Binaural processing model based on contralateral inhibition. I. Model structure

Jeroen Breebaart[a)]
IPO, Center for User–System Interaction, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

Steven van de Par and Armin Kohlrausch
IPO, Center for User–System Interaction, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands and Philips Research Laboratories Eindhoven, Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands

(Received 23 May 2000; revised 18 December 2000; accepted 3 May 2001)

This article 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 two accompanying articles, the model predictions are compared with data obtained with human observers in a great variety of experimental conditions.
[DOI: 10.1121/1.1383297]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Dc [DWG]

I. 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; Stern and Shear, 1996; Bernstein and Trahiotis, 1996), 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 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

[a)]Now at: Philips Research Laboratories Eindhoven, Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands. Electronic mail: jeroen.breebaart@philips.com

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 level 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 previously described 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, Breebaart *et al.* (1999) showed that $\mathrm{NoS}\pi$ 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, Breebaart and Kohlrausch (2001) 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 $\mathrm{NoS}\pi$ condition is difficult to understand in terms of the cross correlation (see Sec. V). 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 accommodations to reduce the detrimental effects of stimulus level variability on detection performance in narrow-band 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 Sec. VII.

II. 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 three-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. A detailed description of the complete model and its various stages are given in Secs. III–VI.

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 two accompanying articles, which focus on simulations of various detection experiments. In particular, Breebaart *et al.* (2001a) focuses on spectral and interaural parameters of the stimuli, and Breebaart *et al.* (2001b) deals with temporal stimulus properties.

III. MODEL OVERVIEW

The model is divided into three stages as shown in Fig. 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. 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.

IV. 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:

(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 article and the two subsequent articles, 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

