusoid (i.e., the multiplied noise becomes an interaurally out-of-phase signal), we obtain a new relation between interaural phase ϕ and the instantaneous value of the low-pass noise A:

$$A(\phi) = S \sin \alpha \tan(\phi - \pi/2) + \beta S \sqrt{\sin^2 \alpha \tan^2(\phi - \pi/2) + 1}. \tag{2.A.9}$$

Again, β equals one for $\phi \geq 1$ and -1 otherwise. The probability density function is then given by Eq. 2.A.3, where A(ϕ) has to be taken from Eq. 2.A.9.

The IID probability density function is given as in Eq. 2.A.8, however with:

$$A(\lambda) = S \frac{1 + 10^{\lambda/10}}{1 - 10^{\lambda/10}} \cos \alpha + \beta S \sqrt{-1 + \left(\frac{1 + 10^{\lambda/10}}{1 - 10^{\lambda/10}} \right)^2 \cos^2 \alpha}, \tag{2.A.10}$$

where β=1 for A≥ 0 and -1 otherwise.

2.B Appendix: interaural correlation with multiplied noise

For a multiplied-noise masker combined with a sinusoidal test signal with $\alpha=0$ in an MoSπ condition, the waveforms arriving at the left and right ears (L(t) and R(t), respectively), are given by

$$L(t) = M\sqrt{2}\frac{N(t)+\mu}{\sqrt{1+\mu^2}}\sin(2\pi ft) + \sqrt{2}S\sin(2\pi ft) \tag{2.B.1}$$

$$R(t) = M\sqrt{2}\frac{N(t)+\mu}{\sqrt{1+\mu^2}}\sin(2\pi ft) - \sqrt{2}S\sin(2\pi ft). \tag{2.B.2}$$

Here, M denotes the RMS value of the masker, S the RMS value of the signal, N(t) denotes the low-pass noise that is used for generating the multiplied noise, μ denotes the static component and f is the carrier frequency. The definition of the normalized interaural correlation is

$$\rho = \frac{<L.R>}{\sqrt{<L^2><R^2>}}, \tag{2.B.3}$$

where $<\,.\,>$ denotes the expected value. Combining Eqs. 2.B.1 to 2.B.3 results in

$$\rho = \frac{<M^2>-<S^2>}{\sqrt{\left(<M^2>+<S^2>+\frac{<2SM\mu>}{\sqrt{1+\mu^2}}\right)\left(<M^2>+<S^2>-\frac{<2SM\mu>}{\sqrt{1+\mu^2}}\right)}}. \tag{2.B.4}$$

CHAPTER 3

The influence of interaural stimulus uncertainty on binaural signal detection[1]

In this chapter, a study of the influence of stimulus uncertainty in binaural detection experiments and the predictions of several binaural models for such conditions is described. Masked thresholds of a 500-Hz sinusoid were measured in an $N\rho S\pi$ condition for both running and frozen-noise maskers using a 3IFC procedure. The nominal masker correlation varied between 0.64 and 1, and the bandwidth of the masker was either 10, 100 or 1000 Hz. The running-noise thresholds were expected to be higher than the frozen-noise thresholds because of stimulus uncertainty in the running-noise conditions. For an interaural correlation close to +1, no difference between frozen-noise and running-noise thresholds was expected for all values of the masker bandwidth. These expectations were supported by our experimental data: for interaural correlations less than 1.0 substantial differences between frozen and running-noise conditions were observed for bandwidths of 10 and 100 Hz. Two additional conditions were tested to further investigate the influence of stimulus uncertainty. In the first condition a different masker sample was chosen on each trial, but the correlation of the masker was forced to a fixed value. In the second condition one of two independent frozen-noise maskers was randomly chosen on each trial. Results from these experiments emphasized the influence of stimulus uncertainty in binaural detection tasks: if the degree of uncertainty in binaural cues was reduced, thresholds decreased towards thresholds in the conditions without any stimulus uncertainty. In the analysis of the data, stimulus uncertainty was expressed in terms of three theories of binaural processing: the interaural correlation, the EC-theory and a model based on the processing of IIDs and ITDs. This analysis revealed that none of the theories tested could quantitatively account for the observed thresholds. In addition, it was found that, in conditions with stimulus uncertainty, predictions based on correlation differ from those based on the EC-theory.

3.1   Introduction

For a period of more than 50 years, the phenomenon of the binaural masking level difference (BMLD) has intrigued psychoacousticians. It has been shown that the interaural correlation of both the masker and the signal are important parameters influencing binaural detection thresholds. For example, when a low-frequency out-of-phase sinusoid is added to an in-phase broadband

---

[1]This chapter is based on Breebaart and Kohlrausch (2001a).

noise masker (NoSπ condition), the threshold of audibility is up to 15 dB lower compared to that for an in-phase sinusoidal signal (i.e., NoSo condition, cf. Hirsh, 1948b; Hafter and Carrier, 1970; Zurek and Durlach, 1987). If the signal has an interaural correlation of +1 and an out-of-phase masker is used (i.e., an NπSo condition), BMLDs of up to 12 dB are reported (Jeffress *et al.*, 1952, 1962; Breebaart *et al.*, 1998).

In experiments where the masker correlation was varied between -1 and + 1 using Sπ signals, Robinson and Jeffress (1963) found a monotonic increase in the BMLD with increasing interaural correlation. Small reductions from +1 of the interaural masker correlation in an NρSπ condition led to a large decrease of the BMLD, while for smaller correlations, the slope relating BMLDs to interaural correlation was shallower. The stimuli used by Robinson and Jeffress (1963) were composed by adding interaurally correlated noise with an interaurally uncorrelated noise. The relative intensities of both sources determined the *mean* interaural correlation. The consequence of this method for generating the stimuli is that for the masking noise alone, the interaural cues (i.e., interaural time- and intensity differences) fluctuate randomly. Moreover, because finite-length masker samples are used, the actual correlation within an observation interval can deviate considerably from the adjusted mean correlation. Thus, in terms of binaural cues, the masker contains uncertainty. The addition of the Sπ signal results in a change in the mean of the interaural cues but does not reduce the randomness of the interaural cues.

Analogous to monaural conditions (Lutfi, 1990), binaural masking can be attributed to two different sources. The first results from the limited resolution of the binaural auditory system and has been termed energetic masking by Lutfi. In models of binaural processing, this source of masking is included as internal noise. For example, the EC-theory summarizes the internal errors of timing and amplitude representation in the factor k, which is directly related to the BMLD. The second source of masking results from the uncertainty associated with the trial-to-trial variation of the binaural cues used to detect the signal (called informational masking by Lutfi). This source of masking has, as far as we are aware, not been modelled adequately (see Colburn *et al.*, 1997, for a discussion). In standard MLD conditions like NoSπ, the masker contains no uncertainty in terms of binaural cues: the interaural correlation is always exactly one, the energy of the difference signal between the right and left masker is zero, and the interaural differences in time and intensity are always exactly zero.

Because it is well known that the auditory system can benefit from the presence of binaural cues in a detection task, it is interesting to study the influence of uncertainty in these cues and the extent to which uncertainty limits detection. One of the possibilities to remove stimulus uncertainty

is by using *frozen* noise. Thresholds for frozen binaural maskers would thus reflect the amount of energetic masking. The difference in detection performance between running noise and frozen noise indicates the amount of informational masking and this is the main topic of this chapter. Data will be presented that were measured under conditions with and without stimulus uncertainty. Three common theories for binaural processing (the change in interaural correlation, the EC-theory and processing of interaural intensity differences, or IIDs, and interaural time differences, or ITDs) will be discussed for their ability to predict these data. We selected these theories because they are often used to explain binaural processing. In addition, they have been discussed for their ability to predict the amount of energetic masking in binaural conditions with non-Gaussian maskers (see Chapter 2 of this thesis). We will provide an analysis of stimulus uncertainty in terms of these models. An important assumption related to this analysis is that we assume that the detection strategy of human listeners is the same for all experimental conditions. Hence informational masking will be discussed assuming that binaural detection is based on only one of the three theories mentioned above.

3.2 Stimulus uncertainty

3.2.1 Interaural correlation

It is often assumed that a change in the interaural correlation induced by adding a signal to a masker can be used as a detection cue in binaural masking experiments. Various mathematical details have been published treating changes in the interaural correlation for different experimental paradigms. For example, Domnitz and Colburn (1976) argued that models based on interaural correlation and models based on interaural differences yield similar predictions for NoSπ conditions with Gaussian noise. Chapter 2 of this thesis presented data with non-Gaussian noise maskers for which this close correspondence between the change in the cross-correlation and the size of the interaural differences is no longer found. Durlach *et al.* (1986) determined an analytical expression for the interaural correlation in an NoSπ condition. Analytical expressions for the interaural waveform correlation and the interaural envelope correlation were derived by van de Par and Kohlrausch (1995) for NoSπ and later also for NoSm (van de Par and Kohlrausch, 1998a). Bernstein and Trahiotis (1996) showed that for NoSπ stimuli, the interaural correlation of the stimuli after peripheral preprocessing did account for their NoSo vs. NoSπ discrimination results for a wide range of center frequencies. Because this correlation approach is widely accepted, we will discuss stimulus uncertainty first in terms of this concept.

An interaurally partially correlated noise can be generated by adding an interaurally correlated noise ($N_0(t)$) and an interaurally out-of-phase noise ($N_\pi(t)$). In the following we assume that these two independent noise sources

have the same RMS value. To end up with a long-term normalized interaural correlation of ρ, the left-ear signal $L(t)$ and the right-ear signal $R(t)$ are combined as follows:

$$\begin{cases} L(t) = \frac{1}{2}\sqrt{2}\sqrt{1+\rho}N_0(t) + \frac{1}{2}\sqrt{2}\sqrt{1-\rho}N_\pi(t) \\ R(t) = \frac{1}{2}\sqrt{2}\sqrt{1+\rho}N_0(t) - \frac{1}{2}\sqrt{2}\sqrt{1-\rho}N_\pi(t) \end{cases}. \tag{3.1}$$

Because both N_0 and N_π stem from random processes, the short-term energy estimates (i.e., integrated over one interval in a 3IFC task) of N_0 and N_π, E_0 and E_π, respectively, can deviate substantially from their expected (i.e., long-term) values provided that the product of time and bandwidth is small. Furthermore, the samples taken from the two noise sources can be partially correlated. Fluctuation of the short-term estimate of the noise energy leads to a variability in the interaural correlation for a finite-length noise interval (see Appendix 3.A, Eq. 3.A.2 and Gabriel and Colburn, 1981; Richards, 1987). The interaural correlation of a finite sample will be referred to as *effective correlation,* ρ_{eff}, while the mean interaural correlation (i.e., the expected value of ρ_{eff}) will be referred to as *reference correlation,* ρ.

We determined the probability distribution for the interaural correlation for an NρSπ condition as a function of the signal-to-noise ratio, the bandwidth and the duration of the masker. From the mathematical expressions for the effective correlation probability distribution as given in Appendix 3.A, we found that three important factors affect the distribution for the effective correlation:

- Bandwidth and duration of the noise. With increasing duration and bandwidth, the variance of the effective correlation, ρ_{eff}, will decrease as a result of the decreasing variances of E_0 and E_π (see Appendix 3.A).

- The reference correlation. For a reference correlation of +1 (and -1), there is no correlation uncertainty, and the effective correlation will always be +1 (-1). On the other hand, for reference correlations between -1 and +1, the effective correlation will follow a distribution rather than have a fixed value. For a reference correlation close to zero, the width of the effective correlation distribution will be widest (i.e., the correlation uncertainty is maximum).

- The presence or absence of the signal. The addition of an Sπ signal results in a shift of the mean interaural correlation towards -1.

To demonstrate the effect of these properties upon the correlation uncertainty, probability density functions (PDFs) for a 300-ms noise and three different combinations of reference correlation and noise bandwidth are shown in Fig. 3.1. Each panel shows two distributions, one for the noise alone (solid line) and one for noise plus signal (dashed line). The signal had a duration of 200 ms, was temporally centered in the noise and had a level of 10 dB

below the masker level (i.e., S/N =-10 dB). It is clear that the *width* of the PDF decreases with increasing bandwidth (10 Hz in the left panel, 100 Hz in the middle and right panels) and with increasing correlation (0.9 in the left and middle panels, 0.8 in the right panel), indicating less correlation uncertainty. Furthermore, the peak of the curve migrates towards higher correlation values for decreasing bandwidth. This results from the fact that the interaural correlation is a nonlinear function of the noise energies E_π and E_o. If the mean interaural correlation is set to 0.9 and the bandwidth is 10 Hz, it can be observed in the left panel of Fig. 3.1 that there is a finite probability for correlation values to be smaller than 0.8, a value that differs by more than 0.1 from the mean interaural correlation. This property is highly asymmetric; correlations higher than +1 cannot occur. If, despite of this asymmetric property, the mean correlation is 0.9, the peak of the curve must occur at a correlation greater than 0.9.
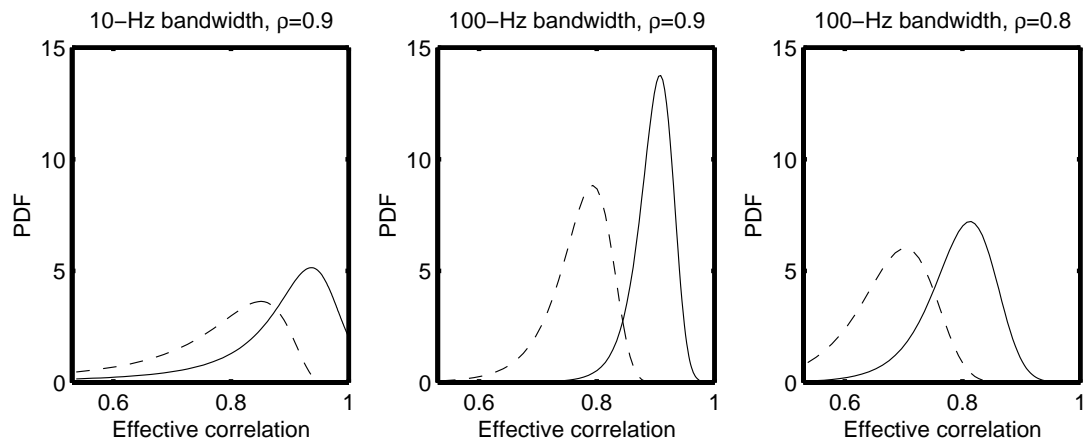
Figure 3.1: *Probability density functions for the effective interaural correlation. The left panel corresponds to a bandwidth of 10 Hz and a reference correlation of 0.9. For the middle and right panels these parameters were 100 Hz, 0.9 and 100 Hz, 0.8, respectively. The solid lines represent a 300-ms masker alone, the dashed lines represent a 300-ms masker with a 200-ms $S\pi$ signal. The distributions were calculated for the complete 300-ms interval. The signal-to-masker ratio was -10 dB.*

The addition of the $S\pi$ signal results in a shift of the curves towards lower correlation values. Furthermore, the distributions show a small increase in their widths. In the left panel (10-Hz bandwidth), the shift of the curve is small compared to the width of the distributions. Thus, from a signal-detection point of view, it is likely that at this signal-to-masker ratio, interaural correlation uncertainty can influence the detection performance. For a bandwidth of 100-Hz and a reference correlation of 0.9 (middle panel), the curves for masker alone and masker plus signal show a smaller overlap. If the reference correlation is reduced to 0.8 (right panel), the amount of overlap is increased. We can conclude that both the bandwidth and the reference correlation of the noise have a strong effect on the detectability of the signal in terms of interaural correlation. If human observers indeed use

the interaural correlation as a decision variable, thresholds should depend on the stimulus parameters that determine the amount of correlation uncertainty.

In fact, experimental data from Gabriel and Colburn (1981) and van der Heijden and Trahiotis (1998) confirm this hypothesis. Gabriel and Colburn (1981) found that if the bandwidth of a noise stimulus is increased from 5 to 1000 Hz, the interaural correlation jnd for a reference correlation of 0 decreases by a factor 2. Moreover, the change in the correlation jnd was largest for bandwidths below the critical bandwidth. For masker bandwidths beyond the critical bandwidth, the correlation jnd did not change by a large amount. This might indicate that the internal interaural correlation is evaluated after filtering in the periphery of the auditory system. For a reference correlation of +1, the correlation jnd remained approximately constant for bandwidths up to the critical bandwidth. However, an increase in the correlation jnd was observed when the bandwidth was increased well beyond this value. Although critical band filtering seems to play a role under these conditions, this increase in thresholds is not yet understood. The data obtained by van der Heijden and Trahiotis (1998) showed that the correlation dependence of thresholds is much stronger at narrow bandwidths (3 Hz) than at large bandwidths (900 Hz). This corresponds to the notion that correlation uncertainty influences detection, because the probability density function for the correlation is wider at narrow bandwidths (see Fig. 3.1).

The consequences of the use of frozen noise upon correlation uncertainty are very simple. If exactly the same noise waveform is used in each trial and each token of a multiple-interval, forced-choice procedure, there is no uncertainty in the masker interval; the interaural correlation always has the same value. The addition of the $S\pi$ signal results in a deviation from this fixed value. The actual value depends on the signal-to-noise ratio: a higher signal level results in a lower correlation.

3.2.2 The EC-Theory

Durlach's EC-theory (Durlach, 1963) is another well-known theory to account for BMLDs. According to this theory the waveforms which arrive at both ears are modified by an interaural time delay and an interaural level adjustment in such a way that the masker waveforms are equalized (the E-process). This process is performed imperfectly as a result of internal errors. Subsequently, the stimulus in one ear is subtracted from the stimulus in the other ear (cancellation, or C-process). In binaural conditions, this process often leads to an improvement of the signal-to-masker ratio and hence to the prediction of a positive BMLD. In an $N\rho S\pi$ condition no equalization is available that yields a signal-to-masker ratio improvement. Hence the improvement in signal-to-masker ratio is obtained by calculating the amount of masker energy

that is removed by the cancellation process. From Eq. 3.1, it can be seen that the common part of the masking noise will be removed and that the N_π masker portion remains. Thus, the amount of stimulus energy that remains after the EC-process equals [2]

$$E_{EC} = 2(1 - \rho)E_\pi + 4E_S, \tag{3.2}$$

where E_S denotes the signal energy and E_{EC} is the energy of the difference signal between the left and right ears. The use of the difference energy as a decision variable was also suggested from the results in chapter two of this thesis for NoSπ stimuli with non-Gaussian maskers. If no signal is present, E_S is simply zero. Assuming that E_{EC} is used as a decision variable, stimulus uncertainty will influence the detection task because E_π is a random variable with a certain mean and standard deviation. Equations 3.A.4 to 3.A.6 in Appendix 3.A give a description of the variability of E_π. A graphical representation of this description is shown in Fig. 3.2. The format is the same as in Fig. 3.1; the left and middle panels correspond to a reference correlation of 0.9, the right panel to 0.8. The bandwidth of the noise is 10 Hz in the left panel and 100 Hz in the other panels. Each panel contains two curves; the solid lines represent the PDF for E_{EC} for a masker alone, the dashed lines for masker plus signal. For simplicity it is assumed that the RMS value of the noise sources equals 1 (arbitrary units) and the signal-to-masker ratio is -10 dB. The masker had a duration of 300 ms. The curves in Fig. 3.2 show a similar behavior as in Fig. 3.1; a wider bandwidth or a higher reference correlation results in a narrower distribution of E_{EC}, and hence a better detectability of the signal.

If a frozen-noise sample is used, E_π has a fixed value. Hence no uncertainty in terms of the EC-theory is present in the stimulus (the power of the difference signal is frozen) and the only limitation for detection is internal noise.

3.2.3 Interaural differences in time and intensity

The interaural differences (IIDs and ITDs) present in an interaurally partially correlated noise fluctuate as a function of time. In a running-noise condition, the random fluctuations can be described in terms of a probability distribution. We determined these probability distributions by computing a partially correlated noise in the digital domain of sufficient duration (3 seconds at a sample rate of 32 kHz). After a Hilbert transform of the left and right signals, the interaural intensity differences and interaural time differences were obtained. From these differences, histograms were computed which are

---

[2]We assume that the correlation between the signal S and the noise N is zero. Although this is mathematically not correct for a finite-length interval, a computational analysis revealed that the effect of these correlations is neglegible in our analysis.
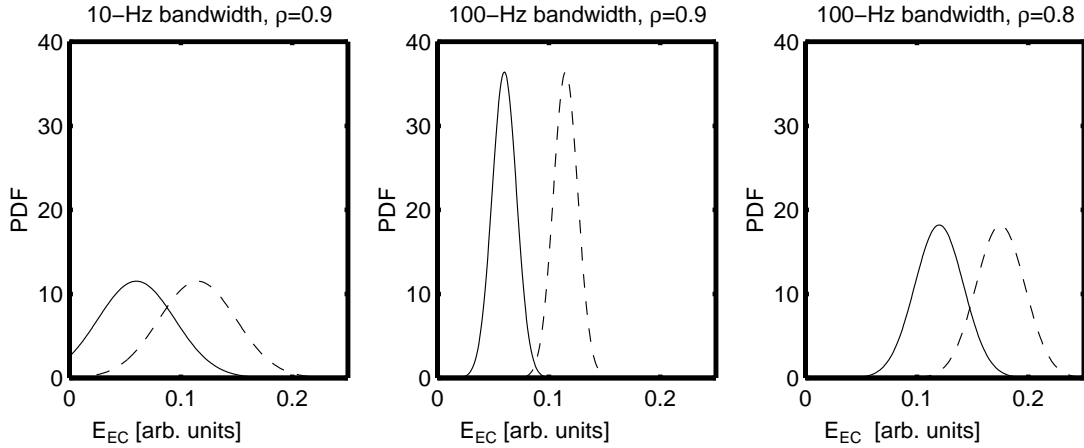
Figure 3.2: *Probability density functions for E_{EC} (see text) for different values of the reference correlation and the masker bandwidth. The format is the same as in Fig. 3.1.*

(close) approximations of the PDFs of the IIDs and ITDs. This procedure was repeated for masker plus signal for a signal-to-masker ratio of -10 dB. The results are shown in Fig. 3.3. The format is the same as Figs. 1 and 2. The left panels correspond to a masker bandwidth of 10 Hz and a reference correlation of 0.9; the middle panels to a bandwidth of 100 Hz and a correlation of 0.9 and the right panels to a bandwidth of 100 Hz and a reference correlation of 0.8. The solid lines represent distributions for the maskers alone, the dashed lines for masker plus signal. The upper panels represent the interaural phase differences (IPD); the lower panels represent the IIDs.

The following facts can be observed in Fig. 3.3. First, if we compare the middle panels to the left panels (i.e., the effect of bandwidth), no difference is observed. Thus, the width of the PDF for the interaural differences does not depend on the bandwidth and the range of variation of the IIDs and ITDs does not change systematically with bandwidth. The *rate* of variation does, however, increase if the bandwidth is increased. This property is important for our hypothesis about stimulus uncertainty. It is often assumed that the binaural auditory system is sluggish in processing binaural cues (cf. Grantham and Wightman, 1978, 1979; Grantham, 1984a; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998). Thus, because the amount of uncertainty is *not* changed by the masker bandwidth, it is expected that thresholds will *increase* with increasing bandwidth as a result of the increase in the rate of fluctuation of the IIDs and ITDs. This is in contrast to the expectations based on the EC-theory or models based on interaural correlation; these models predict a *decrease* with increasing bandwidth.

The solid line in the right panel of Fig. 3.3 demonstrates that a decreasing interaural correlation results in an increase in the width of the PDF, a similar effect as observed in the curves for the correlation and the EC-theory. The addition of the signal has a different effect on the PDFs compared to the two
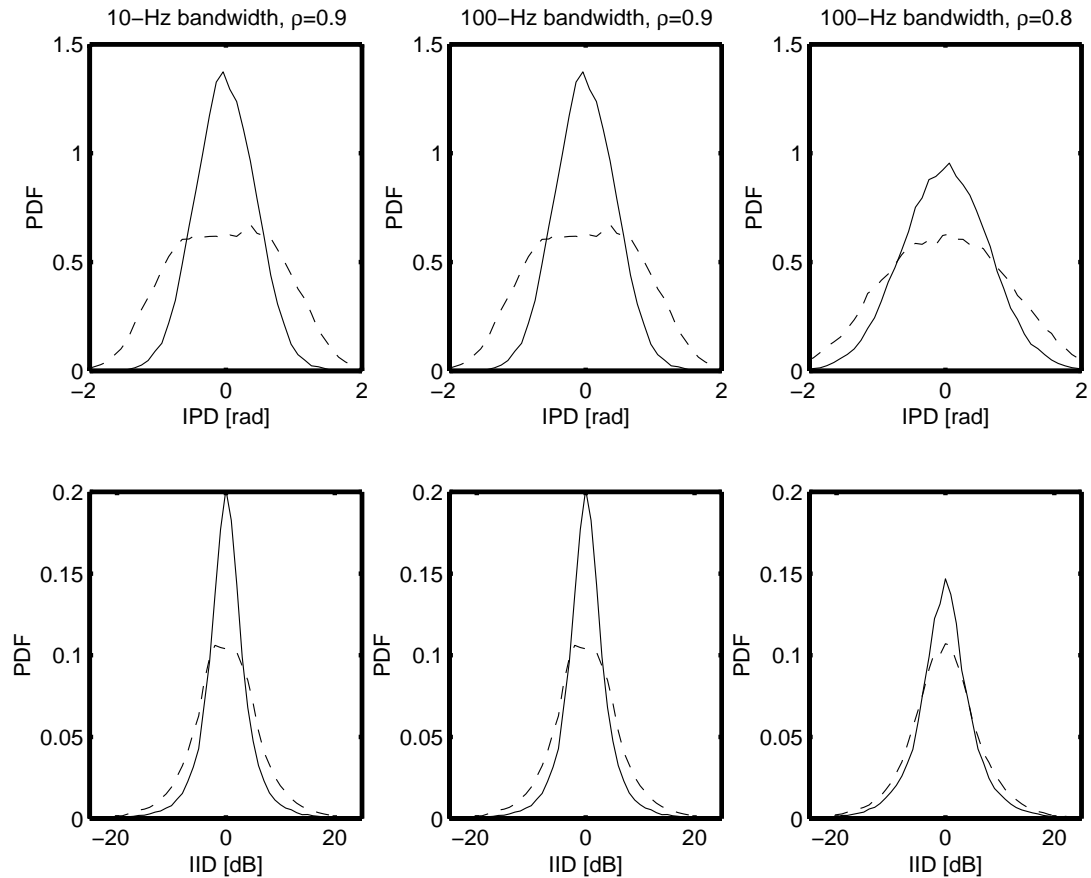
Figure 3.3: *Probability density functions for the IPD (upper panels) and IID (lower panels) in the same format as Figs. 1 and 2.*

other models discussed in this chapter. Instead of a shift of the mean, an increase in the width of the distribution is observed. This property makes it more difficult to analyse these PDFs in terms of detectability. Nevertheless, the observation that a change in the bandwidth results in different expectations for the three theories makes it valuable to discuss our data also in terms of IIDs and ITDs.

3.3 Experiment I

3.3.1 Procedure and stimuli

A 3-interval forced-choice procedure with adaptive signal-level adjustment was used to determine masked thresholds. Three masker intervals of 300-ms duration were separated by pauses of 300 ms. A signal of 200-ms duration was added to the temporal center of one of the masker intervals. Feedback was provided after each response of the subject.

The signal level was adjusted according to a two-down one-up rule (Levitt, 1971), tracking the 70.7% correct score within a 3IFC paradigm. This corresponds to d'=1.26. The initial step size for adjusting the level was 8 dB. The

stepsize was halved after every second track reversal until it reached 1 dB. The run was then continued for another 8 reversals. The median level at these last 8 reversals was used as the threshold value. At least three threshold values were obtained for each parameter value and subject. All stimuli were generated digitally and converted to analog signals with a two-channel, 16-bit D/A converter at a sampling rate of 32 kHz. The stimuli were presented over Beyer Dynamic DT990 headphones.

The 300-ms masker samples were obtained by adding interaurally in-phase noise and interaurally out-of-phase noise with the appropriate weighting factors (Eq. 3.1). For running-noise conditions, the noise samples for each interval were obtained by randomly selecting 300-ms segments from a two-channel 2000-ms bandpass-noise buffer. The 2000-ms noise buffer was created in the frequency domain by selecting the desired frequency range from the Fourier transforms of two independent 2000-ms broadband Gaussian noises. After an inverse Fourier transform, and combination of the two noise signals according to Eq. 3.1, the two-channel (for the left and right ears) band-limited noise buffer with the specified reference correlation was obtained. It is important to note that the specified reference correlation is the correlation of the 2000-ms noise buffer. The correlations of shorter segments (like the 300-ms noise segments used in the running-noise experiments) will in general deviate from this exact value (see Fig. 3.1).

For frozen-noise conditions, only one fixed 300-ms noise sample was used for which the interaural correlation was equal to the reference correlation [3]. This noise sample was generated by adding two independent band-limited noise samples of 300 ms with a *fixed* RMS value. These band-limited noise samples were generated in the same way as the noise buffers for random-noise conditions followed by a normalization of their RMS values. The partially correlated noise was then generated by combining the noises according to

---

[3]The correlation between two finite-length samples from independent noise sources is almost never *exactly* zero. In our analysis and generation of stimuli, however, we assume that this correlation is zero. To justify this assumption, we generated 1000 intervals of interaurally partially-correlated noise and determined the width of the probability density functions for the interaural correlation after (1) combination of the signals according to Eq. 3.1, and (2) combination of the signals according to Eq. 3.1 *after normalizing* the energies of the noise samples to a fixed value (i.e., there was no energy fluctuation for this case). The width of the probability density function for a reference correlation of 0 after normalization of the noise intervals was approximately 10^5 times narrower than without normalization. This indicates that for the way we generated the $N\rho$ stimuli (i.e., with *two* independent noise sources), energy fluctuation is the main cause for correlation fluctuations. Another reason why this assumption is reasonable is the fact that the processing of the cochlea results in phase shifts in the presented stimuli. It is possible to generate a waveform with an exact interaural correlation. But this is only possible for the waveforms arriving at the eardrums. After the processing in the peripheral hearing system, phase shifts result in changes in the correlation. It is therefore not so valuable to take the correlation between waveforms into account, because this property changes by the processing of the cochlea.

Eq. 3.1. The same noise sample was used during one run. To exclude the possibility that the frozen-noise thresholds would depend on the specific waveform of the token, a different frozen-noise sample was used for each run, and the mean threshold from these runs was used as threshold value. All noise maskers were presented at an overall level of 65 dB SPL.

The 200-ms signals were interaurally out-of-phase sinusoids with a frequency of 500 Hz. In order to avoid spectral splatter, the signals and the maskers were gated with 50-ms raised-cosine ramps. Thresholds are expressed as the means of at least three repetitions per condition and subject. Binaural masked thresholds were measured for $N\rho S\pi$ conditions, where the bandwidth of the noise was either 10, 100 or 1000 Hz. The center frequency of the noise masker was always 500 Hz. Reference correlations of ρ=1, 0.98, 0.96, 0.93, 0.87, 0.81 and 0.64 were used. In addition, NoSo thresholds were also obtained. Three well-trained subjects with normal hearing participated in the experiments.

3.3.2 Results

The experimental data are shown in Fig. 3.4 as a function of the reference correlation. The bottom-right panel shows the mean thresholds, while the other panels show individual thresholds for the three subjects. The squares correspond to a masker bandwidth of 10 Hz, the upward triangles correspond to 100 Hz and the downward triangles to 1000 Hz. The open symbols represent running-noise conditions, the filled symbols represent frozen-noise conditions. The NoSo thresholds are plotted in the upper-right corners of each panel. The errorbars denote the standard error of the mean.

For both running- and frozen-noise conditions, the $N\rho S\pi$ thresholds increase with decreasing reference correlation. This increase is strongest for the 10-Hz running-noise masker, which increases by 18.8 dB if the correlation is decreased from +1 to 0.64. For the 100-Hz-wide and 1000-Hz-wide running-noise conditions, the increase amounts to 15.5 dB and 10 dB, respectively. These values are in good agreement with data from van der Heijden and Trahiotis (1998). For frozen-noise maskers, the increase amounts to 12 dB, 13.3 dB and 9.2 dB, for the 10, 100 and 1000-Hz-wide conditions, respectively.

The thresholds for frozen and running-noise maskers are approximately equal for a reference correlation of +1, while for decreasing reference correlations, the difference between frozen and running-noise maskers increases, especially for the narrowband conditions. The reference correlations at which frozen and running-noise thresholds become different are 0.98 for a bandwidth of 10 Hz and 0.93 for a bandwidth of 100 Hz. As interaural correlation decreases the differences between running- and frozen-noise conditions reach 7 dB for the 10-Hz-wide masker, and 4 dB for the 100-Hz-wide condition. For the
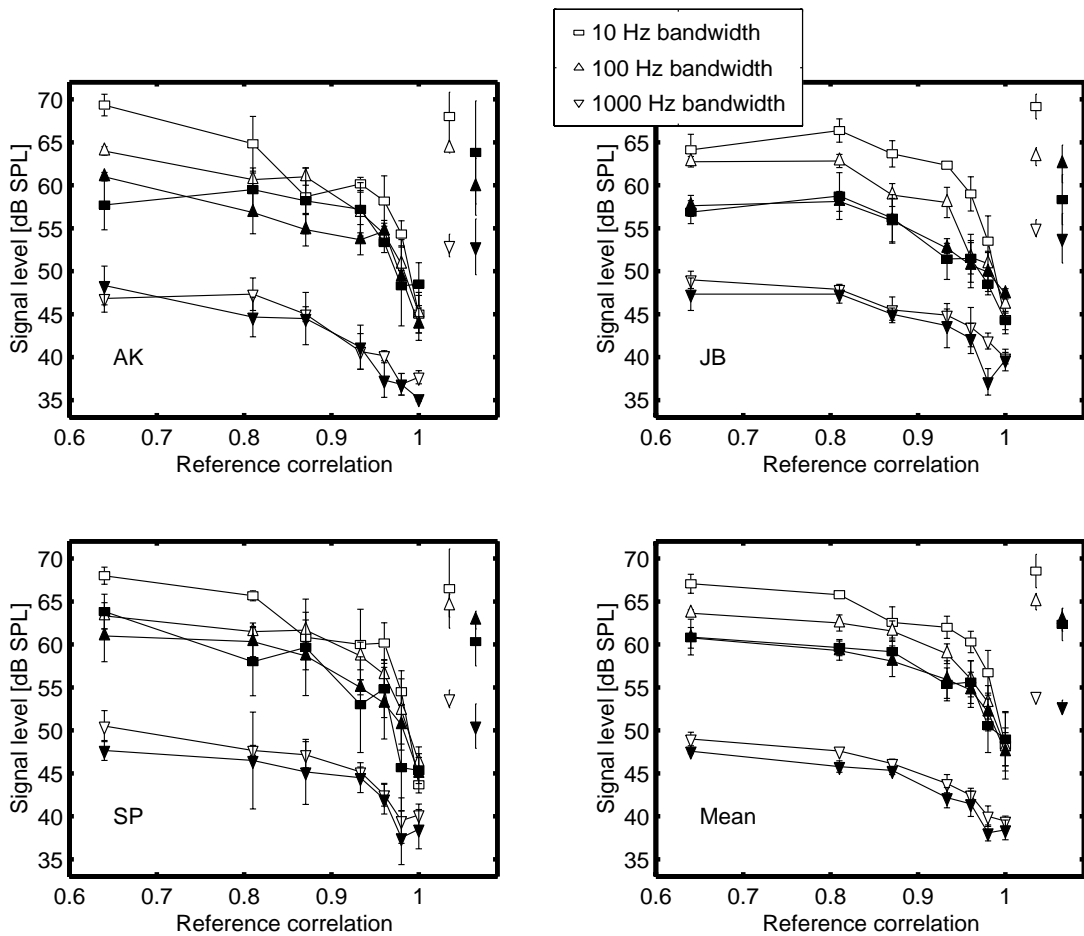
Figure 3.4: *Binaural masked thresholds as a function of the reference correlation. The bottom-right panel shows the mean thresholds, the other panels show individual thresholds for three subjects. The squares correspond to a masker bandwidth of 10 Hz, the upward triangles to 100 Hz and the downward triangels to 1000 Hz. The open symbols represent running-noise conditions, the filled symbols represent frozen-noise conditions. The isolated symbols represent NoSo reference data. Errorbars denote the standard error of the mean. The masker level for all three bandwidths was 65 dB.*

1000-Hz-wide maskers, the thresholds for running and frozen noise are very similar.

The differences between the 100-Hz and 1000-Hz conditions vary considerably across reference correlations, both for running and frozen-noise conditions. For running noise, the difference in thresholds amounts to 9 dB for a reference correlation of +1 and increases up to a value of 14 dB for a reference correlation of 0.64. For frozen noise, these values are 10.6 dB and 13.3 dB, respectively. Because the overall masker level was kept constant, a difference of about 10 dB between 100-Hz and 1000-Hz thresholds would correspond to a constant signal-to-noise ratio at the output of an auditory filter with a bandwidth of 78 Hz (1 ERB at 500 Hz, Glasberg and Moore, 1990).

The NoSo thresholds for *running* noise show a decrease with increasing masker bandwidth. The signal-to-noise ratio decreases from +4 dB at 10 Hz to 0 dB at 100 Hz and finally to -11 dB at 1000 Hz, very similar to experimental data from van de Par and Kohlrausch (1997, 1999). In contrast, for *frozen* noise, the NoSo thresholds are very similar for the 10 and 100-Hz bandwidth (S/N=-2 dB), while for the 1000-Hz bandwidth, the threshold is 10 dB lower. In general, the relation between running- and frozen-noise thresholds in the NoSo condition equals that for $N\rho S\pi$ with $\rho < 0.95$. For the smallest reference correlation (i.e., ρ=0.64), the running-noise BMLD for a bandwidth of 10 and 100 Hz is almost zero (except for subject JB, who has a BMLD of 5 dB for this condition). For the 1000-Hz-wide condition, the BMLD is 6 dB for ρ=0.64, consistent with data from Robinson and Jeffress (1963).

3.3.3 Discussion

Following our hypothesis that stimulus uncertainty influences $N\rho S\pi$ thresholds, the difference between frozen and running-noise conditions should be larger at lower reference correlations, as a result of the fact that stimulus uncertainty increases with decreasing correlation. In addition, frozen- and running-noise thresholds should be equal for a reference correlation of +1, because no uncertainty in terms of binaural cues is present in the masker intervals. These effects are clearly visible in our data (Fig. 3.4) for the bandwidths of 10 and 100 Hz. For a bandwidth of 1000 Hz, there is almost no difference between the running and frozen-noise conditions for all values of the reference correlation. This suggests that for this value of the masker bandwidth, stimulus uncertainty does not influence the detection of the signal and the thresholds are limited by internal noise (similar to the frozen-noise data). Interestingly, the data that are limited by internal errors also show a dependence on the masker correlation. This implies that the net effect of the internal noise must be larger for smaller interaural correlations. One possibility to implement this property in a quantitative binaural model is given in Chapter 4 of this thesis.

For the quantitative analysis of our data in terms of the binaural models described in Section 3.2, we will concentrate on those conditions where, presumably, external variability is dominant over internal noise. In terms of the terminology used by Lutfi (1990), we are interested in conditions with a substantial amount of informational masking. As a measure for this we take the difference between running- and frozen-noise threshold. Substantial differences are observed for a 10-Hz-wide masker and reference correlations at or below 0.98, and for a 100-Hz-wide masker at or below 0.93. For the 1000-Hz data, the difference is small at all correlation values and those data are therefore not included in the analysis.

The influence of external variability in the binaural data is stronger at 10-Hz than at 100-Hz bandwidth. This supports the expectations based on the EC-theory and on correlation uncertainty, because both models predict a stronger difference between frozen and running noise at narrower bandwidths. The data are however not in line with the expectations based on the evaluation of IID and ITD cues. The distributions of these cues do not depend on the bandwidth and hence no effect is expected if uncertainty is considered[4]. Including the effect of binaural sluggishness, this should lead to an increase in threshold with an increase in masker bandwidth. This, however, is not found in the data, which show a decrease of the running-noise thresholds with increasing bandwidth.

In order to verify the hypotheses based on the EC-theory and the interaural correlation quantitatively, we computed the detectability of the running-noise thresholds shown in Fig. 3.4 based on the two models for those conditions in which detection is apparently limited by stimulus uncertainty (i.e., $\rho \leq 0.98$ at 10-Hz bandwidth and $\rho \leq 0.93$ at 100-Hz bandwidth). For the interaural correlation, the detectability was calculated using the distribution of the interaural correlation expressed in terms of Fisher's Z (see Appendix 3.A). The rationale for the transformation from correlation to Z lies in the fact that the correlation probability does not have a Gaussian distribution, while Fisher's Z does have an approximately Gaussian distribution. For both the masker alone and the masker plus test signal at threshold, the probability distributions for Z were computed and from these distributions, the sensitivity index d' was calculated. For the sinusoidal signals, a length of 200 ms including 50-ms ramps was used to calculate the change in correlation. For the masker-alone correlation interval, a duration of 200 ms was assumed, because this corresponds to the signal length and hence the duration from which the binaural system can extract useful information concerning interaural correlation changes. Peripheral preprocessing was simulated by first filtering the signals with a 4th-order gammatone filter with a center frequency of 500 Hz and a bandwidth of 78 Hz (cf. Glasberg and Moore, 1990). The values for d' are shown in Fig. 3.5.

The squares denote the 10-Hz masker condition and the upward triangles the 100-Hz condition. Clearly, most values of d' are higher than the theoretical value of 1.26 that results from the applied procedure. The values of d' across reference correlations are relatively constant for 10-Hz bandwidth and increase systematically towards high reference correlations for 100-Hz bandwidth. Only the 10-Hz-wide condition shows a fair agreement with d' values in terms of the correlation uncertainty (i.e., d' values close to +1). From

---

[4]It should be noted that the uncertainty analysis for the EC theory and the cross-correlation is based on the complete intervals, while the analysis for the ITDs and IIDs relies on changes *within* one interval.
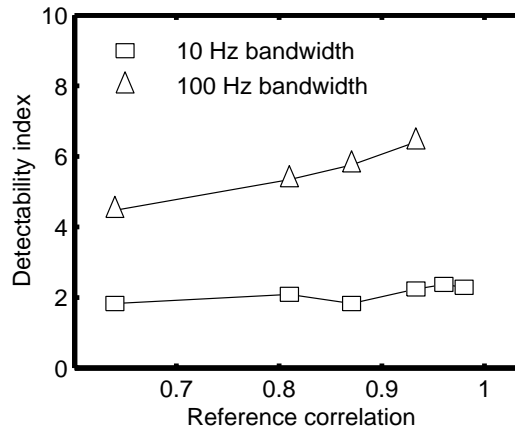
Figure 3.5: *Detectability index d' in terms of interaural correlation for the running-noise $N\rho S\pi$ conditions as a function of the reference correlation. The squares denote the 10-Hz-wide condition and the upward triangles the 100-Hz-wide condition.*

this simulation it appears that the correlation uncertainty is a valid statistic only for the 10-Hz-wide condition.

The large values for d' for the 100-Hz-wide conditions may indicate that in the processing of these stimuli in the auditory system, information is lost. An optimal detector, basing its decision on the correlation change within the 200 ms of signal duration would perform much better than the subjects, given the high values of d' for correlation discrimination. Such a loss of information might be caused by the fact that the subjects are not able to process the whole stimulus but extract a decision variable based on a shorter part of the sample.

Another possiblity is that the correlation hypothesis is not correct and that detection behavior can be better described by another statistic, for example based on the EC-theory. Equation 3.2 gives the relation between the decision variable E_{EC} and the source of stimulus uncertainty, E_π. With the help of equations 3.A.4 to 3.A.6 and 3.A.18 in Appendix 3.A, the detectability index in terms of E_{EC} can be determined for the conditions limited by stimulus uncertainty. These indices are shown in Fig. 3.6, in the same format as in Fig. 3.5.

All values for d' in terms of E_{EC} are much higher than the theoretical value of 1.26. This indicates that E_{EC} is not a valid descriptor for the influence of stimulus variability.

An important remark can be made if the values for d' are compared for the correlation (Fig. 3.5) and the EC-theory (Fig. 3.6). The values are completely different for these theories, the latter being much higher. This observation is particularly of interest given the analysis of Green (1992). He stated that a correlation model leads to identical predictions as an EC-model in an
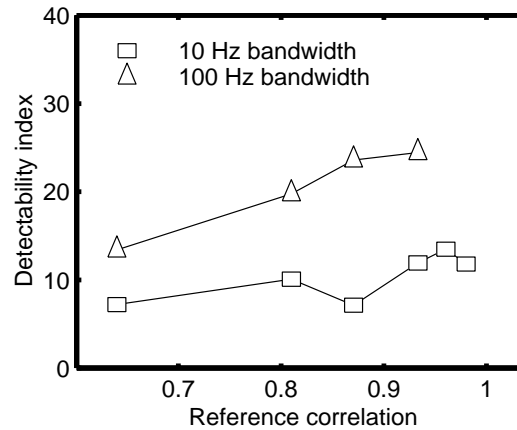
Figure 3.6: *Detectability index d' in terms of E_{EC} for the running-noise $N\rho S\pi$ conditions as a function of the reference correlation. The squares denote the 10-Hz-wide condition and the upward triangles the 100-Hz-wide condition.*

NoSo vs. NoSπ discrimination paradigm. Figures 3.5 and 3.6 show, that this conclusion is not valid for conditions that are dominated by stimulus uncertainty.

In summary, both the 10-Hz-wide and the 100-Hz-wide conditions show large differences between the running and frozen-noise conditions, indicating that stimulus uncertainty changes the detection process. This effect is smaller at 100 Hz than at 10-Hz bandwidth. The thresholds decrease with an increase in masker bandwidth in the running-noise conditions, which is not in line with expectations based on the processing of IID and ITD cues. An uncertainty analysis in terms of the EC-theory revealed that an EC-process fails to account for the thresholds found in the running-noise conditions.

The only close match between experimental data and predictions was found for the 10-Hz-wide conditions based on the interaural cross-correlation. If one assumes that stimulus uncertainty limits the detection and the correlation is a valid statistic for describing thresholds, the psychometric function for an NρSπ condition as a function of the signal level can be predicted. To study to what extent this is true, a second experiment was performed, where predicted and measured psychometric functions were compared.

3.4 Experiment II

3.4.1 Procedure and stimuli

In order to further examine the role of stimulus uncertainty in an NρSπ condition, we determined the psychometric functions for running-noise conditions at bandwidths of 10 and 100 Hz for reference correlations of 0.98, 0.96 and 0.87. Proportions correct were determined in a 2IFC procedure with 50 trials per condition by 3 subjects. The generation of the stimuli and the

method of presentation to the subjects were similar to the method described in experiment I. The signal levels used to determine the subjects' performance were 46 to 66 dB SPL at 10 Hz bandwidth and 42 to 62 dB SPL at 100 Hz bandwidth with a stepsize of 2 dB.