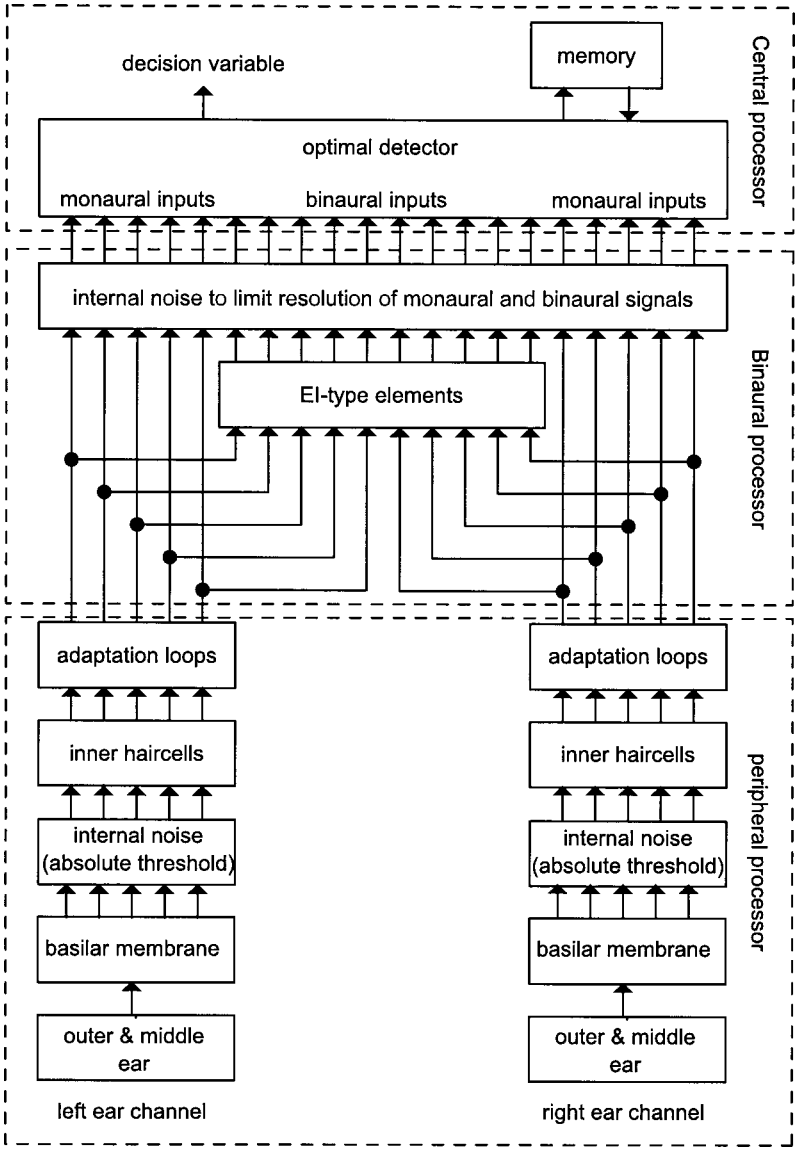

FIG. 1. Successive stages of the model. The signals arriving from both ears are processed by a peripheral preprocessing stage (outer and middle ear transfer function, linear basilar membrane model, additive internal noise, inner hair cell stage, and adaptation loops), followed by a binaural processor. 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.

spacing is two 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 a 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 fifth-order low-pass filter with a cutoff frequency (-3 dB point) of 770 Hz. For frequencies below about 770 Hz, the low-pass 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 low-pass 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.

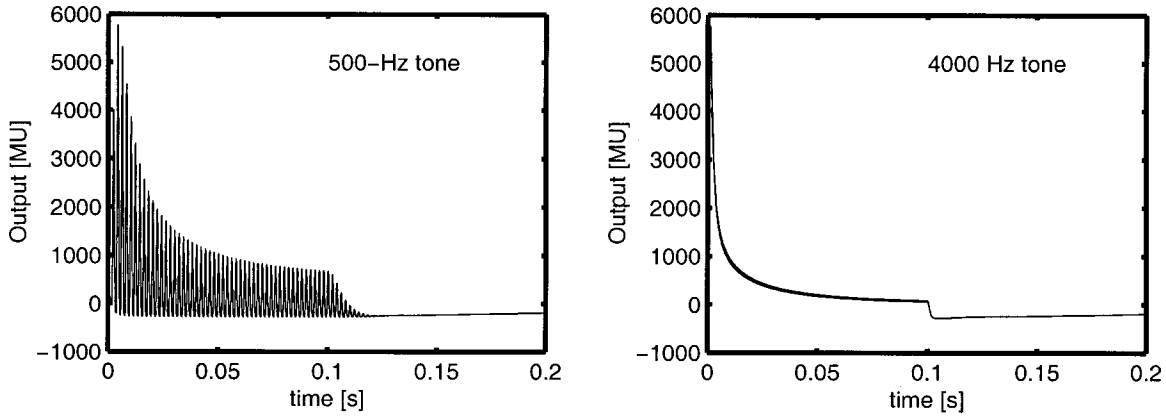

FIG. 2. Output of the peripheral preprocessor for a 500-Hz tone (left panel) and a 4000-Hz tone (right panel) of 100-ms duration. The output was calculated for a filter tuned to the frequency of the tone.

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.

Second, 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 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 Breebaart *et al.* (2001a).

An example of the output of the peripheral preprocessing stage is given in Fig. 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.

V. BINAURAL PROCESSING STAGE

A. 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. 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.

B. 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 is defined as

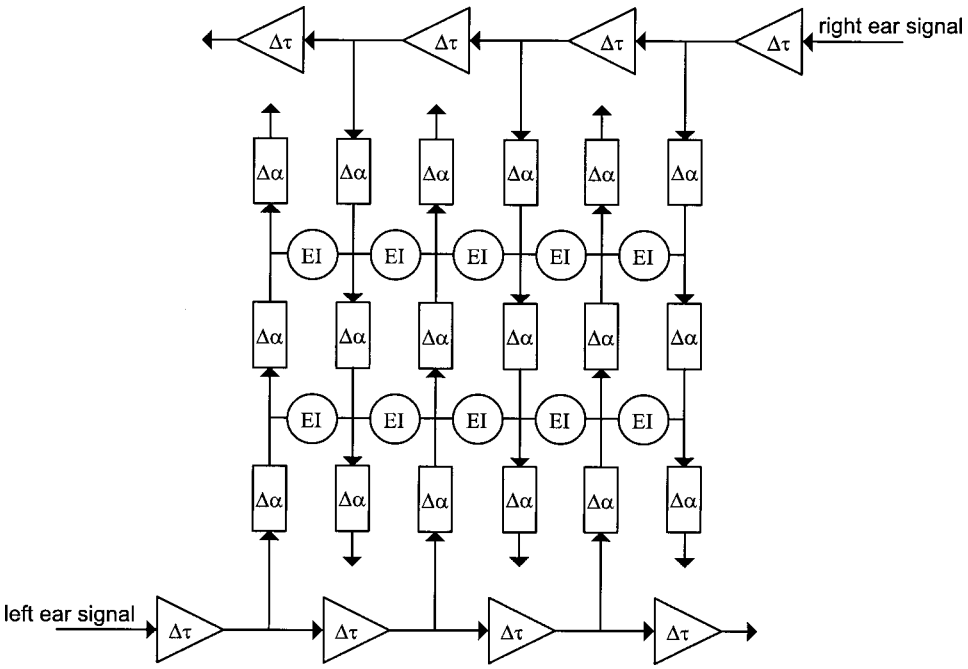

FIG. 3. Structure of the binaural processor. The triangles denote delays ($\Delta\tau$), the blocks are attenuators ($\Delta\alpha$), and the circles denote EI-type elements.

$$E_L(i,t,\tau,\alpha)=\lceil 10^{\alpha/40}L_i(t+\tau/2)-10^{-\alpha/40}R_i(t-\tau/2)\rceil^2, \tag{1}$$

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

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

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 preprocessor 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 ($\lceil\cdot\rceil$) denote a half-wave 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. (1) and (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 $\alpha=10$ dB are realized.[1]

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. (1) and (2). It can be shown that summation of these signals results in an output E given by

$$E(i,t,\tau,\alpha)=(10^{\alpha/40}L_i(t+\tau/2)-10^{-\alpha/40}R_i(t-\tau/2))^2. \tag{3}$$

An important consequence of the above summation is that the EI-type element described in Eq. (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. (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:

$$E'(i,t,\tau,\alpha)=\int_{-\infty}^{\infty}E(i,(t+t_{\text{int}}),\tau,\alpha)w(t_{\text{int}})\,dt_{\text{int}}, \tag{4}$$

with

$$w(t)=\frac{\exp(-|t|/c)}{2c}. \tag{5}$$

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

$$E''(i,t,\tau,\alpha)=ap(\tau)\ \log\ (bE'(i,t,\tau,\alpha)+1)+n(i,t,\tau,\alpha). \tag{6}$$

An internal noise $n(i,t,\tau,\alpha)$ limits the accuracy of internal binaural processing.[2] 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

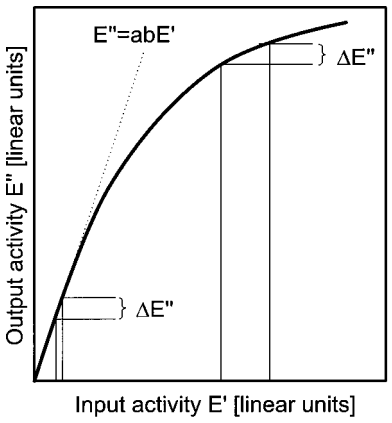

FIG. 4. Input–output characteristic of the EI-type element. The dotted line represents the line $E''=abE'$ (see text).

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

$$p(\tau)=10^{-|\tau|/5}, \tag{7}$$

where the internal delay τ is expressed in ms. This formula resulted from data with $(\mathrm{NoS}\pi)_\tau$ stimuli which are presented in Appendix 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. (6), leaving out the internal noise, is shown in Fig. 4. For small values of E', the input–output function is linear. For higher values of E', the curve converges to a logarithmic function.