3.4.2 Results

The proportions correct for the $N\rho S\pi$ condition as a function of the signal level are shown in Fig. 3.7 for a masker bandwidth of 10 Hz and in Fig. 3.8 for a bandwidth of 100 Hz. The different symbols denote different subjects. The upper-left panel represents data for $\rho=0.98$, the upper-right panel for $\rho=0.96$ and the lower panel for $\rho=0.87$. The data show an increase in the proportion of correct responses from 0.5 to 1 if the signal level is increased from 45 dB SPL to 65 dB SPL at 10-Hz bandwidth and from 40 dB SPL to 60 dB SPL at 100-Hz bandwidth. The solid lines represent the proportions correct based on the correlation probability density functions.
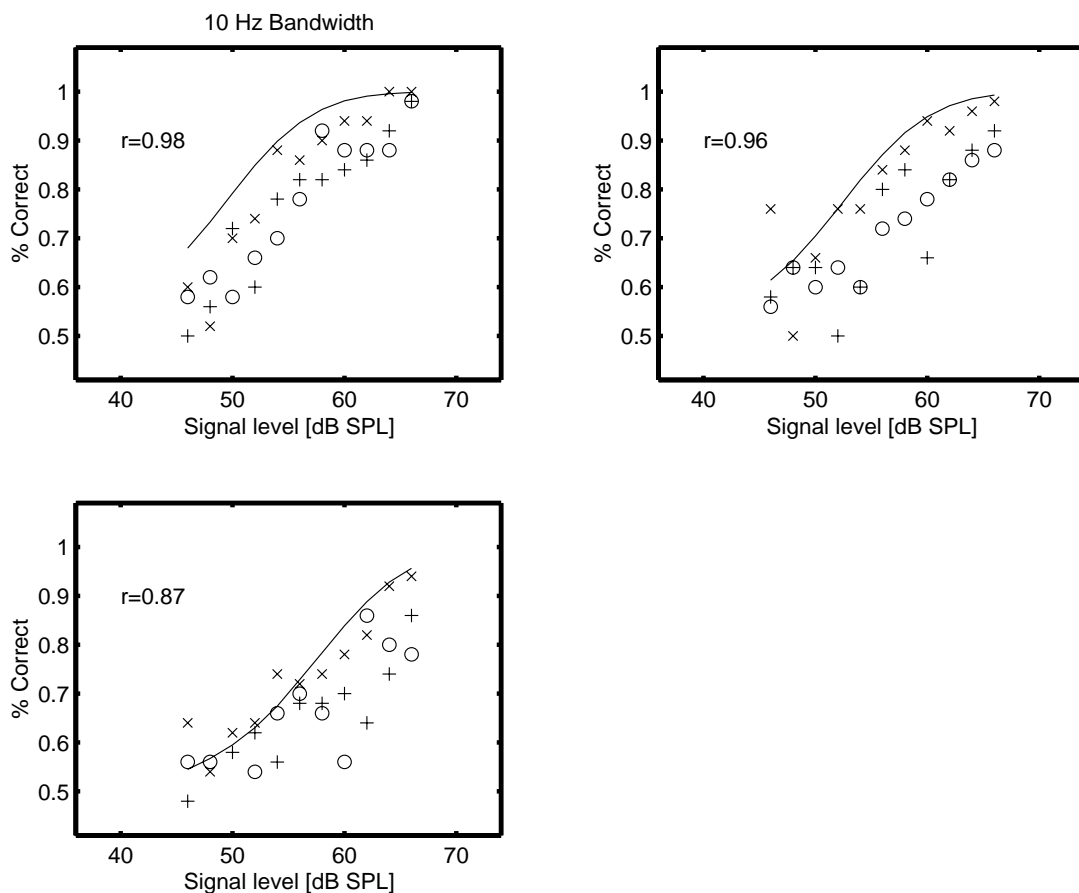
Figure 3.7: *Proportions correct as a function of the signal level for a running-noise $N\rho S\pi$ condition for $\rho=0.98$ (upper left panel), $\rho=0.96$ (upper right panel) and $\rho=0.87$ (lower left panel). The different symbols represent different subjects. The bandwidth of the masker was 10 Hz. The solid line represents the predictions according to a correlation-uncertainty model (see text).*
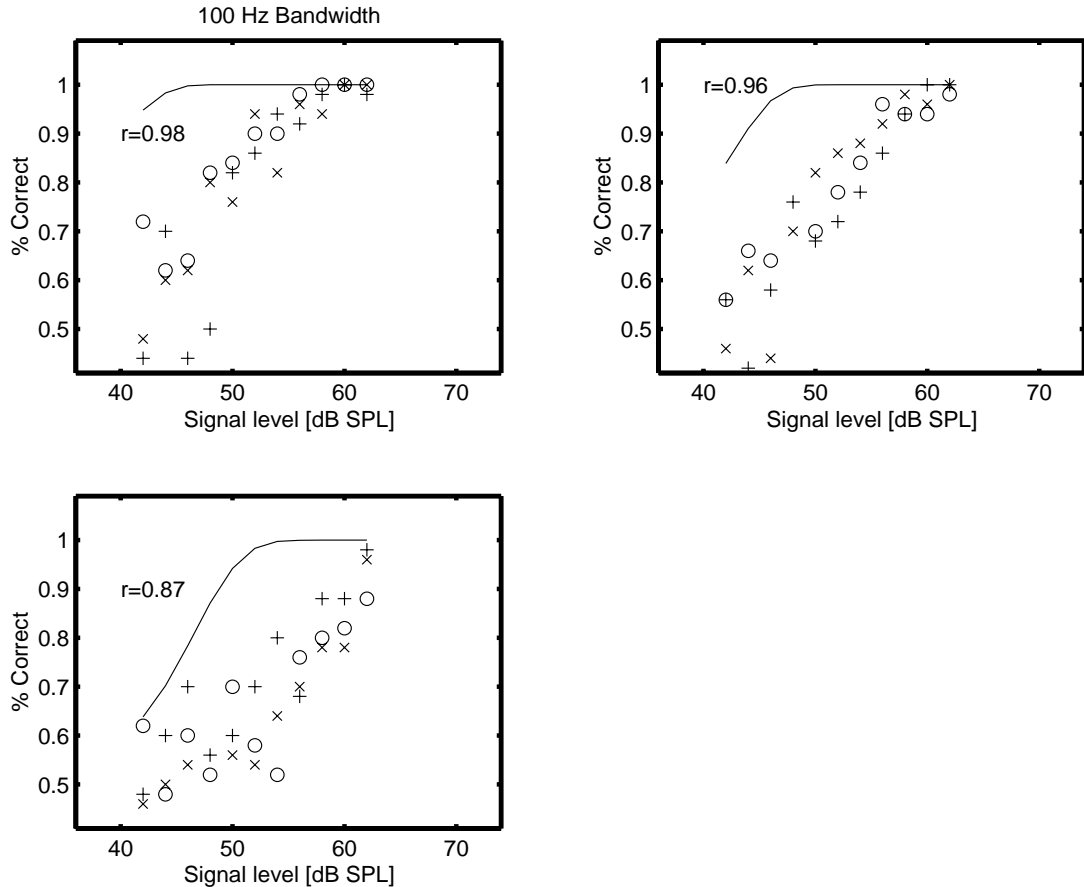
Figure 3.8: *Same as Fig. 3.7 for a masker bandwidth of 100 Hz.*

We calculated the predicted proportions correct as a function of the signal level based on the sensitivity index (d') determined from the correlation uncertainty. The values for d' were converted to proportions correct (p) by computing the area under the normal curve up to $d'/\sqrt{2}$ (see Green and Swets, 1966):

$$p = \int_{-\infty}^{d'/\sqrt{2}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \tag{3.3}$$

The predicted proportions correct are shown by the solid lines in Figs. 3.7 and 3.8. For a bandwidth of 10 Hz, the curves lie on top, close to the subjects' responses, indicating that the data can quite accurately be described (especially for the subject denoted by 'x') by the stimulus uncertainty in the interaural correlation. However, at 100-Hz bandwidth, the subjects perform worse than the predictions based on the correlation uncertainty. This indicates that correlation uncertainty is not a valid descriptor for the 100-Hz data.

3.4.3 Discussion

Because of the close correspondence between the predicted and observed psychometric functions for the 10-Hz-wide maskers, it is likely that stimulus

uncertainty limits the detection and that this uncertainty expressed in terms of the interaural correlation is a valid way to predict thresholds. For the 100-Hz condition, however, such an analysis overestimates the performance of the subjects.

We want to emphasize that although the interaural correlation is a valid detection statistic in *describing* the 10-Hz-wide running-noise conditions, this does not prove that observers indeed use this particular measure. We have shown, however, that stimulus uncertainty can play an important role in binaural detection paradigms. To further investigate the role of the interaural correlation as a detection statistic and the role of stimulus uncertainty, we performed a third experiment. This experiment is a compromise between the running-noise condition (i.e., with stimulus uncertainty) and the frozen-noise condition (i.e., absolutely no stimulus uncertainty). Two conditions were tested, which are referred to as 'fixed-ρ' and 'interleaved'.

3.5 Experiment III

3.5.1 Procedure and stimuli
The following experimental paradigms were used:

1. fixed-ρ. In this condition, each trial consisted of three intervals which contained exactly the same noise sample. To one of these noise samples, the signal was added. For each trial, a different noise sample was calculated according to the frozen-noise algorithm described in Section 3.3.1. This implies that both the interaural correlation and the power of the interaural difference signal was fixed across all intervals of a run, but each noise sample was a different realization under the above constraints. Thus, across trials, the waveforms arriving at both ears were totally different, but the interaural correlation and the power of the interaural difference signal of the masker was fixed.

2. Interleaved. Similar to the fixed-ρ condition, each trial consisted of three identical masker intervals, and again one interval contained the signal. However, the number of masker realizations was reduced to two. Thus, two frozen-noise samples were calculated as described in Section 3.3.1. For each trial, one of these realizations was chosen at random and used as the masker in all three intervals of this trial.

The measurement procedure, the signal durations and levels and the way of presentation to the subjects were the same as described in Section 3.3.1. We measured thresholds for two masker bandwidths (10 and 100 Hz) and two masker correlations (0.93 and 0.87). Three subjects participated in this experiment.

3.5.2 Results

The mean thresholds across subjects are shown in Fig. 3.9. The upper panels show thresholds for a masker bandwidth of 10 Hz, the lower panels correspond to 100-Hz bandwidth. The left and right panels correspond to an interaural correlation of 0.93 and 0.87, respectively. In each panel, 4 threshold values are shown. From left to right, these are thresholds for the running-noise condition of experiment I (labeled 'run'), the thresholds for the fixed-ρ condition ('fix'), the interleaved condition ('int') and the frozen-noise condition of experiment I ('fro').

As described above, these conditions reflect different levels of stimulus uncertainty. The first level corresponds to absolutely no stimulus uncertainty (frozen-noise conditions) and this condition results in uniformly lower thresholds than all other conditions. If the level of stimulus uncertainty is increased by a small step (the interleaved condition), thresholds increase by 1 to 3 dB for all tested conditions. A third level of stimulus uncertainty was to apply only one restriction to the masker stimuli: the overall interaural correlation and hence the power of the difference signal had to be constant. For three out of four conditions, this also resulted in an increase in thresholds. The differences between frozen-noise and fixed-ρ are about 6 dB at 10-Hz bandwidth and 4 dB at 100-Hz bandwidth. Finally, the highest level of stimulus uncertainty in the present experiments (i.e., running-noise) resulted on average in very similar thresholds to those in the fixed-ρ conditions[5].

3.5.3 Discussion

Some striking remarks can be made with respect to the thresholds for stimuli with a fixed interaural correlation (i.e., the fixed-ρ, the interleaved and the frozen-noise conditions). If the binaural auditory system uses the interaural correlation of each token as a decision variable, the processing of the masker alone would result in an internal estimate of the externally presented interaural correlation. This internal value is fixed and only limited by internal noise. The addition of the signal results in a decrease of the interaural correlation and can thus be detected. Based on such an interaural-correlation processing, all the thresholds for the conditions with a fixed interaural correlation should give the same thresholds. This was not found in our data.

---

[5]A MANOVA analysis of the data shown in Fig. 3.9 was performed with the following independent parameters: amount of stimulus uncertainty, stimulus bandwidth, interaural correlation and subject. The analysis revealed that bandwidth, correlation and amount of stimulus uncertainty were statistically significant effects at a 95% confidence interval. A multiple comparison procedure (Fisher's Least Significant Difference method) on the means for the different values of stimulus uncertainty revealed that the contrast for running noise versus fixed-ρ was not statistically significant. On the other hand, the contrasts between fixed-ρ, interleaved and frozen conditions were all statistically significant at a 95% confidence interval.
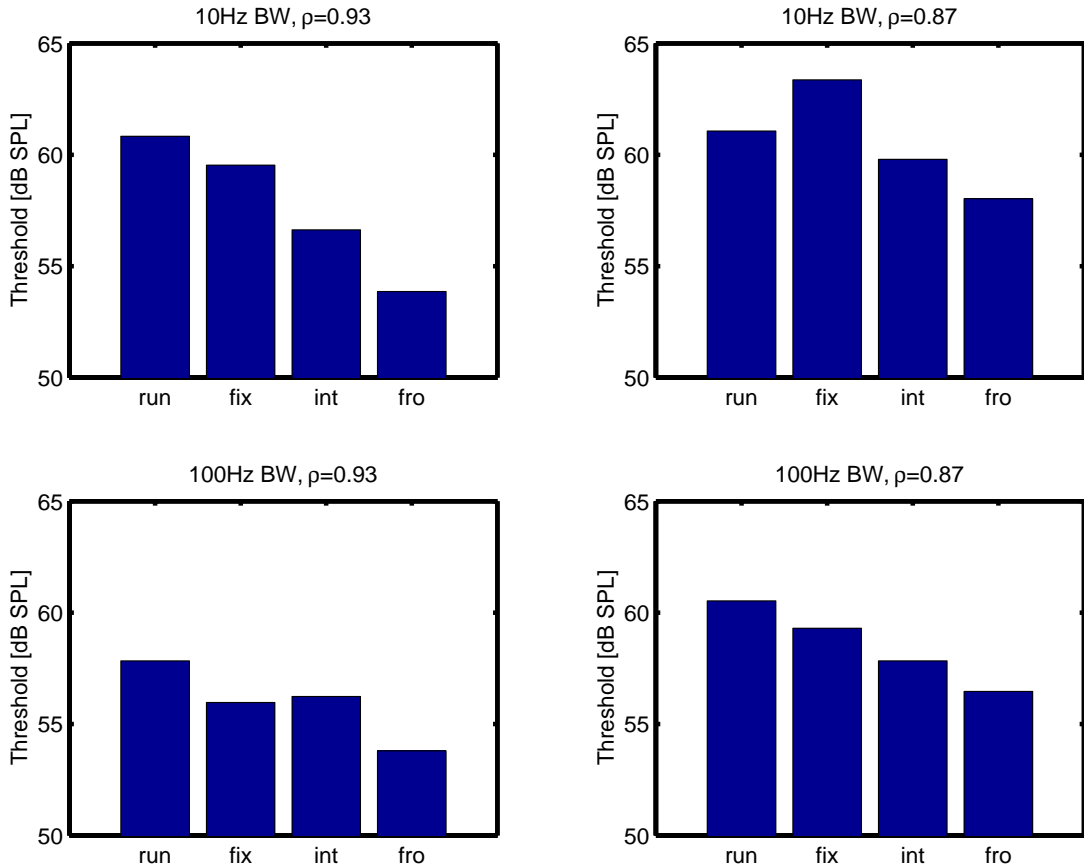
Figure 3.9: *Mean thresholds across 3 subjects for the running-noise (run), fixed-ρ (fix), interleaved (int) and frozen-noise (fro) conditions. The upper panels show thresholds for a bandwidth of 10 Hz, the lower panels for 100-Hz. The left and right panels correspond to an interaural correlation of 0.93 and 0.87, respectively.*

One reason for the differences across these conditions may be peripheral filtering. The externally presented stimulus has a fixed interaural correlation. Peripheral filtering in the cochlea results in frequency dependent phase shifts in the presented waveforms at both ears. These phase shifts result in a change in the interaural correlation. Therefore, for the fixed-ρ condition, the interaural correlation of different tokens after peripheral filtering follows a distribution rather than having a fixed value. To evaluate this hypothesis quantitatively, we computed 1000 partially correlated noises following three different procedures:

1. Running-noise samples which are generated in the same way as described in Section 3.3.1,

2. The same running-noise samples after filtering by a 4th-order gammatone filter with a center frequency of 500 Hz. This filter simulates the effect of peripheral filtering in the inner ear,

3. Fixed-ρ samples generated as described in Section 3.3.1, also after filtering with the gammatone filter.
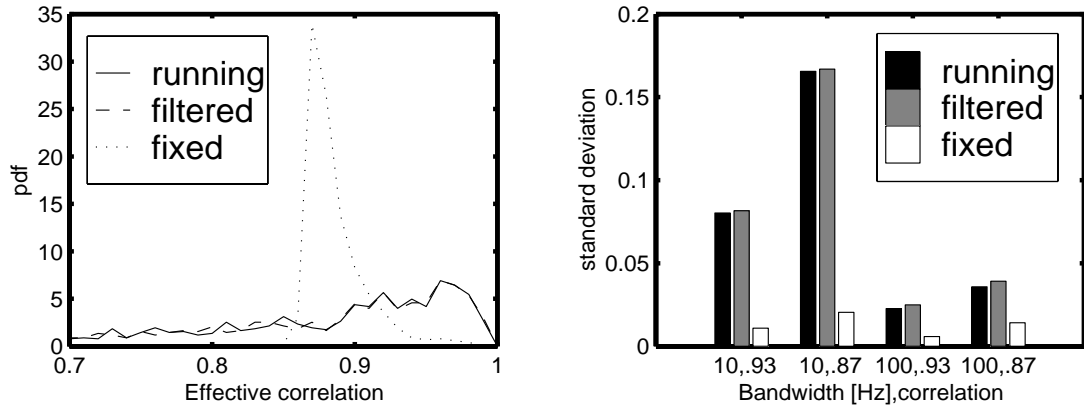
Figure 3.10: *Left panel: simulated correlation distribution for a 10-Hz wide, Nρ stimulus with a reference correlation of 0.87. These distributions were obtained for running noise (running), running noise after peripheral filtering (filtered) and for fixed-ρ conditions after peripheral fitering (fixed). Right panel: standard deviations for the same stimuli as in the left panel for different values for the masker bandwidth and correlation.*

An example of the correlation distributions that were found with this procedure is given in the left panel of Fig. 3.10. In this example, the bandwidth of the noise was 10 Hz and the reference correlation was 0.87. The solid line is the distribution for running noise without filtering (i.e., procedure 1). The dashed line corresponds to procedure 2 (i.e., running noise after peripheral filtering). Clearly, these distributions are very similar, indicating that peripheral filtering does not change the *statistics* of the interaural correlation for running noise. The distribution for fixed-ρ after peripheral filtering is shown by the dotted line. In line with our hypothesis, the distribution has a substantially reduced standard deviation. The values for the standard deviation of the correlation distribution are shown in the right panel of Fig. 3.10. The black bars correspond to the running-noise procedure without filtering (number 1 in the above description); the grey bars denote running noise after peripheral filtering (2) and the white bars denote fixed-ρ samples after peripheral filtering (3). The x-axis indicates the combinations of bandwidth and reference correlation of the noise for each condition. All fixed-ρ conditions have a non-zero amount of correlation uncertainty. This supports the hypothesis that peripheral filtering produces uncertainty in the interaural correlation. The magnitude of this correlation uncertainty is however not sufficient to explain the thresholds of the fixed-ρ condition. We computed the values for the detectability index in terms of interaural correlation for the fixed-ρ condition after peripheral filtering as described in Section 3.3.3. All d' values for the filtered fixed-ρ condition were above 27, indicating that subjects performed worse than expected from correlation uncertainty introduced by peripheral filtering.

An important conclusion that can be drawn from these results is that it is unlikely that the auditory system uses the interaural correlation of the

complete token as a decision variable. These results also show that the overall power of the difference signal of the complete token as a decision variable is not a valid descriptor of how the auditory system processes Nρ stimuli.

A possibility to explain the results qualitatively is based on the idea of internal templates (Dau *et al.*, 1996a,b; Breebaart and Kohlrausch, 1999). Assume that listeners develop an internal representation of the interaural differences that occur as a function of time if a masker alone is presented. Such a template can be obtained from the masker-alone intervals in the 3IFC task. One possible realization would be the running-average of the difference power based on a time constant that is smaller than the duration of the tokens. If such a template exists, then the task of the listener is to match the template to the internal representation of the actually presented stimuli.

For example, in an NoSπ condition, the masker alone contains no interaural differences. Hence the template consists of a sequence of zero interaural differences. The addition of the signal results in changes in the interaural differences which can be detected. In this case, there is no uncertainty in the masker-alone representation. The same holds for the frozen-noise conditions. All masker-alone intervals are identical resulting in the same template. The only task that a listener has to perform is to detect which interval produces an internal representation that differs from the template. This process is in principle limited by internal noise only.

If two different Nρ tokens are used in random order (i.e., the 'interleaved' condition), detection becomes somewhat more complicated. For perfect detection, the listener has to store two templates (one for each token) and must be able to recognize which masker token is used before the templates are compared with the actual stimulus. If the wrong template is matched with the stimulus, all intervals from the trial result in an imperfect match of the template. This increases the probability of choosing the wrong interval and hence detection performance decreases.

For the fixed-ρ condition, finally, it is only possible to derive an averaged template based on many different noise realizations. This explains the increase in thresholds with respect to the interleaved and frozen conditions, in which the template has a close relation to the actual stimulus. In such a view, fixed-ρ and running-noise conditions are equivalent with respect to the detection strategy. The fixed-ρ condition, however, does allow comparison of internal representations across the three intervals within a trial. Given our experimental data, which show no statistically significant difference between fixed-ρ and running-noise conditions, we can conclude that such an across-interval comparison does not give a significant detection advantage.

3.6   General conclusions

The results suggest that for binaural signal detection with partially correlated noises, two factors play an important role:

- The reference correlation. With decreasing masker correlation, the $N\rho S\pi$ thresholds increase.

- Stimulus uncertainty. Our results show that uncertainty in binaural cues reduces detection performance, especially in narrow-band conditions.

An unresolved issue concerning the data presented in this chapter is the method of internal binaural processing. We have shown that the three theories tested (the interaural correlation, the EC-theory or the processing of IIDs and ITDs) cannot account for the results found in this study. The data suggest that the auditory system is able to use internal templates in the process of binaural signal detection. Quantitative tests to support this notion are given in Chapter 6 of this thesis.

3.A Appendix: interaural correlation distribution

For the generation of an $N\rho S\pi$ stimulus, two independent noise sources $N_0(t)$ and $N_\pi(t)$ with the same RMS value are used, which are combined as follows to yield $L(t)$ and $R(t)$ for the left and right ears, respectively:

$$\begin{cases} L(t) &= \frac{1}{2}\sqrt{2}\sqrt{1+\rho}N_o(t) + \frac{1}{2}\sqrt{2}\sqrt{1-\rho}N_\pi(t) + S(t) \\ R(t) &= \frac{1}{2}\sqrt{2}\sqrt{1+\rho}N_o(t) - \frac{1}{2}\sqrt{2}\sqrt{1-\rho}N_\pi(t) - S(t) \end{cases}. \tag{3.A.1}$$

As a result of fluctuations in the energy of a finite-length interval of the Gaussian-noise samples $N_o(t)$ and $N_\pi(t)$, the effective correlation (ρ_{eff}) of the masker sample may deviate from the desired reference correlation (ρ). Because the noise sources $N_o(t)$ and $N_\pi(t)$ are independent, the effective (i.e., physically occurring) interaural correlation (ρ_{eff}) for the $N\rho S\pi$ stimulus can be written as (neglecting the correlation between the two independent noise sources N_o and N_π, see also footnote 3):

$$\rho_{eff} = \frac{(0.5+0.5\rho)E_o - (0.5-0.5\rho)E_\pi - E_s}{(0.5+0.5\rho)E_o + (0.5-0.5\rho)E_\pi + E_s}, \tag{3.A.2}$$

where E_x, with x equals o or π, is defined as the energy of the stimulus of duration T, according to:

$$E_x = \int_{-T/2}^{T/2} N_x^2(t)dt. \tag{3.A.3}$$

From Rice (1959) it is known that for a Gaussian-noise sample, E is distributed normally according to:

$$p(E) = \frac{1}{\sigma_E\sqrt{2\pi}}e^{\frac{-(E-m_E)^2}{2\sigma_E^2}}, \tag{3.A.4}$$

with

$$m_E = T\int_0^\infty \omega(f)df, \tag{3.A.5}$$

and

$$\sigma_E^2 = T\int_0^\infty \omega^2(f)df. \tag{3.A.6}$$

Here, $\omega(f)$ refers to the spectral power density of the noise source. The relation between the energies E_o, E_π and E_s for a certain ρ_{eff} according to Eq. 3.A.2 is given by:

$$E_o = \alpha E_\pi + \beta E_s, \tag{3.A.7}$$

with

$$\alpha = \left(\frac{0.5-0.5\rho}{0.5+0.5\rho}\right)\left(\frac{1+\rho_{eff}}{1-\rho_{eff}}\right), \tag{3.A.8}$$

and

$$\beta = \frac{1 + \rho_{eff}}{(1 - \rho_{eff})(0.5 + 0.5\rho)}.$$ (3.A.9)

One way to realize a correlation of ρ_{eff} is to fix E_π at a certain value and compute the necessary value of E_o according to Eq. 3.A.7. The probability for that realization of ρ_{eff} is then equal to the product of the probabilities $p_E(E_o)$ and $p_E(E_\pi)$. Because there are many possible ways to realize a correlation of ρ_{eff}, we have to sum all possibilities of these realizations:

$$p(\rho_{eff})\Delta\rho_{eff} = \sum_{E_\pi} p_E(E_o)\Delta E_o p_E(E_\pi)\Delta E_\pi,$$ (3.A.10)

which results in:

$$p(\rho_{eff}) = \int_{E_\pi} p_E(\alpha E_\pi + \beta E_s)p_E(E_\pi)\frac{\partial E_o}{\partial \rho_{eff}}dE_\pi,$$ (3.A.11)

and hence

$$p(\rho_{eff}) = \frac{1}{\sigma_E^2 2\pi}\int_{E_\pi} \exp\left(\frac{-(\alpha E_\pi + \beta E_s - m_E)^2 - (E_\pi - m_E)^2}{2\sigma_E^2}\right)\frac{\partial E_o}{\partial \rho_{eff}}dE_\pi,$$ (3.A.12)

with

$$\frac{\partial E_o}{\partial \rho_{eff}} = \frac{0.5 - 0.5\rho}{0.5 + 0.5\rho}\frac{2}{(1 - \rho_{eff})^2}E_\pi + \frac{2}{(1 - \rho_{eff})^2(0.5 + 0.5\rho)}E_s.$$ (3.A.13)

In summary, if the spectral shape of the noise source and the sample duration are known, Eqs. 3.A.5 and 3.A.6 supply values for m_E and σ_E^2. For a given signal energy E_s and a given reference correlation ρ, Eq. 3.A.12 gives the probability density for the occurrence of a certain interaural correlation.

A difficulty arising from the probability density function given by the above equations is that for correlations close to 1, the function becomes skewed. If the distribution for the interaural correlation would be Gaussian, it would be easier to calculate parameters like the detectability index d' for two different distributions. Therefore, the Fisher ρ-to-Z transformation is used. This transformation results in a probability density function that behaves approximately normal, and is given by:

$$Z = 0.5\ln\frac{1 + \rho_{eff}}{1 - \rho_{eff}}.$$ (3.A.14)

Thus, for a given Z, the corresponding interaural correlation becomes

$$\rho_{eff} = \frac{e^{2Z} - 1}{e^{2Z} + 1},$$ (3.A.15)

and hence

$$\frac{d\rho_{eff}}{dZ} = \frac{4e^{2Z}}{(e^{2Z} + 1)^2}.$$ (3.A.16)

The probability density function for Z is then given by:

$$p(Z) = p(\rho_{eff}) \frac{d\rho_{eff}}{dZ}.$$ (3.A.17)

The detectability index for the NρSπ condition is determined by the means and the standard deviations in terms of Z as follows. The mean (μ) and standard deviations (σ) of the distributions of $p(Z)$ are determined for both masker alone and masker plus signal (μ_{N_ρ}, σ_{N_ρ}, $\mu_{N\rho S\pi}$, $\sigma_{N\rho S\pi}$ respectively). The detectability index is then obtained as

$$d' = \frac{\mu_{N\rho S\pi} - \mu_{N\rho}}{\sqrt{(\sigma_{N\rho S\pi}^2 + \sigma_{N_\rho}^2)/2}}.$$ (3.A.18)

'A theory has only the alternative of being right or wrong.
A model has a third possibility: it may be right, but irrelevant.'
Manfred Eigen.

A binaural signal detection model based on contralateral inhibition[1]

*This chapter presents a quantitative binaural signal detection model which extends the monaural model described by Dau et al. [J. Acoust. Soc. Am. **99**, 3615–3622 (1996)]. The model is divided into three stages. The first stage comprises peripheral preprocessing in the right and left monaural channels. The second stage is a binaural processor which produces a time-dependent internal representation of the binaurally-presented stimuli. This stage is based on the Jeffress delay line extended with tapped attenuator lines. Through this extension, the internal representation codes both interaural time and intensity differences. In contrast to most present-day models, which are based on excitatory-excitatory interaction, the binaural interaction in the present model is based on contralateral inhibition of ipsilateral signals. The last stage, a central processor, extracts a decision variable that can be used to detect the presence of a signal in a detection task, but could also derive information about the position and the compactness of a sound source. In three subsequent chapters, the model predictions are compared with data obtained with human observers in a great variety of experimental conditions.*

4.1 Introduction

Over the past decades many models of binaural processing have emerged that address various aspects of binaural hearing. Among other things, these models are able to predict the intracranial locus of a binaural sound (Lindemann, 1985; Raatgever and Bilsen, 1986; Stern *et al.*, 1988; Shackleton *et al.*, 1992; Gaik, 1993) or account for binaural masking level differences (Durlach, 1963; Green, 1966; Colburn, 1977; Culling and Summerfield, 1995; Stern and Shear, 1996; Bernstein and Trahiotis, 1996; Zerbs, 2000), as well as for binaural pitch phenomena (Bilsen and Goldstein, 1974; Bilsen, 1977; Raatgever and Bilsen, 1986; Raatgever and van Keulen, 1992; Culling *et al.*, 1996). The majority of these models rely on the coincidence counter hypothesis following an internal delay line as suggested by Jeffress (1948). The physiological basis for such coincidence counters are the so-called Excitation-Excitation (EE)-type cells (Rose *et al.*, 1966; Goldberg and Brown, 1969; Yin and Chan, 1990; Joris and Yin, 1995; Joris, 1996; Batra *et al.*, 1997a,b). These cells are found

---

[1]This chapter is based on Breebaart, van de Par, and Kohlrausch (2001a).

in the medial superior olive. Their discharge rate in response to binaural stimulation depends on the interaural time difference (ITD) and, at favorable ITDs, i.e., when exhibiting maximum response, typically exceeds the sum of the responses for either ear alone (Goldberg and Brown, 1969). This favorable ITD is referred to as the cell's *best delay*. If a given neuron is activated by different frequencies, the different periodic discharge curves appear to reach a maximum amplitude for the same interaural delay of the stimulus. This delay is referred to as the cell's *characteristic delay* and provides an estimate of the difference in travel time from each ear to the coincidence detector.

In models based on an array of EE-type cells with a range of characteristic delays, the neural discharge rate resulting from the EE-interaction is usually modeled as an interaural cross-correlation function. The intracranial locus of a sound presented with a certain interaural time difference is usually assumed to be based on the locus of the largest neural activity or on the centroid computed along the internal delay line. For a signal without any interaural time disparity, the interaural cross-correlation function is maximum at an internal delay of zero. An interaural time difference results in a shift of the cross-correlation function along the delay axis and hence leads to a predicted lateralization.

Some of these models also allow for the prediction of binaural masking level differences (BMLD). When a broadband noise is presented in phase to both ears, and pure tones are presented out of phase to each ear simultaneously (NoSπ condition), the masked threshold is generally lower than when both the noise and the tone are presented in phase (NoSo condition) (Hirsh, 1948a; Hafter and Carrier, 1970; Zurek and Durlach, 1987). Within the framework of these models, the detection of the Sπ signal is based on the reduction of the cross-correlation value for NoSπ due to the addition of the test signal (Colburn, 1973, 1977; Colburn and Latimer, 1978; Durlach *et al.*, 1986; van de Par and Kohlrausch, 1995; Bernstein and Trahiotis, 1996; Stern and Shear, 1996).

Another important theory of binaural hearing is the Equalization-Cancellation (EC) theory (Durlach, 1963, 1972). The basic idea of the EC theory is that the auditory system attempts to eliminate masking components by first transforming the stimuli presented to the two ears in order to equalize the two masking components (E-process). Possible equalization transformations are interaural level adjustments and internal time delays, but also internal phase shifts have been suggested as part of the transformation repertoire. It is assumed that this E-process is performed imperfectly due to internal errors. Consequently, if the stimulus in one ear is subtracted from the stimulus in the other ear (C-process), part of the energy of the masker cannot be canceled. For many binaural masking conditions, this operation leads to

an improvement in the signal-to-masker ratio and hence to the prediction of a BMLD. The EC theory proposed by Durlach is purely analytical. More recently, time-domain EC models have emerged which besides BMLDs (cf. Culling and Summerfield, 1995; Zerbs, 2000) also account for binaural pitch phenomena (Culling and Summerfield, 1998).

There is some support from physiological data that an EC-like process exists in the mammalian auditory system. A subgroup of cells in the lateral superior olive (LSO) and a subgroup of cells in the inferior colliculus (IC) are excited by the signals from one ear and inhibited by the signals from the other ear (Rose *et al.*, 1966; Boudreau and Tsuchitani, 1968; Kuwada *et al.*, 1984; Joris and Yin, 1995; Batra *et al.*, 1997a,b; Palmer *et al.*, 1997; McAlpine *et al.*, 1998). The cells in the LSO are typically excited by the ipsilateral ear and inhibited by the contralateral ear and are therefore classified as EI-type (Excitation-Inhibition) cells. For neurons situated in the IC the excitatory and inhibitory channels are typically reversed and these cells are classified as IE-type cells. The opposite influence of the two ears makes these cells sensitive to interaural intensity differences (IIDs). With increasing inhibitory level, the neuron's activity decreases up to a certain level where its activity is completely inhibited. The IID necessary to completely inhibit the cell's response varies across neurons (Park *et al.*, 1997; Tsuchitani, 1997; Park, 1998). We refer to the minimum interaural intensity difference needed to completely inhibit the activity as the neuron's *characteristic IID*. Within a phenomenological context we may think of the whole population of EI-type neurons with different characteristic IIDs as multiple 'taps' wherein differences in levels between channels are processed in parallel, very similar to the ITD-sensitivity for EE-type neurons. There are some suggestive data for the LSO (Park *et al.*, 1997; Tsuchitani, 1997) and for the IC (Irvine and Gago, 1990) that the IID sensitivity of EI-type neurons reflects the differences in threshold between the excitatory and inhibitory inputs that innervate each EI-type cell. In addition to IID sensitivity, EI-type cells have been reported to exhibit ITD sensitivity as well (Joris and Yin, 1995; Joris, 1996; Park, 1998). These results suggest that both ITD and IID sensitivity may be understood by considering the outcome of a subtractive mechanism for EI-type neurons with a characteristic IID *and* ITD.

Despite this apparent similarity between EI-type cell properties and the basic mechanism of the EC theory, it is uncertain to what extent EI-type neurons contribute to binaural hearing phenomena in humans. In experimental animals, ITD sensitive IE units only comprise 12% of low-frequency units in the IC (Palmer *et al.*, 1997). Furthermore, anatomical studies revealed that the LSO in humans is much less well developed than in experimental animals (Moore, 1987). Hence the physiological basis for a human binaural processing model based on EI-type neurons is uncertain.

Although the two binaural mechanisms described above are different in their phenomenological properties, this does not necessarily mean that these processes differ in terms of their predictive scope. In fact, Domnitz and Colburn (1976) argued that for an interaurally out-of-phase tonal signal masked by a diotic Gaussian noise, a model based on the interaural correlation and a model based on the distribution of the interaural differences will predict essentially the same thresholds. Furthermore, Colburn and Durlach (1978) and Green (1992) stated that the decision variables based on the correlation and on an EC mechanism are linear functions of one another, hence resulting in equivalent predictions. Consequently, as written in Colburn and Durlach (1978), the effect of interaural parameters of both the masker and signal can be accounted for independently of whether the decision variable is identified with the interaural correlation or with the interaural differences.

Recently, however, it has been shown that in certain other conditions, differences exist between these models. For example, the results of Chapter 2 of this thesis showed that $NoS\pi$ conditions with non-Gaussian noise result in different predictions for the two theories. It was argued that an EC-like model may be favored over a model based on the cross-correlation in that it describes thresholds for static and dynamically varying ITDs and IIDs more satisfactorily: it provides a way to describe sensitivity to IIDs and ITDs, as well as binaural detection data with one single parameter. Second, in Chapter 3 it was demonstrated that uncertainty in the binaural parameters is treated differently by models based on the correlation compared to models based on an EC-like mechanism. A third difference between the two binaural mechanisms is related to the temporal processing. More specifically, the effect of signal and masker duration in an $NoS\pi$ condition is difficult to understand in terms of the cross-correlation (see Section 4.7). A fourth difference is related to stimulus-level variability. The EC-type detection process is not vulnerable to stimulus level variability. Cross-correlation based models on the other hand require specific accomodations to reduce the detrimental effects of stimulus level variability on detection performance in narrowband-noise conditions (see van de Par *et al.*, 2001; Colburn and Isabelle, 2001).

In summary, for many binaural detection conditions, models based on the EC-theory and models based on the cross-correlation are expected to give very similar predictions. In conditions where these predictions are not similar, however, a model based on an EC mechanism may be favored over a cross-correlation model because of its wider predictive scope and the fact that less specific assumptions have to be made. We therefore have chosen to base the binaural interaction in our model on an EI-like interaction. This choice will be further motivated in Section 4.7.

4.2 Model Philosophy

Two different approaches can be pursued when developing a model. An important category of models is purely analytical. This means that the model and its predictions heavily rely on stimulus statistics rather than on explicit waveforms. This class of models provides a powerful means to understand many aspects of data in the literature, independent of details of realization. At the same time the analytical nature presents us with the limitation that it is very difficult to obtain predictions for arbitrary stimulus configurations, like for frozen-noise tokens. This drawback conflicts with the most important objective in our modeling efforts: to develop a model that can simulate a wide variety of binaural detection data *without any restrictions with respect to the stimuli*. In this respect, we followed the philosophy of Dau *et al.* (1996a) to make the model applicable to binaural conditions with stochastic as well as deterministic stimuli, such as frozen noise. Therefore, the model must be able to deal with actual time signals and each processing stage of the model must be described accordingly. The advantage of this approach is that the model can be used as an artificial observer, for example, in a 3-interval, forced-choice procedure with adaptive signal adjustment or for measuring psychometric functions. We tried to combine this philosophy with the requirements of binaural models that were discussed by Colburn and Durlach (1978), who stated that all published models were deficient in at least one of the following areas:

1. Providing a complete quantitative description of how the stimulus waveforms are processed and of how this processing is corrupted by internal noise.

2. Having a sufficiently small number of free parameters in the model to prevent the model from becoming merely a transformation of coordinates or an elaborate curve-fit.

3. Taking account of general perceptual principles in modeling the higher-level, more central portions of the system for which there are no adequate systematic physiological results available.

4. Deriving all the predictions that follow from the assumptions of the model and comparing these predictions to all the relevant data.

5. Relating the assumptions and parameters of the model in a serious manner to known physiological results.

With respect to the first two points, we decided to expand the monaural detection model developed by Dau *et al.* (1996a). This model provides a detailed description of the processing of stimulus waveforms and the extension of this model to a binaural model requires only a few extra model parameters. This extension has the advantage that the large predictive scope of the original monaural model is inherited by the current model.

As noted by Colburn and Durlach (1978) in their third point, it is presently not possible to base the central decision process on systematic physiological data. The analysis stage of the current model is therefore based on mathematical principles rather than physiological knowledge. In particular, an adaptive template-matching procedure was incorporated. The idea of template matching has been used before in modeling monaural auditory perception (cf. Dau, 1992; Dau *et al.*, 1996a) and also for binaural perception (Holube *et al.*, 1995). A new feature that was added is the adaptive nature of the template-matching procedure: if the signal level is changed during a run in a forced-choice detection task with adaptive signal-level adjustment, the model adapts its internal templates accordingly. The advantage of this approach is that the model does not need to 'learn' the stimulus and available detection cues beforehand and hence simulations can start on the fly (as real subjects do).

With respect to the fourth and fifth points raised by Colburn and Durlach (1978), we refer to the three subsequent chapters, which focus on simulations of various detection experiments. In particular, Chapter 5 focusses on spectral and interaural parameters of the stimuli, Chapter 6 deals with temporal stimulus properties, and Chapter 7 discusses 'natural' listening conditions.

4.3   Model overview

The model is divided into three stages as shown in Fig. 4.1. The first stage comprises peripheral preprocessing, including the spectral filtering and hair cell transduction in the cochlea. In the second stage, the binaural processor, the signals from one ear are compared to the corresponding signals from the other ear by means of EI-interactions as a function of the internal characteristic IID and ITD. The third stage is a central processor. This stage can decide whether a signal is present in a masker, but in principle could also extract localization information from the EI-type activity pattern. This central stage receives both the outputs of the binaural processor as well as the direct outputs of the peripheral preprocessor. Each box within the three stages of Fig. 4.1 represents a building block which is a functional or phenomenological model of physiological stages in the mammalian auditory system. Each stage and its building blocks will be specified separately in the following sections.

4.4   Peripheral processing stage

The first stage of the model simulates the effective signal processing of the peripheral auditory system resulting from the outer, middle and inner ear and the auditory nerve. This stage is very similar to the implementation described by Dau *et al.* (1996a). The processing blocks have the following properties:
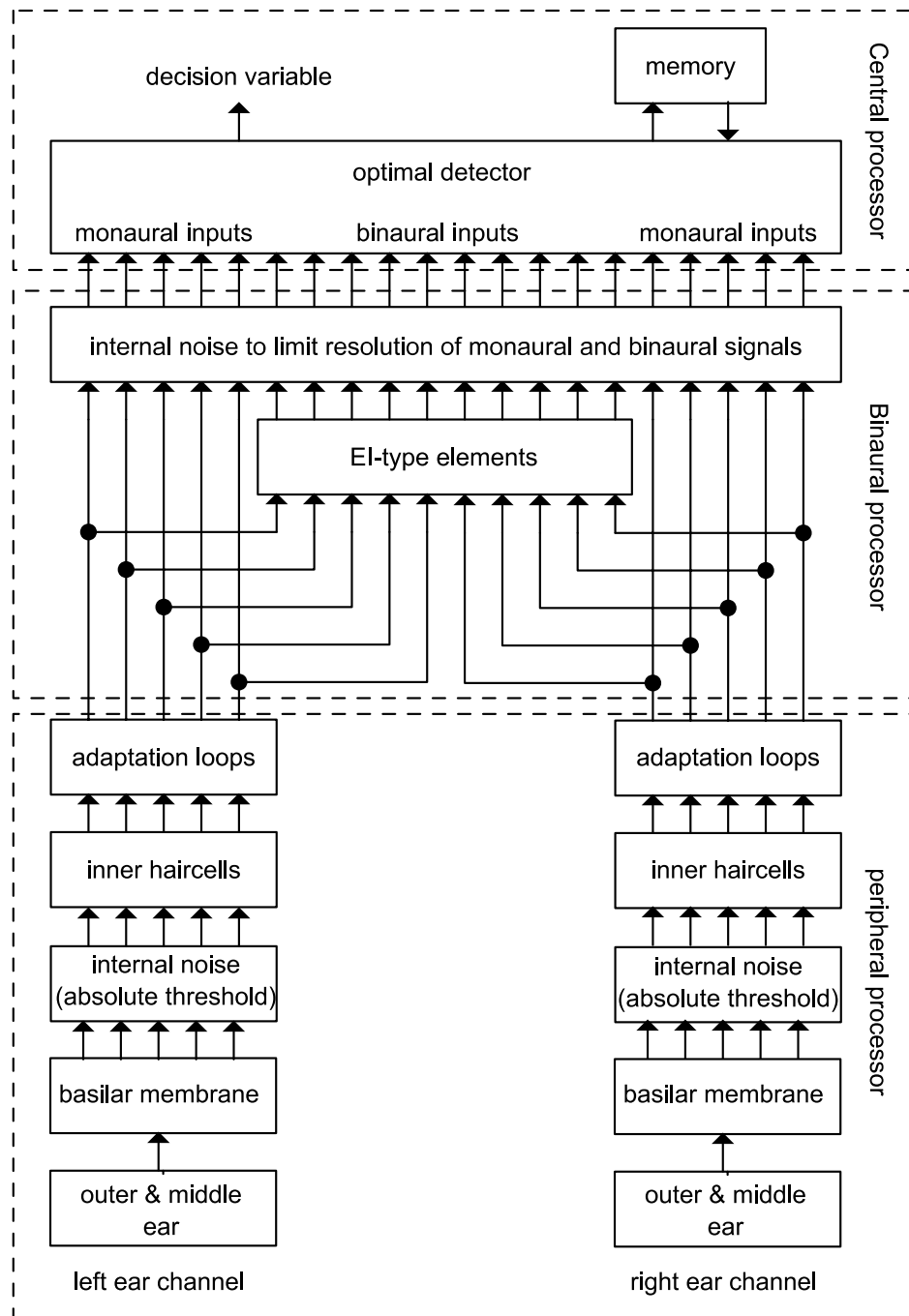
Figure 4.1: *Successive stages of the model. The signals arriving from both ears are processed by a peripheral preprocessing stage (outer and middle ear transfer function, linear basilar membrane model, additive internal noise, inner hair cell stage and adaptation loops), followed by a binaural processor. The signals from the monaural channels and the binaural channels are processed by a central processor, which extracts one decision variable. Internally represented input signals are corrupted by internal noise and are compared to templates stored in memory. This comparison results in a single decision variable.*

1. The combined outer and middle-ear transfer function is modeled by a time-invariant bandpass filter with a rolloff of 6 dB/oct below 1 kHz and -6 dB/oct above 4 kHz. This filter is sufficient to simulate headphone data. For simulations with directional sound, additional convolution with a corresponding head-related transfer function (HRTF) would be necessary. Since we only simulate headphone experiments in this chapter and the two subsequent chapters, HRTF filtering is not included.

2. The cochlea including the basilar membrane is modeled by a third-order gammatone filterbank (Johannesma, 1972; Patterson *et al.*, 1988), using filters with a bandwidth corresponding to the equivalent rectangular bandwidth (ERB) (Glasberg and Moore, 1990). The spectral spacing is 2 filters per ERB. Because of the linear behavior of the gammatone filterbank, basilar membrane nonlinearities such as compression are not included in this stage.

3. To incorporate an absolute threshold (i.e., a noise floor), an independent Gaussian noise is added to each signal originating from the filterbank. The noise is statistically independent for each frequency channel and has a level which corresponds to a sound pressure level of 60 μPa (i.e., 9.4 dB SPL). This value is chosen such that the absolute threshold of a 2-kHz input signal with a level of 5 dB SPL results in an level increase of about 1 dB. In combination with the effect of stage 1, the model thus has a frequency-dependent absolute threshold. For long-duration sinusoidal signals, this threshold is about 5 dB SPL between 1 and 4 kHz.

4. The effective signal processing of the inner hair cells is modeled by a half-wave rectifier, followed by a 5th-order lowpass filter with a cutoff frequency (-3 dB point) of 770 Hz. For frequencies below about 770 Hz, the lowpass filter has (almost) no effect on the output. Hence only the negative phase of the waveform is lost and therefore the timing information in the fine structure of the waveform is preserved at the output. For frequencies above 2000 Hz, (nearly) all phase information is lost after the lowpass filter and only the envelope of the incoming signals is present at the output of this stage. For frequencies in between, a gradual loss of phase information is observed. In this way, the model effectively simulates the decrease of phase locking with increasing frequency observed in the auditory nerve (Kiang, 1975; Johnson, 1980; Weis and Rose, 1988; Bernstein and Trahiotis, 1996).

5. To include the influence of adaptation with various time constants, a chain of five adaptation loops was included (Dau *et al.*, 1996a,b). For a signal with a flat temporal envelope, the input-output characteristic of this chain in steady state is almost logarithmic. The output of these adaptation loops is expressed in model units (MU). These units are scaled in such a way that input levels which correspond to a sound

pressure level of 0 and 100 dB are scaled to 0 and 100 MU, respectively. Fast dynamic changes in the envelope are not compressed by the adaptation loops but are processed almost linearly. These adaptation loops are included at this stage of the model for the following reason. In the first place, the adaptation loops have been successful in predicting detection performance in monaural nonsimultaneous masking conditions (Dau *et al.*, 1996b, 1997). Therefore, the current model has the same capabilities of predicting monaural thresholds, including specific masker waveform dependence and forward- and backward masking. Furthermore, Kohlrausch and Fassel (1997) concluded that adaptation has to *precede* the binaural interaction stage in order to account for binaural forward masking data.

Secondly, it has been shown frequently that for both monaural and binaural detection of signals added to a wideband masker with a variable level, the threshold *signal-to-masker* ratio is approximately constant, as long as the masker level is well above the absolute threshold (cf. McFadden, 1968; Hall and Harvey, 1984). If it is assumed that a certain constant *change* at the output of the adaptation loops is needed to detect a signal, the signal must be equal to a certain *fraction* of the masker level due to the logarithmic compression. Hence the signal-to-masker ratio will be approximately constant at threshold. Thus, by compressing the input signals logarithmically combined with the assumption that a fixed change in the output is necessary for detection, the model can account for the constant signal-to-masker ratio. Hence the adaptation loops work as an automatic gain control exhibiting a monotonic relation between between steady-state input and output levels. To be more explicit, the output waveform is not a simple linearly-scaled version of the input signal. This has implications for binaural conditions with an overall IID, which are discussed in Chapter 5.

An example of the output of the peripheral preprocessing stage is given in Fig. 4.2. The left panel shows the output for a 500-Hz tone with a duration of 100 ms in the auditory channel tuned to the frequency of the tone, while the right panel shows the same for a 4000-Hz tone, both at a level of 70 dB SPL. In this example, it is clear that for high frequencies the fine-structure of the input waveform is lost. Furthermore, effects of peripheral filtering (longer ringing for the 500-Hz signal) and adaptation are clearly visible. Because of the amplitude scaling of the output of the adaptation loops, the fine structure waveform of the output can in principle go negative, to ensure that the *average* steady-state output approximates the rms input in dB SPL. This has no effect on the performance of the model.
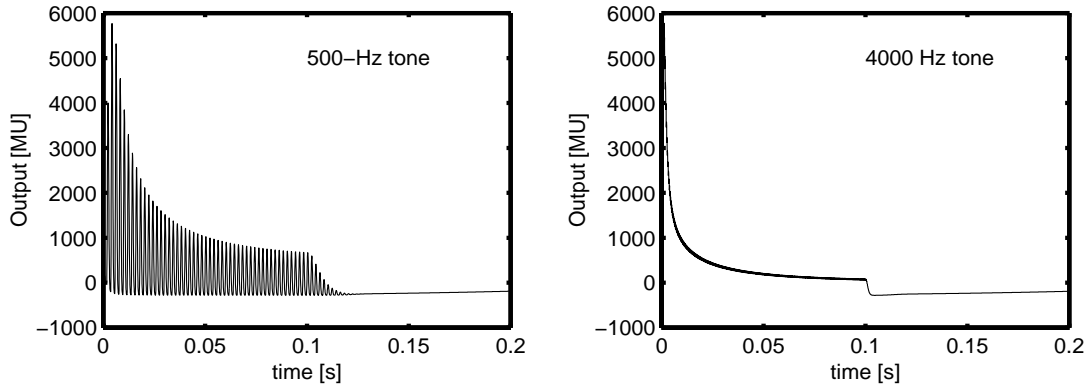
Figure 4.2: *Output of the peripheral preprocessor for a 500-Hz tone (left panel) and a 4000-Hz tone (right panel) of 100-ms duration. The output was calculated for a filter tuned to the frequency of the tone.*

4.5   Binaural processing stage

4.5.1   Structure

In the binaural processor, signals from corresponding auditory channels are compared by EI-type elements. Each EI-type element is described by a characteristic ITD and a characteristic IID. We can think of such a characterization as being the result of an orthogonal combination of the Jeffress' delay line (Jeffress, 1948) with the multiple IID taps of Reed and Blum (1990). This combination is depicted in Fig. 4.3.