The rationale for including the logarithmic transformation in Eq. (6) is as follows. Egan *et al.* (1969) measured psychometric functions for $\mathrm{NoS}\pi$ stimuli as a function of the signal power. They found that the sensitivity index d' was linearly related to the signal power $\langle S^2\rangle$:

$$d'=m\langle S^2\rangle/\langle N^2\rangle. \tag{8}$$

Here, $\langle N^2\rangle$ 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 $\mathrm{S}\pi$ 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. (3) (i.e., E) is linearly related to the signal power $\langle S^2\rangle$. The temporal integrator in Eq. (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. (6) can be described in a first-order approximation by

$$E''(i,t,0,0)\approx abp(\tau)E'(i,t,0,0)+n(i,t,0,0). \tag{9}$$

This relation, without incorporation of the internal noise n, is shown by the dotted line in Fig. 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. (9). If E'' is used as a decision variable in the $\mathrm{NoS}\pi$ detection paradigm, d' is related linearly to the signal power $\langle S^2\rangle$ as found by Egan *et al.* (1969). The fact that the *power* of the signal is used as a decision variable in $\mathrm{NoS}\pi$ paradigms is also supported by the results of Breebaart *et al.* (1999). They proposed the power of the difference signal 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 $\mathrm{N}\rho\mathrm{S}\pi$ condition with $\rho<0.95$, the approximation from Eq. (9) does not hold. For such stimuli, E'' can be approximated by

$$E''(i,t,0,0)\approx ap(\tau)\ \log\ bE'(i,t,0,0)+n(i,t,0,0). \tag{10}$$

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. 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 $\mathrm{N}\rho\mathrm{S}\pi$ conditions, it is expected that our model has similar prediction performance for these stimuli. As can be observed from Eq. (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 0.000 02, respectively. These values resulted from the calibration procedure as described in Breebaart *et al.* (2001a). 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.[3]

C. 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. 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

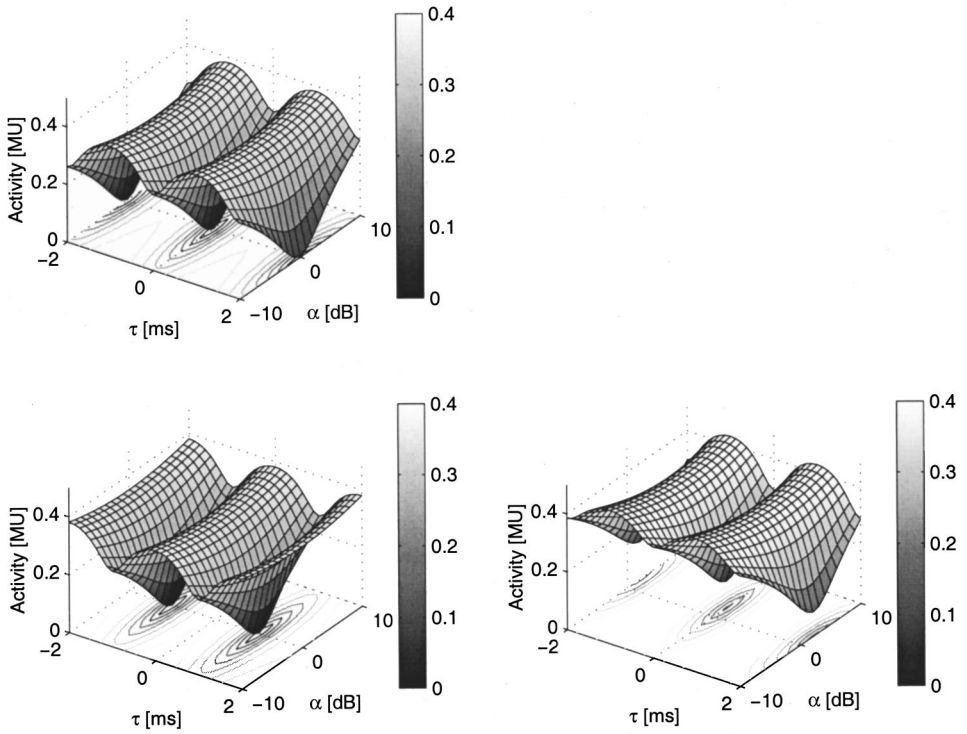

FIG. 5. Idealized [no internal noise, $p(\tau)=1$] EI-activity patterns for a 500-Hz sinusoid as a function of the characteristic IID and ITD of each unit. The upper panel corresponds to a diotic signal (i.e., no external IID or ITD). The signal in the lower-left panel has an ITD of 1 ms and no IID; the signal in the lower-right panel has an IID of 20 dB and no ITD.