The upper and lower horizontal lines carry the time signals from corresponding auditory channels from the right and left ears. The tapped delays (denoted by triangles) combined with the opposite transfer directions of the signals result in a relative interaural delay that varies with the *horizontal* position within the matrix. At the left side, the right-ear signal is delayed compared to the left-ear signal and vice versa. Our extension lies in the fact that each tap of the delay line is connected to a chain of attenuators (depicted by the blocks). The EI-type elements (circles) are connected to these tapped attenuator lines. In a similar way as for the delay line, a relative attenuation occurs which varies with the *vertical* position within the matrix. In this way, the two-parameter characterization of each element which is included for each frequency band results in a three-dimensional time-varying activity pattern if auditory stimuli are presented to the model.

4.5.2   Time-domain description

In principle, two different EI-type elements can be assigned to each auditory filter: one which is excited by the left ear and inhibited by the right ear and a second one with interaurally reversed interaction. The output E_L of the EI-type elements which are excited by the left ear and inhibited by the right ear
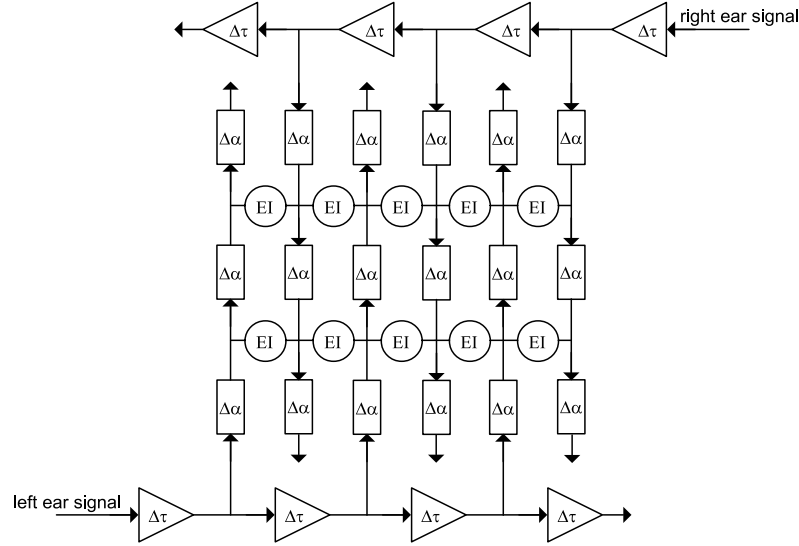
Figure 4.3: *Structure of the binaural processor. The triangles denote delays ($\Delta\tau$), the blocks are attenuators ($\Delta\alpha$) and the circles denote EI-type elements.*

is defined as

$$E_L(i, t, \tau, \alpha) = \lceil 10^{\alpha/40} L_i(t + \tau/2) - 10^{-\alpha/40} R_i(t - \tau/2) \rceil^2, \tag{4.1}$$

while the output of the EI-elements *excited* by the right ear and *inhibited* by the left ear, E_R, is given by

$$E_R(i, t, \tau, \alpha) = \lceil 10^{-\alpha/40} R_i(t - \tau/2) - 10^{\alpha/40} L_i(t + \tau/2) \rceil^2. \tag{4.2}$$

Here, $L_i(t)$ denotes the time-domain output from the left-ear peripheral preprocessor at filter i, $R_i(t)$ the output from the right-ear peripheral pre-processor at filter i and the subscript i refers to auditory channel i. The characteristic IID in dB is denoted by α, the characteristic ITD in seconds by τ. The ceiling brackets ($\lceil . \rceil$) denote a halfwave rectifier: if the inhibitory signal is stronger than the excitatory signal, the output is zero. The fact that the output is squared is explained later. From Eqs. 4.1 and 4.2 we can see that the left and right ear signals undergo a relative delay of τ and a relative level adjustment of α dB. Different values of τ and α correspond to different EI-type elements, resulting in a *population* of elements in the (τ, α) space. It is assumed that all possible combinations of τ and α that may occur in real-life listening conditions are represented by an EI-type element, but that some elements are able to deal with even larger values of τ and α. In the model, internal delays of up to 5 ms and internal intensity differences of α=10 dB are realized[2].

---

[2]If the sound pressure at the ear drums is considered, much larger interaural intensity differences may occur than 10 dB. However, these differences in the acoustic signals are severely reduced by the compression in the adaptation loops. We found that the limit for α of 10 dB is appropriate for all conditions that we tested. The range for the internal delays was chosen such that at very low frequencies (i.e., 100 Hz), a delay of half the period of that frequency (e.g. 5 ms) is available.

We found that it is very convenient to reduce the number of EI-type elements by combining the outputs E_L and E_R given in Eqs. 4.1 and 4.2. It can be shown that summation of these signals results in an output E given by

$$E(i, t, \tau, \alpha) = \left(10^{\alpha/40} L_i(t + \tau/2) - 10^{-\alpha/40} R_i(t - \tau/2)\right)^2. \qquad (4.3)$$

An important consequence of the above summation is that the EI-type element described in Eq. 4.3 does not have a monotonic dependence on the externally presented IID but it shows a *minimum* in its activity if the inputs match the characteristic IID of the element. From this point on, the term EI-type element will refer to the combined elements as described in Eq. 4.3. To incorporate a finite binaural temporal resolution, the EI-activity E is processed by a sliding temporal integrator $w(t)$. This integrator is based on results from Kollmeier and Gilkey (1990) and Holube *et al.* (1998) and consists of a double-sided exponential window $w(t)$ with a time constant c of 30 ms:

$$E'(i, t, \tau, \alpha) = \int_{-\infty}^{\infty} E(i, (t + t_{int}), \tau, \alpha) w(t_{int}) dt_{int}, \qquad (4.4)$$

with

$$w(t) = \frac{\exp(-|t|/c)}{2c}. \qquad (4.5)$$

Finally, a compressive function is applied to the output of the integrator to model saturation effects in the EI cells:

$$E''(i, t, \tau, \alpha) = ap(\tau) \log(bE'(i, t, \tau, \alpha) + 1) + n(i, t, \tau, \alpha). \qquad (4.6)$$

An internal noise $n(i, t, \tau, \alpha)$ limits the accuracy of internal binaural processing[3]. It is assumed that the rms level of this Gaussian-noise source is constant and equals 1 MU, and that the noise is independent of time t, auditory channel i, and is the same for different EI-type elements. The scalars a and b are constants. These constants describe the sensitivity to interaural differences and are fixed and equal for all EI-type elements. By adjusting a and b, the output of the EI-type elements is scaled relative to the internal noise and hence the sensitivity for binaural differences can be adjusted.

The weighting function $p(\tau)$ refers to the fact that cells with larger characteristic interaural delays are less frequent than cells with smaller characteristic delays (Batra *et al.*, 1997a). This corresponds to Jeffress (1948) statement that for coincidence counter neurons, "cells are less dense away from the median plane." In our approach, fewer cells means less accurate precision in processing and hence more internal noise. To include this relative increase in the

---

[3]In Section 4.4, an additive noise was described to implement the absolute threshold of hearing. This is a different noise source from the noise mentioned here which is added at the level of the EI-type elements. The EI-type element noise limits the detection of interaural differences which are present in stimuli with a level above the absolute threshold.

internal noise, the EI-type element is scaled by a weighting function which *decreases* with internal delay. The weighting function is described as follows:

$$p(\tau) = 10^{-|\tau|/5},\qquad(4.7)$$

where the internal delay τ is expressed in ms. This formula resulted from data with $(\text{NoS}\pi)_\tau$ stimuli which are presented in Appendix 4.A. Such a distribution along internal delays has also been included in several other binaural detection and localization models (Colburn, 1977; Stern and Colburn, 1978; Stern *et al.*, 1988; Shackleton *et al.*, 1992; Stern and Shear, 1996).

A graphical representation of Eq. 4.6, leaving out the internal noise, is shown in Fig. 4.4. For small values of E', the input-output function is linear. For higher values of E', the curve converges to a logarithmic function.
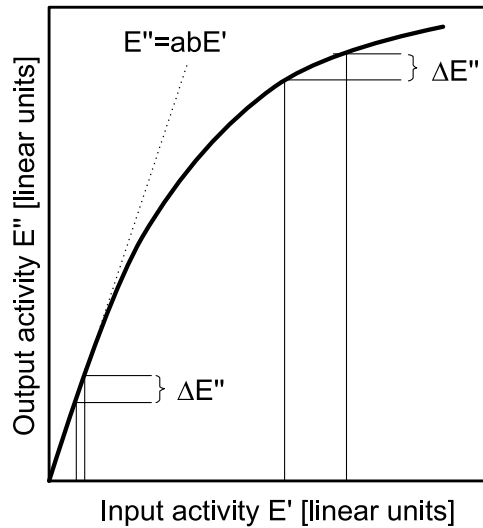
Figure 4.4: *Input-output characteristic of the EI-type element. The dotted line represents the line $E'' = abE'$ (see text).*

The rationale for including the logarithmic transformation in Eq. 4.6 is as follows. Egan *et al.* (1969) measured psychometric functions for $\text{NoS}\pi$ stimuli as a function of the signal power. They found that the sensitivity index d' was linearly related to the signal power $< S^2 >$:

$$d' = m < S^2 > / < N^2 > .\qquad(4.8)$$

Here, $< N^2 >$ denotes the masker power and m is a constant. We will now show that this experimental finding matches our EI-type element input-output function for low signal-to-masker ratios. For an No masker alone, there is no activity E'' for an EI-type element with $\tau=0$ and $\alpha=0$ (if the internal errors are neglected), since the masker is completely canceled. When an $S\pi$ signal is added to the masker, the quadratic input-output characteristic of the EI-type elements results in an output which is related linearly to the *power* of the difference signal between the left and right ear signals. Hence for an interaurally

out-of-phase signal, the result of Eq. 4.3 (i.e., E) is linearly related to the signal power $< S^2 >$. The temporal integrator in Eq. 4.4 does not alter this property. Since for the measurement of psychometric functions, the signal level is low (i.e., near threshold), the result of Eq. 4.6 can be described in a first order approximation by

$$E''(i, t, 0, 0) \approx abp(\tau)E'(i, t, 0, 0) + n(i, t, 0, 0). \tag{4.9}$$

This relation, without incorporation of the internal noise n, is shown by the dotted line in Fig. 4.4. Thus, the change at the output of the EI-type element near threshold as a function of the input can be described by a linear relation, as given in Eq. 4.9. If E'' is used as a decision variable in the NoSπ detection paradigm, d' is related linearly to the signal power $< S^2 >$ as found by Egan *et al.* (1969). The fact that the *power* of the signal is used as a decision variable in NoSπ paradigms is also supported by the results of Chapter 2, where the power of the difference signal was proposed as a detection variable for stimuli which comprise combinations of static and dynamically varying ITDs and IIDs. The slope relating signal power to d' in the model is represented by the product ab. Therefore, this product represents the model's sensitivity to binaural stimuli with a reference correlation near +1.

For maskers which are not perfectly correlated, for example in an NρSπ condition with $\rho < 0.95$, the approximation from Eq. 4.9 does not hold. For such stimuli, E'' can be approximated by

$$E''(i, t, 0, 0) \approx ap(\tau) \log bE'(i, t, 0, 0) + n(i, t, 0, 0). \tag{4.10}$$

Thus, the input-output relation of this curve is logarithmic. If it is assumed that a certain constant change in E'' is needed to detect a signal (this assumption is reflected in the additive noise with a constant rms value), the change in E' must be equal to a certain *fraction* of E'. Thus, for an additive noise n, we need a constant Weber fraction in E' for equal detectability. This Weber fraction is also shown in Fig. 4.4. At higher input levels, the change in the input (E') necessary to produce a fixed change in the output ($\Delta E''$) is larger than at low input levels. This is in essence similar to the EC-theory (Durlach, 1963). Durlach assumed a fixed signal-to-masker ratio after a (partial) cancellation of the masker. Since Durlach's theory is very successful in predicting BMLDs for wideband NρSπ conditions, it is expected that our model has similar prediction performance for these stimuli. As can be observed from Eq. 4.10, the Weber fraction necessary at threshold is determined by the constant a. Thus, a represents the model's sensitivity for binaural signals at reference correlations smaller than +1.

In the following, some basic aspects of the binaural processing stage will be demonstrated. For all examples, the sample rate of the processed stimuli was

32 kHz. The model parameters a and b were set to 0.1 and 0.00002, respectively. These values resulted from the calibration procedure as described in Chapter 5. All output examples given in this section are shown without the incorporation of the internal noise $n(i, t, \tau, \alpha)$ and with $p(\tau)$=1 for all delays to show the properties at hand more clearly[4].

4.5.3 Static ITDs and IIDs

If a 500-Hz pure tone at a level of 70 dB is presented to the model, an activity pattern in the binaural processor occurs as shown in the upper panel of Fig. 4.5. Here, the idealized (i.e., no internal noise and $p(\tau)$=1) activity of EI units (E'') at 500-Hz center frequency is shown as a function of the characteristic ITD and IID of each element. This activity was computed from the stationary part of the response, it covers the range from 250 to 500 ms after the onset of the tone.

The pattern is periodic along the characteristic ITD axis (τ) and shows a sharp minimum along the characteristic IID axis (α). At the minimum ($\tau = \alpha$=0) the signals are perfectly matched and thus are fully canceled. For other characteristic values within the EI array, only partial cancellation occurs resulting in a remaining activity for these units. Due to the periodic nature of the 500-Hz signal, minimum activity will occur at delays of integer amounts of the signal period. If we apply an external interaural time difference of 1 ms to a 500-Hz tone, an activity pattern occurs as shown in the lower-left panel of Fig. 4.5. Basically, the pattern is the same as the pattern shown in the upper panel of Fig. 4.5 except for a shift along the characteristic ITD axis. Thus, externally presented ITDs result in a shift of the pattern along the internal ITD axis. By scanning the minimum in the pattern, the externally presented ITD can be extracted in a similar way as in models based on cross-correlation.

If a sound is presented with a certain external IID, a similar shift along the internal characteristic IID axis occurs. This is shown in the lower-right panel of Fig. 4.5. The externally presented IID was 20 dB. The pattern is shifted towards positive characteristic IIDs. A noteworthy effect is that the activity in the minimum is no longer equal to zero, indicating that the waveforms from the left and right side cannot be canceled completely. This incomplete cancellation results from the nonlinear processing in the peripheral processor: due to the different input levels at both sides the waveforms

---

[4]The effect of internal noise was not included in the graphs because all pictures consist of a *snapshot* of the EI activity at a certain moment in time. The amount of internal noise for such a snapshot is of the same order of magnitude as the output and hence the model properties that are demonstrated would be impossible to see. The fact that the model does not suffer from this internal noise in the same way as the visual observer does is due to the fact that the optimal detector which is present in the central processor (see Section 4.6 for details) is able to strongly reduce the internal noise by temporal integration.
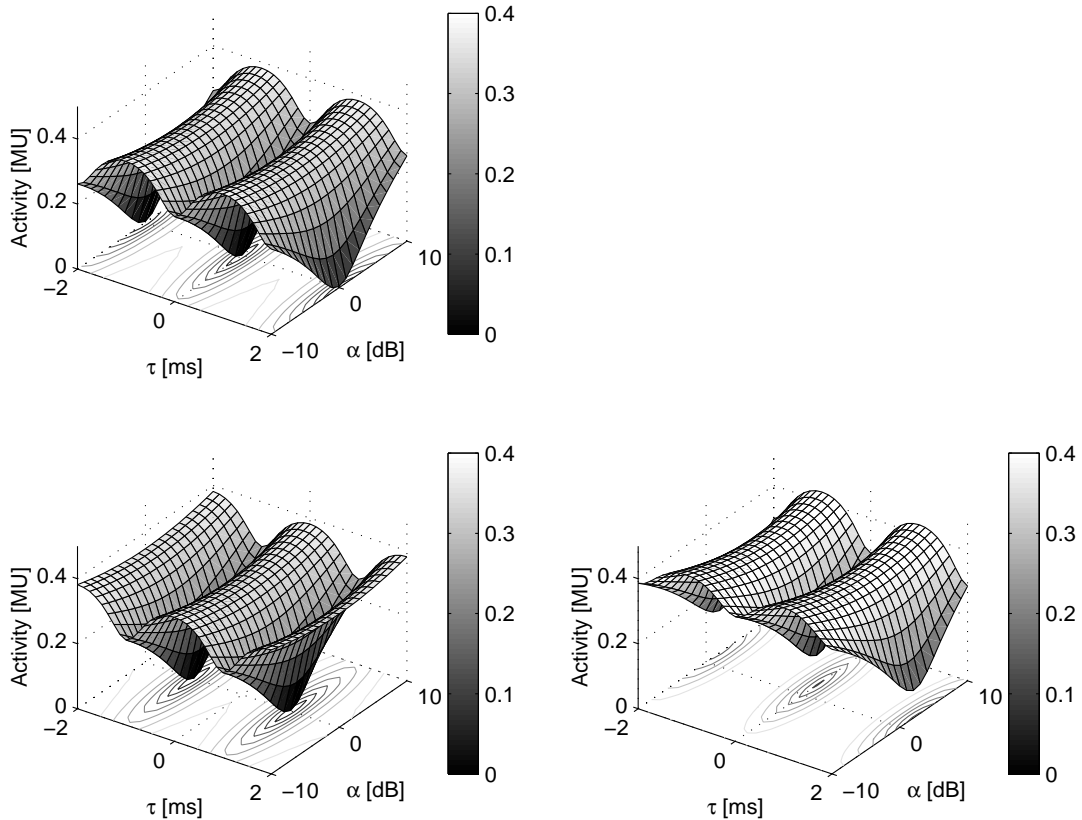
Figure 4.5: *Idealized (no internal noise, p(τ)=1) EI-activity patterns for a 500-Hz sinusoid as a function of the characteristic IID and ITD of each unit. The upper panel corresponds to a diotic signal (i.e., no external IID or ITD). The signal in the lower-left panel has an ITD of 1 ms and no IID, the signal in the lower-right panel has an IID of 20 dB and no ITD.*

cannot be equalized perfectly by applying an internal characteristic IID. Since incomplete cancellation corresponds to a reduced correlation, and this is typically associated with a less compact auditory image, our model's output corresponds to the observation that applying IIDs to a diotic stimulus result in a less compact perceived image (Blauert, 1997, page 170).

Thus, by determining the position of the minimum in the activity pattern, both the externally presented ITD and IID can be extracted. For wideband stimuli the ambiguity of which delay is the delay that corresponds to the location of a sound source can be obtained by combining information across frequency bands (for example a straightness measure) as demonstrated by Stern *et al.* (1988) and Shackleton *et al.* (1992). For narrowband stimuli and pure tones, the ITD can usually be resolved by the headwidth constraint: in daily-life listening conditions the interaural delay is limited to about 0.7 ms by the size of the head.

The ITDs and IIDs are very important when the location of a sound source must be estimated (especially the azimuth). Studies have shown that the perceived locus of a sound source depends on both the IID and the ITD

(Sayers, 1964; Yost, 1981; Schiano *et al.*, 1986). For stimuli presented through headphones, the ITD and IID can be manipulated in such a way that their contributions to the laterality of the perceptual image tend to cancel or reinforce each other. 'Time-intensity tradeability' refers to the extent to which the intracranial locus of a binaural sound depends only on the combined effect of these time and intensity differences, as opposed to the magnitude of these differences considered individually. This trading effect is, however, not perfect. Hafter and Carrier (1972) and Ruotolo *et al.* (1979) found that subjects can discriminate between images that are perceived with the same lateralization but were created by different combinations of IIDs and ITDs. This implies incomplete trading of these interaural parameters. The current model can in principle account for this phenomenon, because IID and ITD estimates of the presented sound source can be extracted independently from the activity pattern and can be combined into one lateralization estimate, for example by weighted addition (e.g., Hafter, 1971).

4.5.4 Time-varying ITDs

In order to analyze the effect of time-varying interaural parameters, consider the internal representation for binaural beats (cf. Perrott and Nelson, 1969; Perrott and Musicant, 1977). The presentation of two identical tones, one to each ear, results in a single fused image centered in the listener's head. If a small interaural frequency difference is introduced (up to 2 Hz), apparent motion is reported. For intermediate frequency differences (i.e., up to 40 Hz), roughness (fast beats) are heard and for large frequency differences, two separate images are perceived. In the model, two tones with the same frequency result in an EI-activity pattern as shown in the upper panel of Fig. 4.5. If the fine-structure waveforms are compared on a short time scale, a *small* interaural frequency difference is equivalent to an interaural phase difference that increases linearly with time. Since this phase difference increases with time, an ongoing shift of the minimum along the characteristic delay axis occurs, and the perceived locus of the sound moves along the line connecting both ears. If the interaural frequency difference is increased (e.g., 10 Hz), the limited temporal resolution of the model becomes increasingly important. During the time span defined by the temporal window, the interaural phase differences will now change considerably. Therefore there is no EI-type element which can cancel the signal completely, resulting in an increase of the EI-activity in the valley and a lowering of the maximum activity. Consequently, there is no sharp minimum within the pattern, indicating that there is no well defined audible locus. Thus, in accordance with psychophysical data, such fast motion is not represented within the binaural display.
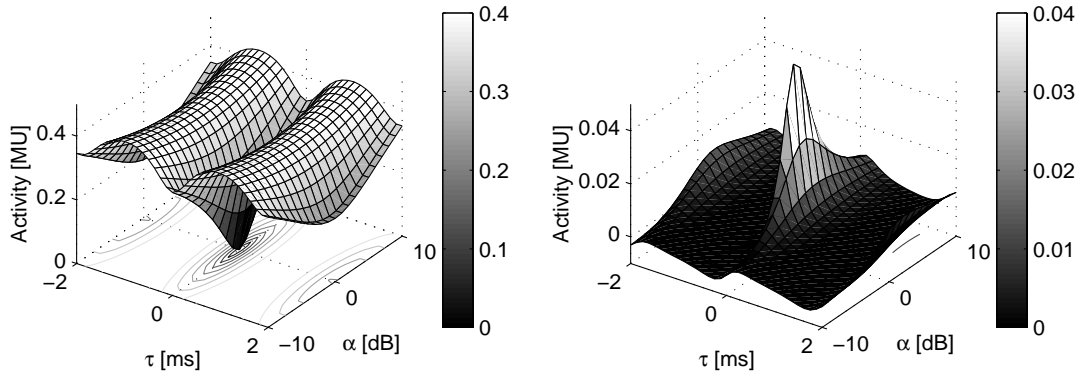
Figure 4.6: *Left panel: Idealized EI-activity for a wideband diotic noise (0-4000 Hz) with an overall level of 70 dB SPL for an auditory filter centered at 500 Hz. Right panel: change in the activity pattern of the left panel if a 500-Hz interaurally out-of-phase signal (Sπ) is added with a level of 50 dB.*

4.5.5 Binaural detection

Human observers are very sensitive to changes in the interaural correlation of binaural signals. This sensitivity reveals itself in the phenomenon of binaural masking level differences (BMLDs). If an interaurally out-of-phase signal is added to an interaurally in-phase noise, the threshold for detecting the signal is up to 25 dB lower than for an in-phase signal (Hirsh, 1948b; Hafter and Carrier, 1970; Zurek and Durlach, 1987). In our modeling framework, the addition of the Sπ signal results in a specific change in the EI-activity pattern. To demonstrate this, the left panel of Fig. 4.6 shows the idealized EI-activity for a diotic wideband noise (0-4 kHz, 70 dB overall level) for an auditory filter centered at 500 Hz.

If a 500-Hz out-of-phase signal with a level of 50 dB SPL is added, the activity pattern changes. The difference between the pattern for the No noise alone and the NoSπ stimulus is shown in the right panel of Fig. 4.6 (note the different scale on the activity axis). Clearly, for a characteristic IID and ITD of zero, there is a substantial change in activity while for other characteristic values, the change is much less. This change in activity can be used as a basis for a decision process in a detection task as will be described in the next section.

4.6   Central processor

The central processor receives both binaural (from the binaural processor) and monaural (directly from the adaptation loops) information. For signal detection purposes, the model can be used as an 'artificial observer', for example in a 3-interval, forced-choice procedure with feedback. The feedback is used by the artificial observer to learn what features of the stimuli have to be used for successful detection. In the 3-IFC procedure, two intervals contain only the masker, while the third interval contains the masker plus signal. The

model's task is to identify which interval contains the test signal. This task is implemented in the following way. We assume that a template, $\bar{E}(i, t, \tau, \alpha)$ is stored in memory, consisting of the mean internal representation of several masker-alone realizations. The ability of listeners to use such a template for detection puposes was suggested before (Dau, 1992; Dau *et al.*, 1996a), and for binaural detection by Holube *et al.* (1995) and in Chapter 3 of this thesis. In our simulations, such a template can be derived in the beginning of a simulated adaptive track, where the large difference between masker-alone and masker-plus-signal intervals allows an easy automatic identification of the masker-alone and the signal intervals. Also the feedback from the simulated adaptive track provides identification of the masker-alone intervals. The task for the detection algorithm is to determine which interval induces an internal representation that differs most from this template. In principle, the differences for all EI-type elements (i.e., as a function of the channel i, time t, characteristic delay τ and characteristic intensity difference α) could be used. However, this results in a considerable complexity due to the large number of dimensions which causes the computing power necessary to compute the output for all relevant EI-type elements to be enormous. We found that for the conditions described in Chapters 5, 6 and 7 it is sufficient to reduce this multidimensional space to only two dimensions, namely time and auditory frequency channel. For each detection experiment, the optimal combination of τ and α is determined for the on-frequency channel. These values are kept constant during that specific experiment for all channels. For example, in a wideband $\text{NoS}\pi$ condition, we already showed that for this specific condition, a maximum change in activity occurs for $\alpha=\tau=0$ (see Fig. 4.6, right panel), while for other values of α and τ, a much smaller effect is observed. It is therefore reasonable to only analyze the position corresponding to minimum activity (which is not necessarily $\alpha=\tau=0$), knowing that not too much information is lost. Conceptually, this would mean that listeners only pay attention to *one* position in space.

The idealized output for one token of an No masker alone as a function of time for the EI-type element with $\tau=\alpha=0$ is shown by the dotted line in the left panel of Fig. 4.7. The masker had a duration of 400 ms, and the 300-ms signal was temporally centered in the masker. Since there is no internal noise and the masker is completely canceled, the output is zero. This result is independent of specific masker realizations, and therefore the template for the masker alone consists also of a zero line. If a signal is added to the masker (with the same parameters as for Fig. 4.6), the output increases. This is shown by the solid line in the left panel of Fig. 4.7 for one realization of an $\text{NoS}\pi$ condition. The peaks and valleys in the output are the result of the adaptation loops in the peripheral preprocessing. If at a certain moment the noise masker has a relatively large amplitude, the adaptation loops will react to this large amplitude and compress the incoming signals more heavily. The result is
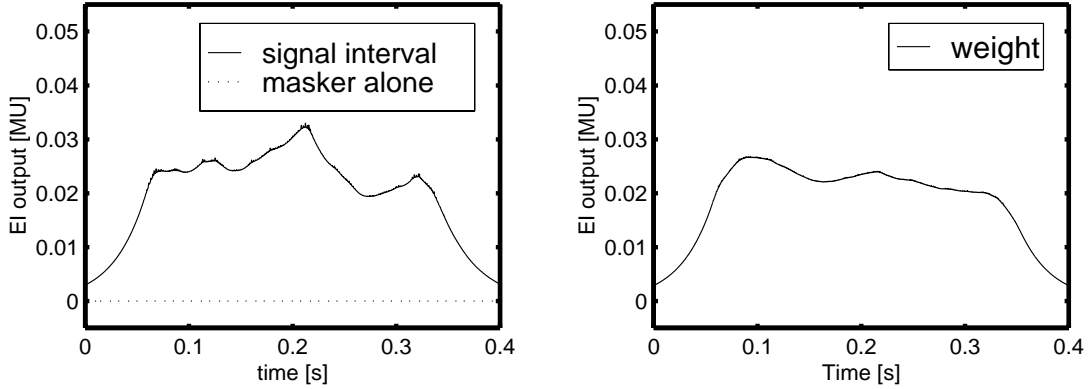
Figure 4.7: *NoSπ EI-activity as a function of time for the EI-type element with α = τ=0 without incorporation of internal noise . The left panel shows the output for a single signal interval (solid line) and for a masker alone (dotted line). The right panel shows the average difference between masker alone and masker-plus-signal. The masker had a duration of 400 ms. The 300-ms signal was temporally centered in the masker. Both signal and masker were gated with 50-ms Hanning windows.*

that the sinusoidal signal, which has a constant envelope, is reduced in level at the output of the adaptation loops and hence the EI-output decreases. Similarly, if a valley occurs in the masker envelope, the EI-output increases. The occurrence of valleys and peaks in the noise masker occurs completely at random; the expected value of the masker amplitude is constant over time. Hence the expected output of the EI-type element in an NoSπ condition is also constant over time. This is demonstrated in the right panel of Fig. 4.7. The solid line (labeled 'weight') represents the mean output for an NoSπ condition averaged over 10 stimulus realizations. These weights inform the model about where in time and frequency the cues for the detection process are present (e.g., the integration window). As expected, the weight is nearly constant, except for the on- and offset of the signal.

An idealized example that has a non-zero output for a masker alone is given in Fig. 4.8. Here the masker and signal have the same properties as in the previous example, except for the fact that the interaural masker correlation was reduced to 0.5 (i.e., an NρSπ condition with ρ=0.5) and the signal level was increased to 60 dB. As in the left panel of Fig. 4.7, the solid line represents the output for a single masker-plus-signal interval, the dashed line represents the mean output for 10 masker-alone intervals (i.e., the template). At the interval between 100 and 350 ms, the signal interval (solid line) results in a larger output than the template (dotted line). This is the cue that the model must detect. In contrast, during the interval from 0 to 100 ms the signal interval actually results in a smaller output than the averaged masker alone. This is the result of the specific fine structure waveform of the current masker realization and is not related to the presence or absence of the signal. This demonstrates the necessity of the weights[5] shown in the right panel of Fig. 4.8. As in Fig. 4.7, the

---

[5]It should be noted that in these examples, the variability in the EI output due to internal
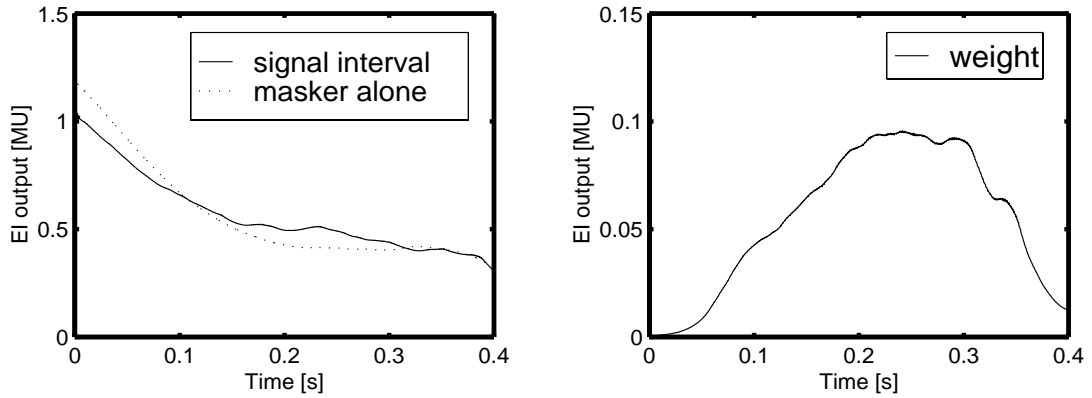
Figure 4.8: *Same as in Fig. 4.7, only for an $N\rho S\pi$ condition with ρ=0.5.*

weights consist of the average difference between the masker-alone intervals and the masker-plus-signal intervals. Since the weights are relatively low during the first 100 ms, the model 'knows' that in this time interval, differences between template and actual stimulus are no reliable cue for the presence of the signal.

To facilitate monaural detection, the output of the adaptation loops is included after being lowpass filtered by a double-sided exponential window with time constants of 10 ms. These lowpass-filtered outputs are multiplied by a constant factor which denotes the monaural sensitivity of the model. The resulting signals are treated as an extra set of signals $E''(i,t)$ which enter the optimal detector. This detector compares the presented stimulus with the average internal representation of the masker-alone stimulus. This average internal representation is referred to as the *template*. All differences across frequency channels and time between the actual stimulus and the template are weighted according to weight functions as shown in, e.g., the right panels of Figs. 4.7 and 4.8 and subsequently combined into one distance measure. This process is described in detail in Appendix 4.B.

4.7 Motivation for EI-based binaural processing

As described in the introduction, basically two binaural interaction processes have been used extensively in binaural models during the last decades. One is based on the interaural cross-correlation, the other on the EC theory. These mechanisms are supported by so-called EE and EI units, respectively, as found in the neurophysiological pathway. In terms of their predictive scope, these mechanisms are very similar (Domnitz and Colburn, 1976; Colburn and Durlach, 1978; Green, 1992). For several reasons it is almost impossible to validate all these models with the same data which have been used for the

---

noise or due to stimulus uncertainty was not taken into account. As shown in Appendix 4.B, the weight that is actually applied by the model consists of the average difference in EI output between masker alone and masker plus signal, *divided* by the variance in the output for a masker alone.

current model. First, a substantial part of the models have not been specified as time-domain models which makes comparisons impossible without additional assumptions. Second, it is difficult to analyze and simulate all of these models including all variations and suggestions for improvements that have been suggested because of the enormous amount of work involved. Third, by describing the current model it is not our intention to demonstrate failures of other models but to show the predictive scope of a time-domain model based on EI interaction. For many of the conditions simulated in Chapters 5 and 6, predictions would be similar if the binaural interaction was based on an EE(correlation)-type interaction instead of an EI-type interaction. There are, however, some conditions where we think that the interaural correlation and EC-based models do *not* give similar results or require different assumptions.

1. A first difference concerns the effect of changes in the duration of the signal and the masker in an NoSπ condition. In principle, two approaches can be applied when using the interaural cross-correlation. The first is to assume that the (normalized) correlation is calculated from the complete duration of the stimulus. The normalized interaural cross-correlation (ρ) for an NoSπ condition is then given by:

$$\rho = \frac{< N^2 > - < S^2 >}{< N^2 > + < S^2 >}.$$

(4.11)

   Here, $< N^2 >$ denotes the masker energy, and $< S^2 >$ denotes the signal energy in the interval over which the correlation is computed, i.e., the duration of the masker burst. For a masker alone, the interaural correlation is 1, because $< S^2 >$ equals zero. The addition of an interaurally out-of-phase signal results in a decrease in the cross-correlation. If the *signal* duration is changed within the interval from which the cross-correlation is computed, a constant signal *energy* will lead to a constant decrease in the cross-correlation. Thus, a doubling in the signal duration can be compensated by a decrease of the signal power by a factor 2 and vice versa. This inverse relation between signal duration and binaural masked thresholds is indeed close to experimental data, which show an effect of 4.5 dB/doubling and 1.5 dB/doubling of signal duration for signal durations below and beyond 60 ms, respectively (cf. Zwicker and Zwicker, 1984; Yost, 1985; Wilson and Fowler, 1986; Wilson and Fugleberg, 1987). According to such a scheme, a doubling in *masker* duration while having a constant short signal duration should lead to a 3 dB increase in threshold. This does, however, not correspond to psychophysical results: NoSπ thresholds for a signal of fixed duration are hardly influenced by the masker duration (McFadden, 1966; Trahiotis *et al.*, 1972; Robinson and Trahiotis, 1972; Kohlrausch, 1986).

   Alternatively, the correlation can be computed only from the stimulus part that contains the signal. In this case, the interaural correlation

would be *independent* of the duration of both signal and masker (as long as the masker duration is at least as long as the signal duration) and hence thresholds would not be influenced by either signal or masker duration, which, again, is in contrast with the experimental results.

The performance of a cross-correlation model could be improved by assuming that an internal noise source is present which accumulates over the signal interval. If the model computes the cross-correlation only from the signal portion of the presented stimulus, both the signal energy and internal noise energy increase equally with signal duration. However, the *variability* of the accumulated internal noise energy *decreases* with signal duration because the number of independent noise samples increases. Since this variability is the limiting factor in the detection process, thresholds are expected to decrease by 1.5 dB/doubling of signal duration. Although this is an improvement of such a model, it still predicts a much shallower slope than found experimentally (see Chapter 6 of this thesis). In the current model, the output of the EI-elements that cancel the masker completely is independent of the masker duration, while an increase of the signal duration results in lower thresholds because the change in the internal activity pattern will be present for a longer period. Chapter 6 demonstrates that the model can quantitatively account for the effect of signal duration in an NoSπ detection task.

2. The interaural cross-correlation is insensitive to static interaural intensity differences. If the relative intensities of the signals arriving at both ears are changed, the product of the waveforms (i.e., the cross-correlation) remains unchanged. Since it is well known that both ITDs and IIDs result in a lateralization of the perceived locus of a sound source (Sayers, 1964), the cross-correlation approach needs additional assumptions to incorporate the processing of IIDs. Some suggestions have been made to incorporate the processing of IIDs, which are based on the incorporation of inhibition of secondary peaks in the cross-correlation function (cf. Lindemann, 1986) or a separate evaluation of the IIDs which is superimposed on the interaural cross-correlation (Stern and Colburn, 1978). Hence it is certainly possible to incorporate IID sensitivity in a cross-correlation-based model. However, we think that the integral IID and ITD sensitivity for static and dynamically varying interaural differences in the current model is a strong point. The common treatment of ITDs and IIDs is a rather restrictive aspect of the model. The internal errors in binaural processing of IIDs and ITDs are characterized by one variable only, the amount of internal noise. In addition, the internal averaging of the binaurally processed stimuli occurs with one temporal window. Thus, the same (internal) temporal resolution is applied to IIDs, ITDs and binaural detection experiments with tones in noise.

3. A third point concerns normalization of the interaural cross-correlation. Several models that have been published are essentially based on the unnormalized cross-correlation, i.e., on the product of the (peripherally filtered) waveforms. However, Breebaart *et al.* (1998) and van de Par *et al.* (2001) noted that unnormalized cross-correlation models cannot account for binaural detection data with narrowband noise maskers because of their inability to cope with fluctuations in the overall masker energy. They argued that the uncertainty in the excitation of the simulated neural activity (i.e., the unnormalized cross-correlation) resulting from a diotic narrowband masker is much larger than the reduction in the excitation due to the addition of an interaurally phase-reversed sinusoid (i.e., NoSπ). This leads to the prediction of very poor binaural performance. Hence cross-correlation based models require specific accommodations to reduce the detrimental effects of stimulus level variability (see van de Par *et al.*, 2001; Colburn and Isabelle, 2001). An often proposed solution is to normalize the inputs to the cross-correlator. However, the accuracy of this normalization must be better than we think is physiologically plausible. Therefore, van de Par *et al.* (2001) suggested that an equalization-cancellation (EC) mechanism may be favored over models based on cross-correlation since this approach is insensitive to overall fluctuations in the masker energy.

4.8   Summary and conclusions

A binaural signal detection model was described that transforms arbitrary stimuli into a three-dimensional internal representation with a minimum of free parameters. This representation is based on Durlach's EC theory instead of the common cross-correlation approach. It was explained that for many experimental conditions, a model based on the EC theory or the cross-correlation give similar predictions, but that in conditions where predictions differ an EC-like mechanism may be favored over the cross-correlation. The internal representation is analyzed by a template-matching procedure which extracts information about the presence or absence of a signal added to a masker. Chapters 5 and 6 provide quantitative predictions for a wide range of binaural signal detection conditions derived with the model described in this chapter. In particular, Chapter 5 discusses the influence of spectral masker and signal parameters on detection thresholds, while Chapter 6 deals with temporal stimulus parameters.

4.A Appendix: Experimental determination of $p(\tau)$

A simple experiment was performed to determine the effect of internal delays upon binaural detection. In particular, an interaural delay was superimposed on an $\mathrm{NoS}\pi$ stimulus. The task was to detect a signal within an interaurally-delayed masker. The signal had a reversed interaural phase plus the same additional delay that was applied to the masker. We refer to this stimulus as $(\mathrm{NoS}\pi)_\tau$. The rationale for this paradigm is that if the binaural system can compensate for the external delay which is present in both masker and signal, the stimulus effectively corresponds to $\mathrm{NoS}\pi$ and a large BMLD should be observed. It is expected, however, that with increasing delays, this compensation results in more internal errors and thresholds will increase. The distribution of errors as a function of τ needed to model these data correctly can be captured in the p(τ) function.

Our data were obtained for one subject only using a 3-interval, forced-choice procedure with adaptive signal-level adjustment. A 400-ms narrowband masker (10-Hz wide) with center frequencies of 125 and 500 Hz was presented at a level of 65 dB SPL. The 300-ms signal was temporally centered in the masker and had a frequency which was equal to the center frequency of the noise. Both masker and signal were gated with 50-ms Hanning windows. Delays up to half the period of the center frequency were used. The results are shown in Fig. 4.A.1. The triangles correspond to a center frequency of 500 Hz, the squares to 125 Hz. The diamonds represent data from a similar experiment performed by Colburn and Latimer (1978). They measured $(\mathrm{NoS}\pi)_\tau$ thresholds for a 500-Hz signal added to a wide-band (20-1000 Hz) Gaussian-noise masker with an overall level of 75 dB SPL.
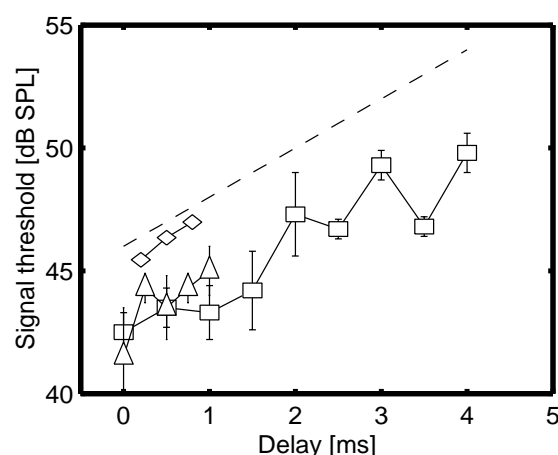
Figure 4.A.1:*(NoSπ)$_\tau$ thresholds as a function of the interaural delay τ. The triangles correspond to a center frequency of 500 Hz, the squares to 125 Hz. The diamonds are data adapted from Colburn and Latimer (1978). The errorbars denote the standard error of the mean. The dashed line indicates a slope of 2 dB/ms.*

As expected, the thresholds increase with increasing delay. The slope of this increase is about 2 dB/ms, which is indicated by the dashed line. The delay dependence of the thresholds is close to linear if the thresholds are expressed in dB. To incorporate a similar threshold dependence in our model, the weighting function must have an exponential decay. A slope of 2 dB/ms means that every 3 ms the signal amplitude is doubled at threshold, which corresponds to a factor 4 in the EI-type element output. Therefore, the $p(\tau)$ function must decrease by a factor 4 every 3 ms, which results in the formulation given in Eq. 4.7.

4.B   Appendix: Optimal detector

A set of channels E'' consisting of binaural and monaural signals (one for each frequency channel) is presented at the input of the optimal detector. Since further processing of the monaural and binaural channels is exactly equal, we will refer to the complete set of channels $E''(i,t)$ as the input of the optimal detector rather than using binaural and monaural channels separately.

For the channel i, the distance $U(i,t)$ between a template $\bar{E}(i,t)$ and the actual output $E''(i,t)$ is given by

$$U(i,t) = E''(i,t) - \bar{E}(i,t). \tag{4.B.1}$$

The variance of $U(i,t)$ resulting from internal noise and masker uncertainty is denoted by $\sigma^2(i,t)$, while the mean difference between masker plus test signal and masker alone near threshold level is denoted by $\mu(i,t)$. A single number which describes the total difference between stimulus and template is assigned to each interval. This difference value, U, is computed by integrating the temporally-weighted difference signal $U(i,t)$:

$$U = \int_i \int_{t=0}^{t=T} \frac{\mu(i,t)}{\sigma^2(i,t)} U(i,t) di\, dt. \tag{4.B.2}$$

T denotes the interval duration. Thus, integration is performed over both time and auditory channels. The weighting function $(\mu(i,t)/\sigma^2(i,t))$ ensures that the model only takes differences between template and actual signal into account at positions where differences are expected. Furthermore, if at a certain position, the difference has a large amount of variability (i.e., a large value of σ), this uncertain output has a smaller weight compared to positions with smaller uncertainty. The weighting function is optimal when the variability represented by σ is Gaussian, an assumption which does not always hold for the internal representation. Still it seems the most reasonable choice to use this weighting function. In a detection task decisions will be based on the value of U. The higher U, the greater the likelihood that a signal is present. Thus, in a 3-IFC procedure the model will choose the interval with the highest value of U. After each trial, the model receives feedback. By storing the internal representations of the three stimuli in memory (i.e., two masker-alone realizations and one masker-plus-signal realization), the model can update its estimate of $\bar{E}(i,t)$, $\sigma^2(i,t)$ and $\mu(i,t)$. $\bar{E}(i,t)$ is updated by averaging the output $E(i,t)$ of all presented masker realisations. In a similar way, the average value of all internal signal representations is computed. Then $\mu(i,t)$ is obtained by subtracting the mean internal representation of masker-alone intervals and of masker-plus-signal intervals. Finally, $\sigma^2(i,t)$ is obtained by computing the variance in the internal masker-alone representations.

4.C   Appendix: Discrete-time implementation

4.C.1   Conventions

In the following, several stages are described that process the incoming signals. Each stage is described by an input-output relation, given in the discrete time domain. All inputs are denoted by $\mathbf{x}[n]$, where n is a sample of the time-domain waveform $x(t)$ for $t = nT$ where T denotes the sampling period. The sampling frequency is denoted by f_s, and is usually set at 32 kHz. In a similar way, the output of each stage is described by $\mathbf{y}[n]$. The signal levels are calibrated as follows: a signal with a rms-value of 1 model unit [MU] corresponds to a sound pressure level of 20 μPa (= 0 dB SPL).

4.C.2   Peripheral preprocessor

4.C.2.1   Outer- and middle ear tranfer function

The outer- and middle ear transfer function is simulated by a simple bandpass filter with cutoff frequencies of 1 kHz and 4 kHz. This filter was implemented as a recursive time-domain filter of which the input-output relation is given by

$$\mathbf{y}[n] = (1 - q)r\mathbf{x}[n] - (1 - q)r\mathbf{x}[n - 1] + (q + r)\mathbf{y}[n - 1] - qr\mathbf{y}[n - 2] \quad \textbf{(4.C.1)}$$

with

$$q = 2 - \cos(2\pi 4000/f_s) - \sqrt{(\cos(2\pi 4000/f_s) - 2)^2 - 1}, \quad \text{(4.C.2)}$$

and

$$r = 2 - \cos(2\pi 1000/f_s) - \sqrt{(\cos(2\pi 1000/f_s) - 2)^2 - 1}. \quad \text{(4.C.3)}$$

4.C.2.2   Basilar membrane filter characteristics

The frequency resolution of the basilar membrane was modeled by a third-order gammatone filterbank. This filterbank consists of a set of bandpass filters at ERB spacing (Glasberg and Moore, 1990). Each filter is characterized by a filter number i and has a center frequency f_c given by

$$f_c = \frac{e^{(0.11i)} - 1}{0.00437}. \quad \text{(4.C.4)}$$

The impulse response $\mathbf{h}[n]$ of the gammatone filter is given by

$$\mathbf{h}[n] = \begin{cases} A(nT)^{\nu-1}e^{-2\pi bnT}\cos(2\pi f_c nT + \theta) & \text{for } n \geq 0 \\ 0 & \text{for } n < 0 \end{cases}. \quad \text{(4.C.5)}$$

Here, ν denotes the order of the filter, b is a parameter that determines the bandwidth, A is an overall scaling parameter and θ is the starting phase.