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 cancelled. 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. 5. Basically, the pattern is the same as the pattern shown in the upper panel of Fig. 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. 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 sides 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 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 results in a less compact perceived image (Blauert, 1997, p. 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 narrow-band 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

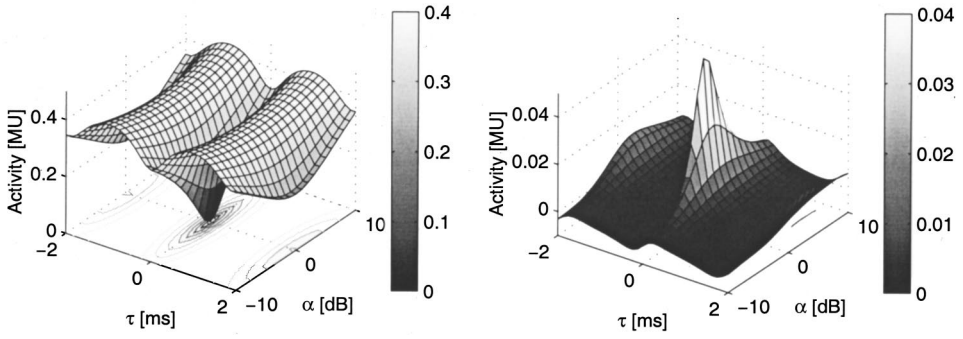

FIG. 6. Left panel: Idealized EI-activity for a wideband diotic noise (0–4000 Hz) with an overall level of 70 dB SPL for an auditory filter centered at 500 Hz. Right panel: change in the activity pattern of the left panel if a 500-Hz interaurally out-of-phase signal ($S\pi$) is added with a level of 50 dB.

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).

D. 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) is 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. 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.

E. 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 $S\pi$ signal results in a specific change in the EI activity pattern. To demonstrate this, the left panel of Fig. 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 $NoS\pi$ stimulus is shown in the right panel of Fig. 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.

VI. 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 three-interval, forced-choice (3-IFC) 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 purposes was suggested before by Dau (1992) and Dau *et al.* (1996a), and for binaural detection by Holube *et al.* (1995) and by Breebaart and Kohlrausch (2001). 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

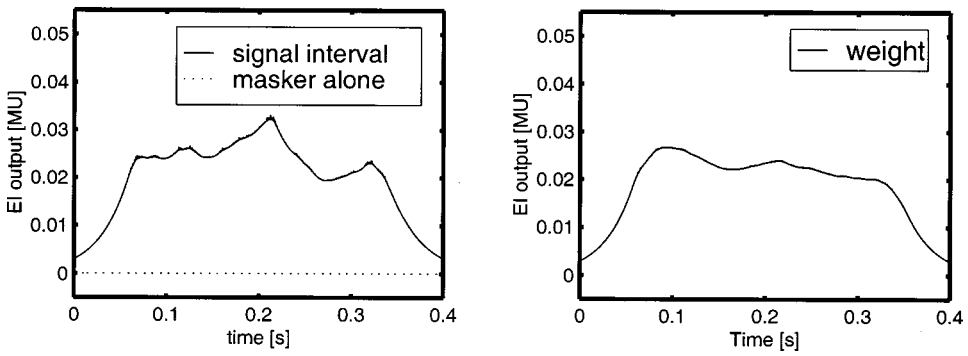

FIG. 7. NoSπ EI-activity as a function of time for the EI-type element with $\alpha=\tau=0$ without incorporation of internal noise. The left panel shows the output for a single signal interval (solid line) and for a masker alone (dotted line). The right panel shows the average difference between masker 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.

power necessary to compute the output for all relevant EI-type elements to be enormous. We found that for the conditions described in the two accompanying articles 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 NoSπ condition, we already showed that for this specific condition, a maximum change in activity occurs for $\alpha=\tau=0$ (see Fig. 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. 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. 6), the output increases. This is shown by the solid line in the left panel of Fig. 7 for one realization of an NoSπ 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 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. 7. The solid line (labeled ''weight'') represents the mean output for an NoSπ condition averaged over ten 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 nonzero output for a masker alone is given in Fig. 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 $\rho=0.5$) and the signal level was increased to 60 dB. As in the left panel of Fig. 7, the solid line represents the output for a single masker-plus-signal interval, the dotted line represents the mean output for ten 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[4] shown in the right panel of Fig. 8. As in Fig. 7, the weights consist of the average difference