The scalefactor A was set to ensure a 0 dB filter response for $f = f_c$, resulting in

$$A = \frac{2(2\pi b)^{\nu}}{(\nu - 1)!}. \tag{4.C.6}$$

The 3-dB bandwidth B_{3dB} of the gammatone filter is given by (Patterson *et al.*, 1988)

$$B_{3dB} = 2b\sqrt{2^{1/\nu} - 1}, \tag{4.C.7}$$

while the equivalent rectangular bandwidth B_{ERB} of the filter is given by

$$B_{ERB} = \frac{\pi b(2\nu - 2)!2^{-(2\nu-2)}}{(\nu - 1)!^2}. \tag{4.C.8}$$

According to Glasberg and Moore (1990), the equivalent rectangular bandwidth B_{human} of the auditory filters as a function of frequency is given by

$$B_{human} = 24.7(0.00437f_c + 1). \tag{4.C.9}$$

The parameter b was set to meet $B_{human} = B_{ERB}$, resulting in

$$b(f_c) = \frac{24.7(0.00437f_c + 1)(\nu - 1)!^2}{\pi(2\nu - 2)!2^{-(2\nu-2)}}. \tag{4.C.10}$$

First, the incoming signal is shifted in frequency by f_c Hz. The frequency-shifted signal $\mathbf{y}[n]$ is given by

$$\mathbf{y}[n] = \mathbf{x}[n]e^{-2\pi j f_c nT}. \tag{4.C.11}$$

Subsequently, the signal is lowpass filtered following

$$\mathbf{y}[n] = (1 - e^{-2\pi bT})\mathbf{x}[n] + e^{-2\pi bT}\mathbf{y}[n - 1]. \tag{4.C.12}$$

To achieve a third-order filter, the above low-pass filtering was applied 3 times. After the lowpass filtering, the signal is shifted up in frequency following

$$\mathbf{y}[n] = 2Re\left(\mathbf{x}[n]e^{2\pi j f_c nT}\right). \tag{4.C.13}$$

The resulting transfer function $H(f)$ can be approximated by (Patterson *et al.*, 1988)

$$H(f) = \left(\frac{1}{1 + j(f - f_c)/b} \right)^{\nu}. \tag{4.C.14}$$

4.C.2.3 Absolute threshold

To incorporate an absolute hearing threshold, an internal noise $N[n]$ is added to each signal originating from the filterbank. This noise is independent of the frequency channel and has a level which corresponds to a sound-pressure level of 60 μPa:

$$\mathbf{y}[n] = \mathbf{x}[n] + N[n]. \tag{4.C.15}$$

The noise $N[n]$ stems from a Gaussian-noise process and has an expected average of 0 and a standard deviation of 3 MU.

4.C.2.4 Inner haircells

The effective signal processing of the inner haircells is modeled by a halfwave rectifier:

$$\mathbf{y}[n] = \begin{cases} \mathbf{x}[n] & \text{if } \mathbf{x}[n] > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.C.16}$$

To include the loss of phaselocking at frequencies above 1 kHz, a 5th order lowpass filter was used. This 5th order filter was realized by cascading 5 1th order lowpass filters with a cutoff frequency of 2000 Hz. The input-output relation of this 1th order filter is given by:

$$\mathbf{y}[n] = (1 - u)\mathbf{x}[n] + u\mathbf{y}[n - 1], \tag{4.C.17}$$

with

$$u = 2 - \cos(2\pi 2000/f_s) - \sqrt{(\cos(2\pi 2000/f_s) - 2)^2 - 1}. \tag{4.C.18}$$

The resulting 5th-order filter has a -3dB cutoff frequency of 770 Hz.

4.C.2.5 Adaptation loops

To include the effects of adaptation, a chain of adaptation loops was included. Each loop has a specific time constant v. The adaptation loop is described as follows:

$$\mathbf{y}[n] = (1 - e^{-1/f_s v})\frac{\mathbf{x}[n]}{\mathbf{y}[n - 1]} + e^{-1/f_s v}\mathbf{y}[n - 1] \tag{4.C.19}$$

Five adaptation loops were used, with time constants of v=5, 129, 253, 376 and 500 ms. After these 5 adaptation loops, the output is scaled in such a way, that an level of 0 results in a mean output of 0, and that an input level which corresponds to 100 dB SPL results in an output of 100 MU in the steady state:

$$\mathbf{y}[n] = 273(\mathbf{x}[n] - 1). \tag{4.C.20}$$

4.C.3 EI-processor

4.C.3.1 EI-type element in the time domain

Each EI-type element is characterized by an internal delay τ, an internal IID α and is attached to a filter i. The output $\mathbf{E}(\tau, \alpha, i)[n]$, given the outputs $\mathbf{L_i}[n]$ and $\mathbf{R_i}[n]$ of the left and right ears, respectively, is given by

$$\mathbf{E}(\tau, \alpha, i)[n] = \left(10^{\alpha/40}\mathbf{L_i}[n + \tau f_s/2] - 10^{-\alpha/40}\mathbf{R_i}[n - \tau f_s/2]\right)^2. \tag{4.C.21}$$

After this EI-interaction, the output $\mathbf{E}(\tau, \alpha, i)[n]$ is lowpass filtered with a timeconstant c = 30 ms following

$$\mathbf{E}'(\tau, \alpha, i)[n] = \frac{1}{f_s} \sum_{\delta=-\infty}^{\delta=\infty} \mathbf{E}(\tau, \alpha, i)[n + \delta]\mathbf{w}[\delta], \tag{4.C.22}$$

with

$$\mathbf{w}[\delta] = \frac{1}{2c} \exp\left(-|\delta/f_s|/c\right). \tag{4.C.23}$$

For finite-length signals, the above lowpass filtering can be calculated in a faster iterative way by first filtering E with

$$\mathbf{E}_\mathbf{1}'[n] = (1 - e^{-1/f_s c})\mathbf{E}[n] + e^{-1/f_s c}\mathbf{E}_\mathbf{1}'[n - 1], \tag{4.C.24}$$

and subsequently using the same filter but in reverse order following

$$\mathbf{E}_\mathbf{2}'[n] = (1 - e^{-1/f_s c})\mathbf{E}[n] + e^{-1/f_s c}\mathbf{E}_\mathbf{2}'[n + 1]. \tag{4.C.25}$$

Then E$'$ is given by

$$\mathbf{E}'[n] = \mathbf{E}_\mathbf{1}'[n] + \mathbf{E}_\mathbf{2}'[n] - (1 - e^{-1/f_s c})\mathbf{E}[n]. \tag{4.C.26}$$

4.C.3.2 EI-type element in the α,τ domain

The above algorithm is suitable to compute the output for one or a few EI-type elements as a function of time. If one requires a snapshot within the α,τ domain at a fixed time, the following approximation algorithm may be favoured. Assume that a snapshot is required at a certain time n. Then $\mathbf{E}'(\tau, \alpha, i)[n]$ is given by

$$\mathbf{E}'(\tau, \alpha, i)[n] = \frac{1}{f_s} \sum_{\delta=-\infty}^{\delta=\infty} \mathbf{E}(\tau, \alpha, i)[n + \delta]\mathbf{w}[\delta], \qquad (4.C.27)$$

which results in

$$\mathbf{E}'(\tau, \alpha, i)[n] =$$
$$\frac{1}{f_s} \sum_\delta \left(10^{\alpha/40}\mathbf{L_i}[n + \tau f_s/2 + \delta] - 10^{-\alpha/40}\mathbf{R_i}[n - \tau f_s/2 + \delta]\right)^2 \mathbf{w}[\delta]. \qquad (4.C.28)$$

We now make the following substitutions:

$$\mathbf{L}'[\delta + \tau f_s/2] = \mathbf{L_i}[n + \delta + \tau f_s/2]\sqrt{\mathbf{w}[\delta]} \qquad (4.C.29)$$

$$\mathbf{R}'[\delta - \tau f_s/2] = \mathbf{R_i}[n + \delta - \tau f_s/2]\sqrt{\mathbf{w}[\delta]}. \qquad (4.C.30)$$

It can be shown that for $\tau \ll c$, $\mathbf{E}'(\tau, \alpha, i)[n]$ can be approximated by

$$\mathbf{E}'(\tau, \alpha, i)[n] =$$
$$\frac{10^{\alpha/20}}{f_s} \sum_{\delta=-\infty}^{\delta=\infty} \mathbf{L_i}'^2[\delta] + \frac{10^{-\alpha/20}}{f_s} \sum_{\delta=-\infty}^{\delta=\infty} \mathbf{R_i}'^2[\delta] - 4\mathbf{R_{L'R'}}[\tau f_s], \qquad (4.C.31)$$

where $\mathbf{R_{L'R'}}[\tau f_s]$ denote the unnormalized cross-correlation function between the signals $\mathbf{L}'$ and $\mathbf{R}'$:

$$\mathbf{R_{L'R'}}[\tau/f_s] = \frac{1}{f_s} \sum_{\delta=-\infty}^{\delta=\infty} \mathbf{L_i}'[\tau/(2f_s) + \delta]\mathbf{R_i}'[-\tau/(2f_s) + \delta], \qquad (4.C.32)$$

which can be obtained in the frequency domain by using Fast-Fourier Transform (FFT) algorithms. Thus, $\mathbf{E}'(\tau, \alpha, i)[n]$ for all values of τ and α can be obtained as follows:

1. windowing the left and right-ear signals according to Eqs. 4.C.29 and 4.C.30,

2. calculating the squared sum of these windowed signals,

3. calculating the cross correlation between the windowed signals using FFTs,

4. the EI activity is then obtained as a linear combination of these values as shown in Eq. 4.C.31.

4.C.3.3 Logarithmic compression

The following stage comprises a logarithmic input-output curve, which is given by

$$\mathbf{E}''(\tau, \alpha, i)[n] = ap(\tau) \log\left(b\mathbf{E}'(\tau, \alpha, i)[n] + 1\right). \tag{4.C.33}$$

Here, a and b denote constants that modify the model's sensitivity to binaural parameters. These values were set to 0.1 and 0.00002, respectively. $p(\tau)$ denotes a weighting function across internal delays (τ, in ms) which is given by

$$p(\tau) = 10^{-|\tau|/5}. \tag{4.C.34}$$

4.C.4   Central processor

The central processor receives monaural inputs $\mathbf{L_i}$ and $\mathbf{R_i}$ directly from the adaptation loops and binaural inputs $\mathbf{E}''(\tau, \alpha, i)[n]$ from the binaural processor. The number of binaural inputs is restricted to only one per auditory filter. The values for τ and α were chosen in such a way that for a masker alone, these values corresponded to the minimum in the EI-activity pattern. We write these combined inputs to the central processor as $\mathbf{E}''(i)[n]$. In principle, all these inputs are corrupted by an additive internal noise which is independent across i and n. This internal noise has a fixed rms value (σ_N) and an average of zero. However, the internal templates converge much faster if these are computed without internal noise. Therefore, the following equations contain expressions for the internal noise, but in the model implementation, the actual addition of noise occurs at a later stage (after the matched filter operation). Hence each input $\mathbf{E}''(i)[n]$ is corrupted by internal noise $\mathbf{N}(i)[n]$ in the following way:

$$\mathbf{E}'''(i)[n] = \mathbf{E}''(i)[n] + \mathbf{N}(i)[n]. \tag{4.C.35}$$

From the combined inputs $\mathbf{E}'''(i)[n]$, the central processor develops a template $\bar{\mathbf{E}}(i)[n]$ which consists of the average internal representation of all masker-alone intervals β that have been processed:

$$\bar{\mathbf{E}}(i)[n] = \frac{1}{\beta} \sum_{\beta} \mathbf{E}'''(i)[n]. \tag{4.C.36}$$

Because the average value for the internal noise is zero, for large values of β this converges to

$$\bar{\mathbf{E}}(i)[n] = \frac{1}{\beta} \sum_{\beta} \mathbf{E}''(i)[n]. \qquad (4.\text{C}.37)$$

Furthermore, the model estimates the variance in $\mathbf{E}'''(i)[n]$, $\sigma^2(i)[n]$, given by:

$$\sigma^2(i)[n] = \frac{1}{\beta} \sum_{\beta} \left(\mathbf{E}'''(i)[n]\right)^2 - \bar{\mathbf{E}}^2(i)[n]. \qquad (4.\text{C}.38)$$

Because the internal noise is independent of the actual values of the input signals, this converges for large β to

$$\sigma^2(i)[n] = \sigma_N^2 + \frac{1}{\beta} \sum_{\beta} \left(\mathbf{E}''(i)[n]\right)^2 - \bar{\mathbf{E}}^2(i)[n]. \qquad (4.\text{C}.39)$$

Thus, σ is the combined effect of stimulus variability and the internal noise. Subsequently, the average distance between the average masker alone representation and the average signal interval computed over all signal intervals γ is given by

$$\mu(i)[n] = \frac{1}{\gamma} \sum_{\gamma} \left(\mathbf{E}'''(i)[n]\right) - \bar{\mathbf{E}}(i)[n]. \qquad (4.\text{C}.40)$$

For large values of γ this converges to

$$\mu(i)[n] = \frac{1}{\gamma} \sum_{\gamma} \left(\mathbf{E}''(i)[n]\right) - \bar{\mathbf{E}}(i)[n]. \qquad (4.\text{C}.41)$$

Then the weighted distance U between template and actual stimulus $\mathbf{E}'''(i)[n]$ is given by (see also Eq. 4.B.2)

$$U = \sum_i \sum_n \frac{\mu(i)[n]}{\sigma^2(i)[n]} \left(\mathbf{E}'''(i)[n] - \bar{\mathbf{E}}(i)[n]\right). \qquad (4.\text{C}.42)$$

In this equation, only $\mathbf{E}'''(i)[n]$ is corrupted by internal noise, given the fact that the templates are completely converged. Hence U can be written using Eq. 4.C.39 as

$$U = \sum_i \sum_n \frac{\mu(i)[n]}{\sigma^2(i)[n]} \left(\mathbf{E}''(i)[n] - \bar{\mathbf{E}}(i)[n]\right) + N_U, \qquad (4.\text{C}.43)$$

where N_U denotes the noise generator that generates all internal noise on U, of which the variance is given by

$$\sigma_{N_U}^2 = \sigma_N^2 \sum_i \sum_n \frac{\mu^2(i)[n]}{\sigma^4(i)[n]}, \qquad (4.\text{C}.44)$$

with $\sigma_N^2 = 1$.

CHAPTER 5 _____

Predictions as a function of spectral stimulus parameters[1]

In this chapter, the model described in Chapter 4 is tested and validated by compar-
ing its predictions with experimental data for binaural discrimination and masking
conditions as a function of the spectral parameters of both masker and signal. For
this purpose, the model is used as an artificial observer in a 3-interval, forced-choice
adaptive procedure. All model parameters were kept constant for all simulations
described in this and the subsequent chapter. The effects of the following experi-
mental parameters were investigated: center frequency of both masker and target,
bandwidth of masker and target, the interaural phase relations of masker and tar-
get and the level of the masker. Several phenomena that occur in binaural listening
conditions can be accounted for. These include the wider effective binaural critical
bandwidth observed in band-widening NoSπ conditions, the different masker-level
dependence of binaural detection thresholds for narrow and for wideband maskers,
the unification of IID and ITD sensitivity with binaural detection data and the de-
pendence of binaural thresholds on frequency.

5.1 Introduction

This chapter describes and discusses simulations of binaural detection tasks with a binaural processing model which is described in detail in Chapter 4. This model basically consists of three stages. The first stage simulates the effective signal processing of the basilar membrane and the inner haircells and includes adaptation by means of adaptation loops (Dau *et al.*, 1996a). Binaural interaction is modeled in the second stage by means of a contralateral inhibition mechanism: the model computes the squared difference signal between the left and right ears as a function of time, frequency channel, internal interaural delay (τ in seconds) and internal interaural level adjustment (α in dB). These binaural signals are corrupted by internal noise and subsequently analyzed by the third stage in the model, the central processor.

_____

[1]This chapter is based on Breebaart, van de Par, and Kohlrausch (2001b).

The model is used as an artificial observer in a 3-interval, forced-choice procedure, in which the central processor matches the representations of the presented stimuli to templates (derived during previous presentations), and on this basis the model indicates which interval contains the signal.

The scope of the simulations of this chapter is restricted to binaurally "stationary" stimuli. For these cases, the spectral parameters and the interaural phase relations of the stimuli were not varied as a function of time, and the stimuli had a duration which was long in comparison to the temporal resolution of the auditory system. First, the ability of the model to capture some basic properties of binaural hearing is demonstrated. These include IID and ITD sensitivity and binaural detection performance of tones in noise as a function of the bandwidth, center frequency and interaural phase relationships of the stimuli.

The major focus is the apparently wider critical bandwidth in binaural conditions which was found in a number of studies. If an interaurally out-of-phase signal (Sπ) must be detected against an No masker of variable bandwidth, the estimate of the critical bandwidth is 2 to 3 times the estimate which is found in monaural experiments (Sever and Small, 1979; Zurek and Durlach, 1987; van de Par and Kohlrausch, 1999). Hall *et al.* (1983) found that this discrepancy between monaural and binaural estimates is largest at high masker levels. Furthermore, the wider effective bandwidth is only observed if the interaural cross-correlation of the masker is very close to +1 (van der Heijden and Trahiotis, 1998). On the other hand, experiments that use a masker with frequency-dependent interaural phase relations reveal a critical band estimate that basically agrees with the monaural estimate (Sondhi and Guttman, 1966; Kohlrausch, 1988; Kollmeier and Holube, 1992; Holube *et al.*, 1998). Furthermore, data that were obtained as a function of the bandwidth of the *signal* also show a monaural bandwidth behavior (Langhans and Kohlrausch, 1992; Breebaart *et al.*, 1999, and Chapter 2 of this thesis). In this chapter, it is demonstrated that the current model, which includes peripheral filters with a bandwidth based on monaural estimates, can account for all observations described above. Depending on the experimental paradigm, the model shows different bandwidth dependencies. It is also explained that the wider effective critical bandwidth does not result from a poorer spectral resolution of the binaural auditory system compared to the monaural system, but is related to the ability of the model to *integrate* information across filters.

In Chapter 6, temporal properties of the model are discussed, which include the effect of signal and masker duration, phase transitions in the time domain and forward masking.

5.2 Method

5.2.1 Relevant stages of the model

In the introduction, a coarse description of the general model setup was given. In this section, the stages of the model that are relevant for the simulations described in this chapter (i.e., spectral behavior) are discussed in more detail. For a detailed description of the complete model, see Chapter 4 of this thesis.

- Filtering of the gammatone filterbank. The filterbank present in the peripheral processing stage determines the spectral resolution of the model, in line with the ERB estimates published by Glasberg and Moore (1990).

- Inner hair cell model. This stage consists of a half-wave rectifier followed by a fifth-order lowpass filter with a cutoff frequency (-3 dB) of 1000 Hz. Hence below 1000 Hz, both the ITDs and IIDs are preserved at the output of this stage. However, above 2 kHz, the output approximates the envelope of the incoming signals and hence only IIDs and ITDs present in the envelope are preserved. Between 1 and 2 kHz, the ITD in the fine structure waveforms is gradually lost.

- Adaptation loops. The chain of adaptation loops in the peripheral processor has an almost logarithmic input-output characteristic in steady state and is a non-linear device. These properties have two consequences. First of all, it has been shown frequently that for both monaural and binaural detection of signals added to a wideband masker with a variable level, the threshold *signal-to-masker* ratio is approximately constant, as long as the masker level is well above the absolute threshold (cf. McFadden, 1968; Hall and Harvey, 1984). If it is assumed that a certain constant *change* at the output of the adaptation loops is needed to detect a signal, the signal must be equal to a certain *fraction* of the masker level due to the logarithmic compression. Hence the signal-to-masker ratio will be approximately constant at threshold. Second, due to the nonlinear behavior, large interaural intensity differences at the input cannot be canceled completely by a linear level adjustment at the output.

- Compressive input-output characteristic of EI-type elements. The temporally-smoothed difference signal of the EI-type elements is compressed logarithmically. In combination with an additive internal noise, this stage results in thresholds of interaural differences that depend on the interaural cross-correlation of the reference stimuli.

- Weighting as a function of the internal delay. The model includes a weighting function that decreases with the internal delay of the EI-type elements. Consequently, the relative amount of internal noise increases with internal delay.

- Optimal detector in the central processor. The EI-type element outputs are corrupted by an additive internal noise. Subsequently, the internal representations of the external stimuli are compared to a template that consists of the average masker-alone representation. The differences between the actual stimulus and the template are weighted and integrated both in the time and the frequency domain according to an optimal criterion. This enables the optimal detector to reduce the internal signal-to-noise ratio for stimuli that have a valuable detection cue in more than one auditory filter.

5.2.2 Procedure

Masked thresholds were simulated using an adaptive 3-interval forced-choice (3IFC) procedure. The masker was presented in three consecutive intervals. One of the intervals contained the signal. The model's task was to indicate which interval contained the signal. The level of the signal was adjusted according to a two-down one-up algorithm (Levitt, 1971). The initial stepsize for adjusting the level was 8 dB. The stepsize was halved after every second track reversal until it reached 1 dB. The run was then continued for another 8 reversals. The median level at these last 8 reversals was used as the threshold value. At least 5 repetitions were performed for each parameter value. All thresholds are plotted as mean values of all repetitions, and the errorbars denote the standard deviation of the repetitions.

5.2.3 Stimuli

All stimuli were generated digitally at a sampling frequency of 32 kHz. The maskers used in the different experiments had a duration of 400 ms unless stated differently. They were presented interaurally in-phase (No), interaurally out-of-phase (Nπ), or with a specific interaural correlation (Nρ) obtained by combining No and Nπ noises (see Chapter 3 of this thesis). The signals were presented interaurally in-phase (So), interaurally out-of-phase (Sπ) or to one ear only (Sm). Bandpass Gaussian noises were generated by computing the Fourier transform of white noise and setting the amplitude coefficients outside the desired frequency range to zero. After an inverse Fourier transform, the bandpass noise was obtained.

In all simulations, the level, bandwidth, on- and offset ramps and onset delay of both the maskers and signals were set to the values used in the corresponding experiments with human subjects. If more data sets from various authors with different experimental settings were used, the experimental settings from one of these studies were used for determining the model simulations. Comparison with the other datasets was possible because in such conditions, the BMLDs were calculated or thresholds were expressed relative to the spectrum level of the masking noise.

5.2.4 Model calibration

As described in Chapter 4, the model effectively calculates the difference signal between the left and right ears as a function of an internal delay and internal level adjustment. The subtraction is performed by so-called EI (excitation-inhibition) elements. The sensitivity to interaural differences of these EI-type elements is determined by two (fixed) model parameters a and b. By changing these parameters, the EI-output is scaled relative to the internal noise which has a fixed level. These sensitivity parameters were determined as follows. Detection thresholds were simulated for a 500-Hz interaurally out-of-phase sinusoid in a Gaussian low-pass masker (cutoff frequency of 1 kHz) which had an interaural correlation of 0.64[2] and an overall level of 65 dB SPL. In this condition only the parameter a determines the detectability of the tone. This parameter was adjusted to reach a threshold of 46 dB SPL. Subsequently, the interaural correlation of the masker was set to +1 and the parameter b was adjusted in order to reach a threshold of 38 dB SPL. These thresholds were adapted from frozen-noise $N\rho S\pi$ data given in Chapter 3 of this thesis. The resulting values of a and b are 0.1 and 0.00002, respectively. The monaural sensitivity of the model was adjusted such that the just noticeable difference in intensity of a 500-Hz, 400-ms sinusoid with a level of 65 dB SPL was 1 dB.

Note that during all simulations, all model parameters and procedures were kept constant. This restriction has the consequence that sometimes overall differences between model predictions and experimental data sets occur. However, similar differences exist between experimental data sets from different publications. For such conditions, much better predictions could have been obtained by calibrating the model's BMLD separately for each experiment. Nevertheless, all parameters were kept constant in order to demonstrate to what extent the current model can account for different experimental findings.

5.3 Simulations

5.3.1 Detection of static interaural differences

The first simulations comprised the detection of static interaural intensity and time differences. Hence in these experiments only pure tones were presented to the model in the absence of any noise masker. The pure tones had a duration of 400 ms and were gated with 50-ms Hanning ramps. The presentation level was 65 dB SPL. The reference stimuli were presented diotically

---

[2]The value of 0.64 for the interaural correlation was chosen because this was the lowest correlation that was used in a recent study measuring $N\rho S\pi$ thresholds (see Chapter 3 of this thesis). Data from this study were used because these thresholds were obtained with *frozen* noise. The advantage of using frozen noise is that these thresholds are not influenced by stimulus uncertainty, but are only determined by internal noise.
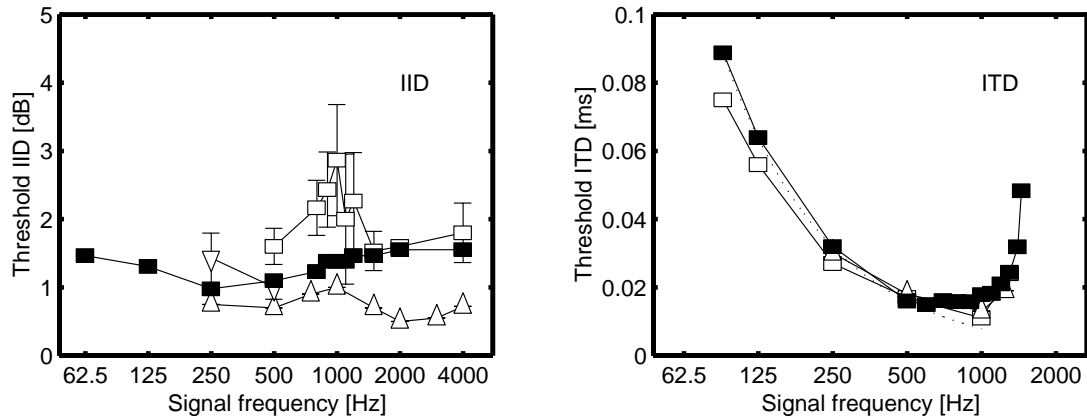
Figure 5.1: *IID thresholds (left panel) and ITD thresholds (right panel) for tones as a function of frequency. The open symbols denote data adapted from literature, the filled symbols are model predictions. The dotted line in the right panel denotes a constant interaural phase difference of 0.05 rad. Legend left panel: squares Grantham (1984b), upward triangles Mills (1960), downward triangles McFadden et al. (1971). Legend right panel: squares Klumpp and Eady (1956), upward triangles Zwislocki and Feldman (1956).*

and the target interval contained an IID in the first set of simulations and an ITD in the second set. The size of the ITD or IID was varied adaptively, similar to the procedure described in Section 5.2.2. For IID-discrimination simulations, various frequencies between 62.5 and 4000 Hz were tested. For ITD-discrimination simulations, frequencies with octave spacing were used below 500 Hz and a linear spacing of 100 Hz was used above 500 Hz. In the left panel of Fig. 5.1, the IID thresholds (filled symbols) of the model are presented as a function of the frequency of the tone together with experimental data (open symbols). The experimental data were adapted from literature: squares denote Grantham (1984b), upward triangles Mills (1960), and downward triangles denote McFadden *et al.* (1971). The predicted IID thresholds do not depend systematically on the frequency and lie between 1 and 1.6 dB. This is well in the range of the experimental data. The remarkable bump at 1 kHz seen in one set of the experimental data is, however, lacking in the predictions.

The right panel of Fig. 5.1 shows ITD thresholds as a function of frequency. The open squares are data adapted from Klumpp and Eady (1956), the upward triangles from Zwislocki and Feldman (1956) and the filled symbols are model predictions. For frequencies up to 500 Hz, the ITD threshold decreases with increasing frequency. This ITD threshold curve can be characterized by a constant phase sensitivity, as shown by the dotted line. This line represents a constant phase difference of 0.05 rad. For frequencies above 1 kHz, the ITD threshold increases sharply due to the decrease of phase locking in the inner hair cell stage. Above 1.5 kHz, the model is not sensitive to static ITDs in the fine structure of the presented waveforms.

The predictions for IID thresholds and ITD thresholds below 1000 Hz only depend on the model parameters a and b which were derived from $N\rho S\pi$ and $NoS\pi$ detection experiments (see Section 5.2.4). In these experiments, both IIDs and ITDs are present in the stimulus, which fluctuate as a function of time. In the simulations shown in Fig. 5.1, the stimuli contained only static IIDs or static ITDs. Thus, this model is able to unify IID and ITD sensitivity with binaural detection data. The use of a cancellation mechanism to achieve the correct sensitivity for both IIDs and ITDs was already suggested in Chapter 2 of this thesis, where it was found that for stimuli containing several different probability distributions of the IIDs or ITDs, a model based on subtraction may be favored over models based on the interaural cross-correlation or models based on the direct evaluation of the interaural differences.

5.3.2 Dependence on frequency and interaural phase relationships in wideband detection conditions

A lowpass noise with a cutoff frequency of 8 kHz and a spectral level of 40 dB/Hz was used as masker. The following binaural conditions were tested: $NoS\pi$, $N\pi So$, $NoSm$ and $N\pi Sm$. Thresholds were determined as a function of the frequency of the signal (125, 250, 500, 1000, 2000 and 4000 Hz). The upper and middle panels in Fig. 5.2 represent the results of the four different conditions that were tested. The open symbols are experimental data extracted from different studies (see figure caption for a description), and the filled symbols are the model predictions. The thresholds are plotted relative to the spectral level of the masker to compensate for differences in masker spectral level.

For the $NoSm$ and the $NoS\pi$ conditions (upper panels), the thresholds decrease slightly with frequency between 125 and 500 Hz and they increase towards higher frequencies with a slope of 4 to 5 dB/oct. In the model, this slope partly results from the increased bandwidth of the auditory filters towards high frequencies because thresholds are expressed relative to the masker spectral level. This predicts a threshold increase by about 3 dB/oct[3].

A second reason why thresholds increase above 1 kHz center frequency is related to the loss of phase locking in the inner hair-cell model. In the stimuli that are considered here, both interaural intensity and time differences are present (cf. Zurek, 1991). As described in Section 5.3.1, the model is insensitive

---

[3]Note that in the peripheral preprocessing stage of the model, a chain of adaptation loops is included. For an input signal with a constant envelope, the input-output characteristic of the chain of adaptation loops is almost logarithmic. Due to this compression, the signal level must be a certain *fraction* of the masker level to produce a fixed change in an EI-type-element output which is necessary to exceed the internal noise level. Thus, if the masker energy within one auditory channel increases (which is the case towards higher frequencies), the thresholds increase by the same amount.
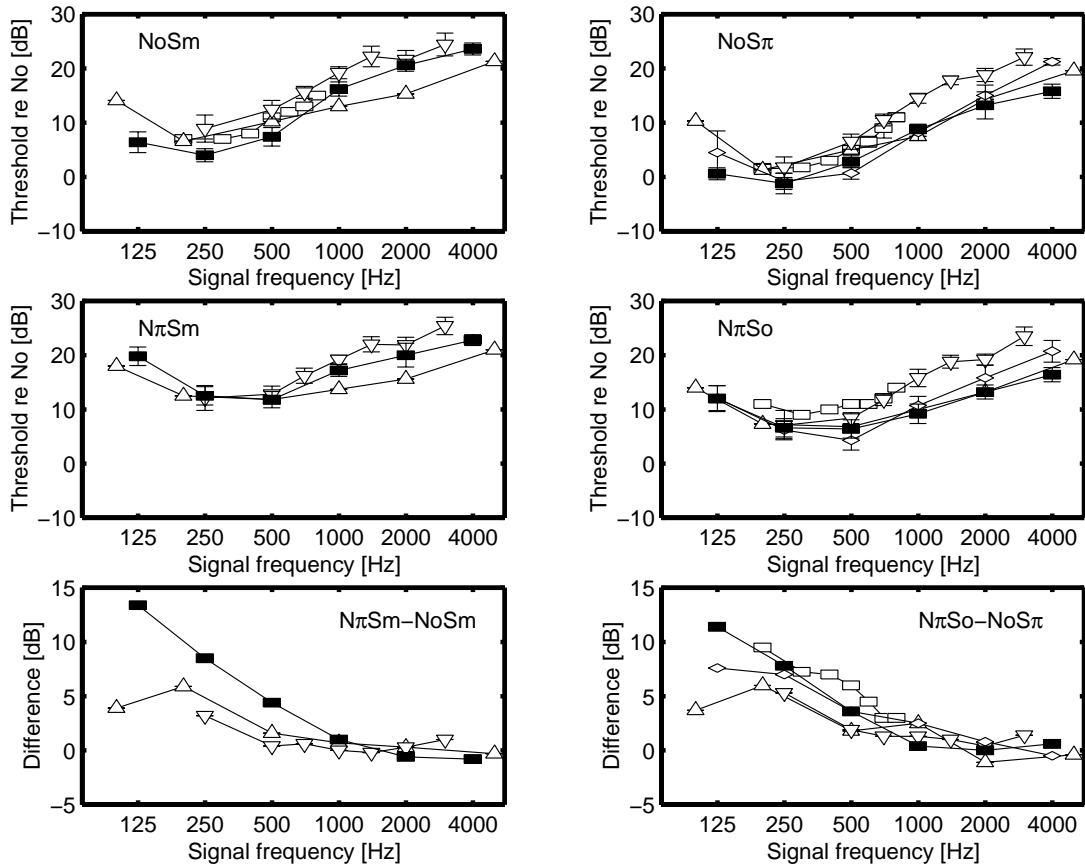
Figure 5.2: *Masked thresholds for wideband NoSm (upper left panel), NoSπ (upper right panel), NπSm (middle left panel) and NπSo (middle right panel) conditions as a function of the frequency of the signal. The lower left panel represents the difference in threshold between the NπSm and NoSm conditions, the lower right panel between NπSo and NoSπ. The open symbols denote experimental data adapted from literature, the filled symbols are model predictions. Squares Kohlrausch (1988), upward triangles Hirsh (1948b), downward triangles Hirsh and Burgeat (1958), diamonds van de Par and Kohlrausch (1999).*

to interaural time differences within the fine structure of the waveforms for frequencies above 1.4 kHz, because of the loss of phase locking in the inner haircell stage. Hence a part of the cues that are available at low frequencies are lost at high frequencies resulting in higher thresholds.

A third reason for a threshold increase with frequency results from peripheral compression. Above 1.4 kHz, only the envelope of the incoming waveforms is present at the output of the inner haircell stage. These envelopes are compressed by the nonlinear adaptation loops that follow the inner haircell stage. Such a compression results in a decrease of the stimulus IID (van de Par, 1998; van de Par and Kohlrausch, 1998b). Furthermore, compression makes the model less sensitive to interaural time differences present in the envelopes, because the envelopes are flattened. Hence compression results in higher thresholds at frequencies above 1.4 kHz. Low-frequency detection is not affected much by compression because the model can use ITDs in the fine structure waveforms which are not affected by compression.

The major difference between the NoSπ and NoSm conditions is an overall difference of 6 dB. This difference can be understood by considering the fact that an Sπ signal results in twice as much increase in the EI-type element output compared to an Sm signal. To compensate for this difference, the Sm signal level must be raised by 6 dB.

The NπSo and NπSm conditions show similar thresholds as the NoSπ and the NoSm conditions for frequencies beyond 1 kHz, while for frequencies below 1 kHz, the interaurally out-of-phase maskers result in higher thresholds than the in-phase maskers. This difference is depicted in the lower panels of Fig. 5.2. The left panel shows the difference in thresholds between the NπSm and NoSm conditions, the right panel between NπSo and NoSπ. The model predictions in these two panels are very similar for the monaural and dichotic signal, having differences of 12 to 14 dB at 125 Hz center frequency which decreases to 0 dB around 1 or 2 kHz. These frequency effects are the result of two model properties which are included in most EC-like models (cf. Durlach, 1963; Rabiner *et al.*, 1966; Metz *et al.*, 1968) and in models based on coincidence detectors (Colburn, 1977; Stern and Shear, 1996). The first comprises a limited repertoire of internal delays. A lower center frequency corresponds to a larger internal delay necessary to compensate for the phase shift of the masking noise and hence less sensitivity to changes in the EI-type element output. The second is the fact that the phase shift can only be compensated by an internal delay, resulting in imperfect cancellation of the noise masker due to damping of the autocorrelation function of the noise.

5.3.3 NoSπ masker-bandwidth dependence

If an interaurally out-of-phase signal is masked by an interaurally in-phase noise of variable bandwidth, a remarkable phenomenon is observed which is usually referred to as the wider effective binaural critical bandwidth: the critical bandwidth estimate from binaural band-widening experiments is often a factor 2 to 3 higher than monaural estimates (cf. Bourbon and Jeffress, 1965; Sever and Small, 1979; Hall *et al.*, 1983; Zurek and Durlach, 1987; van de Par and Kohlrausch, 1999). In order to show that the model can account for this phenomenon, NoSπ thresholds were determined as a function of the bandwidth of the masker at center frequencies of 250, 500, 1000, 2000 and 4000 Hz. The bandwidth was varied between 5 Hz and twice the center frequency. The overall masker level was kept constant at 65 dB SPL. Both model predictions (filled symbols) and experimental data (open symbols) are shown in Fig. 5.3, where the 6 panels correspond to center frequencies of 125, 250, 500, 1000, 2000 and 4000 Hz, respectively. The data can be characterized as staying fairly constant up to a certain bandwidth and then declining with 3 dB/oct. The bandwidth at which this decline starts is much larger than
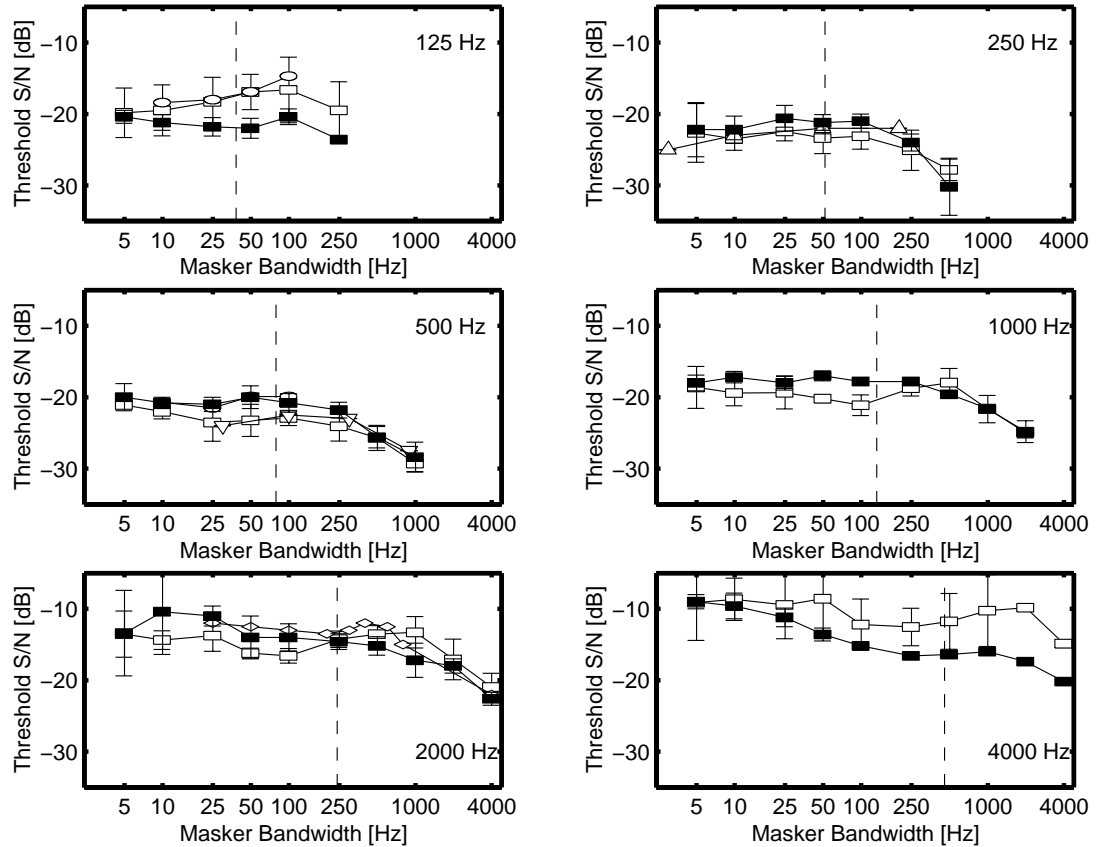
Figure 5.3: *NoSπ thresholds as a function of the masker bandwidth for a constant overall level of the masker. The six panels represent center frequencies of 125, 250, 500, 1000, 2000 and 4000 Hz, respectively. The dashed line indicates the ERB value at the signal frequency. The filled symbols are model predictions. The open symbols are data adapted from literature. Squares van de Par and Kohlrausch (1999), upward triangles Wightman (1971), downward triangles van der Heijden and Trahiotis (1998), diamonds Sever and Small (1979), circles Breebaart* et al. *(1998).*

what is expected from monaural notched-noise experiments (Glasberg and Moore, 1990), as shown by the dashed lines in Fig. 5.3. The estimated critical bandwidths of the model are also larger by a factor up to 2 or 3, in line with the experimental data.

At 4 kHz, the simulated thresholds decrease as a function of bandwidth for bandwidths between 10 and 250 Hz, an effect which is not observed in the experimental data. The threshold decrease with bandwidth at 4 kHz of the model can be understood by considering the properties of the adaptation loops in the peripheral processor. Due to the loss of phase locking at high frequencies, the outputs of the inner haircell model only contain IIDs and some ITD information present in the envelope of the waveforms. The IIDs are reduced in magnitude due to the compressive nature of the adaptation loops. The amount of compression depends on the bandwidth of the stimulus. If the bandwidth is small (i.e., 10 Hz), the envelope of the masker varies slowly with time. Hence the adaptation loops which adapt with certain

time constants easily follow these envelope fluctuations. Consequently, the envelopes are compressed logarithmically. If the masker bandwidth is increased, the speed of fluctuation in the envelopes increases accordingly. As described in Chapter 4, fast fluctuations are processed more linearly, while slow fluctuations in the envelope are processed logarithmically. Hence the magnitude of the IID after the adaptation loops depends on the bandwidth of the stimulus: a very large bandwidth results in larger IID after adaptation and hence lower NoSπ thresholds[4].

At all center frequencies the increased effective bandwidth is captured by the model. The explanation for this result in this specific experimental paradigm relies on the across-frequency integration of information according to an optimal detector, as described by van de Par and Kohlrausch (1999). This spectral integration is an integral part of the central processor in this model (see Chapter 4 of this thesis). For a narrowband masker (i.e., below the monaural critical bandwidth), the on-frequency filter has the largest stimulus power. For off-frequency channels, the entire stimulus resides at the skirts of the filters and is therefore reduced in its power. However, the relative amount of masker and signal energy is hardly changed. Since the signal-to-masker ratio within an auditory channel determines the detectability, information about the presence of the signal is not only available in the on-frequency channel, but also in several off-frequency channels. The only limitation for this extended availability is the absolute threshold: if the stimulus is attenuated too much it becomes undetectable in that channel. This extended availability of information is depicted in Fig. 5.4. In this figure, the output for a 500-Hz NoSπ stimulus of an EI-type element with $\alpha=\tau=0$ without internal noise is shown as a function of the masker bandwidth and of the auditory-filter number. The masker had a fixed level of 65 dB while the signal-to-masker ratio was -25 dB.

If a masker alone were presented, the output of these EI-type elements would be zero (neglecting the internal noise) since the masker can be canceled completely. Thus, any increase in the activity can be used as a cue for the presence of the signal. This increase by the presence of the Sπ signal is shown in Fig. 5.4. If the masker bandwidth is very small (10-Hz wide), a whole range of EI-type elements shows a considerable amount of activity. The fact that in this condition the cue for detection is available in several channels enables reduction of the internal error in the following way. The internal error which is added to the EI-type elements is *independent* across elements. Thus, individual noise sources add up by their intensities. On the other hand,

---

[4]This mismatch is another example of the same disadvantage caused by the strong overshoot of the adaptation loops (see Dau *et al.*, 1996b, for a discussion). Despite several years of simulations, neither we or our colleagues in Oldenburg have so far found a satisfying solution for this problem which would reduce the overshoot and preserve the major advantage of this stage: its ability to predict nonsimultaneous masking.
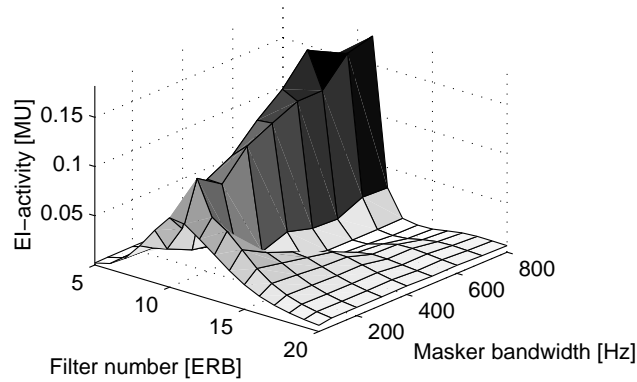
Figure 5.4: *EI-activity without internal noise as a function of masker bandwidth and peripheral filter number for an NoSπ condition with a signal-to-masker ratio of -25 dB and a fixed masker level of 65 dB.*

the increase in the EI-type elements by the addition of the signal is available in several filters and is *correlated* and therefore adds up linearly. Thus, if the model uses the sum of activities across several elements instead of using the output of only one element, the internal signal-to-noise ratio is increased. This results in lower thresholds for narrowband maskers.

For bandwidths that just exceed the critical bandwidth, the masker energy in the on-frequency channel starts to be reduced by the bandpass filter. This results in an increase in the signal-to-masker ratio in the on-frequency channel. This can be observed from Fig. 5.4 by the increasing EI-activity with increasing bandwidth for filter number 10. However, the signal-to-masker ratio in the *off-frequency* channels starts to be *reduced*. This is clearly visible for filters 13 to 20; the activity *decreases* with increasing masker bandwidth. Therefore, the ability of the model to reduce the internal error by integrating across filters is diminished. Both processes influence the internal error about equally but in opposite directions, resulting in constant thresholds for bandwidths between the critical bandwidth and 2 to 3 times that value. For even larger bandwidths, all off-frequency channels are masked. Only the on-frequency channel provides useful information and due to filtering thresholds decrease with 3 dB/oct of masker bandwidth. In summary, the wider critical bandwidth in the model is the result of an uncorrelated noise source in each auditory filter combined with an optimal detector. Other implementation issues are relatively unimportant, a notion which is supported by the results of Zerbs (2000). He developed a binaural signal detection model which is also based on an EC-like process but with a different implementation of the binaural processing stage. This model does account in a very similar way for the wider effective critical bandwidth.
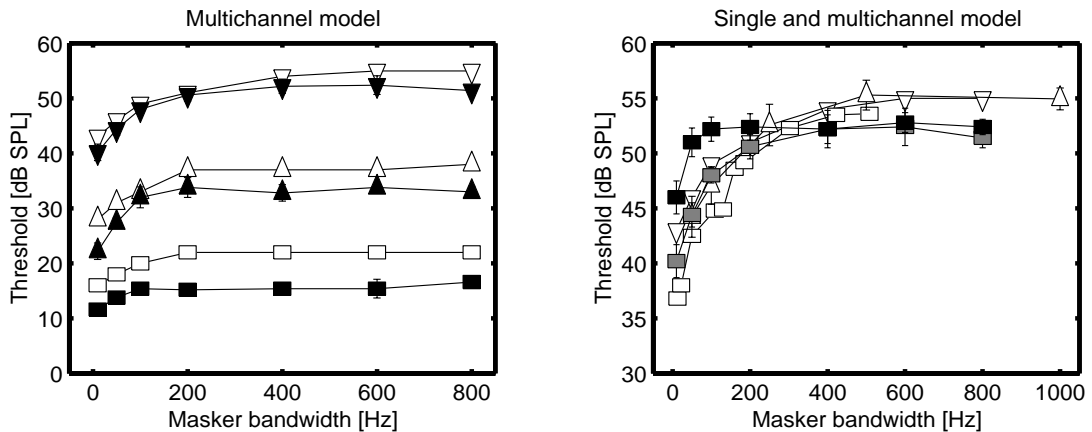
Figure 5.5: *NoSπ thresholds as a function of masker bandwidth. Left panel: predictions of a multichannel model. In all conditions, the spectral energy density of the masker was kept constant at 10 dB/Hz (squares), 30 dB/Hz (upward triangles) and 50 dB/Hz (downward triangles). The open symbols are data adapted from Hall* et al. *(1983), the filled symbols are model predictions. Right panel: similar to the left panel for a masker level of 50 dB/Hz. The squares denote data adapted from Bourbon (1966), the upward triangles denote Cokely and Hall (1991) and the downward triangles Hall* et al. *(1983). The gray symbols represent the multichannel model, the black symbols the single-channel model.*

Another set of data that can be explained by this across-channel integration hypothesis has been published by Hall *et al.* (1983), who also performed band-limiting NoSπ measurements. In their study, the spectral *energy* density of the masker was kept constant and they estimated the binaural critical bandwidth at three noise levels, namely 10, 30 and 50 dB/Hz. Their critical-bandwidth estimate increased as the noise level increased. To test whether the model can account for this observation, the same experiment was simulated with masker bandwidths of 10, 50, 100, 200, 400, 600 and 800 Hz and a center frequency of 500 Hz. Both model predictions and data adapted from Hall *et al.* (1983) are shown in the left panel of Fig. 5.5.

The squares denote a spectral energy density of 10 dB/Hz, the upward triangles of 30 dB/Hz and the downward triangles 50 dB/Hz. At band-widths below the monaural critical bandwidth, the thresholds increase with increasing bandwidth. This is the result of the increasing amount of masker energy within the auditory filter. At a certain wider bandwidth the thresholds remain constant. Hall et al. used the bandwidth that corresponds to a threshold 3 dB below this constant threshold level as an estimate of the critical bandwidth. For the 10 dB/Hz condition, the estimate was 58 Hz, very close to the monaural estimate of 79 Hz at 500 Hz (Glasberg and Moore, 1990). At 50 dB/Hz, however, the estimate was 220 Hz, which was close to three times the monaural estimate. Our explanation for this level dependence of the critical band estimate also relies on across-frequency integration. Consider the experiment with a narrowband masker with a spectral level of 10 dB/Hz. In this case, the excitation pattern across auditory channels is very narrow

due to the low stimulus level. Off-frequency channels do not provide useful information since the stimulus level in these channels is below absolute threshold. Therefore, the bandwidth dependence of thresholds will depend only on processing of the on-frequency channel and consequently reflects the critical bandwidth of this on-frequency channel. For high stimulus levels (i.e., 50 dB/Hz) the same argument for the wider critical bandwidth can be given as described earlier.

The differences in predictions that are obtained between a model that uses across-frequency integration and a model with only single-channel processing are shown in the right panel of Fig. 5.5. Here, experimental data for a spectral level of 50 dB/Hz are shown (open symbols) combined with model simulations for a single-channel model (on-frequency channel only, filled symbols) and for the multichannel model (gray symbols). Clearly, the bandwidth dependence of the single-channel model resembles monaural behavior instead of the wider binaural bandwidth. Furthermore, the multi-channel model has lower thresholds for bandwidths up to 400 Hz. The difference is as large as 6 dB. Thus, below a bandwidth of 400 Hz, the model can improve its detection performance by integrating information across filters. At larger bandwidths, all off-frequency channels are masked and the performance for both the single and multichannel models is equal. This demonstrates that in the model, the wider critical bandwidth results from across-frequency integration.

A general observation regarding the experiments from Hall *et al.* (1983) is that the model predictions are up to 5 dB lower than the experimental data. One reason for this difference may be the fact that the model used a 2-down, 1-up procedure to vary the signal level, while in the original experiment, a 3-down, 1-up procedure was used. Hence the thresholds for the experimental data are somewhat higher than for the model predictions.

In summary, critical aspects of the model necessary to explain the wider effective critical bandwidth are the uncorrelated internal noise in each auditory filter combined with an optimal integration of information across frequency.

5.3.4   NπSo masker-bandwidth dependence

In this section, interaurally out-of-phase maskers of variable bandwidth combined with a diotic signal (i.e., NπSo) are discussed. Thresholds for different center frequencies (125, 250, 500 and 1000 Hz) and bandwidths (from 5 Hz up to twice the center frequency) were simulated. The overall masker level was kept constant at 65 dB SPL. The experimental data for the four different center frequencies (adapted from van de Par and Kohlrausch, 1999) and model simulations are shown in Fig. 5.6.
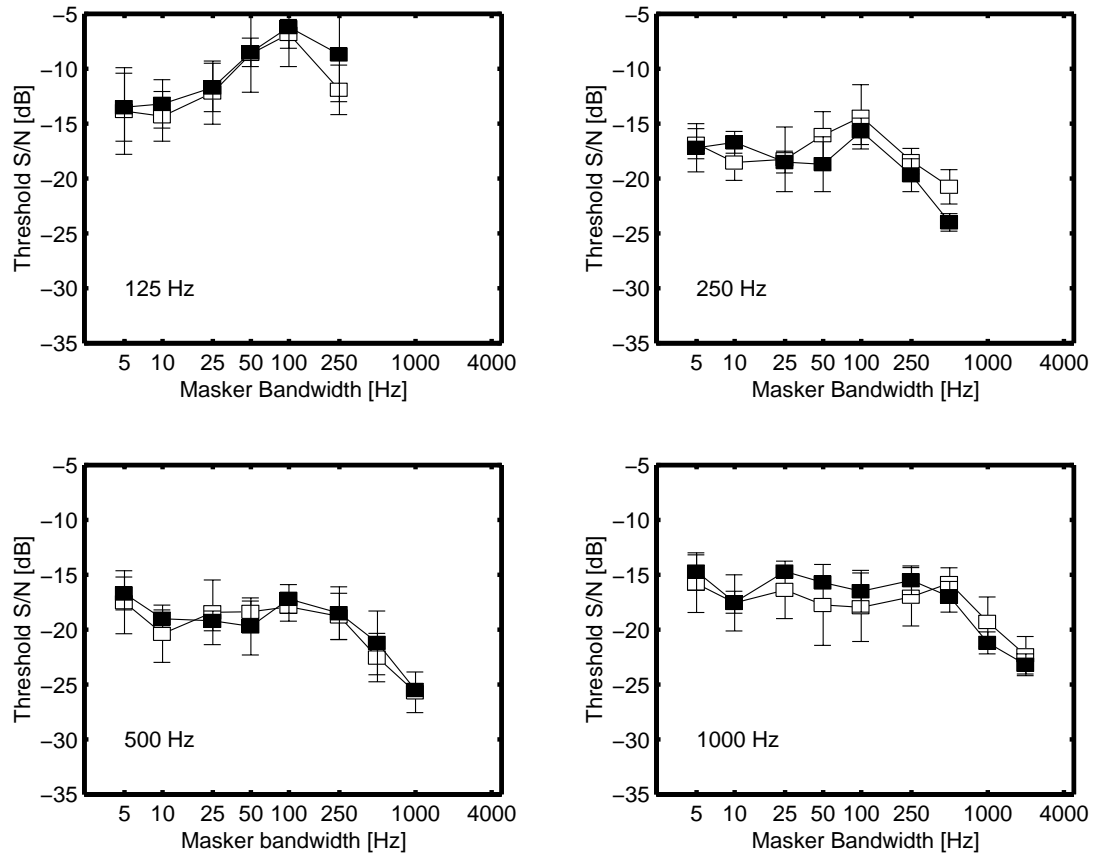
Figure 5.6: *NπSo thresholds as a function of masker bandwidth for 125 Hz (upper-left panel), 250 Hz (upper-right panel), 500 Hz (lower-left panel) and 1000 Hz center frequency (lower-right panel). The open symbols are data adapted from van de Par and Kohlrausch (1999), the filled symbols are model predictions. The masker had the same overall level at all bandwidths.*

If the NπSo thresholds are compared to the NoSπ thresholds shown in Fig. 5.3, one can see that at 125 and 250 Hz, the slope relating bandwidth to threshold at subcritical bandwidths is significantly different, while at higher center frequencies, no difference is observed. This observation holds for both model predictions and experimental data. In our model, the differences between the two experimental conditions are the result of the same model properties that were mentioned in Section 5.3.2: the fact that the interaural phase shift in the masker can only be compensated by an internal delay. Thus both a *lower* center frequency and a *wider* masker bandwidth result in more masker energy that cannot be canceled and a relative increase in the internal noise. In addition, the limited range of internal delays becomes more important for lower frequencies. This is supported by the data: the slope relating threshold to bandwidth is steeper at 125 Hz center frequency than at higher frequencies. Moreover, this slope is practically zero for the 500-Hz and the 1000-Hz condition. At these frequencies, the effect of damping of the autocorrelation function of the noise is so small that the thresholds are not influenced by this effect. Hence the thresholds approach the NoSπ
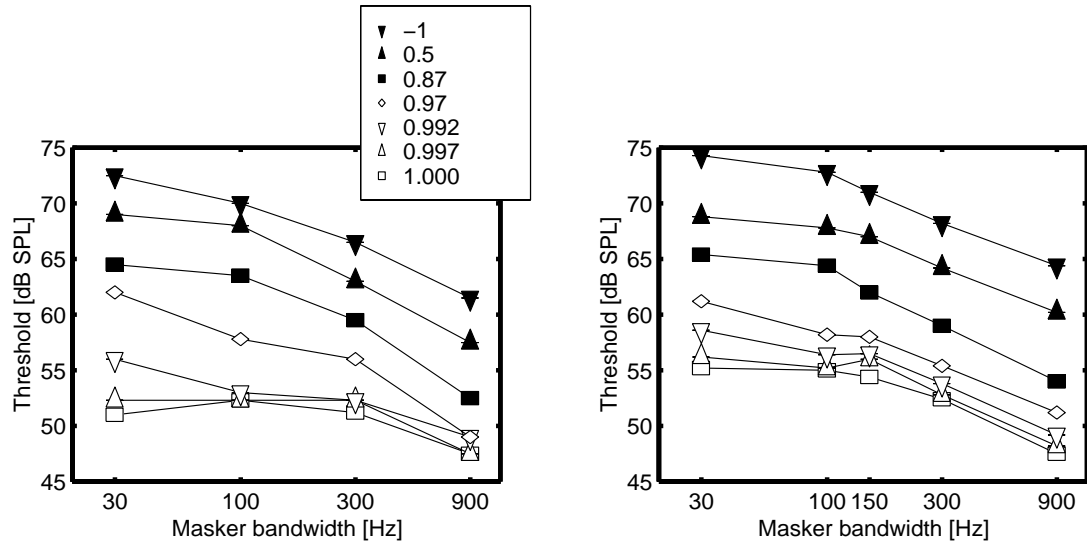
Figure 5.7: *NρSπ thresholds as a function of masker bandwidth for a constant overall level of the masker. The left panel shows experimental data adapted from van der Heijden and Trahiotis (1998); the right panel shows model predictions. Different curves denote different interaural masker correlations.*

thresholds (see Fig. 5.3). Furthermore, overall differences in thresholds occur across center frequencies, which are most clearly visible at 125 and 250 Hz.

5.3.5 NρSπ masker-bandwidth dependence

NρSπ detection thresholds were measured as a function of the bandwidth of the masker by van der Heijden and Trahiotis (1998). The overall masker level was kept constant. They used several values of interaural correlation (ρ) ranging from -1 to +1. Their results show that the wider effective critical bandwidth is only observed for interaural masker correlations > 0.97 and that for smaller correlations, the effective critical bandwidth is similar to the monaural estimate. Their results and the model simulations are shown in Fig. 5.7. The left panel shows the experimental data, the right panel shows model predictions.

The empirical curves for the largest interaural correlations (i.e., 1 and 0.997) show flat thresholds for masker bandwidths between 30 and 300 Hz, indicating a wider critical bandwidth. In these conditions, the *internal* noise limits the detection and an internal error reduction scheme is applied as described in the NoSπ bandwidening condition (see Section 5.3.3). For an interaural masker correlation of -1, however, the thresholds are similar to the monaural thresholds since no binaural advantage is present in an NπSπ condition. These monaural thresholds show a bandwidth dependence well in line with the view that the decision variable that is used in the detection process has the same statistics as the energy at the output of an auditory filter. For bandwidths below the critical bandwidth, thresholds decrease with 1.5 dB per

octave of masker bandwidth. This slope results from the sample-by-sample variability in the masker energy (Bos and de Boer, 1966). Thus, in these conditions, the *external stimulus variability* limits the detection of the tone instead of the internal noise. Since this external variability (or masker energy fluctuation) is highly correlated across auditory channels, the model *cannot reduce this variability* by combining information across frequency. The data presented in Chapter 3 suggested that a similar argument holds for $N\rho S\pi$ conditions. $N\rho S\pi$ thresholds were measured as a function of both the masker correlation and the masker bandwidth. It was found that variability of the interaural differences has a strong effect on the binaural performance and that reduction of this variability results in a decrease in thresholds. The current model supports this hypothesis in a qualitative way. For interaural correlations below 0.97, the fluctuations in the EI-element output are dominated by external stimulus fluctuations. The amount of fluctuation increases with masker correlation reduction or masker bandwidth reduction. This is also seen in the model predictions: the data decrease with both a bandwidth increase or a correlation increase. This experiment shows that our model can account both for external stimulus fluctuations and for internal errors as limitations for detection.

5.3.6 $NoS\pi$ masker-level dependence

If the across-frequency hypothesis as stated above is correct, different influences of masker level on $NoS\pi$ are expected for a narrowband masker compared to a broadband masker. For a broadband masker, off-frequency channels cannot contribute to the detection of the signal because these channels are masked by the broadband noise. On the other hand, off-frequency channels can contribute in the case of a narrowband masker as long as the stimulus level in these channels is above the absolute threshold. The number of off-frequency channels that can be used depends on the stimulus level: at a higher level the excitation pattern along the frequency axis is larger and hence more auditory filters can contribute to the reduction of the internal error. Thus, the increase in the $NoS\pi$ threshold with masker level should be shallower for a narrowband masker than for a broadband masker.

For the broadband condition, a broadband (0-8000 Hz) Gaussian noise served as masker. Its spectral energy density varied between -15 and 60 dB/Hz. The signal was a 500-Hz interaurally out-of-phase tone. The upper-left panel of Fig. 5.8 shows the experimental data (open symbols) and the model predictions (filled symbols) as a function of the noise level.

For spectral levels below 0 dB/Hz, the thresholds are approximately constant. For these conditions, the masker energy within an auditory filter is too low to influence the detectability of the signal: the threshold is determined by the absolute hearing threshold. If the spectral level of the masker is increased, the
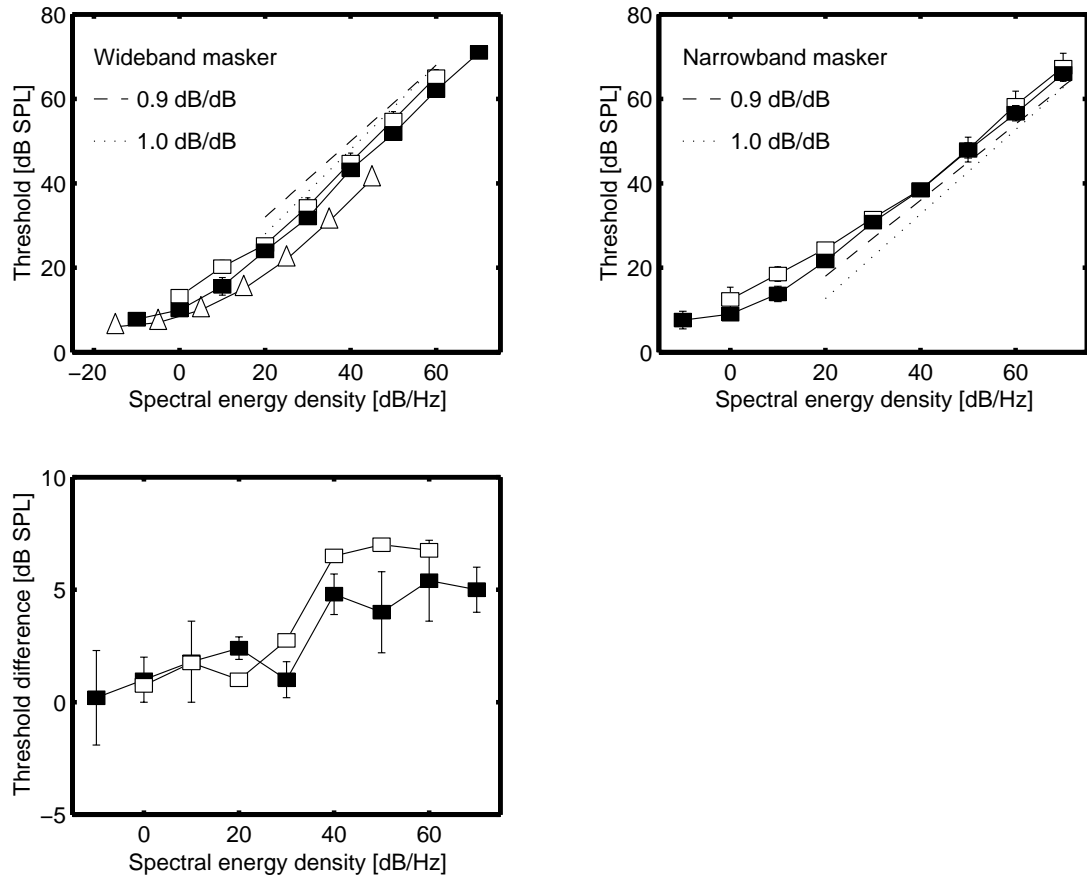
Figure 5.8: *NoSπ thresholds as a function of the spectral energy density of the masker for broadband noise (upper-left panel) and narrowband noise (upper-right panel). Open symbols are data adapted from literature: squares from Hall and Harvey (1984), upward triangles from McFadden (1968). The filled symbols are model predictions. The dashed lines have a slope of 0.9 dB/dB, the dotted lines have a slope of 1 dB/dB. The lower panel contains the difference in thresholds between narrowband and broadband conditions in the same format as the upper panels.*

amount of masker energy within the on-frequency auditory filter increases also, resulting in higher thresholds. The slope relating masker level and threshold equals 1 dB/dB (indicated by the dotted line) for masker spectral levels above about 20 dB/Hz.

The upper-right panel of Fig. 5.8 shows NoSπ thresholds for a narrowband masker (50 Hz wide) spectrally centered around the signal. The data are very similar to the data for the wideband condition except for the fact that the slope relating signal threshold to spectral noise level is lower. This is indicated by the dashed line, which has a slope of 0.9 dB/dB. As expected, the availability of off-frequency channels in the narrowband condition results in a shallower slope compared to the broadband condition where off-frequency channels do not provide useful information about the presence of the signal. The difference in thresholds between broadband and narrowband conditions is depicted in the lower panel of Fig. 5.8, for both the experimental data (open symbols) and model predictions (filled symbols).
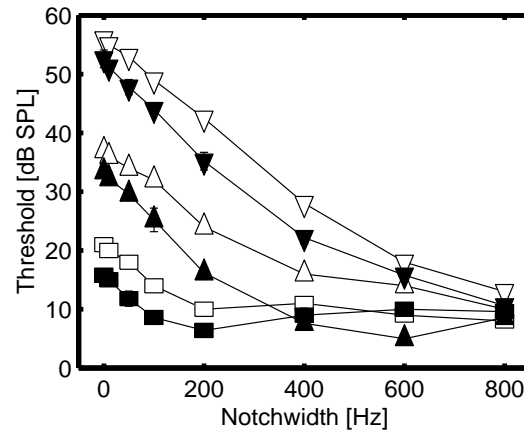
Figure 5.9: *NoSπ thresholds as a function of masker notchwidth. For each set of connected points, the spectral energy density of the masker was kept constant at 10 dB/Hz (squares), 30 dB/Hz (upward triangles) and 50 dB/Hz (downward triangles). The open symbols are data adapted from Hall* et al. *(1983), the filled symbols are model predictions.*

5.3.7 NoSπ notchwidth dependence

In the context of across-frequency integration it is interesting to consider additional conditions where the use of off-frequency channels is disabled. Such a paradigm was presented by Hall *et al.* (1983). Instead of a band-limited masker, they used a notched noise and varied the notchwidth. The out-of-phase signal was spectrally centered in the gap. This experiment was performed at three spectral noise levels, 10, 30 and 50 dB/Hz. The data (open symbols) and the model predictions (filled symbols) are shown in Fig. 5.9. The data show a decrease with increasing notchwidth until the absolute threshold is reached. Hall *et al.* (1983) used the notchwidth corresponding to a 3-dB threshold improvement compared to thresholds at a notchwidth of 0 Hz as an estimate of the critical bandwidth. These estimates for the three masker levels were close to the monaural estimate and did not depend on the masker level. This finding is also supported by the model predictions. The slope relating threshold to notchwidth is very similar and hence the 3-dB estimates of the critical bandwidth are similar too. Thus, as expected, the wider effective critical bandwidth is not observed for these stimuli. An overall difference of up to 5 dB between the model predictions and the experimental data adapted from Hall *et al.* (1983) is observed in Fig. 5.9. This difference was already discussed in Section 5.3.3, where for a different data set from the same study the same systematic difference was observed.

5.3.8 Maskers with phase transitions in the spectral domain

Another experimental paradigm to measure the spectral resolution of the binaural auditory system was used by Kohlrausch (1988). He measured the detectability of an interaurally out-of-phase signal in a masker which had a frequency-dependent phase difference: for frequencies below 500 Hz, the masker was in phase, while above 500 Hz the masker was interaurally out-
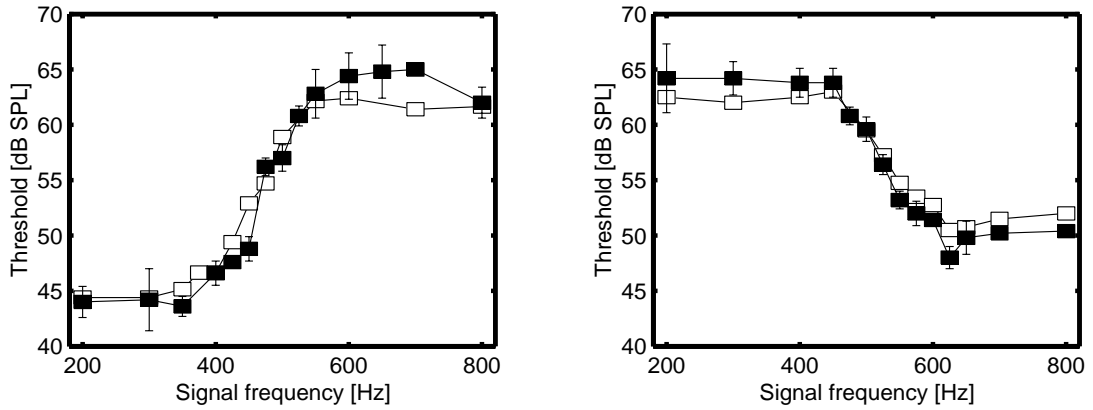
Figure 5.10: *Frequency dependence of Sπ thresholds for a masker with a frequency-dependent interaural phase difference. The left panel shows thresholds for an NoπSπ stimulus configuration (masker interaurally in-phase below 500 Hz and out-of-phase above 500 Hz), the right panel for NπoSπ (masker interaurally out-of-phase below 500 Hz and in-phase above 500 Hz). The open symbols are data adapted from Kohlrausch (1988), the filled symbols are model predictions.*

of-phase. This condition is denoted NoπSπ. The spectral energy density of the masker was 43 dB/Hz. When the signal frequency was sufficiently below 500 Hz, the effective stimulus configuration was NoSπ and a large BMLD was observed. For frequencies well above 500 Hz, the stimulus corresponded to NπSπ, and no BMLD could be measured. For frequencies near 500 Hz, a gradual increase in thresholds was observed, indicating a limited spectral resolution of the auditory system. The data combined with model predictions are shown in the left panel of Fig. 5.10. The right panel shows thresholds for an NπoSπ condition, where the masker is interaurally out-of-phase below 500 Hz and in-phase above 500 Hz.

The gradual change in the thresholds near 500 Hz results from the limited spectral resolution of the filterbank. For example, if the signal has a frequency of 200 Hz in the NoπSπ condition, the channel tuned to 200 Hz effectively contains an NoSπ condition. The cue for detection is most salient for an EI-type element with $\alpha=\tau=0$. If the signal frequency is increased, an increasing amount of the antiphasic masker energy is present in a channel tuned to the signal frequency. This part of masker energy cannot be canceled by the EI-type element. Hence the input level of the EI-type element for a masker alone increases with increasing signal frequency. Therefore, a gradual increase of the thresholds is observed near the frequency of the masker phase transition. This gradual increase reflects the monaural critical bandwidth because spectral integration cannot occur.

A modified version of such a measurement of the spectral resolution of the auditory system uses a masker that has an inverted interaural phase relationship within a passband of the masker. Basically four conditions can be used in this way which are referred to as NoπoSπ, Nπoπsπ, NoπoSo and
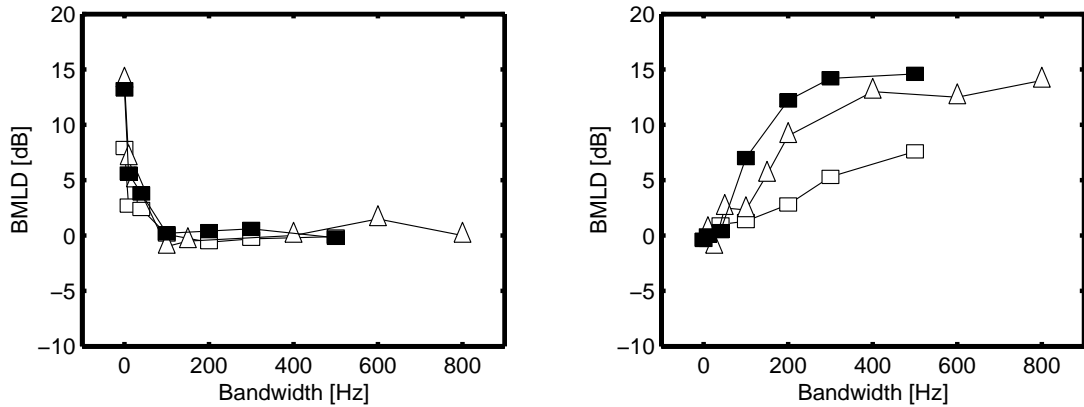
Figure 5.11: *NoπoSπ MLDs (left panel) and NπoπSπ MLDs (right panel) as a function of the bandwidth of the central band (see text for details). The open squares are experimental data adapted from Sondhi and Guttman (1966), the triangles are adapted from Holube et al. (1998). Filled symbols are model predictions.*

NπoπSo. The subscript of S denotes the interaural phase of the signal, the subscript of N denotes the frequency-dependent interaural phase relationship of the masker. Thus, an Noπo masker is interaurally in-phase except for a certain inner passband which is interaurally out-of-phase. This passband is centered around the signal. Thus, in case of NoπoSπ, the condition is effectively NπSπ if the passband is wider than the critical band, while it is NoSπ if the passband has a bandwidth of 0 Hz. By studying the bandwidth dependence of such conditions, the frequency resolution of the binaural auditory system can be estimated. Such an experiment was performed by Sondhi and Guttman (1966) and also by Holube *et al.* (1998). Their most striking result was that NoπoSπ and NπoπSπ thresholds reveal a completely different bandwidth dependence. This can be observed in Fig. 5.11. The left panel shows the BMLD as a function of the bandwidth of the inner band for the NoπoSπ condition, the right panel for NπoπSπ. The filled symbols are the model predictions, the open symbols are experimental data.

The NoπoSπ condition (left panel) has a BMLD which is large for a very small bandwidth but decreases quickly with bandwidth. This strong decrease in the BMLD is expected for the following reason. If the bandwidth of the out-of-phase passband is equal to half the equivalent rectangular bandwidth of the auditory filter, the amount of masker energy from the in-phase noise and the out-of-phase noise is approximately equal in the on-frequency filter. Consequently, the effective interaural correlation of the masker in the on-frequency filter is about zero. For such a low correlation value, the BMLD is reduced to only 3 dB. Thus, if the bandwidth of the passband is 40 Hz, the BMLD should be significantly reduced. This is supported by the data in Fig. 5.11: for this bandwidth the BMLD is only a few dB.
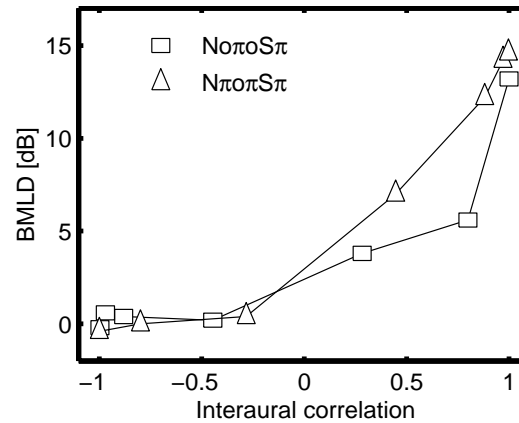
Figure 5.12: *BMLDs of the model as a function of the (computed) interaural correlation after peripheral filtering. The squares correspond to the NoπoSπ condition, the triangles to the NπoπSπ condition.*

In the NπoπSπ condition (right panel of Fig. 5.11), the bandwidth dependence of the thresholds is completely different. If the bandwidth of the passband is very large (i.e., 500 Hz), the condition is effectively NoSπ and a large BMLD is observed. If the bandwidth is decreased, the amount of masker energy in the on-frequency filter that is interaurally out-of-phase increases. Since these parts of the masker reside at the skirt of the filter, this masker energy is strongly attenuated. Consequently, the interaural masker correlation in the on-frequency channel decreases, but not very much. Only if the bandwidth is equal to about half the equivalent rectangular bandwidth[5], the BMLD is reduced to about 3 dB for the same reason as in the NoπoSπ condition. Therefore, a much more gradual decrease in the BMLD is observed if the bandwidth of the passband is decreased. This is supported by both model predictions and experimental data. This interpretation is also supported by Fig. 5.12, which shows the BMLDs of the model as a function of the computed interaural cross-correlation of the masker after peripheral filtering. The squares denote the NoπoSπ condition, the triangles the NπoπSπ condition. For both conditions, the BMLD for an interaural correlation of zero is very close to 3 dB. Furthermore, the BMLD as a function of the correlation is very similar for both conditions.

A substantial difference between the two datasets is observed in the maximum BMLD: the data of Sondhi and Guttman (1966) have a maximum BMLD of about 7 dB, while the data of Holube *et al.* (1998) show BMLDs of up to 14 dB. The reason for these differences is unclear but the model could accommodate these differences by changing the parameters a and b.

---

[5]For the filter at 500 Hz in the model, a bandwidth of 42 Hz (i.e., 0.53 times the ERB) resulted in a correlation of zero.
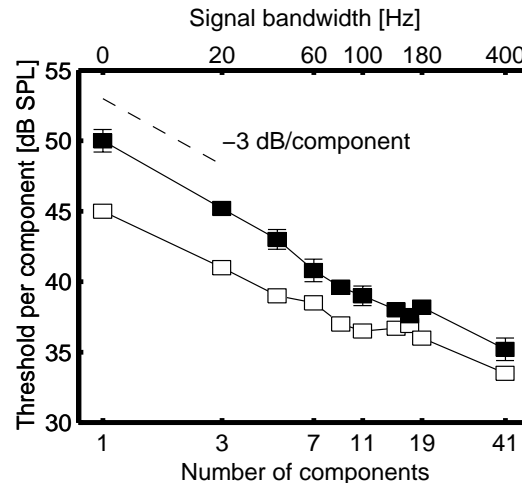
Figure 5.13: *Threshold level per component for an out-of-phase harmonic complex in an in-phase noise masker as a function of the number of signal components. The signal components had a spectral spacing of 10 Hz and were centered around 400 Hz. The open symbols are data adapted from Langhans and Kohlrausch (1992), the filled symbols are model predictions.*

5.3.9 NoSπ signal-bandwidth dependence

Langhans and Kohlrausch (1992) measured NoSπ thresholds for target signals of variable bandwidth. In this experiment, the masker consisted of a band-limited diotic noise (0-2kHz, No=47dB/Hz), while the signal consisted of harmonic complexes with a flat amplitude spectrum, a spectral spacing between the components of 10 Hz, and a center frequency of 400 Hz. The upper and lower frequency boundaries of the harmonics were varied in order to generate signals of different bandwidths. The total number of components in the complex was 1,3, 5, 7, 9, 11, 15, 17, 19 or 41. The masked thresholds of the harmonic complex tones as a function of the number of components is shown in Fig. 5.13. The thresholds are expressed as level per component.

Figure 5.13 shows a decrease of the masked threshold with increasing number of components. This can be understood as follows. If the number of components is increased from 1 to 3, the total signal level is 4.7 dB higher than the level of the individual components. The bandwidth of a signal which consists of 3 components equals 20 Hz, hence the signal has a smaller bandwidth than the auditory filter. Therefore, a threshold decrease of 4.7 dB per component is expected between 1 and 3 components in order to keep the signal power in the auditory filter constant. This is supported by the data and the model predictions. As long as the signal bandwidth is below the critical bandwidth, it is expected that an increase in the number of harmonics results in a decrease in the masked threshold level per component due to the increase in the total signal level. If the auditory filters had a rectangular shape (ideal bandpass filter), a doubling of the number of harmonics would result in a decrease of 3 dB in the level per component as long as the signal bandwidth is below the auditory filter bandwidth (9 harmonics). This is indicated by the

dashed line. Of course, a rectangular filter is not a valid description of the auditory filters. Therefore, the decrease in threshold is a little bit less than 3 dB (see Langhans and Kohlrausch, 1992, for a detailed analysis of the slope in their data). A striking result is, however, that the thresholds still decrease for more than 10 harmonics. In this case, the bandwidth of the signal exceeds the auditory filter bandwidth. To account for these results, it is necessary to include several filters in the detection process rather then only the center channel since the cue for detection is available in several filters. As described in Section 5.3.3, the availability of the cue for detection in several filters enables the improvement of the internal signal-to-noise ratio which results in lower signal thresholds for broadband signals. As can be observed from Fig. 5.13, the model accounts for this across-channel processing of binaural cues. However, the efficiency of this process in the model seems to be a little bit too high because the model predictions decrease stronger with increases in signal bandwidth than the experimental data.

5.3.10 NoSπ including spectral flanking bands

Cokely and Hall (1991) measured narrowband NoSo and NoSπ thresholds with a fixed-frequency masker (50-Hz wide centered around 500 Hz, No=50 dB/Hz) combined with an interaurally in-phase flanking noise band (30 Hz wide, No also 50 dB/Hz) of variable frequency. They found that for monaural detection (i.e., masker and signal both interaurally in phase), the flanking band had only a small effect on the masked thresholds if presented spectrally close to the signal (thresholds increased by less than 1 dB, which is expected on the basis of the increase in the masker energy in the on-frequency channel). For the NoSπ condition, however, a larger effect was observed (up to 2.5 dB), which is difficult to understand in terms of stimulus properties within the on-frequency channel. The experimental data combined with model predictions are given in Fig. 5.14, where the thresholds are shown as a function of the center frequency of the flanking band. The squares denote the monaural (NoSo) condition, the triangles refer to the binaural (NoSπ) condition. The open symbols are experimental data, the filled symbols are model predictions.

Although there is an overall difference of about 5 dB between experimental data and model predictions for the NoSπ condition, the effect of the flanking band is very similar. If the flanker has a center frequency that is close to the fixed-frequency masker, the thresholds of the model increase up to 2.4 dB. As described in Section 5.3.3, a narrowband NoSπ condition allows the reduction of the internal error in the model. The addition of an additive noise band at a higher or lower frequency results in (partial) masking of the off-frequency channels. Therefore, the efficiency of internal error reduction is decreased and an increase in thresholds is observed. In the NoSo condition, this effect is not present, because in this condition, the sample-by-sample variability of
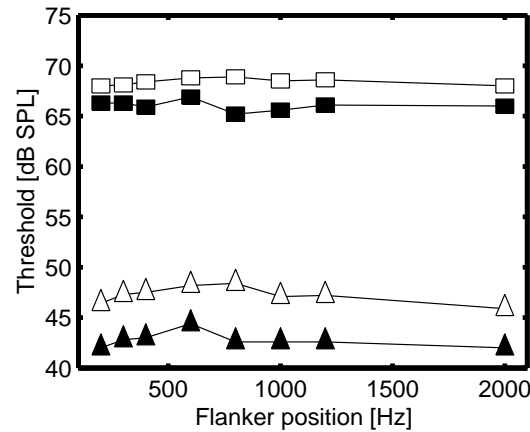
Figure 5.14: *Detection thresholds for a 500-Hz signal added to a 50-Hz-wide noise masker with a center frequency of 500 Hz as a function of the center frequency of a 30-Hz-wide flanking band. Squares correspond to thresholds for an interaurally in-phase signal, triangles to an out-of-phase signal. Open symbols are experimental data adapted from Cokely and Hall (1991), filled symbols are model predictions.*

noise energy limits the detection. As described in Section 5.3.5, this external variability is correlated across auditory channels and cannot be reduced by the model.

5.3.11 NoSπ with interaural disparities in stimulus intensity

In this experiment, performed by McFadden (1968), the masker consisted of an interaurally in-phase wideband noise with a spectral density of 45 dB/Hz; the signal was a 400-Hz interaurally out-of-phase tone. The total stimulus at one ear was attenuated by a variable amount (i.e., both masker and signal had an equal contralateral attenuation, hence the signal-to-masker ratio was the same in both ears). The results are shown in the left panel of Fig. 5.15. The abscissa shows the disparity in interaural stimulus intensity, the ordinate shows the signal threshold in terms of the level at the ear without attenuation.

Clearly, for a contralateral attenuation between 0 and 10 dB, the thresholds remain constant. In this region, the model can fully compensate for the externally presented IID. If the external IID is increased, however, the thresholds increase. This is the result of the nonlinear preprocessing stage. As described Chapter 4, the adaptation loops which are present in the peripheral processor are highly nonlinear and show a compressive behavior. Therefore, the externally presented IID is reduced and the EI-type element that optimally compensates for the external IID has a characteristic IID that is much smaller than the external IID. Despite the availability of EI-type elements that can compensate for the *mean* level difference at the output of the adaptation loops, these cells cannot cancel the masker noise completely due to the nonlinear processing which results in different waveforms at the output of the adaptation loops from the left and right sides. Hence parts of
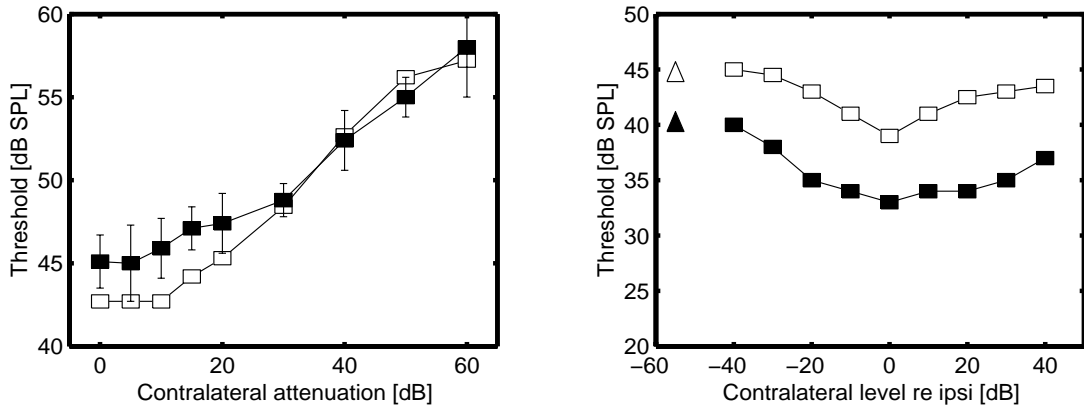
Figure 5.15: *Left panel: NoSπ thresholds as a function of the attenuation of the stimuli at one ear. The thresholds are expressed in terms of the level at the non-attenuated ear. The open symbols are data adapted from McFadden (1968), the filled symbols are model predictions. Right panel: NoSm thresholds as a function of the noise level in the non-signal ear relative to the signal ear. The open symbols are adapted from Weston and Miller (1965), the filled symbols are model predictions. The squares denote the NoSm conditions, the triangles are the monaural NmSm reference conditions.*

the noise masker are present in the output of the EI-type element, resulting in increasing thresholds. Moreover, for an IID of 60 dB, the attenuated signal lies below the absolute threshold; the model cannot cancel any part of the masking noise. Hence the thresholds are determined by the monaural masker level in the non-attenuated ear.

A similar experiment was performed by Weston and Miller (1965). They measured NoSm detection thresholds at 500 Hz as a function of the noise level in the non-signal ear. The noise level in the signal ear was 26 dB/Hz. Besides attenuating, Weston and Miller (1965) also increased the noise level in the non-signal ear. Their results are shown in the right panel of Fig. 5.15 as a function of the relative noise level in the non-signal ear compared to the signal ear. The open symbols denote their experimental data, the filled symbols are model predictions. The triangles refer to the monaural reference conditions. Interestingly, both for a decrease and an *increase* in the contralateral noise level, thresholds increase. This effect is also captured by the model. Both curves show an increase of 3 to 4 dB if the contralateral level is increased by 40 dB. In the model, this increase is caused by the nonlinear processing of the peripheral adaptation loops as described above.

The predictions and experimental data for NoSm differ by about 5 to 6 dB, the latter being higher. The maximum BMLD found by Weston and Miller (1965) is about 5.5 dB, while the model predicts a maximum BMLD of about 7 dB. These values are in line with other experimental data, showing BMLDs between 5 and 10 dB for NoSm (Hirsh, 1948b; Hirsh and Burgeat, 1958; Kohlrausch, 1988).
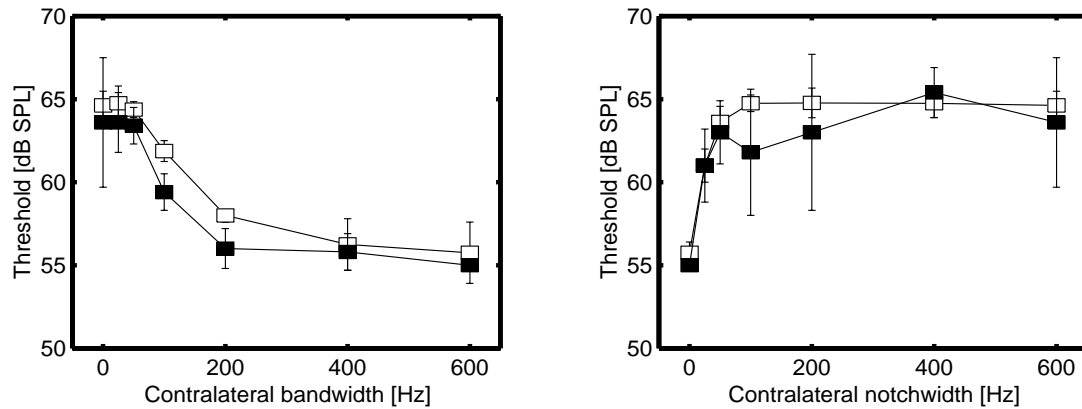
Figure 5.16: *NoSm thresholds as a function of the bandwidth in the non-signal ear (left panel) and as a function of the notchwidth in the non-signal ear (right panel). The open symbols are data adapted from Hall and Fernandes (1984), the filled symbols are model predictions.*

5.3.12 NoSm as a function of the notchwidth and bandwidth in the non-signal ear

Hall and Fernandes (1984) measured NoSm detection thresholds for stimuli with a variable masker bandwidth or notchwidth in the non-signal ear. A 500-Hz pure-tone signal was presented with a 600-Hz-wide band of noise to the signal ear. Bands of noise ranging in width from 25 to 600 Hz, or notched noises (bandwidth also 600 Hz) ranging in notchwidth from 0 to 600 Hz were presented to the nonsignal ear. The effects of varying the bandwidth were different from those of varying the notchwidth. If the bandwidth was varied, the thresholds decreased over a range of 400 Hz, while for the notched experiment, significant threshold changes only occurred for notchwidths between 0 and 50 Hz. These results are shown in Fig. 5.16. The left panel corresponds to thresholds as a function of the bandwidth of the masker in the non-signal ear, the right panel to the notchwidth.

Both model and experimental data show a different behavior as a function of bandwidth for a notched or band-limited masker. This can be understood as follows. If a narrowband masker is presented in the non-signal ear, the masker cannot be canceled by a simple subtraction, since the masker waveforms in both ears are completely different. In fact, for a bandwidth of 0 Hz, the stimulus corresponds to NmSm (i.e., masker and signal are presented to one ear only), and no binaural advantage can be achieved. If the bandwidth in the non-signal ear is increased, the similarity (or cross-correlation) between the maskers at both ears increases which enables the model to cancel the masker more efficiently. This efficiency increases even beyond the critical bandwidth. As a result of this, the signal thresholds decrease with increasing bandwidth, as observed in the data. For the notched noise, a similar process occurs. In the absence of a notch (i.e., a notchwidth of 0 Hz), the masker can be canceled completely and a large binaural advantage is observed. With increasing

notchwidth, the cancellation of the masker is less successful and thresholds increase. The essential reason for the different slopes relating threshold to bandwidth or notchwidth lies in the fact that the masker cancellation is performed after peripheral filtering.

As described above, the non-signal ear contains only *part* of the masker of the signal ear; some spectral components are removed. If more spectral components are removed, the amount of masker energy that remains after cancellation increases and hence the signal thresholds increase. Thus, the amount of masker energy *after peripheral filtering* that is present in the signal ear but *not* in the non-signal ear determines the detection threshold. If the notchwidth is increased from 0 to 25 Hz, there exists a large difference between the maskers at both ears since this part of the masker is in the center of the auditory filter. Therefore, an increase in the notchwidth has a strong effect on the thresholds. Moreover, for notchwidths beyond 50 Hz, the thresholds are equal to the monaural thresholds indicating that the binaural system cannot increase the detection performance.

In the band-limited case, however, this process is reversed. For a bandwidth of 600 Hz, the maskers in both ears are equal and a large binaural advantage is observed. If the bandwidth is decreased to 400 Hz, the thresholds do not increase since the parts of the masker that are removed in the non-signal ear are filtered out by the bandpass filter of the inner ear. Only at a bandwidth of 300 Hz do the maskers at both ears become somewhat decorrelated after peripheral filtering and the thresholds show a slight increase. For smaller bandwidths, the correlation between the maskers after filtering decreases and thresholds show a gradual increase. Since the part of the masker that is removed in the non-signal ear resides in the filter skirt, a reduction in the bandwidth has less effect than an increase in the notchwidth at the center of the filter.

5.4   Conclusions

Predictions for binaural detection performance were shown as a function of the spectral parameters and interaural phase relationships of both maskers and signals. Although some overall differences exist between the model predictions and the experimental data, most of the experimental results can be accounted for with the present model. All stages that are included in the model contribute in some way to the exactness of the predictions. For example, the loss of phase-locking in the inner haircells at frequencies above 1 kHz is necessary to account for the ITD-thresholds shown in Fig. 5.1 and the increase of binaural thresholds above 1 kHz (see Fig. 5.2). For some simulations, almost all stages have to be taken into account to obtain a good fit. Examples are the experiments discussed in Section 5.3.4 (i.e., NπSo as a function of the bandwidth and center frequency). The combined effect

of peripheral filtering, compressive behavior of the EI-type elements, the distribution of internal delays and the chain of adaptation loops in the peripheral preprocessing stage results in good predictions, which cannot be achieved if any of these elements is removed. The price one has to pay is a more complex model than those used and described so far.

In summary, the current model accounts for many binaural detection phenomena in a quantitative way. These include:

- The wider effective critical bandwidth in bandwidening $NoS\pi$ conditions.

- The unification of IID and ITD sensitivity with binaural detection data.

- The level dependence of binaural thresholds.

- The frequency dependence of binaural detection thresholds.

- The effect of frequency and bandwidth on the difference between $NoS\pi$ and $N\pi So$ thresholds.

- The combination of both external stimulus fluctuations and internal errors into one decision variable.

- The influence of interaural level differences on binaural thresholds.

'The most exciting phrase to hear in science, the one that heralds new discoveries,
is not "Eureka!" but "That's funny ..."'
Isaac Asimov.

Predictions as a function of temporal stimulus parameters[1]

Simulations of binaural masking experiments were performed as a function of temporal stimulus parameters and compared to psychophysical data adapted from literature. For this purpose, the model as described in Chapter 4 was used as an artificial observer in a 3-interval, forced-choice procedure. All model parameters were kept constant for all simulations. Model predictions were obtained as a function of the interaural correlation of a masking noise and as a function of both masker and signal duration. Furthermore, maskers with a time-varying interaural correlation were used. Predictions were also obtained for stimuli with time-varying interaural time or intensity differences. Finally, binaural forward masking conditions were simulated. The results show that the combination of a temporal integrator followed by an optimal detector in the time domain can account for all conditions that were tested, except for those using periodically-varying ITDs and those measuring interaural correlation jnds as a function of bandwidth.

6.1 Introduction

This chapter describes ability of the model to predict binaural detection thresholds as a function of temporal stimulus properties. The model consists of three stages. The first stage simulates the effective signal processing of the basilar membrane and the inner haircells and includes adaptation by means of adaptation loops (Dau *et al.*, 1996a). Binaural interaction is modeled in the second stage by means of a contralateral inhibition mechanism: the model computes the squared difference signal between the left and right ears as a function of time, frequency channel, internal interaural delay and internal interaural level adjustment. These binaural signals are corrupted by internal noise and subsequently analyzed by the third stage in the model, the central processor. The model is used as an artificial observer in a 3-interval, forced-choice procedure, and the central processor matches the representations of the presented stimuli to templates (derived during previous presentations) and on this basis the model indicates which interval contains the signal.

---

[1]This chapter is based on Breebaart, van de Par, and Kohlrausch (2001c).

In Chapter 5 of this thesis, model predictions for binaural detection were discussed as a function of the spectral parameters of the stimuli, keeping the temporal parameters constant. All stimuli had a duration which was long compared to the temporal resolution of both the monaural and binaural stages of the model (i.e., 200 ms or longer). We demonstrated that the model is very successful in describing the threshold dependence on spectral stimulus parameters and that this success can for a large extent be attributed to an optimal combination of information across auditory channels.

In the current chapter, we focus on the temporal properties of the stimuli, keeping the spectral parameters constant. Two important temporal properties are studied intensively. The first concerns temporal *integration*. It has been shown that the binaural system is able to integrate binaural cues temporally if such a process enhances a detection task. For example, an increase of the signal duration in an $NoS\pi$ condition results in a decrease of the signal threshold for signal durations up to 300 ms (Zwicker and Zwicker, 1984; Wilson and Fowler, 1986; Wilson and Fugleberg, 1987; Bernstein and Trahiotis, 1999). The second property is related to the temporal *resolution* of the binaural auditory system. Several studies have revealed that the auditory system is sluggish in its processing of interaural differences. For example, the minimum audible angle of a sound source strongly depends on its velocity (Perrott and Musicant, 1977). Experiments using time-varying IIDs revealed that IID detection shows a lowpass behavior with a cutoff frequency of about 20 Hz (Grantham, 1984a). The detection of dynamic ITDs seems to be even worse; Grantham and Wightman (1978) showed that ITD detection has a lowpass response with a cutoff frequency of 2 to 5 Hz. Detection experiments performed with a masker which has a time-varying interaural correlation show that the binaural auditory system has a time constant between 44 and 243 ms (Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999), which is rather high compared to the 4 to 44 ms for monaural processing (Kollmeier and Gilkey, 1990; Plack and Moore, 1990). The aim of the current study is to demonstrate that the model presented in Chapter 4 can also account for these temporal phenomena.

6.2 Method

6.2.1 Relevant stages of the model

In the introduction, a coarse description of the general model setup was given. In this section, the stages of the model that are relevant for the simulations described in this chapter (i.e., temporal behavior) are discussed in more detail. For a detailed description of the complete model, see Chapter 4 of this thesis.

- Filtering of the gammatone filterbank. The filterbank present in the peripheral processing stage determines the spectral resolution of the model, in line with the ERB estimates published by Glasberg and Moore (1990). Due to the limited bandwidth of the filters in the gammatone filterbank, ringing occurs which influences forward-masking thresholds for very short signal delays.