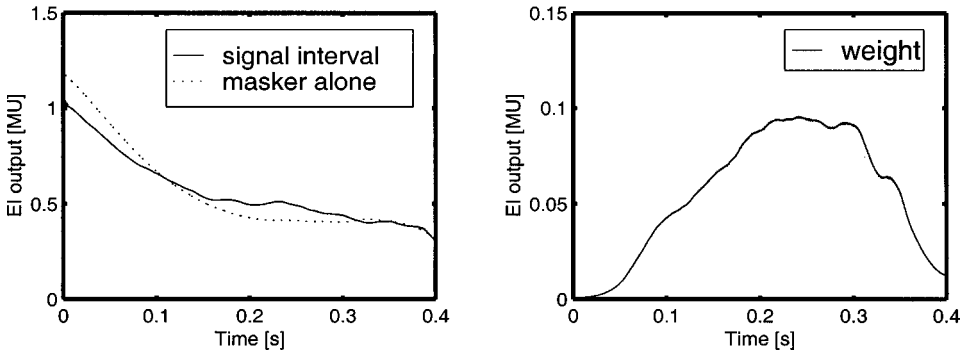

FIG. 8. Same as in Fig. 7, only for an NρSπ condition with $\rho=0.5$.

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 low-pass filtered by a double-sided exponential window with time constants of 10 ms. These low-pass-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. 7 and 8 and subsequently combined into one distance measure. This process is described in detail in Appendix B.

VII. 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 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 the accompanying articles (Breebaart *et al.*, 2001a, b), 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

$$\rho = \frac{\langle N^2 \rangle - \langle S^2 \rangle}{\langle N^2 \rangle + \langle S^2 \rangle}. \tag{11}$$

Here, $\langle N^2 \rangle$ denotes the masker energy and $\langle S^2 \rangle$ 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 $\langle S^2 \rangle$ 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 of 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 not, however, 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 Breebaart *et al.*, 2001b). 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. The third article in this series (Breebaart *et al.*, 2001b) 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 nor-

malized 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 narrow-band 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 narrow-band 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.

VIII. 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, models 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. The two accompanying articles (Breebaart *et al.*, 2001a, b) provide quantitative predictions for a wide range of binaural signal detection conditions derived with the model described in this article. In particular, Breebaart *et al.* (2001a) discusses the influence of spectral masker and signal parameters on detection thresholds, while Breebaart *et al.* (2001b) deals with temporal stimulus parameters.

ACKNOWLEDGMENTS

The investigations were supported by the Research Council for Earth and Life-sciences (ALW) with financial aid from the Netherlands Organization for Scientific Research (NWO). We want to thank the Associate Editor Wesley Grantham, Steve Colburn, and two anonymous reviewers for their very valuable comments and suggestions for improving the original manuscript.

APPENDIX A: 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 NoSπ 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 NoSπ 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 $p(\tau)$ function.

Our data were obtained for one subject only using a three-interval, forced-choice procedure with adaptive signal-level adjustment. A 400-ms narrow-band 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. A1. 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 wideband (20–1000 Hz) Gaussian-noise masker with an overall level of 75 dB SPL.

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 de-

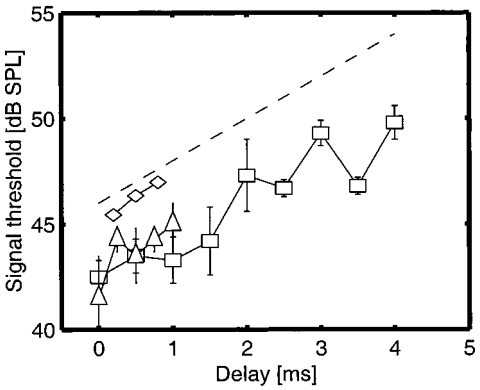

FIG. A1. $(\mathrm{NoS}\pi)_\tau$ thresholds as a function of the interaural delay τ. The triangles correspond to a center frequency of 500 Hz, the squares to 125 Hz. The diamonds are data adapted from Colburn and Latimer (1978). The error bars denote the standard error of the mean. The dashed line indicates a slope of 2 dB/ms.

cay. A slope of 2 dB/ms means that every 3 ms the signal amplitude is doubled at threshold, which corresponds to a factor of 4 in the EI-type element output. Therefore, the $p(\tau)$ function must decrease by a factor of 4 every 3 ms, which results in the formulation given in Eq. (7).