- A chain of 5 adaptation loops is included in the peripheral preprocessor. These adaptation loops limit the detectability of short low-level signals presented shortly after the offset of a high-level masker. Due to this limitation, the monaural detection model by Dau *et al.* (1996a) has been successful in predicting detection performance in monaural nonsimultaneous masking conditions (Dau *et al.*, 1996b, 1997). Because the current model includes the same stages as the model by Dau *et al.* (1996a), this predictive scope is inherited by our model. Furthermore, because the binaural interaction follows the peripheral adaptation (cf. Kohlrausch and Fassel, 1997), also binaural forward masking will be limited in its steepness through the presence of the adaptation loops.

- Central temporal window. In the binaural processor, EI-type elements calculate the squared-difference signal between the outputs of the peripheral processor for each auditory filter. These difference signals are convolved with a double-sided exponential window with an equivalent rectangular duration (ERD) of 60 ms to account for a limited binaural temporal resolution. Because this window operates on the difference signal, the same window is used to analyze IIDs and ITDs as well as binaural detection data.

- Compressive input-output characteristic of EI-type elements. The temporally-smoothed difference signal of the EI-type elements is compressed logarithmically. In combination with an additive internal noise, this stage results in thresholds of interaural differences that depend on the interaural cross-correlation of the reference stimuli.

- Optimal detector in the central processor. The EI-type element outputs are corrupted by an additive internal noise. Subsequently, the internal representations of the external stimuli are compared to a template that consists of the average masker-alone representation. The differences between the actual stimulus and the template are weighted and integrated both in the time and the frequency domain according to an optimal criterion. This enables the optimal detector to reduce the influence of the internal noise, and to accumulate information about the signal by adapting its observation interval (matched temporal integrator).

6.2.2 Procedure and stimuli

The procedure, the method of generation of stimuli and the model calibration were the same as those described in Chapter 5. In particular, all model parameters were kept fixed for all simulations described in this chapter and were the same as in the previous chapter. In all simulations, the duration, level, on- and offset ramps, bandwidth and onset delay of both the maskers and signals equaled the values used in the experiments with human subjects. If more data sets from various authors with different experimental settings were used, the experimental settings from one of these studies were used for determining the model simulations. Comparison with the other datasets was possible because in such conditions, we either calculated BMLDs or normalized the thresholds with the spectral level of the masking noise.

6.3 Simulations

6.3.1 $N\rho S\pi$ and $N\rho Sm$ correlation dependence for wideband noise

This section deals with the detection of a signal in the presence of a masker with various (fixed) values of the interaural correlation (ρ). Similar stimuli were discussed in Chapter 5 focussing mainly on the bandwidth dependence of the masker. In this section, the non-stationary behavior of interaural differences in such conditions will be discussed. By 'non-stationary' we mean that the expected values of the statistical properties, such as the interaural correlation, are constant, but that these properties evaluated on a short-time basis change as a function of time within each interval. It is therefore valuable to discuss $N\rho S\pi$ thresholds in the current chapter, since these experiments reflect the detection of a change in the *distribution* of interaural differences rather than the detection of the *presence* of interaural cues.

Robinson and Jeffress (1963) measured thresholds for a wideband $N\rho S\pi$ condition. They used a 150-ms, 500-Hz tone as signal presented in a 150-ms noise masker. The masker had a spectral level of 50 dB/Hz. Their data are shown in the left panel of Fig. 6.1 (open symbols), together with the model predictions (filled symbols). The binaural masking level difference (BMLD) is shown as a function of the interaural masker correlation. For a correlation of +1, the condition corresponds to $NoS\pi$ and a large BMLD is observed, which decreases with decreasing correlation. If the masker is interaurally uncorrelated (ρ=0), a BMLD of only 2 to 3 dB is present, which completely disappears if the correlation is decreased further towards -1.

The right panel shows data for a monaural signal (i.e., $N\rho Sm$), adapted from Wilbanks and Whitmore (1968). In this experiment, a 200-ms signal was used, also with a frequency of 500 Hz. The spectral masker level was 33 dB/Hz. The data show a similar curve as the $N\rho S\pi$ condition, with two important differences. First, the BMLD at ρ=+1 is 6 dB smaller for the $N\rho Sm$ condition.
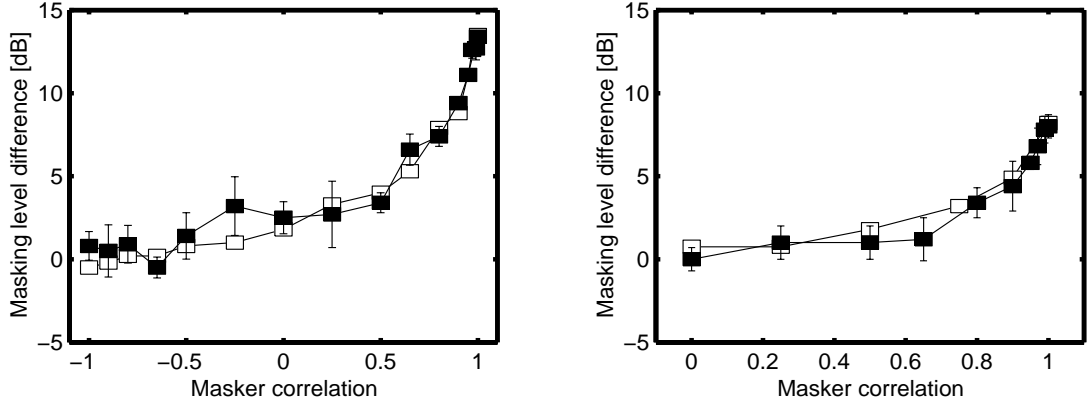
Figure 6.1: *NρSπ (left panel) and NρSm (right panel) BMLDs as a function of the interaural masker correlation. The white symbols are experimental data adapted from Robinson and Jeffress (1963) for NρSπ and from Wilbanks and Whitmore (1968) for NρSm. The black symbols are model predictions.*

Second, almost no BMLD is observed in the NρSm condition with ρ=0, while the NρSπ condition still shows a BMLD of a few dB at this masker correlation.

To understand why the BMLDs decrease with a decrease in the masker correlation, it is useful to first have a closer look at the way these partially correlated maskers are generated. Usually, the Nρ masker is obtained by combining the waveforms of two or three independent noise sources. We will focus on the method using 2 noise sources, keeping in mind that the method using three noise sources is in principle similar (cf. van der Heijden and Trahiotis, 1997). If two independent noise sources, having time-domain waveforms given by $N_1(t)$ and $N_2(t)$ are used to generate a noise with an interaural correlation of ρ, the left and right channels $L(t)$ and $R(t)$ consist of the following linear combination of these noises:

$$\begin{cases} L(t) = \frac{1}{2}\sqrt{2}\sqrt{1+\rho}N_1(t) + \frac{1}{2}\sqrt{2}\sqrt{1-\rho}N_2(t) \\ R(t) = \frac{1}{2}\sqrt{2}\sqrt{1+\rho}N_1(t) - \frac{1}{2}\sqrt{2}\sqrt{1-\rho}N_2(t) \end{cases} . \qquad (6.1)$$

If such a stimulus is presented to the model and if we neglect the processing of the peripheral preprocessor, the waveforms $L(t)$ and $R(t)$ enter an EI-type element that optimally cancels the masker (no internal delay or level adjustment, i.e., τ=α=0). In our implementation (see Chapter 4), the output (E) of the EI-type element is then given by

$$E(t) = (L(t) - R(t))^2 . \qquad (6.2)$$

Substitution of Eq. 6.1 into Eq. 6.2 results in

$$E(t) = (2 - 2\rho)N_2{}^2(t). \qquad (6.3)$$

In a very similar way, it can be shown that the addition of an interaurally out-of-phase signal ($S_\pi(t)$) to the same masker results in

$$E(t) = (2 - 2\rho)N_2{}^2(t) + 4S_\pi^2(t). \qquad (6.4)$$
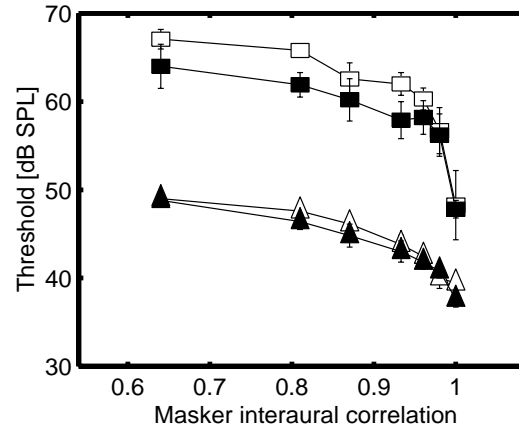
Figure 6.2: *Running-noise NρSπ thresholds for a masker bandwidth of 10 Hz (squares) and 1000 Hz (triangles). The overall masker level was 65 dB SPL for both bandwidths. The black symbols are model predictions, the white symbols are data adapted from Chapter 3.*

The processes in the model that follow the EI-type element basically consist of time averaging resulting in a running average E of the energy of $(2 - 2\rho)N_2(t)$, followed by the logarithmic input-output relation of the EI-type elements. As can be observed from Eq. 6.3, a decreasing interaural correlation ρ results in an increasing amount of masker energy that cannot be canceled by the EI-type elements. As described in Chapter 4, this results in higher signal thresholds due to the logarithmic input-output behavior. The above explanation also holds for the NρSm condition, except for the fact that the amount of signal energy in Eq. 6.4 is decreased by a factor of 4. Thus, to achieve a similar change in E as for an Sπ signal, the signal level must be increased by 6 dB, an effect that is clearly found in the data for $\rho > 0.7$. For lower correlation values, the signal level in the NρSm condition approaches the monaural threshold and hence thresholds remain constant if the correlation is reduced further.

6.3.2 NρSπ thresholds for narrowband noise

In Chapter 3, NρSπ thresholds were measured as a function of both the correlation and the masker bandwidth. The masker duration was 300 ms. A 500-Hz sinusoid with a duration of 200 ms was used as the signal. The results showed that for a narrowband masker with a bandwidth of 10 Hz, thresholds varied more with the interaural correlation than for a wideband masker. The narrowband masker resulted in a much steeper curve for correlations between 0.8 and +1 compared to the wideband case. This is depicted in Fig. 6.2. Both the 10-Hz-wide (squares) and the 1000-Hz-wide (triangles) data are shown as a function of the masker correlation. The black symbols denote the model predictions, the white symbols are experimental data. The separation between the narrowband and the broadband data is due to the choice of a constant overall masker level of 65 dB SPL which gives a higher spectral level for the 10-Hz-wide masker.

In Chapter 3 it was argued that the differences between the 10-Hz curve and the 1000-Hz curve are due to the fact that two different factors limit the detection process: internal errors and external variability. For the wideband masker, the thresholds are determined by the internal errors. In our model, the specific relation between threshold and interaural correlation results from the logarithmic input-output curve present in the EI-type elements which was explained in Section 6.2.1.

For the narrowband condition, other factors are important. The temporal sluggishness filter effectively calculates a running average of the output of Eq. 6.3 if a masker alone is present. This output increases if the signal is present. Because this output serves as a decision variable, the model must look for fluctuations in this variable that are attributable to the addition of the signal. Since the squared waveform of $N_2(t)$ is present in Eq. 6.3, a running average of the energy of $N_2(t)$ is obtained, multiplied with a scalar which depends on the interaural correlation. If the bandwidth of N_2 is very small, the energy estimate E shows large fluctuations due to the limited number of degrees of freedom in the noise. If the standard deviation of these fluctuations is larger than the *change* in E due to the addition of the signal with a certain level, it is very unlikely that the model is able to detect the signal because of the large amount of *stimulus variability*. This is exactly what happens in the narrowband $N\rho S\pi$ condition. Instead of being limited by internal errors, the thresholds are limited by external stimulus fluctuations if $\rho < 1$. Since the amount of energy E increases with decreasing correlation (as can be observed from Eq. 6.3), the amount of fluctuations in E also increases with decreasing correlation, hence resulting in increasing thresholds. This property is not altered by the logarithmic input-output function of the EI-type elements, since both the fluctuations and the change in the output due to the signal are transformed by the same process. Thus for the narrowband condition with ρ at or below 0.98, thresholds are relatively high due to the 'external' variability in E. Only when $\rho = 1$ does external variability play no role, and thresholds are only determined by the internal error and are therefore relatively low.

6.3.3 Interaural cross-correlation discrimination

Gabriel and Colburn (1981) measured just-noticeable differences (jnds) in interaural cross-correlation from two reference correlations (0 and +1) at several bandwidths. The total noise level was kept constant at 75 dB SPL and the stimuli were spectrally centered at 500 Hz. At a reference correlation of +1, their results indicated that for bandwidths less than 115 Hz the correlation jnd was equal to about 0.004, while for larger bandwidths, the jnd increased monotonically with the noise bandwidth. On the other hand, at a reference correlation of 0, the jnd *decreased* with increasing bandwidth, having a value of 0.7 for narrowband stimuli (3 Hz) and 0.3 for the broadband case (4500 Hz).
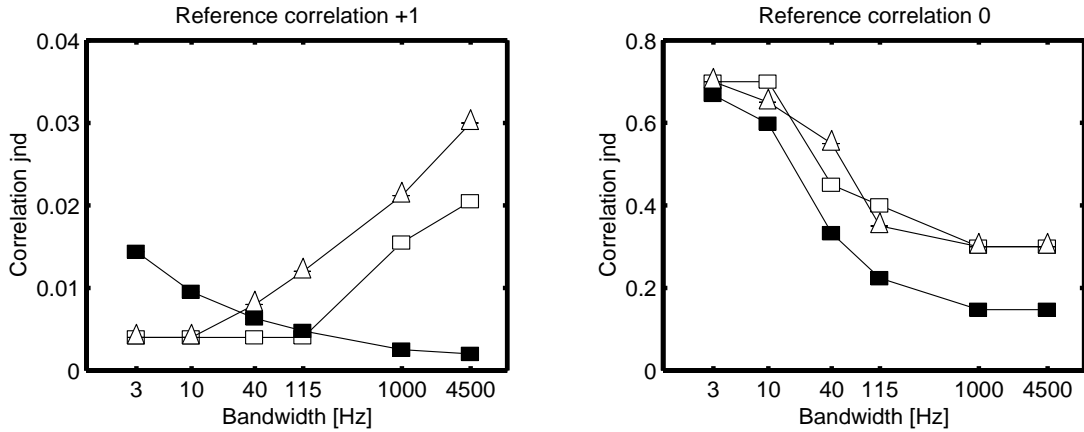
Figure 6.3: *Just noticeable differences in the interaural correlation at a reference correlation of +1 (left panel) and 0 (right panel) as a function of the bandwidth of the stimulus. The white symbols are experimental data adapted from Gabriel and Colburn (1981) for two different subjects. The filled symbols are model predictions.*

Their results are summarized in Fig. 6.3. The left panel shows the correlation jnds at a reference correlation of +1, the right panel at a reference correlation of 0. The white symbols are the experimental data for different subjects, the black symbols denote model predictions.

The model predictions for a reference correlation of +1 (left panel) show a completely different behavior to the experimental data: the experimental data increase with increasing bandwidth, while the model predictions show a monotonic decrease with increasing bandwidth. Only for the data at 40 and 115 Hz is there a close resemblence between model predictions and experimental data. The decrease in correlation jnd with increasing bandwidth for the model can be explained as follows. For all bandwidths, the reference condition has a correlation of +1. Consequently, the stimulus can be eliminated completely by the model. Thus, the reference intervals result in no internal activity for EI-type elements that are optimally tuned to this detection task. If the interaural correlation of the stimulus is reduced, the masker cannot be canceled completely which results in some activity in the model. If the stimulus bandwidth is small (i.e., 3 to 10 Hz), this cue for detection is available in the on-frequency filter and in some adjacent filters due to spread of excitation. If the bandwidth is increased, the number of auditory filters that contains information about the change in the correlation increases since the change in the interaural correlation occurs at the complete spectrum of the stimulus. This enables the model to integrate information across auditory filters resulting in an effective reduction of the internal noise. This in turn results in smaller thresholds, as observed in the model predictions. This effect is not observed in the experimental data, however.

The right panel of Fig. 6.3 shows data for the reference correlation of 0. Both model predictions and experimental data show a decrease in the correlation

jnd with increasing bandwidth, although the model is more sensitive to changes in the correlation at bandwidths beyond 40 Hz. This decrease in the correlation jnd can be explained by considering stimulus uncertainty. If the bandwidth is very small (3 Hz), the fluctuations in the output of the EI-type elements are very large (see Section 6.3.2). Increases in the bandwidth result in more degrees of freedom in the masker and hence less uncertainty in the output of the EI-elements. This decreased uncertainty is reflected by the decrease in the correlation jnd. For bandwidths beyond 115 Hz, increases in the bandwidth have almost no effect on the stimulus uncertainty because these parts of the stimulus fall outside the auditory filter.

6.3.4 NoSπ signal duration

An important property that has a very strong effect on thresholds in an NoSπ condition is the duration of the signal. The threshold behavior in this experimental paradigm basically reflects the ability of the binaural auditory system to *integrate* information over time. Several studies showed an increase in detection performance of up to 25 dB if the signal duration is increased from 2 to 250 ms (cf. Yost, 1985; Wilson and Fowler, 1986; Wilson and Fugleberg, 1987; Bernstein and Trahiotis, 1999). The results of these studies are shown in Fig. 6.4 (white symbols) for a center frequency of 500 Hz (left panel) and for 4 kHz (right panel). Since in these studies, different noise levels were used ranging from 26.3 to 47 dB/Hz, we expressed thresholds as the ratio between signal level and spectral masker density. The thresholds decrease with a slope of 4.5 dB/oct for durations up to about 60 ms, while for longer durations, this slope is shallower (1.5 dB/oct). The model predictions are shown by the black symbols in Fig. 6.4. They were derived for a 500-ms wideband noise masker with a spectral energy density of 36.2 dB/Hz (similar to Wilson and Fugleberg, 1987).

The model predictions show a very similar signal-duration dependence to the experimental data. These results can be explained as follows. First consider a signal of very short duration (2 ms). In this case, the signal interval contains interaural differences which are present within a short period of time compared to the duration of the binaural window (about 60 ms). As described in Chapter 4, the output of each EI-type element is averaged over time by a temporal integrator. Consequently, the cue for detection at the *averaged* output of such a temporal window is strongly reduced for a very short signal. Therefore, the signal must have a relatively high level to elicit a change in the averaged output that can be detected by the model. If the duration of the signal is increased but does not exceed the duration of the temporal window, the average output of the temporal window increases and hence the signal level decreases at threshold. Since the output of the temporal averager is proportional to the signal energy, this process accounts for a
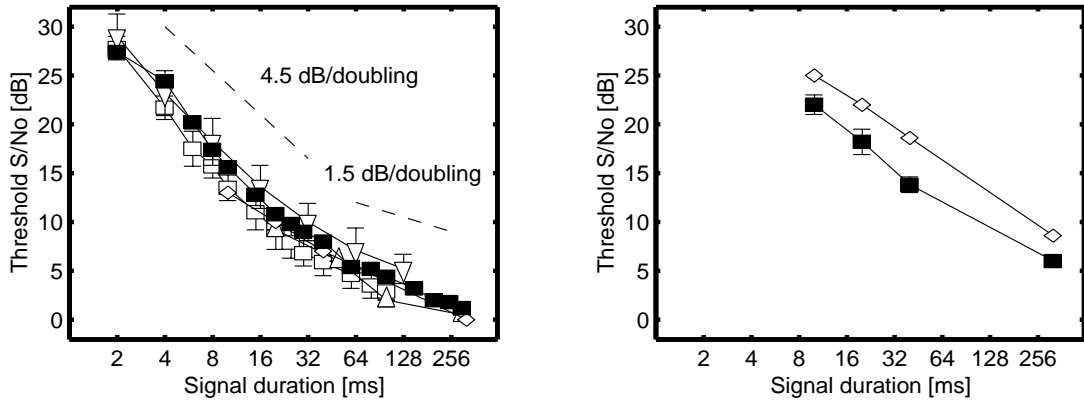
Figure 6.4: *NoSπ thresholds as a function of the signal duration. Data in the left panel are for a signal frequency of 500 Hz, in the right panel for a signal of 4 kHz. The white squares are data adapted from Wilson and Fugleberg (1987), the upward triangles from Yost (1985), the downward triangles from Wilson and Fowler (1986) and the diamonds from Bernstein and Trahiotis (1999). The black symbols are model predictions. Thresholds are given as signal level re masker spectrum level.*

decrease of 3 dB for each doubling of the of signal duration, as long as the signal duration does not exceed the time constant of the temporal averager. For durations exceeding 60 ms, this process does not influence the detection process.

An additional effect of 1.5 dB per duration doubling results from the reduction of the internal error with increasing signal duration by the optimal detector. A longer signal duration means that the average output is available for a longer time. This enables the model to reduce the internal error in the decision variable in a similar way as was described for spectral integration of information (see Chapter 4). After the temporal averager, an additive noise is combined with the output of the EI-type elements, followed by an optimal detector. If, after the temporal integrator, the cue for detection is available for a long time, the optimal detector can decrease the amount of noise in its decision variable by integrating the EI-element output over the total signal duration. In this way, a doubling in the signal duration results in a doubling in the overall difference in output between masker and masker-plus-signal, while the amount of noise increases with the square-root of 2. Hence the detectability of the signal is increased, which results in a lower threshold. Thresholds are expected to decrease with 1.5 dB per doubling of signal duration. Thus, the combined effect of the processes described above accounts for the 4.5 dB per doubling for signal durations below 60 ms and 1.5 dB beyond 60 ms. The slope of 4.5 dB/doubling for signal durations below 60 ms and 1.5 dB/doubling beyond 60 ms is also present in the model simulations at 4 kHz (right panel of Fig. 6.4). Except for an overall difference of about 4 dB, these model predictions are very close to the experimental data, although the difference in slope below and beyond 60 ms cannot be deduced from the experimental data.

6.3.5 NoSπ masker duration

Besides reports on the duration of the signal, several studies have been published on the effect of the duration of the masker. Basically three configurations have been tested. The first uses a forward noise fringe of variable length (Robinson and Trahiotis, 1972; Zwicker and Zwicker, 1984; Yost, 1985), while in the second configuration, a backward noise fringe is used (Trahiotis *et al.*, 1972; Zwicker and Zwicker, 1984; Yost, 1985). A third condition includes a signal which is temporally centered in the masker (Kohlrausch, 1986). The duration of the fringe in these studies varied between 10 and 256 ms, the center frequency was always 500 Hz (except for Zwicker and Zwicker, 1984, using 400 Hz). Most experiments were performed with relatively short signals (10 to 32 ms). The experimental results are shown in Fig. 6.5. The upper-left panel shows data for a forward fringe, the upper-right panel for a backward fringe. The lower panel shows 4 different curves for 4 different masker durations (25, 50, 100 and 500 ms) as a function of center frequency. The white symbols are data from different datasets, the black symbols are the model predictions for a 20-ms, 500-Hz signal added to a broadband masker with a level of 50 dB/Hz (similar to Trahiotis *et al.*, 1972). Thresholds are expressed as the ratio between signal energy and spectral masker density to enable comparison between datasets[2].

The predicted thresholds hardly change if the masker duration is varied. This is in contrast to what is observed if the *signal* duration is changed, as discussed in Section 6.3.4. A maximum decrease of 5 dB is observed experimentally if the forward masker fringe is increased (top-left panel of Fig. 6.5). The model predictions show no effect of the masker duration at all, except for very short forward fringes. The absence of a distinct effect of masker duration is expected, since the output of an EI-type element optimally tuned for this condition has no activity for a masker alone and has some activity during the presence of the signal. Since this activity is used as a cue for the detection and since it does not depend on the duration of the masker, the model's predictions do not, in principle, depend on the masker duration. The decrease in thresholds for very short forward fringes is a consequence of the monaural adaptation loops (Dau *et al.*, 1996a). These loops are not completely charged during the first 25 ms of the stimulus and a substantial overshoot in the output of the adaptation loops exists just after the onset. If interaural differences are presented within this short period of time, they will result in a stronger change at the output of EI-type elements than if they were presented with a longer forward masker fringe. Therefore, the thresholds are up to 5

---

[2]The data adapted from Zwicker and Zwicker (1984) were measured at 400-Hz center frequency instead of 500 Hz. In constrast to the other studies, their masking noise was not spectrally flat but had a spectral energy distribution that is referred to as *uniform masking noise*. The amplitude spectrum of this noise type is spectrally flat from 0 to 500 Hz and decreases with 10 dB/dec above 500 Hz. To be able to use their data, we calculated the spectral energy density of their masking noise at 400-Hz center frequency and used this value to plot the data.
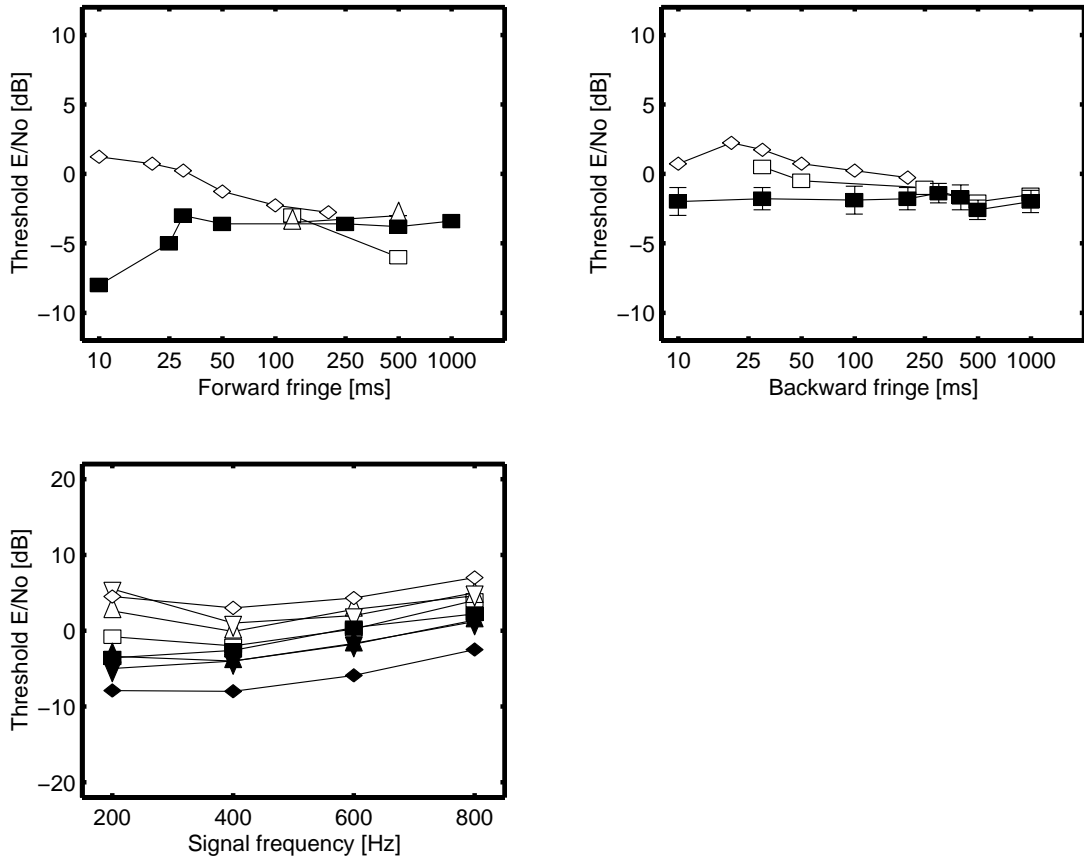
Figure 6.5: *NoSπ thresholds as a function of the masker duration. The upper-left panel denotes experimental data with a forward fringe, the upper-right panel with a backward fringe. The lower panel shows thresholds for a signal temporally centered in a masker for 4 different masker durations (squares denote 500 ms, upward triangles 100 ms, downward triangles 50 ms and diamonds 25 ms) as a function of the center frequency. The white squares in the upper panels are experimental data for a 500-Hz, 32-ms signal adapted from Robinson and Trahiotis (1972), the upward triangles are from the same study with a 256-ms signal, the diamonds are adapted from Zwicker and Zwicker (1984) for a 400-Hz signal. The black symbols are model predictions for a 500-Hz, 20-ms signal.*

dB lower during the first 25 ms. This effect is clearly visible in the upper-left panel of Fig. 6.5 for short forward fringes and in the lower panel (with the centered signal). In the latter case, simulated thresholds are more than 10 dB lower than the experimental data for a 25-ms masker (2.5 ms forward fringe). For masker durations of 50 ms and beyond, the model predicts nearly constant thresholds, while the experimental data show a slight decrease with increasing masker duration. On the other hand, the backward fringe has no influence on the state of the adaptation loops and hence the thresholds do not depend on the duration of the backward fringe, as demonstrated in the right panel of Fig. 6.5.

6.3.6 **Maskers with phase-transitions in the time-domain**

The previous sections dealt with the ability of the binaural auditory system to *integrate* information over time. Besides integration, another very important temporal property that can be measured is the temporal *resolution* of the system. In analogy to the frequency-domain phase transition which we discussed in Chapter 5, an interaural phase transition can be applied in the time domain. As the spectral phase transition enabled the estimation of the spectral resolution, the time-domain equivalent enables estimation of the temporal resolution. One possible realization of such a phase transition is a masker which is first interaurally in-phase and then interaurally out-of-phase. This condition is referred to as NoπSπ if an interaurally out-of-phase signal is used. In a similar way, NπoSπ refers to a masker that is interaurally out-of-phase first, followed by an in-phase noise. If the signal is centered within the in-phase masker portion, the effective condition is NoSπ and a large BMLD is observed. On the other hand, if the signal is presented during the out-of-phase masker portion, no BMLD is expected. Experimental data have shown that for signal positions close to the masker phase transition, a gradual change of the threshold is observed (Kollmeier and Gilkey, 1990; Holube *et al.*, 1998). The experimental data of individual subjects from Kollmeier and Gilkey (1990) are shown in Fig. 6.6. These data were measured with a 500-Hz, 20-ms signal added to a broadband noise with a spectral energy density of 40 dB/Hz. The thresholds are plotted as a function of the signal center relative to the occurence of the phase transition of the masker. The upper-left panel corresponds to an NπoSπ condition, the upper-right panel to NoπSπ. In both conditions, a gradual change in threshold is observed near the phase transition (0 ms). The model predictions (black symbols) have a similar gradual change in threshold as the experimental data and follow the lower bound of the four subjects.

The gradual change as observed in Fig. 6.6 is thought to reflect the *temporal resolution* of the binaural auditory system. In the model, this resolution is limited by the temporal averager at the output of all EI-type elements; the stepwise masker correlation change is heavily smoothed by the averager and hence thresholds show a gradual change instead of a stepwise one.

The lower panels in Fig. 6.6 show data obtained for a corresponding 'monaural' condition. Both the signal and the two masker portions were presented interaurally out-of-phase (NπNπSπ), and one of the masker portions was decreased in level by 15 dB. The data in the lower-left panel were obtained for a masker that drops by 15 dB at the second half of the total stimulus (this condition is referred to as Nπ(-15dB)NπSπ), while the data in the lower-right panel were obtained for a 15-dB increase at the stimulus center. The open symbols denote data from different subjects adapted from Kollmeier and Gilkey (1990), the filled symbols are model predictions. Clearly, the time
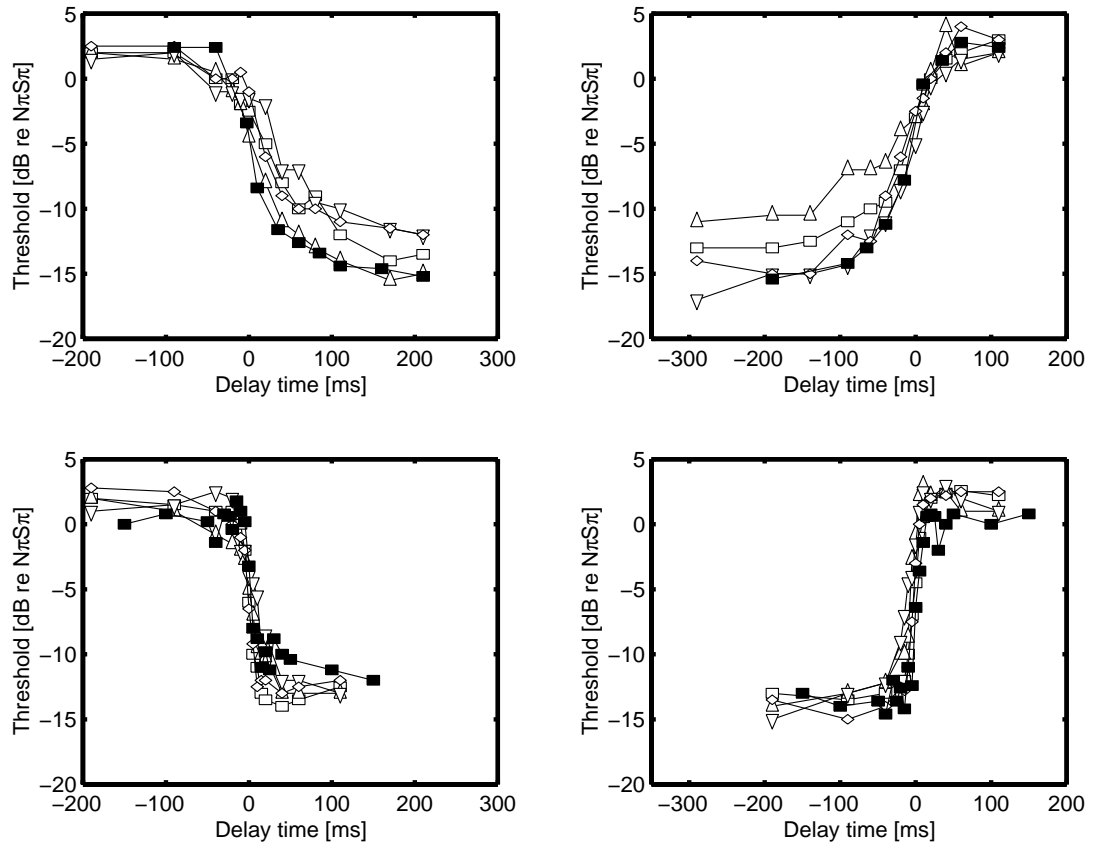
Figure 6.6: *$N\pi oS\pi$ (upper-left panel) and $No\pi S\pi$ (upper-right panel) thresholds as a function of the temporal position of the signal center relative to the masker-phase transition. The lower-left and lower-right panels correspond to the monaural $N\pi(-15dB)N\pi S\pi$ and (-15dB)$N\pi N\pi S\pi$ conditions, respectively (see text). The 0-dB point on the ordinate denotes the nontransient $N\pi S\pi$ thresholds. White symbols are experimental data for different subjects adapted from Kollmeier and Gilkey (1990), the black symbols are model predictions.*

constant for processing monaural cues is much shorter than for the processing of binaural cues. The ability of the present model to also predict monaural forward and backward masking relies on a completely different process from that involved in predicting the binaural temporal resolution effects. The monaural data are predicted due to the presence of the adaptation loops prior to any binaural interaction. Since the model incorporates all the stages of the monaural model described by Dau *et al.* (1996a,b), it inherits the predictive power of that model for all cases where no binaural interaction is needed.

An extension of the experiment with one masker phase transition in the time domain is obtained by using two phase transitions. Culling and Summerfield (1998) measured detection thresholds of a 500-Hz 20-ms $S\pi$ signal which was added to a broadband in-phase noise masker (No) of variable duration, preceded and followed by 400 ms of interaurally uncorrelated noise. The spectral energy density of the noise was 40 dB/Hz. We refer to this condition as NuouSπ. Culling and Summerfield (1998) found that thresholds decrease by up to 12 dB with increasing No duration from 20 to 960 ms. The data are
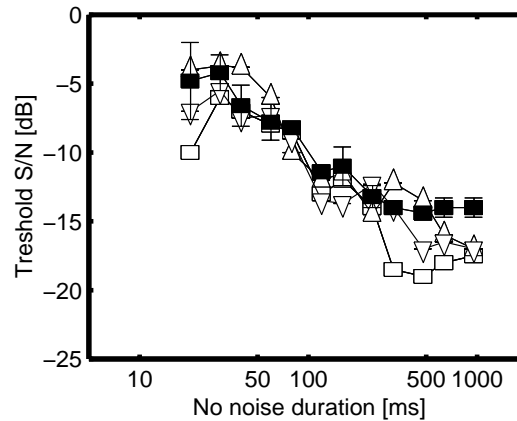
Figure 6.7: *NuouSπ detection thresholds as a function of the duration of the No noise. White symbols are experimental data for different subjects adapted from Culling and Summerfield (1998), black symbols are model predictions.*

shown in Fig. 6.7. The white symbols are data for 3 different subjects, the black symbols are model predictions. For No durations between 20 and 400 ms, the modelled thresholds agree well with the subjects' data. For further increases in No duration, there is a discrepancy, because the experimental data tend to decrease further, while the modelled thresholds remain constant. This indicates that the temporal *resolution* of the human binaural auditory system is very well represented by the model, but that some really long-term processes with a temporal extension of 500 ms and more are not covered by our present model structure.

Besides using stepwise correlation changes in the masker, experiments have also been performed with a sinusoidally changing interaural masker correlation. Holube *et al.* (1998) measured the detectability of a 500-Hz, 20-ms signal as a function of the correlation modulation period. The signal was always centered at a temporal position where either an No or Nπ noise was present. The masker duration was 2500 ms for the modulation periods of 2 and 1 seconds and 750 ms for shorter periods. A band-limited masker (0.1 to 1 kHz) was used with an overall level of 75 dB SPL. The results (white symbols) combined with model predictions (black symbols) are shown in Fig. 6.8. The left panel corresponds to a signal presentation at a position where the noise was interaurally in-phase, the right panel where it was out-of-phase. Thresholds are expressed relative to the monaural (NoSo) thresholds.

The thresholds for the signal presentation centered on No decrease with increasing modulation period. This is the result of the decreasing amount of Nπ noise at the input of the temporal window of the EI-type element during the presentation of the signal. A similar argument holds for the presentation during the Nπ masker phase. With decreasing period, an increasing amount of Nπ noise is present in the EI-type element during signal presentation resulting in higher thresholds.
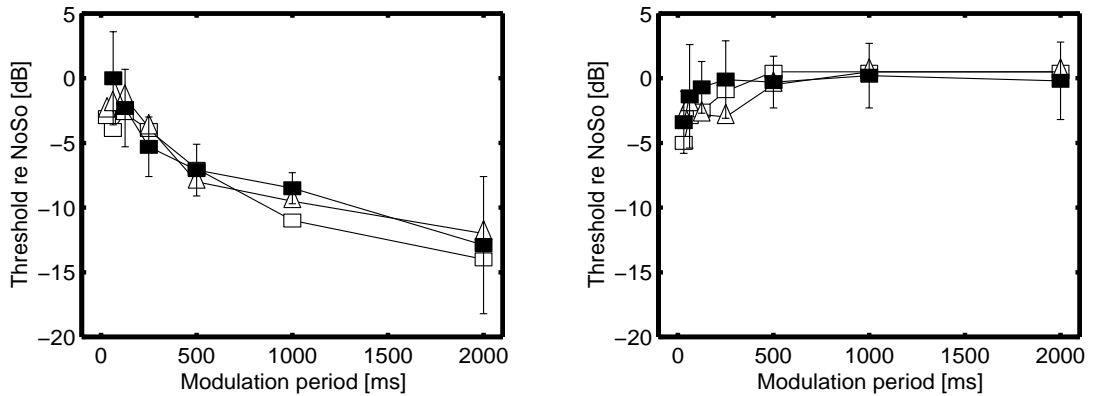
Figure 6.8: *Detection thresholds for an Sπ signal presented in a noise masker with a sinusoidally-modulated interaural correlation. Thresholds are plotted relative to the monaural (NoSo) threshold. The left panel corresponds to a signal presentation at a point where the interaural masker correlation was +1, the right panel to a masker correlation of -1. The white symbols are data adapted from Holube* et al. *(1998), the black symbols are model predictions.*

In contrast to earlier modelling approaches (Kollmeier and Gilkey, 1990; Holube *et al.*, 1998), our implementation does explain data for stepwise and sinusoidal correlation modulation with the *same* temporal window. We will come back to this observation in the general discussion.

6.3.7 Discrimination of dynamic interaural intensity differences

Grantham (1984a) measured observers' ability to detect time-varying interaural intensity differences. The stimuli consisted of interaurally uncorrelated noises of which the envelopes were sinusoidally modulated. The task was to discriminate between a modulation which was interaurally in-phase and a modulation which was interaurally out-of-phase, the latter resulting in interaural intensity differences. The noise used in this experiment had a bandwidth of 0.4 octaves centered at 500 Hz and had a level of 75 dB SPL. The stimulus duration was 1000 ms. The results of Grantham (1984a) (open symbols) combined with the model predictions (filled symbols) are shown in Fig. 6.9.

As can be observed in Fig. 6.9, the thresholds increase with increasing modulation rate, indicating a low-pass modulation function of the binaural auditory system. Although the data show large variations across subjects, the model predictions are a good representation of the subject denoted by the downward triangles. The general trend is that of a lowpass filter; the modulation index required at threshold increases with increasing modulation frequency. In our model, this can be understood as follows. The unmodulated noise that was used had an interaural correlation of 0. If a sinusoidal modulation is superimposed on the noise waveform, the interaural correlation of the noise remains unchanged. Now consider the output of an EI-type element
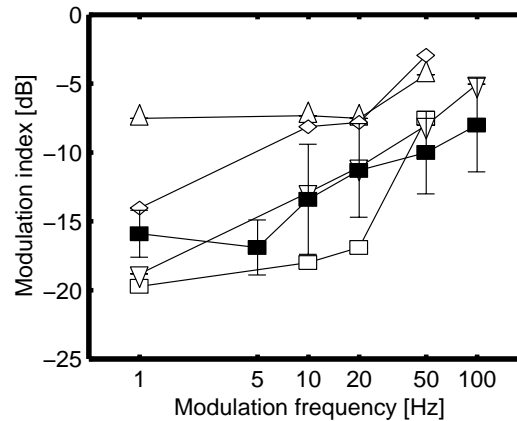
Figure 6.9: *Modulation depth (in dB) required for discrimination of interaural in-phase modulation from out-of-phase modulation as a function of the modulation frequency. The white symbols denote results for different subjects adapted from Grantham (1984a), the black symbols are model predictions.*

centered at the stimulus center frequency and $\alpha=\tau=0$. For the interaurally in-phase modulator (i.e., the reference stimuli), the output of the EI-type element has a similar pattern as the modulator: during positive modulator phases the masker energy is increased and hence the uncorrelated masker results in an increased amount of activity. On the other hand, a negative modulator phase results in a decrease in the EI-activity. These modulations are, however, only present for low modulation frequencies (< 10 Hz); for higher modulation frequencies the EI-output modulation depth decreases due to the temporal averaging. Thus, as long as the modulation period is beyond the time constant of the temporal averager of the EI-type elements, the externally presented monaural modulation is reflected by a modulation of the EI-type element output. The out-of-phase modulator results in hardly any modulation in the EI-type element output: every attenuation (i.e., negative modulator phase) of the left-ear signal is accompanied by an amplification of the right-ear signal (i.e., a positive modulator phase) and vice versa, resulting in only a very small effect on the EI activity. Thus, as long as the modulations in the EI-type output due to the in-phase modulator are clearly visible (i.e., slow modulations), the model shows a low modulation threshold which increases with increasing modulation frequency.

6.3.8 Discrimination of dynamic interaural time differences

Grantham and Wightman (1978) measured the detectability of sinusoidally time-varying interaural time differences present in a low-pass noise. The spectra of the noise stimuli were approximately flat between 10 and 3000 Hz. The presentation level was 70 dB SPL, the duration 440 ms. Grantham and Wightman (1978) found that the peak ITD required for detection strongly depends on the modulation frequency, having a value of about 30 μs at a modulation rate of 0 Hz which increases to 90 μs at 20 Hz. Interestingly, the
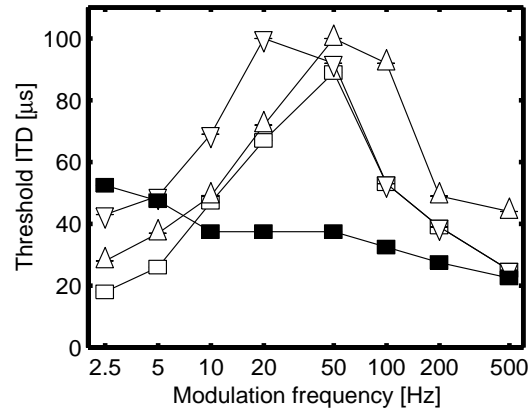
Figure 6.10: *Peak interaural time difference in microseconds at threshold as a function of the modulation frequency for the detection of sinusoidally-varying interaural time differences. The open symbols are experimental data adapted from Grantham and Wightman (1978) representing different subjects, the black symbols are model predictions.*

thresholds decrease again for higher modulation rates, reaching a value of about 30 μs at a modulation rate of 500 Hz. The results are shown in Fig. 6.10. The white symbols denote the experimental data, the black symbols are model predictions.

The bell-shaped curve which is seen in the experimental data is not observed in the model predictions. Considering the properties of the model, this is expected, since the average or peak interaural difference that occurs in this stimulus does not depend on the modulation rate. Therefore the model predictions do not show a bell-shaped curve but decrease monotonically by a factor of about 2.5. This decrease is related to the fact that the ITD at the onset of the signal is set to 0 and changes sinusoidally during the stimulus. If a modulation rate of 10 Hz is used, the first maximum in the ITD occurs 25 ms after the stimulus onset. At a modulation rate of 20 Hz, the maximum occurs at 12.5 ms et cetera. As discussed in Section 6.3.5, interaural differences closer to the onset of the stimulus results in lower thresholds due to the overshoot in the peripheral adaptation loops. Therefore, the ITD thresholds shown in Fig. 6.10 decrease with increasing modulation rate.

6.3.9 Binaural forward masking

In the previous experiments, the signal was always presented simultaneously with the masker. If a short signal is presented *after* the masker, a phenomenon which is referred to as forward masking is observed. For signals that are presented at increasing delays with respect to the masker offset, the thresholds decrease gradually towards the absolute threshold instead of showing a stepwise change (cf. Punch and Carhart, 1973; Yama, 1992; Kohlrausch and Fassel, 1997). This elevation is observed for signal delays of up to 200 ms. Moreover, a binaural release of masking can be observed if the signal
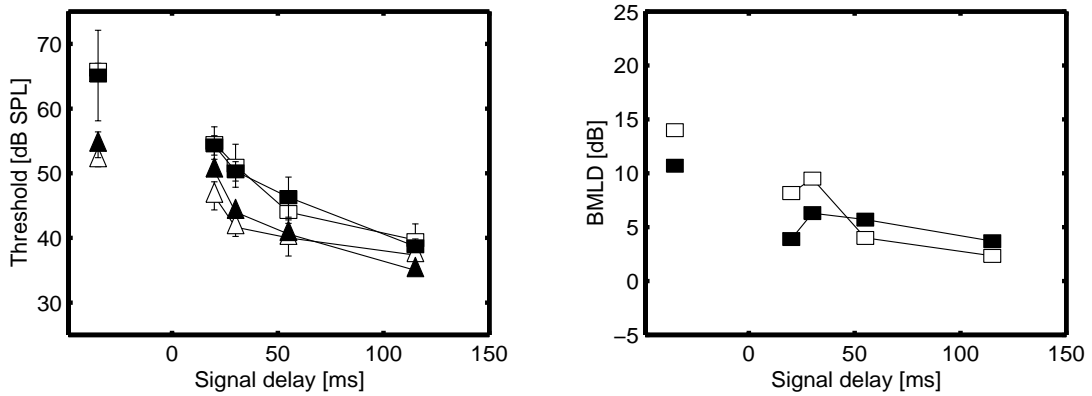
Figure 6.11: *Signal thresholds for NoSπ (triangles) and NoSo conditions (squares) as a function of the time difference between masker and signal offset (left panel) and corresponding BMLDs (right panel). White symbols are experimental data adapted from Yama (1992), black symbols are model predictions. The signal duration was 10 ms. The symbols at the left in both panels denote thresholds for simultaneous masking.*

is presented interaurally out-of-phase compared to an in-phase signal. For example, Yama (1992) measured forward-masking thresholds for a 10-ms, 250-Hz signal combined with a lowpass (0-2.5 kHz, overall level of 70 dB SPL), 500-ms running noise. Linear ramps of 5-ms duration were used to gate both signal and masker. The results show a BMLD of about 14 dB for simultaneous masking which decreases to a few dB for signal delays of 100 ms as can be observed in the left panel of Fig. 6.11. Thresholds are shown as a function of the time difference between signal and masker offset. The squares denote monaural (i.e., NoSo) thresholds, the triangles denote binaural (NoSπ) thresholds. The right panel shows the corresponding BMLDs, for both the model and the experimental data.

As can be observed from Fig. 6.11, the model (black symbols) shows a similar threshold behavior as the experimental data. For simultaneous masking, a BMLD of 16 dB is observed in the experimental data and a few dB less for the model predictions. Both the binaural and monaural forward masking thresholds show a decrease towards the absolute threshold, which is about 35 dB for both the So and Sπ signal. In the region of 0 to 100 ms, a substantial BMLD can be observed which is, however, smaller than the BMLD for simultaneous masking.