APPENDIX B: 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

$$U(i,t)=E''(i,t)-\bar{E}(i,t). \tag{B1}$$

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 $U(i,t)$:

$$U=\int_i\int_{t=0}^{t=T}\frac{\mu(i,t)}{\sigma^2(i,t)}U(i,t)di\,dt, \tag{B2}$$

where 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.

[1]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.

[2]In Sec. IV, 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.

[3]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 Sec. VI for details) is able to strongly reduce the internal noise by temporal integration.

[4]It should be noted that in these examples, the variability in the EI output due to internal noise or due to stimulus uncertainty was not taken into account. As shown in Appendix 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.

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

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (**1997b**). ''Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. 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.

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.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

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

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

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. (**2001a**). ''Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters,'' J. Acoust. Soc. Am. **110**, 1089–1104.
Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001b**). ''Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters,'' J. Acoust. Soc. Am. **110**, 1105–1117.
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, New York), Vol. IV, pp. 467–518.
Colburn, H. S., and Isabelle, S. K. (**2001**). ''Physiologically based models of binaural detection,'' in *Physiological 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., 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. (**1996a**). ''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. (**1996b**). ''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, New York), 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.
Egan, J. P., Lindner, W. A., and McFadden, D. (**1969**). ''Masking-level differences and the form of the psychometric function,'' Percept. Psychophys. **6**, 209–215.
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,'' Hear. Res. **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.
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.
Hafter, E. R. (**1971**). ''Quantitative evaluation of a lateralization model of masking-level differences,'' J. Acoust. Soc. Am. **50**, 1116–1122.
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 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.
Hirsh, I. (**1948a**). ''Binaural summation and interaural inhibition as a function of the level of masking noise,'' Am. J. Psychol. **61**, 205–213.
Hirsh, I. (**1948b**). ''The influence of interaural phase on interaural summation and inhibition,'' J. Acoust. Soc. Am. **20**, 536–544.
Holube, I., Colburn, H. S., van de Par, S., and Kohlrausch, A. (**1995**). ''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.
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.
Jeffress, L. A. (**1948**). ''A place theory of sound localization,'' J. Comp. Physiol. Psychol. **41**, 35–39.
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**). ''Stimulus representation in the discharge patterns of auditory neurons,'' in *The Nervous System* (Raven, New York), Vol. 3.
Kohlrausch, A. (**1986**). ''The influence of signal duration, signal frequency and masker duration on binaural masking level differences,'' Hear. Res. **23**, 267–273.
Kohlrausch, A., 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 Assoc., Mahwah, NJ), 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.
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.
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.
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.
Moore, J. K. (**1987**). ''The human auditory brain stem: A comparative view,'' Hear. Res. **29**, 1–32.

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, 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.
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.
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.
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.
Robinson, D. E., and Trahiotis, C. (**1972**). ''Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions,'' Percept. Psychophys. **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. **79**, 1563–1570.
Shackleton, T., Meddis, R., and Hewitt, M. (**1992**). ''Across frequency integration in a model of lateralization,'' J. Acoust. Soc. Am. **91**, 2276–2279.
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., 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., 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.
Weis, T. F., and Rose, C. (**1988**). ''A comparison of synchronization filters in different auditory receptor organs,'' Hear. Res. **33**, 175–180.
Wilson, R., and Fowler, C. (**1986**). ''Effects of signal duration on the 500-Hz masking-level difference,'' Scand. Audiol. **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.
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. (**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.
Zerbs, C. (**2000**). ''Modeling the effective binaural signal processing in the auditory system,'' Ph.D. thesis, Oldenburg University, Germany.
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, U. T., and Zwicker, E. (**1984**). ''Binaural masking-level difference as a function of masker and test-signal duration,'' Hear. Res. **13**, 215–220.