Results that seem to be in contradiction with those found by Yama (1992) were obtained by Kohlrausch and Fassel (1997). Their forward masking experiments only showed BMLDs for signal delays up to 20 ms instead of the 100 ms found by Yama (1992). The data were obtained with a 300-ms frozen-noise masker which was spectrally flat between 20 and 1000 Hz. The overall masker level was 70 dB SPL. A 20-ms, 500-Hz signal was used. The same values were used to obtain the model predictions. However, the frozen-noise sample was different from the one used in the experiments. The results
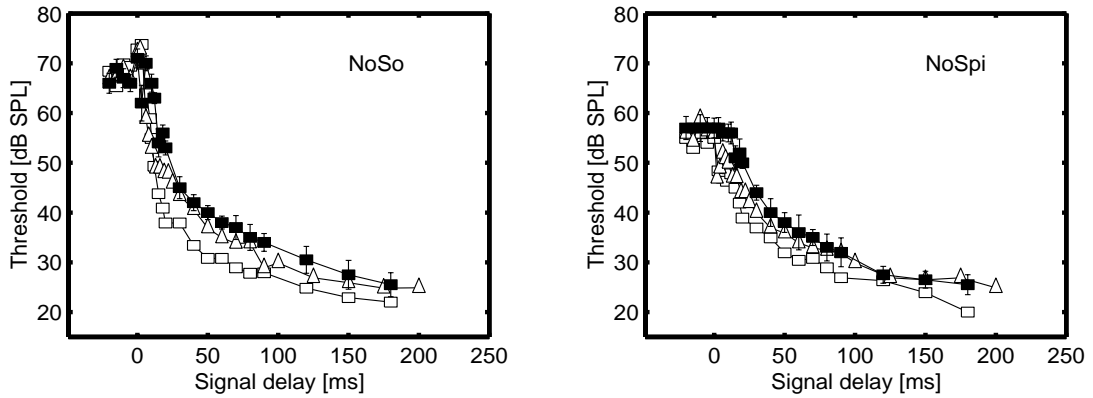
Figure 6.12: *Forward masking thresholds for a 20-ms So signal (left panel) and an Sπ signal (right panel) as a function of the time difference between masker and signal offset. A 300-ms frozen-noise (No) served as the masker. The white symbols are experimental data adapted from Kohlrausch and Fassel (1997), the black symbols are model predictions.*
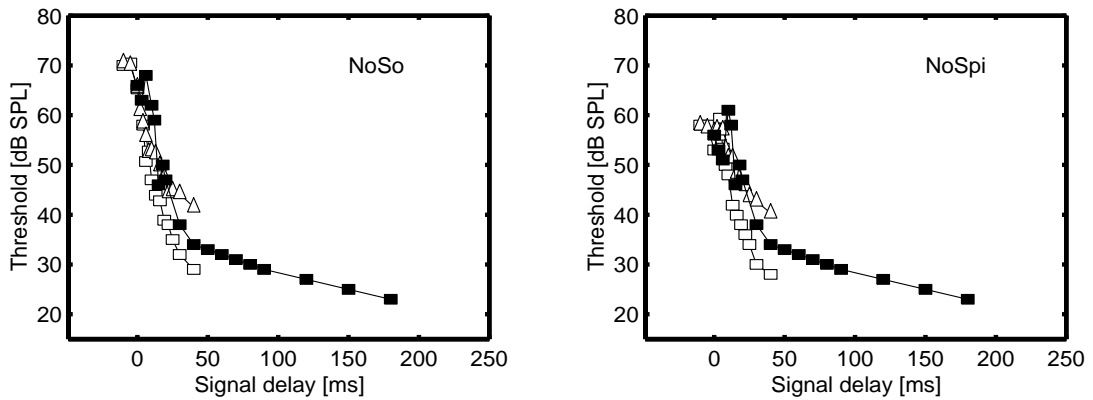
Figure 6.13: *Forward masking thresholds for a 20-ms So signal (left panel) and an Sπ signal (right panel) as a function of the time difference between masker and signal offset. In this case, the masker had a duration of only 20 ms. The white symbols are experimental data adapted from Kohlrausch and Fassel (1997), the black symbols are model predictions.*

and model predictions are shown in Fig. 6.12. The left panel corresponds to monaural conditions (NoSo), the right panel to binaural conditions (NoSπ).

In the NoSo condition, thresholds start to decrease as soon as the offset of the signal occurs after the masker offset. In contrast, NoSπ thresholds remain constant for signal delays up to about 10 ms. For larger signal delays, the thresholds gradually decrease with signal delay towards the absolute threshold. This decrease is stronger for the monaural (NoSo) condition than the binaural (NoSπ) condition. Hence a substantial BMLD is only found for signal delays up to 20 ms. Kohlrausch and Fassel (1997) also measured forward masking thresholds for a 20-ms masker. The waveform of this short masker was identical to the last 20 ms of the long masker. The results are shown in Fig. 6.13 in the same format as Fig. 6.12.

The main difference between the threshold behavior of 300-ms and 20-ms maskers is the slope of the forward-masking curve, which is steeper for the short masker. In the model, this steeper curve is the result of the fact that the adaptation loops which are part of the peripheral preprocessor are not completely 'charged' if a masker of only 20 ms is used (Dau *et al.*, 1996b).

Also in this condition, the model shows BMLDs only for signal delays of up to 20 ms, perfectly in line with the experimental data. The difference in BMLD behavior between the conditions shown Fig. 6.11 on the one hand and in Figs. 6.12 and 6.13 on the other hand is related to the difference in signal duration that was used. Yama (1992) used a relatively short signal (10 ms), while Kohlrausch and Fassel (1997) used a 20-ms signal. If a short signal is presented to the model, the onset of the signal will result in an increase of the output of the peripheral adaptation loops compared to the output in the absence of the signal. On the other hand, if the signal is turned off, the adaptation loops are (at least partially) adapted to the (higher) input signal and hence the signal offset results in a decrease of the output. Moreover, due to the adaptation of the system, the activity after the offset will be *less* then if *no signal* was present. An example of this property can be seen in the lower panel of Fig. 7 of Dau *et al.* (1996a): the presence of the signal results in both an overshoot at the signal onset and an undershoot at the signal offset. If the duration is sufficiently long compared to the temporal resolution of the monaural system, the model can use both the overshoot at the onset of the signal and the undershoot at the offset of the signal to detect the signal's presence. If a very short signal is used, however, the temporal averager at the output of the adaptation loops partially cancels the undershoot by the overshoot, resulting in a smaller overall effect at the output of the temporal averager. In the binaural case, the temporal window does not reduce the detectability of the signal because the window is applied *after* the computation of the *squared difference* between the left and right channels. Therefore, monaural thresholds are elevated more strongly if the signal duration is decreased from 20 to 10 ms than binaural thresholds. This is also observed in the model predictions. If a 20-ms signal is used (Figs. 6.12 and 6.13), both the monaural and binaural cues are about equally strong and no BMLD is observed for signal delays beyond 20 ms. When using a 10-ms signal, however, the monaural thresholds are elevated more than the binaural thresholds resulting in a BMLD which is still present even for signal delays up to 100 ms.

6.4   General discussion

We have shown that our binaural model is quite successful in describing the dependence of binaural detection thresholds on temporal stimulus properties. These properties include the effect of signal and masker duration, forward masking and detection against stimuli with a time-varying interaural correlation. By means of a temporal integrator followed by an optimal detector in the time domain, the model accounts both for temporal resolution and for temporal integration properties within a single framework. It is interesting to remark that a similar framework in the spectral domain is present in the model (i.e., a set of bandpass filters followed by an optimal detector in the frequency domain), also leading to very good predictions as a function of spectral stimulus parameters (see Chapter 4).

Although many of the simulations shown in this chapter and in the previous chapter show a good fit between the data and the predictions, very similar results would probably be obtained if the basic EI interaction in the model was replaced by an EE (or cross-correlation) interaction. There are, however, some specific experimental conditions that may give rise to some modelling difficulties. In Chapter 4 of this thesis, we expected that models based on the interaural cross-correlation may not account for the effect of both signal and masker duration. Two methods of computing the interaural correlation were discussed. The first comprised computation over the complete stimulus (i.e., masker duration). We argued that this method would result in a strong increase of detection thresholds with an increase of masker duration, which is not found in experimental data. The second method comprised computation of the correlation only for the stimulus part that contains the signal. In this case, a maximum effect of 1.5 dB/oct of signal duration is expected, which is not in line with experimental results showing a stronger influence of signal duration for durations below about 60 ms. It is therefore difficult to explain the effects of both masker and signal duration with a model based on cross-correlation. The simulation results in Sections 6.3.4 and 6.3.5 show that a model based on EI-type interaction in combination with an optimal detector shows a performance which is more in line with the experimentally obtained results.

The simulations as a function of interaural correlation and bandwidth revealed that the detection performance of the model can in principle be limited by two sources of errors, namely stimulus uncertainty in the externally presented signals and errors in the internal accuracy (internal noise).

The data in Figs. 6.6 and 6.7 (stepwise correlation change in the time domain) revealed that a double exponential window with time constants of 30 ms (equivalent rectangular duration, or ERD, of 60 ms) can account for the limited temporal resolution of the binaural system which is observed in different

experiments, which often reveal different underlying temporal windows. The shape and ERD of the window were chosen to fit the experimental results of Kollmeier and Gilkey (1990). A somewhat larger value for the ERD of 100 ms was found by Culling and Summerfield (1998). Their estimate was based on a Gaussian window. They also fitted their data with a double-exponential window, resulting in ERDs between 48 and 117 ms, which is much closer to the temporal window we used throughout this thesis. These results support the fact that the ERD itself is not a very valuable property to discuss without mentioning the window shape from which it is derived.

The data in Fig. 6.8 were obtained for a sinusoidally changing masker correlation. The ERD found by Holube *et al.* (1998) that fitted these data (using a double-exponential window) was 91 to 122 ms. A similar experiment by Grantham and Wightman (1979) revealed an ERD of 139 to 189 ms. Despite these rather large ERDs compared to the ERD of our model, the simulations give a good fit to the data. This suggests that not only the shape of the temporal window, but also the stimulus configuration has an influence on the ERD that is estimated from experimental data: experiments with stepwise correlation changes result in lower estimates of the time constants than sinusoidal correlation changes (Kollmeier and Gilkey, 1990; Holube *et al.*, 1998).

A possible explanation for these differences in the estimate of the ERD was given by Holube *et al.* (1998). They stated that "the reason for this mismatch seems to be the different detection strategies employed for the various tasks that are affected by the consistency of binaural information across frequency and time". In their fitting procedure, Kollmeier and Gilkey (1990) and Holube *et al.* (1998) obtained the predicted thresholds by computing the weighted integration of the instantaneous interaural cross-correlation *at the temporal center of the signal*. For the sinusoidal changes in the correlation, it is likely that this detection strategy results in the highest signal-to-masker ratio, given the fact that both the temporal window and the correlation modulation are symmetric around that moment. It is not obvious, however, that this strategy is also optimal for the stepwise correlation changes. In fact, an analysis of the optimal detector in our model revealed that within the framework of our model, the optimal position for detecting the signal is about 10 ms further away from the masker phase transition (i.e., off-time listening). Hence by analyzing the correlation slightly away from the temporal center of the signal, lower thresholds are obtained. Consequently, the fitting procedure used by Kollmeier and Gilkey (1990) and Holube *et al.* (1998) for stepwise correlation changes *underestimates* the time constants present in the system. This is perfectly in line with their experimental results: the time constants for stepwise correlation changes were about a factor 2 lower than for the sinusoidal correlation changes.

One of the experimental findings that the present model could not account for is the bandwidth dependence of interaural correlation jnd's for a reference correlation of +1 (see Fig. 6.3). The experimental data show a substantial decrease in performance with bandwidth, while the model's performance increases. A possible explanation for this discrepancy is that the binaural auditory system can only integrate cues across frequency if these cues *are highly correlated* across auditory filters. Since the data show the strongest increase in the correlation jnd for bandwidths beyond the ERB of the auditory filters, it is likely that the increase in the thresholds results from across-frequency effects instead of within-filter stimulus properties. If a stimulus with a correlation of +1 is presented, the stimulus can be canceled completely. The reduction of the correlation can thus in the model be detected by an increase in the residual noise after optimal cancellation. If the noise is broadband, this residue is in principle independent across peripheral filters. Our model does not incorporate the correlation of cues across frequency. However it could be possible that the binaural auditory system does.

Another experimental result that cannot be accounted for by the model is the low-pass characteristic that is obtained with dynamically-varying ITDs (Fig. 6.10). A model that can account for these data is the position-variable model of Stern and Bachorski (1983). In their model, the intracranial locus of the stimulus is estimated by computing a weighted centroid of the running cross-correlation function. The running cross-correlation function is computed using an exponentially-decaying averaging window with a fixed time constant. If the ITD is modulated with a period that is longer than the duration of the temporal averager, the peak of the running cross-correlation function follows the externally presented ITD. Consequently, the weighted centroid modulates similarly. If the ITD modulation frequency is increased, the lowpass characteristic of the running cross-correlation averages the ITD over time. This results in a lower but wider peak of the cross-correlation function. Moreover, the maximum displacement of the peak decreases. Consequently, the amplitude of the time-varying centroid of the cross-correlation *decreases.* Since this centroid is used as a decision variable, thresholds increase with increasing modulation rate. In fact, the model of Stern and Bachorski (1983) was very successful in predicting the data shown in Fig. 6.10 for the left side of the bell-shaped curve.

It would in principle be possible to modify our model in such a way that it does not use increases in the activity in the EI-type pattern but rather an estimate of the position of the sound source as a decision variable. This can be facilitated by scanning the position of the minimum within the EI-type element activity pattern. The reason that we did not base the model's decision process on a position variable is that such an approach has a detrimental effect on the fits of other data. For example, in Chapter 5, NoSπ thresholds were

discussed as a function of the bandwidth of the masking noise. The predicted thresholds as well as the experimental data show approximately constant thresholds for bandwidths up to twice the ERB of the peripheral filters. We expect that a position-variable model, independently of whether the binaural interaction is based on EE or EI processing, cannot account for this result. The addition of an Sπ signal to a diotic masker results in the presence of dynamically-varying IIDs and ITDs in the stimulus. The rate of fluctuation of these differences depends on the bandwidth; a larger bandwidth corresponds to faster fluctuations. Consequently, the NoSπ thresholds for a position-variable model are expected to increase with increasing bandwidth, which is not in line with the experimental data. In summary, we do not see how the absence of an effect of the ITD modulation rate in band-limited NoSπ conditions and the clear effect of ITD modulation shown by Grantham and Wightman (1978) can be explained with the same detection mechanism.

Finally, the strong overshoot in the peripheral adaptation loops which results in lower detection thresholds for a binaural signal presented during the first 25 ms of the masker is an unwanted effect. In a certain way, we can end this series of binaural modelling chapters with a similar remark to that made at the end of their discussion by Dau *et al.* (1996b). The temporal dynamics and nonlinear compression effects of the adaptation loops are useful in understanding a number of binaural effects like binaural forward masking and the influence of overall interaural level differences on binaural unmasking and on lateralization, while for some specific conditions, their temporal dynamic is too strong. Obviously, we so far have not found the optimal realization of this stage. Therefore we will, together with our colleagues in Oldenburg, continue in our efforts to improve the understanding of this part of our monaural and binaural models.

CHAPTER 7

Perceptual (ir)relevance of HRTF phase and magnitude spectra[1]

This chapter discusses the perceptual consequences of smoothing of anechoic HRTF phase and magnitude spectra. The smoothing process is based on a binaural perception model, in which interaural cues in the auditory system are rendered at a limited spectral resolution. This limited resolution is the result of the filterbank present in the peripheral auditory system (i.e., the cochlea). Listening tests with single and multiple virtual sound sources revealed that both the phase and magnitude spectra of HRTFs can be smoothed with gammatone filters which equal estimates of the spectral resolution of the cochlea without audible artifacts. The amount of smoothing was then increased by decreasing the order of the gammatone filters. If the filter order is reduced by a factor 3, subjects indicate spectral and positional changes in the virtual sound sources. The binaural detection model described in Chapter 4 was used to predict the audibility of the smoothing process. A comparison between model predictions and experimental data showed that the threshold at which subjects start to hear smoothing artifacts can be predicted accurately. Moreover, a high correlation exists between the model output and the amount of stimulus degradation reported by subjects.

7.1 Introduction

Two important features of the waveforms arriving at both ears that determine the lateral location of a sound source are the interaural intensity differences (IIDs) and the interaural time differences (ITDs). Stimuli with specified values of the ITD and IID can be presented over headphones, resulting in a lateralization of the sound source which depends on the magnitude of the ITD or IID (Sayers, 1964; Hafter and Carrier, 1970; Yost, 1981). The usual result of these experiments is that the source images are located inside the head, somewhere between the left and right ear. The reason for the fact that these stimuli are not externalized is that the single frequency-independent IID or ITD is a poor representation of the acoustical signals in the real world. The waveforms of sound sources in the real world are filtered by the pinna, head and torso of the listener, resulting in an intricate frequency dependence of the ITD and IID (Wightman and Kistler, 1989b). The filtering

---

[1]This chapter is based on Breebaart and Kohlrausch (2001b)

can be described by head-related transfer functions (HRTFs), which describe the position-dependent change in the phase and magnitude spectra of a sound source. One of the major difficulties in using HRTFs is that these filters are both position and subject dependent (Wightman and Kistler, 1989b). Usually HRTFs are measured as a function of both elevation and azimuth, but there is evidence that spatial cues also depend on the distance of a sound source (Brungart *et al.*, 1999; Shinn-Cunningham *et al.*, 2000). If individualized HRTFs are used, subjects are not able to discriminate between real and virtual sound sources presented over headphones (Wightman and Kistler, 1989a; Hartmann and Wittenberg, 1996; Langendijk and Bronkhorst, 2000). If nonindividualized HRTFs are used, however, subjects report poor elevation accuracy and front-back confusions (Wenzel *et al.*, 1993; Wightman and Kistler, 1999). Some attempts have been made to increase localization performance with nonindividualized HRTFs by emphasizing the pinna effects (Zhang *et al.*, 1998) or the interaural differences (Durlach and Pang, 1986).

Because of the large amount of data present in individual HRTF sets that is normally required to generate externalized virtual sound sources, researchers have tried to reduce the information in several ways. For example, attempts have been made to only measure HRTF sets for a limited range of source positions and to interpolate HRTF impulse responses for positions in between (Wenzel and Foster, 1993). Other studies described HRTFs in terms of principal components by deriving a small set of basis spectra with individual, position-dependent weights (Kistler and Wightman, 1992; Cheung *et al.*, 1998). Although this method is valid in physical terms, there is a risk that the basis functions that are very important in terms of the least-squares error of the fit are not so relevant in terms of human auditory perception. An other approach consisted of determining the role of spectral and interaural phase cues present in the HRTFs. Wightman and Kistler (1992) showed that low-frequency interaural time differences dominate in sound localization, while if the low frequencies are removed from the stimuli, the apparant direction is determined primarily by interaural intensity differences and pinna cues. Hartmann and Wittenberg (1996), Kulkarni and Colburn (1998) and Kulkarni *et al.* (1999) showed that the frequency-dependent ITD of anechoic HRTFs can be simplified by a frequency-independent delay without perceptual consequences and that the fine structure of the HRTF is relatively unimportant for auditory spatial attributes.

But also more psychoacoustically-motivated methods of HRTF reduction have been suggested. For example, Huopaniemi and Zacharov (1999) discussed three methods to reduce HRTF information. The first entailed smoothing of the HRTF magnitude spectra by a rectangular smoothing filter with a bandwidth equal to the equivalent rectangular bandwidth (ERB) (Glasberg and Moore, 1990). The second embodied weighting of the errors

in an HRTF approximation with the inverse of the ERB scale as weighting function. The third method used frequency warping to account for the non-uniform frequency resolution of the auditory system.

From many of the studies described above, it can be concluded that although a frequency-independent IID and/or ITD does not result in an externalized image, the complex magnitude and phase spectra which are present in HRTFs can be simplified to some extent without deteriorating the externalization. In this chapter, we will investigate the relaxation of anechoic HRTF accuracy based on smoothing of the phase and magnitude spectra. The method of smoothing is derived from a binaural detection model described in Chapter 4. Although smoothing has been proposed before, our efforts differ in two aspects from other studies:

- the method of smoothing aims at a minimized *perceptual degradation* of the sound image. This is achieved by minimizing the changes in the internal representation of a binaural detection model rather than minimizing a norm of the HRTF impulse response errors.

- because we are interested in a generalized theory of describing HRTF data, we do not discuss any filter structure that may achieve the desired smoothing, because we do not want to be limited by implementation issues.

7.2 HRTF smoothing

It is generally accepted that the auditory system splits the incoming waveforms in several band-limited signals. The bandwidth of these filters depends on the center frequency (Glasberg and Moore, 1990) and can be seen as a limit of the spectral accuracy of (binaural) processing. We hypothesize that *the HRTF phase and magnitude spectra do not need a higher resolution than the spectral resolution of the filterbank in the peripheral auditory system*. This hypothesis is supported by the binaural detection model presented in Chapter 4 of this thesis. This model consists of 3 consecutive stages which are described in more detail in Section 7.4. The first stage comprises a peripheral preprocessing model, which among other things simulates the spectral filtering of the cochlea by applying a gammatone filterbank. It has been shown that with the correct choice of its bandwidth parameter, the spectral resolution of a 4th-order gammatone filter closely matches the spectral resolution of the human cochlea (Johannesma, 1972; Patterson *et al.*, 1988). The consecutive stages explore monaural properties (such as spectral content) and binaural properties (ITDs and IIDs) *after* the gammatone filterbank. Hence this filterbank limits the spectral resolution for the binaural auditory system. We will therefore use the same gammatone filter to explore the perceptual consequences of HRTF phase and magnitude smoothing.

7.2.1 HRTF magnitude smoothing

The gammatone filter has an impulse response for $t \geq 0$ given by (Patterson *et al.*, 1988)

$$h(t) = t^{n-1}e^{-2\pi bt}\cos(2\pi f_c t + \phi), \qquad (7.1)$$

where n denotes the order of the filter, b determines the bandwidth, f_c is the center frequency of the filter and ϕ the initial phase. The resulting transfer function $H(f, f_c)$ for ϕ=0 can be approximated by (Patterson *et al.*, 1988)

$$H(f, f_c) = \left(\frac{1}{1 + j(f - f_c)/b} \right)^n, \qquad (7.2)$$

and the 3-dB bandwidth B_{3dB} is given by

$$B_{3dB} = 2b\sqrt{2^{1/n} - 1}. \qquad (7.3)$$

This 3-dB bandwidth was set to the equivalent rectangular bandwidth (ERB) estimate[2] of the human auditory filters given by Glasberg and Moore (1990), resulting in

$$b(f_c) = \frac{24.7(0.00437f_c + 1)}{2\sqrt{2^{1/n} - 1}}. \qquad (7.4)$$

Then the smoothed magnitude $|Y(f_c)|$ of HRTF $X(f)$ is given by

$$|Y(f_c)| = \sqrt{\frac{\int_0^\infty |X(f)|^2 |H(f, f_c)|^2 df}{\int_0^\infty |H(f, f_c)|^2 df}}. \qquad (7.5)$$

The numerator denotes the product of the original magnitude spectrum $|X(f)|$ with the smoothing function $|H(f, f_c)|$, while the denominator compensates for a spectral tilt resulting from the changing bandwidth with center frequency. The explicit form given in Eq. 7.5 has some important advantages:

1. because a gammatone filter is used, the amount of crosstalk between adjacent filters is closer to that in the human auditory system than for a rectangular smoothing window, as suggested by Huopaniemi and Zacharov (1999).

---

[2]The equivalent rectangular bandwidth of a bandpass gammatone filter is always larger than the 3-dB bandwidth. For a filter order of 3, the ERB is about 13% larger than the 3-dB bandwidth. Hence the ERB of our smoothing filters is a bit larger than the ERB estimate of the auditory filters, even at a filter order of 3. Our method of smoothing encompasses decrements of the filter order and hence increments of the ERB of the smoothing filter. However, the 3-dB bandwidth is always kept constant and was equal to the ERB-estimate of the human auditory filters.
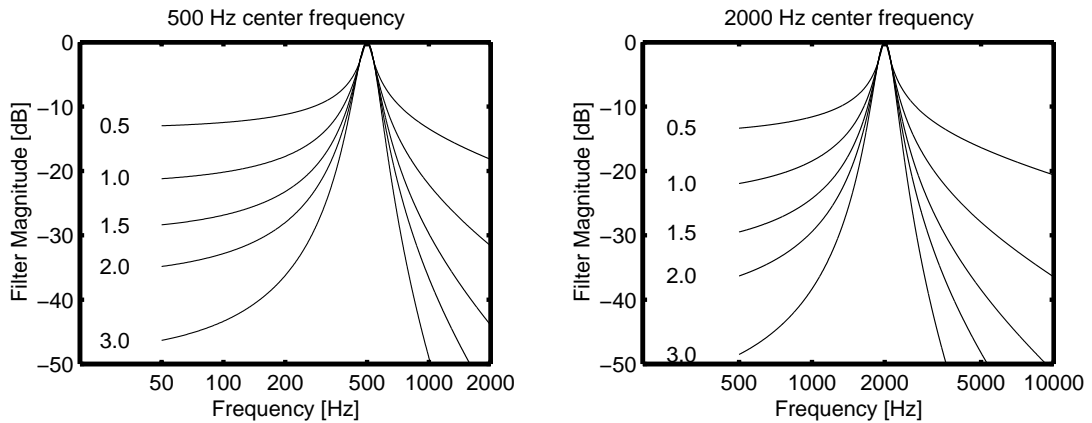
Figure 7.1: *Examples of the smoothing functions* $|H(f, f_c)|$ *as a function of frequency for two center frequencies (500 Hz for the left panel and 2000 Hz for the right panel) and different values for the filter order, ranging from 0.5 to 3.*

2. the binaural detection model described in Chapter 4 uses the *energy* of the difference signal of the waveforms arriving at the two ears after the peripheral filterbank as a decision variable to detect interaural differences. It can be shown that smoothing of the power spectrum of the HRTF magnitude spectra instead of smoothing the linear magnitude spectra gives a better fit in terms of the binaural model.

The parameter that was used to change the effect of the smoothing process is the order of the filter n. If n is decreased, the skirts of the smoothing filter become less steep while keeping the 3-dB bandwidth constant. Hence processing an HRTF with a *lower* filter order leads to *more* smoothing. This is demonstrated in Fig. 7.1. The left panel shows the magnitude of the smoothing filter $|H(f, 500)|$ at a center frequency of 500 Hz for different values of the filter order, ranging from 0.5 to 3. The right panel corresponds to a center frequency of 2000 Hz.

The result of the smoothing process upon the magnitude of the HRTF can be observed in Fig. 7.2. Here, the magnitude of the HRTFs for a sound source at an elevation of $0°$ and an azimuth of $30°$ is shown. The left panel corresponds to the ipsilateral ear, the right panel to the contralateral ear. The solid line denotes the original (i.e., unprocessed) HRTF. The dashed line is a smoothed magnitude spectrum for n=1. Clearly, sharp peaks and dips disappear through the smoothing operation.

7.2.2 HRTF phase smoothing

The phase spectra of HRTF pairs usually consist of interaural phase differences that are not linear with frequency (i.e., an overall delay of the contralateral ear). From an engineering point of view, however, it would be very attractive if linear phase or minimum phase filters could be used for the generation of virtual sound sources due to their lower complexity. We therefore decided
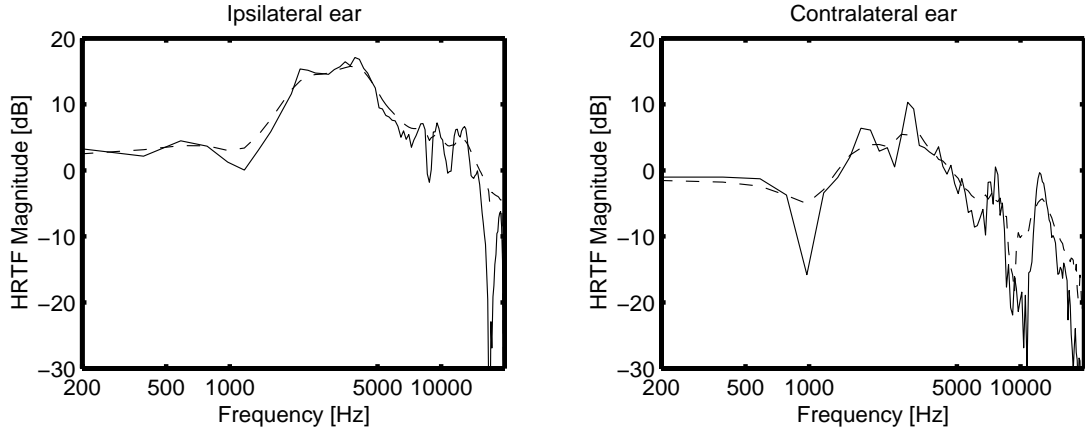
Figure 7.2: *Unprocessed (solid lines) and smoothed (dashed lines) HRTF magnitude spectra for a sound source at 0° elevation and 30° azimuth. The order of the smoothing filter n equals 1. The left panel shows the spectra for the ipsilateral ear, the right panel for the contralateral ear.*

not to smooth the phase spectra themselves, but to use a smoothing that eventually (for low n) results in linear phase HRTFs, and hence in a frequency-independent ITD. This time smoothing is obtained by first dividing the phase spectrum by $2\pi f$. Given a HRTF $X(f_c)$, the smoothed phase spectrum, $Y(f_c)$, is given by

$$\arg\left\{Y(f_c)\right\} = 2\pi f_c \frac{\int_0^\infty \frac{\arg\{X(f)\}}{2\pi f}|H(f, f_c)|df}{\int_0^\infty |H(f)|df} \tag{7.6}$$

An example of the resulting ITD is given in Fig. 7.3. The left panel shows the ITD for a sound source at 30° azimuth, the right panel for 120° azimuth. The solid line denotes the original (i.e., unprocessed) ITD, the dashed line the smoothed ITD for n=1. The following sections describe psycho-acoustic listening tests to reveal the audibility of the smoothing operation described above.

7.3   Perceptual evaluation

7.3.1   Stimuli

Six different wideband CD-quality stereo musical fragments were used to create virtual loudspeakers. These fragments had a duration of about 2.5 seconds. The fragments cover a wide variety of musical genres and sonic attributes: some fragments only contained one instrument (voice or piano), while other fragments created various phantom sources when played through a stereo sound set (orchestra and rock band). The normalized cross-correlation between the left and right channels ranged from 0.05 to 0.94.
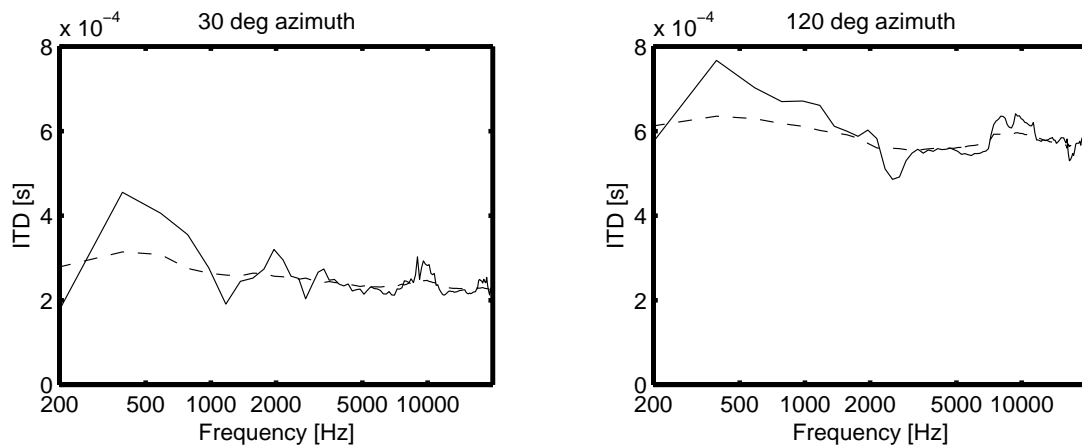
Figure 7.3: *ITD as a function of frequency for an unprocessed HRTF pair (solid line) and a smoothed HRTF pair (dashed line). The left panel corresponds to a sound source at 30° azimuth, the right panel to 120° azimuth.*

Anechoic HRTFs were taken from the AUDIS CDROM (Blauert *et al.*, 1998) for one person only. Each subject in our experiments listened to the same (non-individual) HRTF set. The six audio fragments were filtered with original (unprocessed) HRTFs and smoothed HRTFs. Smoothing was applied for HRTF phase only, HRTF magnitude only, or both. The smoothing order n ranged from 3 (little smoothing) to 0.1 (very close to linear phase or a flat spectrum). Two distinct cases were investigated. The first comprised only one virtual sound source at an elevation of 0° and an azimuth of 0°, 30° or 120°. The second condition comprised 2 virtual loudspeakers, at ± 30° or at ± 120° azimuth. In this condition, the signals consisted of the left channel and the right channel of the original stereo fragment, respectively. All stimuli were presented to the subjects over headphones (Beyer Dynamic DT990) in an isolated listening booth at a level between 70 and 80 dB SPL (depending on the fragment).

7.3.2 Procedure

Three trained subjects participated in the experiments. Each trial consisted of the presentation of a fragment filtered through unprocessed HRTFs, followed by 300 ms silence and the same fragment filtered through processed HRTFs. All combinations of fragment, smoothing parameters and number of virtual loudspeakers were presented once in random order. Subjects had to judge the difference between the two fragments by giving one out of three possible answers: no audible differences, small audible differences (subjects could hear some subtle changes) or large audible differences (subjects could clearly hear the effect of the smoothing operation).
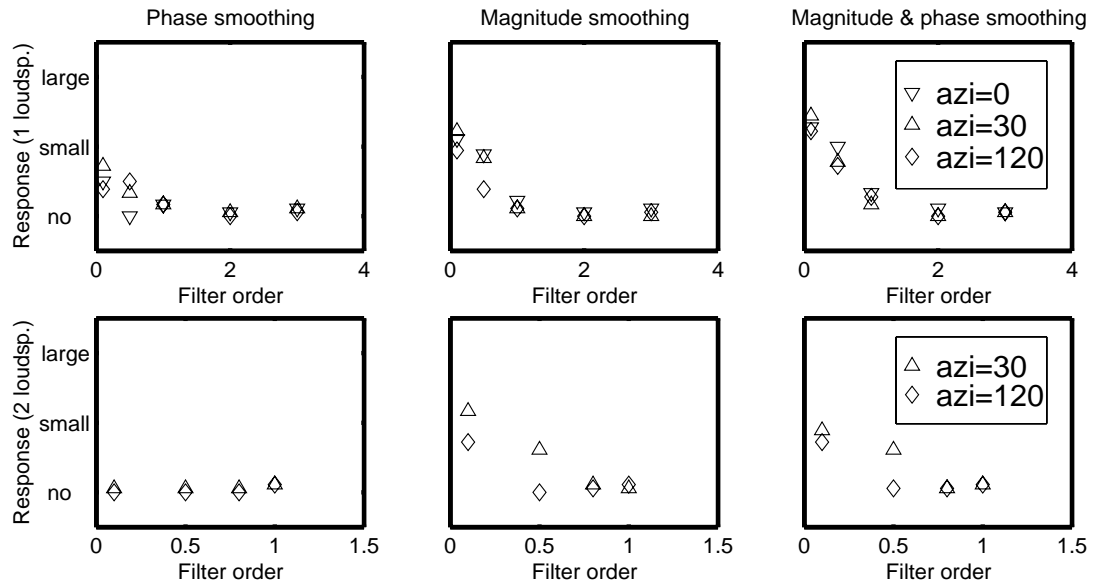
Figure 7.4: *Subjects' average responses as a function of the smoothing filter order for a single virtual loudspeaker (upper panels) and for two simultaneous virtual loudspeakers (lower panels). The left panels show data for phase smoothing only, the middle panels for magnitude smoothing only, and the right panels for combined magnitude and phase smoothing. The downward triangles correspond to virtual loudspeakers at an azimuth of 0°, and the upward triangles and diamonds to 30 and 120° azimuth, respectively.*

7.3.3 Results

The subjects' responses averaged over fragment and subject are shown in Fig. 7.4. The left panels show data for phase smoothing only, the middle panels for magnitude smoothing only. The right panels correspond to combined magnitude and phase smoothing. The upper panels correspond to a single virtual loudspeaker, the lower panels correspond to two simultaneous virtual loudspeakers. If the data for a single virtual loudspeaker are considered, filter orders of 1 and higher do not result in audible artifacts for phase smoothing or magnitude smoothing, independent of the position of the virtual loudspeaker. Below filter order 1, subjects indicate very small audible differences for phase smoothing, small audible artifacts for magnitude smoothing and clear effects of the combined smoothing.

The lower panels show the data for 2 simultaneous loudspeakers, at $\pm$ 30 or $\pm$ 120° azimuth (triangles and diamonds, respectively). Clearly, in these conditions, phase smoothing does not result in audible changes (see lower-left panel). Magnitude smoothing results in audible artifacts if the filter order is at or below 0.5 for sources at $\pm$ 30° azimuth. For sources at $\pm$ 120° azimuth audible artefacts are only reported for filter orders below 0.5. The combined magnitude and phase smoothing results in very similar data as for magnitude smoothing alone. Overall, it can be concluded that for two virtual loudspeakers more smoothing can be allowed to result in similar perceptual stimulus degradation as for a single virtual loudspeaker.

7.4   Model Predictions

The binaural detection model described in this thesis was used to simulate the perceptual consequences of HRTF smoothing. The reader is referred Chapter 4 for the complete details of the model. Only the general model setup will be discussed here.

The model consists of three stages. The first stage comprises a peripheral pre-processing stage. The three most prominent features of this stage are:

- Filtering of the gammatone filterbank. The filterbank present in the peripheral processing stage determines the spectral resolution of the model, in line with the ERB estimates published by Glasberg and Moore (1990).

- Inner haircell model. This stage consists of a half-wave rectifier followed by a fifth-order lowpass filter with a cutoff frequency (-3dB) of 770 Hz. Hence below 770 Hz, both the ITDs and IIDs are preserved at the output of this stage. However, above 2 kHz, the output approximates the envelope of the incoming signals and hence only IIDs and ITDs present in the envelope are preserved. For frequencies in between, the ITD in the fine structure waveforms is gradually lost.

- Adaptation loops. The chain of adaptation loops in the peripheral processor has an almost logarithmic input-output characteristic in steady state and is a non-linear device. It has been shown frequently that for both monaural and binaural detection of signals added to a wideband masker with a variable level, the threshold *signal-to-masker* ratio is approximately constant, as long as the masker level is well above the absolute threshold (cf. McFadden, 1968; Hall and Harvey, 1984). If it is assumed that a certain constant *change* at the output of the adaptation loops is needed to detect a signal, the signal must be equal to a certain *fraction* of the masker level due to the logarithmic compression. Hence the signal-to-masker ratio will be approximately constant at threshold.

The second stage comprises binaural interaction based on an Equalization-Cancellation (EC) mechanism (Durlach, 1963, 1972). For each frequency channel, the squared difference between the waveforms from the left and right peripheral preprocessors is computed *as a function of an internal interaural delay τ (in seconds) and an internal interaural level adjustment α (in dB)* by so-called EI-type (Excitation-Inhibition) elements. These squared-difference signals are then fed through a temporal averager to account for a limited binaural temporal resolution. This process is performed for all center frequencies of the gammatone filterbank, resulting in a set of 3-dimensional activity patterns which usually have a minimum somewhere in these patterns. At this minimum, the externally presented IID and ITD at that frequency are compensated optimally by the internal delay and level adjustments. Hence
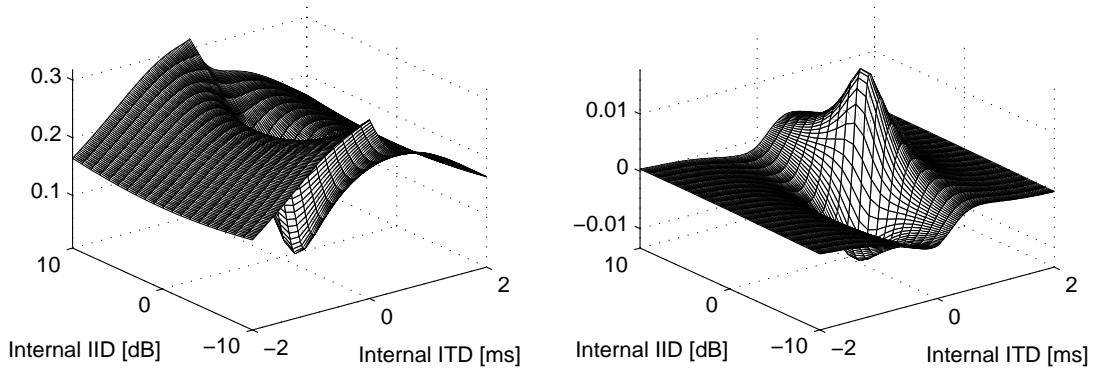
Figure 7.5: *Example of the effect of the combined phase and magnitude smoothing on the internal representation of the model. The left panel shows the internal activity pattern at 250-Hz center frequency as a function of the internal delay and internal intensity difference for an unprocessed HRTF pair corresponding to 30° azimuth. The right panel shows the* change *in the pattern due to HRTF smoothing for* n=1.

the position of the minimum depends on the interaural time and intensity difference that was present in the stimulus at that center frequency.

The third stage, the central processor, receives both binaural (from the binaural processor) as well as monaural (directly from the adaptation loops) information. These inputs are all corrupted by additive internal noise to limit their resolution. The task of this stage is to compute an overall difference measure between two different internal representations.

The model predictions were obtained in the following way. First, an audio fragment which was filtered by an unprocessed HRTF set was fed through the model. This stimulus resulted in a certain internal representation. Such an internal representation can be divided into a monaural component and a binaural component, shown by the direct connections from peripheral pre-processor to the central processor and the outputs of the binaural processor to the central processor, respectively. These channels represent monaural cues, such as timbre or overall power changes. On the other hand, the outputs of the binaural processor supply binaural properties of the presented waveforms, such as the IID or ITD present in the stimulus. An example of the computed binaural activity pattern generated by the binaural processor is shown in the left panel of Fig. 7.5 as a function of the internal ITD and IID. The picture was generated for a virtual loudspeaker radiating white noise at 30° azimuth and 0° elevation. The center frequency of the model was set to 250 Hz. There is a clear minimum at a certain small internal delay, which just compensates for the ITD present in the stimulus due to HRTF filtering.

If the stimuli are filtered by smoothed HRTFs and subsequently fed through the model, the internal representation of that stimulus will be slightly different from the internal representation of stimuli filtered by the original HRTFs.

This change is the cue for detection by the model. This cue can be purely monaural, for example the smoothing of peaks and dips in the magnitude spectrum, but may also be binaural, such as changes in the interaural phase spectrum. The change in activity between the two internal representations was computed for each frequency and as a function of time, and both for monaural and binaural channels. As an example, the right panel shows the *change* in the binaural activity pattern due to combined phase and magnitude HRTF smoothing with order n=1, again at a center frequency of 250 Hz.

To result in a single measure of distance or difference between two internal representations, the changes in the internal representations (for both monaural and binaural cues at all center frequencies) are combined into one difference measure. This difference measure is obtained as follows. For each combination of sound source position and audio fragment, the internal representation of the corresponding stimulus was computed. This resulted in a time-varying model activity for each frequency channel and α, τ combination. However, only two monaural (from the left and right ears) and one binaural output (i.e., one α, τ combination) per frequency channel was used in the detection process. The α, τ values were obtained by computing the time-averaged binaural output per frequency channel and selecting α and τ that corresponded to the minimum average activity. The difference in internal representations between stimuli filtered by smoothed and unprocessed HRTFs was subsequently computed for each filter. To obtain a single time-varying distance measure, these differences were combined across frequency channels according to an optimal criterion (see Chapter 4). To account for the (limited) temporal integration ability of human listeners, this output was smoothed by a double-exponential averaging window with an equivalent rectangular duration of 300 ms. The maximum of this smoothed output was used as the overall distance measure. Because all internal channels are corrupted by internal noise, the overall distance measure will also be corrupted by internal errors. We therefore use the detectability index d', defined as the mean value of the difference (i.e., without noise), divided by the standard deviation of the noise on the decision variable. Thus, d' serves as a measure of detectability of the change in the internal representation due to HRTF smoothing. A low value of d' ($\leq$1) denotes inaudible changes or changes near threshold, while large values of d' >1 correspond to clearly audible artifacts. Values for d' were computed for each stimulus and virtual source position. The results are given in Fig. 7.6. The format is the same as in Fig. 7.4, except for the fact that the subject responses are replaced by the output of the model.
As expected, the model output increases monotonically with a decrease in filter order. Similar to the results obtained with human listeners, phase smoothing results in less audible artifacts than magnitude smoothing. Furthermore, more smoothing is necessary to obtain a similar model output for two virtual loudspeakers than for one virtual loudspeaker. To make a more quantitative
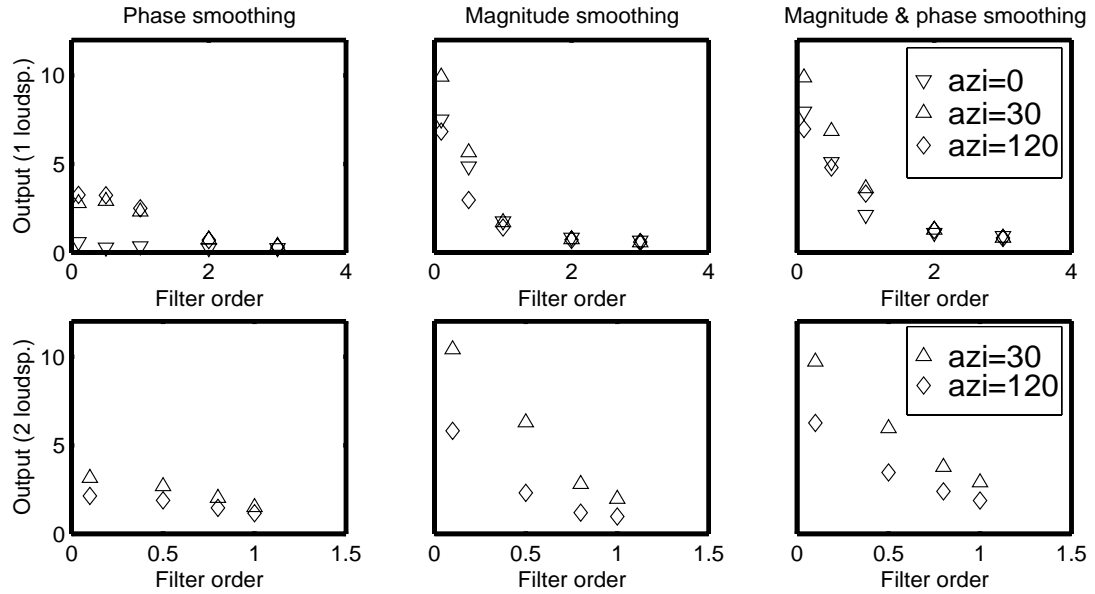
Figure 7.6: *Model output as a function of the smoothing filter order. The format is the same as in Fig. 7.4.*

comparison between model output and subject responses, we compared the responses of the listeners with the model output for corresponding stimuli. The results of this comparison are shown in Fig. 7.7. The abscissa gives the model output (d'), the ordinate the corresponding response of the subjects. The values for d' were averaged across audio fragments and plotted for each parameter combination (i.e., smoothing order, number of virtual loudspeakers and position of the virtual loudspeakers). The left panel corresponds to a single virtual loudspeaker, the right panel for 2 virtual loudspeakers. The solid lines are linear fits to the data. Several remarks can be made with respect to Fig. 7.7.

- A high correlation exists between model output and subjects' response, both for single and multiple virtual sources.

- Predictions for phase smoothing, magnitude smoothing and the combined processing results in a similar relationship between model output and subjects' response. This relation can be successfully described by a linear fit. This indicates that the model can predict the perceptual consequences of different types of smoothing and transform differences in stimuli into one single difference measure.

- For single and multiple sources, the linear relation is not exactly equal, but very similar. In Fig. 7.3, it has been shown that subjects tolerate more smoothing for multiple sources than for a single virtual source. A similar relation for the model was demonstrated in Fig. 7.6. Thus, the model correctly predicts that more smoothing is allowed for multiple sources compared to a single source to result in similar stimulus degradation.
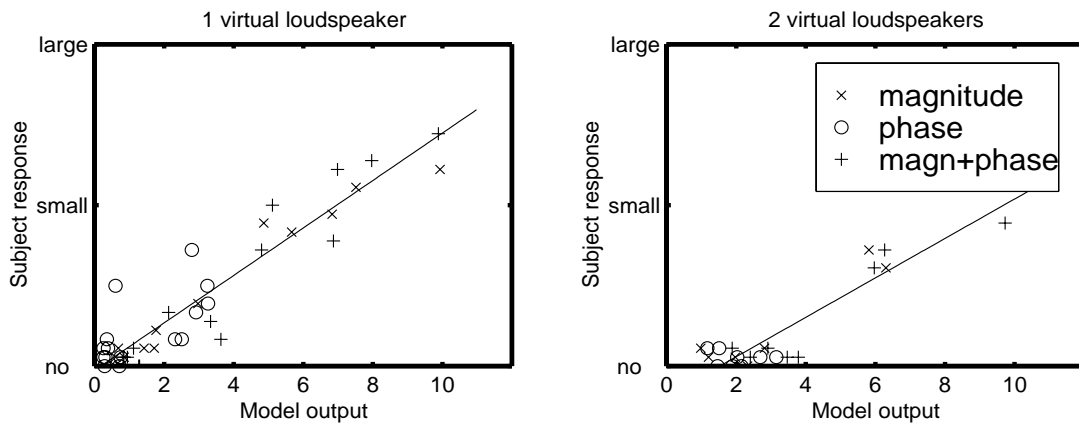
Figure 7.7: *Subjects' responses of the experiments described in Section 7.3 as a function of the model output in terms of d'. The left panel shows data for a single virtual loudspeaker, the right panel for two simultaneous virtual loudspeakers. The crosses denote magnitude smoothing, the circles phase smoothing and the plus signs combined magnitude and phase smoothing. The solid lines are linear fits.*

- The values for d' which correspond to small audible artifacts amount about 8. This is significantly higher than the d' value of about +1 which is used to define the threshold of detectability in critical listening tests (see Chapters 5 and 6 of this thesis). This difference may be the result of the experimental procedure that was used in our experiments. Obviously, the procedure in our experiments does not yield maximum detection performance. Pilot studies revealed that experiments with noise instead of musical fragments combined with many repetitions of the same stimulus resulted in somewhat higher reported artifacts. Furthermore, subjects were asked to rate the audibility of the artifacts and were not instructed to optimally detect the presence of artifacts. A third rationale for high d' values may be related to the fact that the subjects did not listen through their own ears, i.e., nonindividual HRTFs were used to generate virtual speakers. The use of individualized HRTFs may result in larger reported smoothing artifacts. A fourth reason for high d' values may be related to the method of deriving the model predictions. The maximum d' value in time occurring for each audio fragment was used as model prediction. It may well be the case that listeners do not base their response on the maximum audible artifact that occurs during one interval but that they form some kind of average distortion measure. If the model predictions were based on an average output rather than the maximum output, lower model outputs are expected.

7.5 Discussion and conclusions

The results of the experiments demonstrate that the complex phase and magnitude spectra present in anechoic HRTFs can be simplified by the assumption that their spectral resolution does not have to exceed the spectral resolution of the cochlea. Specifically, a first-order gammatone filterbank

with bandwidths of 1 ERB is sufficient to describe the frequency dependence of both the phase and magnitude spectra. The amount of crosstalk between filters of a first-order filter is substantially more than for a 4th-order filterbank, which is usually used to describe the spectral resolution of the auditory system. Furthermore, when the smoothing is strong enough to lead to audible differences, these are stronger for mangitude than for phase smoothing. Even if the phase spectra are almost linear (n=0.1), subjects indicate only very small differences, or do not report differences at all. This result is in line with the data from Kulkarni and Colburn (1998) and Kulkarni *et al.* (1999) where it is shown that linear phase HRTFs are not discriminable from unprocessed HRTFs and that the fine structure of the HRTFs is relatively unimportant for auditory spatial attributes. Our results also suggest that anechoic HRTFs can successfully be described by a linear-phase filter with a magnitude spectrum with a limited resolution, as long as the interaural delay matches the average delay found in the original HRTF set.

If more virtual loudspeakers are combined which have a partially overlapping spectrum, even more smoothing is allowed than for a single virtual source. This is due to masking across channels which results in a decreased sensitivity for interaural parameters of the different sources.

The model can successfully describe the perceptual degradation of the smoothing process, both for phase and magnitude smoothing. It can also account for the difference between one and two virtual loudspeakers. The transformation between physical differences between stimuli to the distance measure provided by the model results in a metric which is highly correlated with the scaled perceptual differences reported by subjects.

Conclusions

8.1 Summary of findings

The main part of this thesis describes and validates a binaural detection model. The binaural interaction stage of this model is in part based on the results given in Chapter 2. In this chapter, the relative contribution of various interaural cues to binaural unmasking was investigated. The results revealed that for combinations of static and dynamically-varying ITDs and IIDs, the energy of the difference signal between the left and right ears is a detection statistic that describes these thresholds satisfactorily. Furthermore, the data did not show any effect of binaural sluggishness: the rate of fluctuation of the interaural differences did not influence the detection thresholds. These results indicate that the detection of dynamically-varying ITDs and IIDs against a masker without interaural differences is not influenced by the limited temporal resolution of the binaural auditory system.

In Chapter 3, the influence of uncertainty in the binaural cues was investigated. $N\rho S\pi$ thresholds for running and frozen-noise maskers suggested that in these conditions, the thresholds can not be described by the overall interaural cross correlation or the energy of the difference signal per se, but that the interaural cues are evaluated by a template-matching procedure.

The idea of template matching based on the energy of the difference signal was incorporated in a time-domain detection model (Chapter 4). This model transforms arbitrary stimuli into an internal representation. This representation comprises four dimensions: time, frequency channel, internal interaural delay and internal interaural level adjustment. The simulated activity as a function of these dimensions entails both binaural and monaural properties of the presented stimulus. The resolution of these properties is limited by the addition of internal noise. It is assumed that the model can detect changes in the externally presented stimuli if these changes in time and frequency result in differences in the internal representation that are at least as large as the internal noise. However, the model is able to increase the signal-to-internal noise ratio by integrating information across time and frequency.

The model's ability to account for detection thresholds as a function of spectral stimulus parameters was demonstrated in Chapter 5. Good fits were obtained for changes in the frequency, bandwidth, level and interaural phase of both masker and signal. The most prominent finding was that the model can account for the wider effective critical bandwidth observed in band-widening $NoS\pi$ experiments. In our model, this phenomenon is the result of the fact that the cue for detection is available in a range of filters if the masker bandwidth is sufficiently small. For bandwidths increasing from the critical bandwidth to two or three times that value, the model's ability to integrate information across frequency is gradually lost. This results in increasing thresholds, even though the amount of masker energy within the on-frequency channel hardly changes. The increased effective bandwidth does therefore not reflect a worse spectral resolution but it is the result of the ability to integrate information across frequency.

The influence of temporal stimulus properties on the model's predictions were described in Chapter 6. Both the effect of temporal integration and temporal resolution was investigated. Similar to the results in the spectral domain, the temporal averager followed by an optimal detector results in good fits for most of the data. Good examples for the support of an optimal detector in the time domain are the predictions for sinusoidally and stepwise interaural correlation changes. If these data are analyzed without an optimal detector (i.e., at the temporal center of the signal), the estimates of the time constant of the binaural system for both experimental procedures differ by a factor 2 to 3. In contrast, the optimal detector predicts very similar time constants in both conditions. Three experiments did not give satisfactory predictions. The first comprised interaural correlation jnds as a function of the bandwidth of the stimulus. The second condition comprised the detection of time-varying ITDs in broadband noise. Finally, the strong overshoot in the peripheral adaptation loops resulted in lower detection thresholds for a binaural signal presented during the first 25 ms of the masker.

Chapter 7 described the perceptual evaluation of the simplification of HRTF phase and magnitude spectra. The results of this study support the hypothesis that monaural and binaural properties of anechoic HRTF pairs do not require a higher spectral resolution than the resolution of the filterbank present in the model. Moreover, the model could predict the perceptual degradation for supra-threshold smoothing, both for single and dual sound sources.

8.2 Future work

The temporal dynamics and nonlinear compression effects of the adaptation loops are useful in understanding a number of binaural effects like binaural forward masking and the influence of overall interaural level differences on binaural unmasking and on lateralization, while for some specific conditions, their temporal dynamic is too strong. Obviously, the current implementation of the adaptation loops is not the optimal realization of this stage.

Although it has often been suggested that models based on an EC mechanism and models based on the cross correlation are essentially similar, we have explained that in some conditions, these binaural interactions either differ in their predictions or require different assumptions to account for specific phenomena. We have shown that from a neurophysiological point of view, there is a strong support for models based on the cross correlation. On the other hand, neurophysiological studies also show that EI-like interactions exist in the mammalian auditory system. It is therefore interesting to investigate the possibility and necessity of combining these binaural interactions into one model.

The model has so far mainly been evaluated for binaural signal detection purposes. It would be worthwhile to include the ability to estimate the direction and compactness of a sound source, given a certain internal representation. In principle, the necessary cues are available in the model (i.e., the binaural parameters IID and ITD and the spectral content). One of the arising problems is that the model does not know the spectral properties of the sound source itself at forehand. Hence it is difficult to discriminate between spectral features that are part of the sound source itself and features induced by the transfer function from a certain position to the eardrums. Furthermore, the relation between the physical cues for localization and the direction of the sound source cannot be given analytically. Hence to be able to determine the direction of the sound, an auditory feature detector, a neural network (Janko *et al.*, 1997) or pattern recognition algorithm (Hartung and Sterbing, 1998; Langendijk and Bronkhorst, 2001) is required in addition to the monaural and binaural processing stages included in the present binaural detection model.

BIBLIOGRAPHY

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). "A binaural analog of gap detection," J. Acoust. Soc. Am. **105**, 2807–2820.

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (**1997**a). "Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary comples. I. Heterogenity of responses," J. Neurophysiol. **78**, 1222–1236.

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (**1997**b). "Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary comples. II. Coincidence detection," J. Neurophysiol. **78**, 1237–1247.

Bernstein, L. R., and Trahiotis, C. (**1996**). "The normalized correlation: accounting for binaural detection across center frequency," J. Acoust. Soc. Am. **100**, 3774–3787.

Bernstein, L. R., and Trahiotis, C. (**1997**). "The effects of randomizing values of interaural disparities on binaural detection and on discrimination of interaural correlation," J. Acoust. Soc. Am. **102**, 1113–1120.

Bernstein, L. R., and Trahiotis, C. (**1999**). "The effects of signal duration on NoSo and NoSπ thresholds at 500 Hz and 4 kHz," J. Acoust. Soc. Am. **105**, 1776–1783.

Bilsen, F. A. (**1977**). "Pitch of noise signals: Evidence for a "central spectrum"," J. Acoust. Soc. Am. **61**, 150–161.

Bilsen, F. A., and Goldstein, J. L. (**1974**). "Pitch of dichotically delayed noise and its possible spectral basis," J. Acoust. Soc. Am. **55**, 292–296.

Bilsen, F. A., and Raatgever, J. (**2000**). "On the dichotic pitch of simultaneously presented interaurally delayed white noises. Implications for binaural theory," J. Acoust. Soc. Am. **108**, 272–284.

Blauert, J. (**1997**). *Spatial hearing: the psychophysics of human sound localization*, (the MIT Press, Cambridge, Massachusetts).

Blauert, J., Brüggen, M., Hartung, K., Bronkhorst, A., Drullmann, R., Reynaud, G., Pellieux, L., Krebber, W., and Sotteck, R. (**1998**). "The audis catalog of human hrtfs," in *Proc. 16th Int. Congr. Acoust.*, ICA, Inst. of Physics, USA-NY, CD-ROM.

Bos, C. E., and de Boer, E. (**1966**). "Masking and discrimination," J. Acoust. Soc. Am. **39**, 708–715.

Boudreau, J. C., and Tsuchitani, C. (**1968**). "Binaural interaction in the cat superior olive S segment," J. Neurophysiol. **31**, 442–454.

Bourbon, W. T. (**1966**). *Effects of bandwidth and level of masking noise on detection of homophasic and antiphasic tonal signals*, Ph.D. thesis, University of Texas, Austin, Texas.

Bourbon, W. T., and Jeffress, L. A. (**1965**). "Effect of bandwidth of masking noise on detection of homophasic and antiphasic tonal signals," J. Acoust. Soc. Am. **37**, 1180–1181.

Breebaart, J., and Kohlrausch, A. (**1999**). "A new binaural detection model based on contralateral inhibition," in *Psychophysics, physiology and models of hearing*, edited by T. Dau, V. Hohmann, and B. Kollmeier, (World Scientific, Singapore), pp. 195–206.

Breebaart, J., and Kohlrausch, A. (**2001**a). "The influence of interaural stimulus uncertainty on binaural signal detection," J. Acoust. Soc. Am. **109**, 331–345.

Breebaart, J., and Kohlrausch, A. (**2001**b). "Perceptual (ir)relevance of HRTF magnitude and phase spectra," J. Audio Eng. Soc. **Submitted**.

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

Breebaart, J., van de Par, S., and Kohlrausch, A. (**1999**). "The contribution of static and dynamically varying ITDs and IIDs to binaural detection," J. Acoust. Soc. Am. **106**, 979–992.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**a). "Binaural processing model based on contralateral inhibition. I. Model setup." J. Acoust. Soc. Am. **Accepted for publication**.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**b). "Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters." J. Acoust. Soc. Am. **Accepted for publication**.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**c). "Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters." J. Acoust. Soc. Am. **Accepted for publication**.

Bronkhorst, A. W. (**2001**). "Effect of stimulus properties on auditory distance perception in rooms," in *Psyiological and psychophysical bases of auditory function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. Prijs, and R. Schoonhoven, (Shaker Publishing BV, Maastricht), pp. 184–191.

Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources. II. Localization of a broadband source," J. Acoust. Soc. Am. **106**, 1956–1968.

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

Cheung, N., Trautmann, S., and Horner, A. (**1998**). "Head-related transfer function modeling in 3-D sound systems with genetic algorithms," J. Audio Eng. Soc. **46**, 531–539.

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

Colburn, H. S. (**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," J. Acoust. Soc. Am. **61**, 525–533.

Colburn, H. S., and Durlach, N. I. (**1978**). "Models of binaural interaction," in *Handbook of Perception*, edited by E. Carterette, and M. Friedman, (Academic Press, New York, San Francisco, London), vol. IV: Hearing, pp. 467–518.

Colburn, H. S., and Isabelle, S. K. (**2001**). "Physiologically based models of binaural detection," in *Psyiological and psychophysical bases of auditory function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. Prijs, and R. Schoonhoven, (Shaker Publishing BV, Maastricht), pp. 161–168.

Colburn, H. S., Isabelle, S. K., and Tollin, D. J. (**1997**). "Modeling binaural detection performance for individual masker waveforms," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey, and T. Anderson, (Lawrence Erlbaum Ass.), chap. 25, pp. 533–555.

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 in discrimination and detection," J. Acoust. Soc. Am. **64**, 95–106.

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

Culling, J. F., and Summerfield, Q. (**1998**). "Measurements of the binaural temporal window using a detection task," J. Acoust. Soc. Am. **103**, 3540–3553.

Culling, J. F., Summerfield, Q., and Marshall, D. H. (**1996**). "Dichotic pitches as illusions of binaural unmasking," J. Acoust. Soc. Am. **99**, 2515–2529.

Dau, T. (**1992**). *Der optimale Detektor in einem Computermodell zur Simulation von psychoakustischen Experimenten*, Master's thesis, Universität Göttingen, Göttingen.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation: II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**a). "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**b). "A quantitative model of the 'effective' signal processing in the auditory system: II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631.

Domnitz, R. H., and Colburn, H. S. (**1976**). "Analysis of binaural detection models for dependence on interaural target parameters," J. Acoust. Soc. Am. **59**, 598–601.

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 *Foundations of modern auditory theory*, edited by J. Tobias, (Academic Press, New York, London), vol. II, pp. 369–462.

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination: II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548–1557.

Durlach, N. I., and Pang, X. D. (**1986**). "Interaural magnification," J. Acoust. Soc. Am. **80**, 1849–1850.

Egan, J. P., Lindner, W. A., and McFadden, D. (**1969**). "Masking-level differences and the form of the psychometric function," Perc. Psychoph. **6**, 209–215.

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

Gabriel, K. J., and Colburn, H. S. (**1981**). "Interaural correlation discrimination: I. Bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394–1401.

Gaik, W. (**1993**). "Combined evaluation of interaural time and intensity differences: psychoacoustic results and computer modeling," J. Acoust. Soc. Am. **94**, 98–110.

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

Goldberg, J. M., and Brown, P. B. (**1969**). "Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization," J. Neurophysiol. **32**, 613–636.

Grantham, D. W. (**1984**a). "Discrimination of dynamic interaural intensity differences," J. Acoust. Soc. Am. **76**, 71–76.

Grantham, D. W. (**1984**b). "Interaural intensity discrimination: insensitivity at 1000 Hz," J. Acoust. Soc. Am. **75**, 1191–1194.

Grantham, D. W., and Robinson, D. E. (**1977**). "Role of dynamic cues in monaural and binaural signal detection," J. Acoust. Soc. Am. **61**, 542–551.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

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. (**1992**). "On the similarity of two theories of comodulation masking release," J. Acoust. Soc. Am. **91**, 1769.

Green, D. M. (**1966**). "Signal-detection analysis of equalization and cancellation model," J. Acoust. Soc. Am. **40**, 833–838.

Green, D. M., and Swets, J. A. (**1966**). *Signal detection theory and psychophysics*, (Wiley, New York).

Hafter, E. R. (**1971**). "Quantitative evaluation of a lateralization model of masking-level differences," J. Acoust. Soc. Am. **50**, 1116–1122.

Hafter, E. R., Bourbon, W. T., Blocker, A. S., and Tucker, A. (**1969**). "A direct comparison between lateralization and detection under conditions of antiphasic masking," J. Acoust. Soc. Am. **46**, 1452–1457.

Hafter, E. R., and Carrier, S. C. (**1970**). "Masking-level differences obtained with pulsed tonal maskers," J. Acoust. Soc. Am. **47**, 1041–1047.

Hafter, E. R., and Carrier, S. C. (**1972**). "Binaural interaction in low-frequency stimuli: the inability to trade time and intensity completely," J. Acoust. Soc. Am. **51**, 1852–1862.

Hall, J. W., and Fernandes, M. A. (**1984**). "The role of monaural frequency selectivity in binaural analysis," J. Acoust. Soc. Am. **76**, 435–439.

Hall, J. W., and Harvey, A. D. G. (**1984**). "NoSo and NoSπ thresholds as a function of masker level for narrow-band and wideband masking noise," J. Acoust. Soc. Am. **76**, 1699–1703.

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.

Hartmann, W. M., and Wittenberg, A. (**1996**). "On the externalization of sound images," J. Acoust. Soc. Am. **99**, 3678–3688.

Hartung, K., and Sterbing, S. J. (**1998**). "A physiology-based computational model for sound localization," in *Psychophysics, physiology and models of hearing*, edited by T. Dau, V. Hohmann, and B. Kollmeier, (World Scientific), pp. 179–184.

Hirsh, I. (**1948**a). "Binaural summation and interaural inhibition as a function of the level of masking noise," Am. J. Psychol. **61**, 205–213.

Hirsh, I. (**1948**b). "The influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.

Hirsh, I., and Burgeat, M. (**1958**). "Binaural effects in remote masking," J. Acoust. Soc. Am. **30**, 827–832.

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

Holube, I., Colburn, H. S., van de Par, S., and Kohlrausch, A. (**1995**). "Model simulations of masked thresholds for tones in dichotic noise maskers," J. Acoust. Soc. Am. **97**, 3411–3412.

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

Huopaniemi, J., and Zacharov, N. (**1999**). "Objective and subjective evaluation of head-related transfer function filter design," J. Audio. Eng. Soc. **47**, 218–239.

Irvine, D. R. F., and Gago, G. (**1990**). "Binaural interaction in high-frequency neurons in inferior colliculus of the cat: effects of variations in sound pressure level on sensitivity to interaural intensity differences," J. Neurophysiol. **63**, 570–591.

Janko, J. A., Anderson, T. R., and Gilkey, R. H. (**1997**). "Using neural networks to evaluate the viability of monaural and interaural cues for sound localization," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey, and T. Anderson, (Lawrence Erlbaum Ass.), chap. 26, pp. 557–570.

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

Jeffress, L. A., Blodgett, H. C., and Deatherage, B. H. (**1952**). "The masking of tones by white noise as a function of the interaural phases of both components. I. 500 cycles," J. Acoust. Soc. Am. **24**, 523–527.

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., Blodgett, H. C., Sandel, T. T., and Wood, C. L. (**1956**). "Masking of tonal signals," J. Acoust. Soc. Am. **28**, 1416–426.

Jeffress, L. A., and McFadden, D. (**1968**). "MLD's and the phase angle, alpha," J. Acoust. Soc. Am. **43**, 164.

Jeffress, L. A., and McFadden, D. (**1971**). "Differences of interaural phase and level in detection and lateralization," J. Acoust. Soc. Am. **49**, 1169–1179.

Johannesma, P. I. M. (**1972**). "The pre-response stimulus ensemble of neurons in the cochlear nucleus," in *Proceedings of the Symposium of Hearing Theory*, IPO, Eindhoven, The Netherlands.

Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.

Joris, P. X. (**1996**). "Envelope coding in the lateral superior olive. II. Characteristic delays and comparison with responses in the medial superior olive," J. Neurophysiol. **76**, 2137–2156.

Joris, P. X., and Yin, T. C. T. (**1995**). "Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences," J. Neurophysiol. **73**, 1043–1062.

Kiang, N. Y. S. (**1975**). *The Nervous System*, (Raven Press, New York), vol. 3, chap. Stimulus representation in the discharge patterns of auditory neurons.

King, W. G., and Laird, D. A. (**1930**). "The effect of noise intensity and pattern on locating sounds," J. Acoust. Soc. Am. **2**, 99–102.

Kistler, D. J., and Wightman, F. L. (**1992**). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91**, 1637–1647.

Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.

Kohlrausch, A. (**1986**). "The influence of signal duration, signal frequency and masker duration on binaural masking level differences," Hearing Research **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. II: Influence of frequency and interaural-phase uncertainty," J. Acoust. Soc. Am. **88**, 1749–1756.

Kohlrausch, A., and Fassel, R. (**1997**). "Binaural masking level differences in nonsimultaneous masking," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey, and T. Anderson, (Lawrence Erlbaum Ass.), chap. 9, pp. 169–190.

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 conditions." J. Acoust. Soc. Am. **92**, 1889–1901.

Kulkarni, A., and Colburn, H. S. (**1998**). "Role of spectral detail in sound-source localization," Nature **396**, 747–749.

Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (**1999**). "Sensitivity of human subjects to head-related transfer-function phase spectra," J. Acoust. Soc. Am. **105**, 2821–2840.

Kuwada, S., Yin, T. C. T., Syka, J., Buunen, T. J. F., and Wickesberg, R. E. (**1984**). "Binaural interaction in low-frequency neurons in inferior colliculus of the cat. IV. Comparison of monaural and binaural response properties," J. Neurophysiol. **51**, 1306–1325.

Langendijk, E. H. A., and Bronkhorst, A. W. (**2000**). "Fidelity of three-dimensional-sound reproduction using a virtual auditory display," J. Acoust. Soc. Am. **107**, 528–537.

Langendijk, E. H. A., and Bronkhorst, A. W. (**2001**). "Contribution of spectral cues to human sound localization," J. Acoust. Soc. Am. p. Under review.

Langhans, A., and Kohlrausch, A. (**1992**). "Spectral integration of broadband signals in diotic and dichotic masking experiments," J. Acoust. Soc. Am. **91**, 317–326.

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. **20**, 150–159.

Lindemann, W. (**1985**). *Die Erweiterung eines Kreuzkorrelationsmodells der binauralen Signalverarbeitung durch kontralaterale Inhibitionsmechanismen*, Ph.D. thesis, Ruhr-Universität Bochum, Bochum.

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

Lutfi, R. A. (**1990**). "How much masking is informational masking?" J. Acoust. Soc. Am. **88**, 2607–2610.

Makoes, J. C., and Middlebrooks, J. C. (**1990**). "Two-dimensional sound localization by human listeners," J. Acoust. Soc. Am. **87**, 2188–2200.

McAlpine, D., Jiang, D., Shackleton, T. M., and Palmer, A. R. (**1998**). "Convergent input from brainstem coincidence detectors onto delay-sensitive neurons in the inferior colliculus," J. Neurosci. **18**, 6026–6039.

McFadden, D. (**1966**). "Masking-level differences with continuous and with burst masking noise," J. Acoust. Soc. Am. **40**, 1414–1419.

McFadden, D. (**1968**). "Masking-level differences determined with and without interaural disparities in masker intensity," J. Acoust. Soc. Am. **44**, 212–223.

McFadden, D., Jeffress, L. A., and Ermey, H. L. (**1971**). "Difference in interaural phase and level in detection and lateralization: 250 Hz," J. Acoust. Soc. Am. **50**, 1484–1493.

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

Mills, A. (**1960**). "Lateralization of high-frequency tones," J. Acoust. Soc. Am. **32**, 132–134.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Moore, J. K. (**1987**). "The human auditory brain stem: a comparative view," Hear. Res. **29**, 1–32.

Osman, E. (**1971**). "A correlation model of binaural masking level differences," J. Acoust. Soc. Am. **50**, 1494–1511.

Palmer, A. R., McAlpine, D., and Jiang, D. (**1997**). "Processing of interaural delay in the inferior colliculus," in *Acoustical signal processing in the central auditory system*, edited by J. Syka, (Plenum Press, New York), pp. 353–364.

Park, T. J. (**1998**). "IID sensitivity differs between two principal centers in the interaural intensity difference pathway: the LSO and the IC," J. Neurophysiol. **79**, 2416–2431.

Park, T. J., Monsivais, P., and Pollak, G. D. (**1997**). "Processing of interaural intensity differences in the LSO: role of interaural threshold differences," J. Neurophysiol. **77**, 2863–2878.

Patterson, R. D., Holdsworth, J., Nimmo-Smith, I., and Rice, P. (**1988**). "Svos final report: The auditory filterbank," Tech. Rep. APU report 2341.

Perrett, S., and Noble, W. (**1997**). "The effect of head rotations on vertical plane sound localization," J. Acoust. Soc. Am. **102**, 2325–2332.

Perrott, D., and Nelson, M. (**1969**). "Limits for the detection of binaural beats," J. Acoust. Soc. Am. **46**, 1477–1480.

Perrott, D. R., and Musicant, A. D. (**1977**). "Minimum auditory movement angle: binaural localization of moving sound sources," J. Acoust. Soc. Am. **62**, 1463–1466.

Plack, C. J., and Moore, B. C. J. (**1990**). "Temporal window shape as a function of frequency and level," J. Acoust. Soc. Am. **87**, 2178–2187.

Punch, J., and Carhart, R. (**1973**). "Influence of interaural phase on forward masking," J. Acoust. Soc. Am. **54**, 897–904.

Raatgever, J., and Bilsen, F. (**1986**). "A central spectrum theory of binaural processing. Evidence from dichotic pitch," J. Acoust. Soc. Am. **80**, 429–441.

Raatgever, J., and van Keulen, W. (**1992**). "Binaural time processing at high frequencies: The central spectrum model extended," in *Proceedings of the 14th International Congress on Acoustics*, pp. H6–3.

Rabiner, L., Laurence, C., and Durlach, N. (**1966**). "Further results on binaural unmasking and the EC model," J. Acoust. Soc. Am. **40**, 62–70.

Recanzone, G. H., Makhamra, S. D. D. R., and Guard, D. C. (**1998**). "Comparisons of relative and absolute sound localization ability in humans," J. Acoust. Soc. Am. **103**, 1085–1097.

Reed, M. C., and Blum, J. J. (**1990**). "A model for the computation and encoding of azimuthal information by the lateral superior olive," J. Acoust. Soc. Am. **88**, 1442–1453.

Rice, S. O. (**1959**). *Selected papers on noise and stochastic processes*, (Dover publications, New York), chap. Mathematical analysis of random noise.

Richards, V. M. (**1987**). "Monaural envelope correlation perception," J. Acoust. Soc. Am. **82**, 1621–1630.

Robinson, D., and Jeffress, L. (**1963**). "Effect of varying the interaural noise correlation on the detectability of tonal signals," J. Acoust. Soc. Am. **35**, 1947–1952.

Robinson, D. E., Langford, T. L., and Yost, W. A. (**1974**). "Masking of tones by tones and of noise by noise," Percept. Psychophys. **15**, 159–167.

Robinson, D. E., and Trahiotis, C. (**1972**). "Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions," Percept. Psychphys. **12**, 333–334.

Rose, J. E., Gross, N. B., Geisler, C. D., and Hind, J. E. (**1966**). "Some neural mechanisms in the inferior colliculus of cat which may be relevant to localization of a sound source," J. Neurophysiol. **29**, 288–314.

Ruotolo, B. R., Stern, R. M., and Colburn, H. S. (**1979**). "Discrimination of symmetric time-intensity traded binaural stimuli," J. Acoust. Soc. Am. **66**, 1733–1737.

Sayers, B. M. (**1964**). "Acoustic image lateralization judgments with binaural tones," J. Acoust. Soc. Am. **36**, 923–926.

Schiano, J. L., Trahiotis, C., and Bernstein, L. R. (**1986**). "Lateralization of low-frequency tones and narrow bands of noise," J. Acoust. Soc. Am. **91**, 1563–1570.

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

Shackleton, T., Meddis, R., and Hewitt, M. (**1992**). "Across frequency integration in a model of lateralization," J. Acoust. Soc. Am. **92**, 2276–2279.

Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N. (**2000**). "Tori of confusion: binaural localization cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636.

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

Stern, R. M., and Bachorski, S. J. (**1983**). "Dynamic cues in binaural perception," in *Hearing - physiological bases and psychophysics*, edited by R. Klinke, and R. Hartmann, (Springer-Verlag, Berlin), pp. 209–215.

Stern, R. M., and Shear, G. D. (**1996**). "Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay," J. Acoust. Soc. Am. **100**, 2278–2288.

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). "Lateralization of complex binaural stimuli: A weighted-image model," J. Acoust. Soc. Am. **84**, 156–165.

Stern, R. M. J., 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.

Trahiotis, C., Dolan, T. R., and Miller, T. H. (**1972**). "Effect of "backward" masker fringe on the detectability of pulsed diotic and dichotic tonal signals," Percept. Psychophys. **12**, 335–338.

Tsuchitani, C. (**1997**). "Input from the medial nucleus of trapezoid body to an interaural level detector," Hear. Res. **105**, 211–224.

van de Par, S. (**1998**). *A comparison of binaural detection at low and high frequencies,* Ph.D. thesis, Eindhoven University of Technology, Eindhoven.

van de Par, S., and Kohlrausch, A. (**1995**). "Analytical expressions for the envelope correlation of certain narrow-band stimuli," J. Acoust. Soc. Am. **98**, 3157–3169.

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 de Par, S., and Kohlrausch, A. (**1998**a). "Analytical expressions for the envelope correlation of narrow-band stimuli used in CMR and BMLD research," J. Acoust. Soc. Am. **103**, 3605–3620.

van de Par, S., and Kohlrausch, A. (**1998**b). "Diotic and dichotic detection using multiplied-noise maskers," J. Acoust. Soc. Am. **103**, 2100–2110.

van de Par, S., and Kohlrausch, A. (**1998**c). "Further evidence for the influence of peripheral compression on binaural detection," in *Proceedings of the joint Acoust. Soc. Am. meeting and the International Congress on Acoustics, Seattle June 1998,* pp. 853–854.

van de Par, S., and Kohlrausch, A. (**1999**). "Dependence of binaural masking level differences on center frequency, masker bandwidth and interaural parameters," J. Acoust. Soc. Am. **106**, 1940–1947.

van de Par, S., and Kohlrausch, A. (**2001**). "The influence of basilar-membrane compression on binaural detection," J. Acoust. Soc. Am. **Under review**.

van de Par, S., Trahiotis, C., and Bernstein, L. R. (**2001**). "A consideration of the normalization that is typically included in correlation-based models of binaural detection," J. Acoust. Soc. Am. **109**, 830–833.

van der Heijden, M., and Trahiotis, C. (**1997**). "A new way to account for binaural detection as a function of interaural noise correlation," J. Acoust. Soc. Am. **101**, 1019–1022.

van der Heijden, M., and Trahiotis, C. (**1998**). "Binaural detection as a function of interaural correlation and bandwidth of masking noise: Implications for estimates of spectral resolution," J. Acoust. Soc. Am. **103**, 1609–1614.

von Hövel, H. (**1984**). *Zur Bedeutung der Übertragungseigenschaften des Aussenohres sowie des binauralen Hörsystems bei gestörter Sprachübertragung*, Ph.D. thesis, Technical University, 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.

Weis, T. F., and Rose, C. (**1988**). "A comparison of synchronization filters in different auditory receptor organs," Hear. Res. **33**, 175–180.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.

Wenzel, E. M., and Foster, S. H. (**1993**). "Perceptual consequences of interpolating head-related transfer functions during spatial synthesis," in *Proceedings of the 1993 workshop on applications of signal processing to audio and acoustics*, New York.

Weston, P., and Miller, J. (**1965**). "Use of noise to eliminate one ear from masking experiments," J. Acoust. Soc. Am. **37**, 638–646.

Wightman, F. (**1969**). "Binaural masking with sine-wave maskers," J. Acoust. Soc. Am. **45**, 71–78.

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

Wightman, F. L., and Kistler, D. J. (**1989**b). "Headphone simulation of free-field listening. I. Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1989**a). "Headphone simulation of free-field listening. II: psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.

Wightman, F. L., and Kistler, D. J. (**1999**). "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am. **105**, 2841–2853.

Wilbanks, W., and Whitmore, J. (**1968**). "Detection of monaural signals as a function of interaural noise correlation and signal frequency," J. Acoust. Soc. Am. **43**, 785–797.

Wilson, R., and Fowler, C. (**1986**). "Effects of signal duration on the 500-Hz masking-level difference," Scand. Audio. **15**, 209–215.

Wilson, R., and Fugleberg, R. (**1987**). "Influence of signal duration on the masking-level difference," J. Speech Hear. Res. **30**, 330–334.

Yama, M. (**1992**). "Effects of temporal separation and masker level on binaural analysis in forward masking," J. Acoust. Soc. Am. **91**, 327–335.

Yin, T. C. T., and Chan, J. C. K. (**1990**). "Interaural time sensitivity in medial superior olive of cat," J. Neurophysiol. **64**, 465–488.

Yost, W. A. (**1972**a). "Tone-in-tone masking for three binaural listening conditions," J. Acoust. Soc. Am. **52**, 1234–1237.

Yost, W. A. (**1972**b). "Weber's fraction for the intensity of pure tones presented binaurally," Percept. Psychophys. **11**, 61–64.

Yost, W. A. (**1981**). "Lateral position of sinusoids presented with interaural intensive and temporal differences," J. Acoust. Soc. Am. **70**, 397–409.

Yost, W. A. (**1985**). "Prior stimulation and the masking-level difference," J. Acoust. Soc. Am. **78**, 901–906.

Yost, W. A., Nielsen, D. W., Tanis, D. C., and Bergert, B. (**1974**). "Tone-on-tone binaural masking with an antiphasic masker," Percept. Psychophys. **15**, 233–237.

Zerbs, C. (**2000**). *Modeling the effective binaural signal processing in the auditory system*, Ph.D. thesis, Oldenburg University, Germany.

Zhang, M., Tan, K., and Er, M. H. (**1998**). "Three-dimensional sound synthesis based on head-related transfer functions," J. Audio. Eng. Soc. **146**, 836–844.

Zurek, P. M. (**1991**). "Probability distributions of interaural phase and level differences in binaural detection stimuli," J. Acoust. Soc. Am. **90**, 1927–1932.

Zurek, P. M., and Durlach, N. I. (**1987**). "Masker-bandwidth dependence in homophasic and antiphasic tone detection," J. Acoust. Soc. Am. **81**, 459–464.

Zwicker, E., and Henning, G. B. (**1985**). "The four factors leading to binaural masking-level differences," Hearing Research **19**, 29–47.

Zwicker, U. T., and Zwicker, E. (**1984**). "Binaural masking-level difference as a function of masker and test-signal duration," Hearing Research **13**, 215–220.

Zwislocki, J., and Feldman, R. S. (**1956**). "Just noticeable differences in dichotic phase," J. Acoust. Soc. Am. **28**, 860–864.

Summary

With the advent of multimedia technology and powerful signal processing systems, audio processing and reproduction has gained renewed interest. Examples of products that have been developed are audio coding algorithms to efficiently store and transmit music and speech, or audio reproduction systems that create virtual sound sources. Usually, these systems have to meet the high audio quality of e.g. the compact-disc standard. Engineers have become aware of the fact that signal-to-noise ratios and distortion measures do not tell the whole story when it comes to sound quality. As a consequence, new algorithms have to be evaluated by extensive listening tests. Drawbacks of this method of evaluation are that these tests are expensive and time consuming. Moreover, listening tests usually do not give any insight *why* a specific algorithm does or does not work. Hence there is a demand for objective and fast evaluation tools for new audio technologies. One way to meet these demands is to develop a model of the auditory system that can predict the perceived distortion and which can indicate the nature of these distortions.

This thesis describes and validates a model for the binaural hearing system. In particular, it aims at predicting the audibility of changes in arbitrary binaural stimuli. Two important properties for binaural hearing are interaural intensity differences (IIDs) and interaural time differences (ITDs) present in the waveforms arriving at both ears. These interaural differences enable us to estimate the position of a sound source but also contribute to our ability to detect signals in noisy environments. Hence one of the most important objectives for a comprehensive model is its ability to describe the sensitivity for interaural differences in a large variety of conditions.

The basis of the model relies on psychoacoustic experiments that were performed with human listeners. In one series of experiments, subjects had to detect the presence of interaural cues for various statistical distributions of the IIDs and ITDs. The results revealed that the energy of the difference of the signals arriving at both ears following a peripheral filtering stage can successfully describe the sensitivity for interaural time and intensity differences. This approach is very similar to Durlach's EC theory. Furthermore, other listening

experiments with varying degrees of stimulus uncertainty revealed that the detection process of the binaural auditory system may well be simulated as a template-matching procedure.

The idea of template matching based on the energy of the difference signal was incorporated in a time-domain detection model. This model transforms arbitrary stimuli into an internal representation. This representation comprises four dimensions: time, frequency channel, internal interaural delay and internal interaural level adjustment. The internal model activity as a function of these dimensions entails both binaural and monaural properties of the presented stimuli. The accuracy of these properties is limited by the addition of internal noise and by the limited frequency and time resolution incorporated in various stages of the model. An important model feature is the 'optimal detector'. This optimal detector analyzes the internal representation of the presented waveforms and extracts information from it, for example the presence or absence of a signal added to a masker. This process entails a strategy that optimally reduces the internal noise by integrating information across time and frequency channels.

The model was tested for its ability to predict thresholds as a function of spectral and temporal stimulus parameters. During all simulations, all model parameters were kept constant. The results revealed that the model can account for a large variety of experimental data that are described in the literature. The most prominent finding was that the model can quantitatively account for the wider effective critical bandwidth observed in band-widening NoSπ experiments. This wider effective bandwidth is found if the threshold of audibility is measured for interaurally out-of-phase signals (Sπ) added to band-limited interaurally in-phase noise (No). In our model, this phenomenon is the result of the fact that the cue for detection is available in a range of filters if the masker bandwidth is sufficiently small. The increased effective bandwidth does therefore not reflect a worse binaural spectral resolution compared the monaural spectral resolution but it follows from the ability to integrate information across frequency. It was also shown that the optimal detector can account for effects found by manipulating temporal stimulus properties. To be more precise, the model can account for the phenomenon that the temporal resolution of the binaural auditory system obtained from stimuli with time-varying interaural correlations seems to be worse for sinusoidally-varying cross correlation than for rectangular correlation modulations.

To extend the model's predictive scope towards more natural listening conditions, experiments were performed with virtual sound sources, which were generated by using head-related transfer functions (HRTFs). The complexity of these impulse responses was gradually decreased by a spectral smoothing

operation. During listening tests, subjects had to rate the audibility of this operation. The results revealed that the fine structure of HRTF phase and magnitude spectra is relatively unimportant for the generation of virtual sound sources in the horizontal plane. The same experiment was subsequently simulated with the model. Comparisons between subject data and model predictions showed that the model could not only predict whether the HRTF smoothing was audible or not, but that it could also predict the amount of perceptual degradation for supra-threshold HRTF smoothing.

SAMENVATTING

Door de snel toenemende technische mogelijkheden op het gebied van multimedia en digitale signaalbewerking is de interesse voor nieuwe geluidsbewerkingen en reproductietechnieken toegenomen. Voorbeelden van recent ontwikkelde producten zijn algoritmen om spraak en muziek op een informatie-zuinige manier te representeren en systemen die virtuele geluidsbronnen creëren. Over het algemeen worden aan deze nieuwe technieken hoge eisen gesteld wat de geluidskwaliteit betreft. Omdat het besef is gegroeid dat geluidskwaliteit zich niet laat beschrijven door fysische maten zoals bijvoorbeeld signaal-ruis verhoudingen, moeten nieuwe applicaties uitvoerig worden getest door middel van luistertesten. Deze tests hebben als nadeel dat ze tijdrovend en arbeidsintensief zijn. Bovendien geven ze vaak geen reden *waarom* een bepaald systeem wel of niet werkt. Er is daarom een toenemende vraag naar snellere en goedkopere evaluatiemethoden. Een mogelijke oplossing hiervoor is het gebruik van modellen van het menselijk ruimtelijk gehoorsysteem die de hoorbaarheid van vervormingen in aangeboden geluidsfragmenten kunnen voorspellen.

Dit proefschrift beschrijft en test een dergelijk model dat de hoorbaarheid van veranderingen in willekeurige ruimtelijke stimuli kan voorspellen. In principe zijn er daarbij 2 eigenschappen van de signalen die door de trommelvliezen worden opgevangen van belang: interaurale tijdsvertragingen (Engels: interaural time differences, ITDs) en interaurale intensiteitsverschillen (Engels: interaural intensity differences, IIDs). Deze eigenschappen stellen luisteraars in staat de horizontale positie van een geluidsbron te schatten en helpen bij het detecteren van signalen in lawaaierige omgevingen. Het is daarom van groot belang dat het model de gevoeligheid voor deze interaurale verschillen nauwkeurig kan beschrijven onder verschillende experimentele omstandigheden.

De basis van het model berust op experimenteel behaalde resultaten met proefpersonen. In een serie van experimenten moesten proefpersonen de aanwezigheid van interaurale verschillen detecteren waarbij verschillende statistische verdelingsfuncties voor deze verschillen werden aangeboden.

Uit de resultaten bleek dat de energie van het verschilsignaal tussen het linker en rechter oor na perifere filtering een beslissingsvariabele is die de gevoeligheid van de proefpersonen voor interaurale verschillen goed kan beschrijven. Deze beschrijving komt overeen met de EC theorie van Durlach. Een andere serie experimenten bracht aan het licht dat de detectiestrategie van proefpersonen gemodelleerd kan worden met een 'template-matching' proces.

Dit template-matching proces gebaseerd op de energie van het interaurale verschilsignaal is verwezenlijkt in een tijdsdomein model. Dit model transformeert aangeboden stimuli naar een interne representatie, die bestaat uit vier dimensies: tijd, frequentie, interne interaurale vertraging en intern interauraal intensiteitsverschil. De interne activiteit als functie van deze dimensies bevat zowel ruimtelijke als monaurale informatie over geluidsbronnen. De resolutie van deze informatie is gelimiteerd door de toevoeging van interne ruisbronnen en door het beperkt oplossend vermogen van tijds- en frequentie-informatie op verschillende plaatsen in het model. Een belangrijk kenmerk van het model is de aanwezigheid van een zgn. 'optimale detector'. Deze detector analyseert de interne representatie van de aangeboden stimuli en extraheert daaruit informatie, zoals de positie van een geluidsbron of de aanwezigheid van een signaal in een maskeerexperiment. Hierbij wordt informatie op een dusdanige manier geïntegreerd over de tijd en de frequentiekanalen dat de interne ruis wordt geminimaliseerd.

De voorspellingen van het model zijn uitvoerig vergeleken met experimentele resultaten waarbij de temporele en spectrale eigenschappen van de stimuli zijn gevarieerd. Alle modelparameters zijn constant gehouden. Uit deze vergelijkingen blijkt dat het model veel experimentele resultaten die in de literatuur zijn beschreven kwantitatief kan voorspellen. Het meest prominente resultaat is dat het model de klaarblijkelijk grotere binaurale bandbreedte van de perifere auditieve filters kan verklaren. Dit fenomeen wordt gevonden bij het bepalen van de kritieke bandbreedte met behulp van interauraal uit-fase signalen (Sπ) gecombineerd met bandbeperkte in-fase ruis (No), en kan begrepen worden door informatie over het gehele excitatiepatroon optimaal te combineren. De schijnbaar grotere binaurale kritieke bandbreedte is daarom niet het gevolg van een slechtere spectrale resolutie maar juist van het efficiëntere gebruik van informatie van verschillende frequentiebanden. De optimale detectiestrategie bleek ook temporele effecten te kunnen verklaren. Wanneer de temporele resolutie van het binaurale gehoorsysteem wordt geschat aan de hand van experimenten met een in de tijd gevarieerde interaurale correlatie, is de schatting van de tijdsconstanten van het systeem groter bij sinusvormige correlatieveranderingen dan bij rechthoekige correlatieveranderingen.

Om het model te testen in natuurlijke luistercondities zijn er experimenten gedaan met virtuele geluidsbronnen, welke gegenereerd zijn met behulp van zgn. hoofd-gerelateerde overdrachtsfuncties (Engels: head-related transfer functions, HRTFs). De complexiteit van deze functies is stapsgewijs verminderd door de spectrale resolutie van deze impulsresponsies te verlagen. Proefpersonen kregen de opdracht de hoorbaarheid van deze bewerking te indiceren. Uit de resulaten bleek dat de spectrale details van HRTFs niet van belang zijn voor het genereren van virtuele geluidsbronnen in het horizontale vlak. Uit modelsimulaties met dezelfde stimuli bleek er een hoge correlatie te bestaan tussen de waarneembaarheid van veranderingen in het geluid en de modelvoorspellingen.

CURRICULUM VITAE

- 13 juli 1970: Geboren te Haarlem.

- 1982-1984: Stedelijk gymnasium te Arnhem.

- 1984-1989: Voorbereidend Wetenschappelijk Onderwijs (VWO), Lorentz Scholengemeenschap Arnhem.

- 1989-1990: Technische Natuurkunde, Technische Universiteit Eindhoven.

- 1990-1992: Werktuigbouwkunde, Technische Universiteit Eindhoven.

- 1992-1997: Werktuigkundige Medische Technologie, bovenbouwstudie Werktuigbouwkunde, Technische Universiteit Eindhoven.

- 1997-2001: Onderzoeker in Opleiding (NWO), IPO, Center for User-System Interaction, Technische Universiteit Eindhoven. Promotie onderzoek naar het modelleren van binaurale detectie.

DANKWOORD

Graag wil ik een aantal mensen bedanken voor hun steun en bijdragen bij het tot stand komen van dit proefschrift.

Allereerst Armin Kohlrausch. Volgens het woordenboek betekent promotor 'die het tot stand komen van iets bevordert'. Dit heeft Armin in de breedste zin van het woord opgevat: hij is altijd een uitstekende begeleider geweest bij het doen van onderzoek, het documenteren daarvan, het incasseren van reviews en alle aanverwante zaken die men zoal tijdens een promotie onderzoek tegenkomt. Een betere promotor is niet te vinden ($p < 0.01$).

Verder wil ik Steven van de Par bedanken voor alle diepgaande discussies (vakinhoudelijk en niet-vakinhoudelijk), zijn onuitputtelijke geduld bij het doorlezen en corrigeren van allerlei schrijfsels en vooral voor de vlijmscherpe analyses waarmee hij mij steeds weer weet te verrassen.

Ook wil ik Mark Houben bedanken voor het delen van zijn gedetailleerde Matlab kennis en het zijn van een gezellige kamergenoot. En natuurlijk alle andere collega's, die, ondanks alle politieke perikelen, van het IPO een hele fijne werkplek hebben gemaakt.

Tot slot wil ik Elise bedanken, die altijd klaar gestaan heeft als steun en toeverlaat bij het schrijven van dit proefschrift maar zeker ook daarbuiten.

STELLINGEN

behorend bij het proefschrift

MODELING BINAURAL SIGNAL DETECTION

1. De stelling dat modellen gebaseerd op binaurale verschillen in tijd en intensiteit en modellen gebaseerd op de interaurale correlatie gelijke voorspellingen geven geldt niet voor stimuli met gemultipliceerde ruis.

   - ```Hoofdstuk 2 van dit proefschrift.```

2. In termen van stimulus onzekerheid geldt niet dat de EC-theorie equivalent is met een model gebaseerd op de interaurale correlatie.

   - ```Hoofdstuk 3 van dit proefschrift.```

3. De geschatte temporele resolutie van het binaurale gehoorsysteem volgend uit een analyse van de interaurale correlatie over een interval gecentreerd op een kort signaal is hoger dan de werkelijke resolutie van het systeem.

   - ```Hoofdstuk 6 van dit proefschrift.```

4. Het feit dat het menselijk gehoor niet gevoelig is voor de spectrale fijnstructuur in HRTFs geeft aan dat men meer oog dan oor voor detail heeft.

   - ```Hoofdstuk 7 van dit proefschrift.```

5. De huidige statistische evaluatiemethoden werken ontkoppeling tussen statistische significantie en maatschappelijke relevantie in de hand.

   - ```http://ourworld.compuserve.com/homepages/rajm/jspib.htm.```

6. ABN (Algemeen Beschaafd Nederlands) is een uitspraaksvorm die bijna iedereen herkent maar bijna niemand meer beheerst.

   - ```Nederlandse Taalkunde, jaargang 5 (2000), nr.3, blz. 272-279.```

7. Dat digitale piano's niet zo klinken als echte piano's ligt voor een groot deel in het feit dat men niet een echte piano probeert te simuleren, maar een opname daarvan.

8. Het feit dat de DVD-audiostandaard bemonsteringsfrequenties tot 192 kHz ondersteunt lijkt er op te wijzen dat men vleermuizen ook tot de doelgroep van DVD-gebruikers rekent.

9. Het verschil tussen bodemprijs en grondprijs zit in de portemonnee van projectontwikkelaar en gemeentebestuur.

10. Verkiezingsuitslagen in de VS zouden betrouwbaarheidsintervallen moeten bevatten.

11. Preventief ruimen is het toppunt van zinloos geweld.

12. Gelukkig is de AH-erlebnis niet gereserveerd voor mensen met een bonuskaart.

Jeroen Breebaart
Eindhoven, 11 mei 2001.