A NEW BINAURAL DETECTION MODEL BASED ON CONTRALATERAL INHIBITION

JEROEN BREEBAART (1) AND ARMIN KOHLRAUSCH (1,2)

(1) IPO-Center for Research on User-System Interaction, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

(2) Philips Research Laboratories Eindhoven, Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands

Binaural models attempt to explain binaural phenomena in terms of neural mechanisms that extract binaural information from acoustic stimuli. In this paper, a model setup is presented that can be used to simulate binaural detection tasks. In contrast to the most often used cross correlation between the right and left channel, this model is based on contralateral inhibition. The presented model is applied to a wide range of binaural detection experiments. It shows a good fit for changes in masker bandwidth or masker correlation, static and dynamic cues and level and frequency dependencies.

1 Introduction

One of the challenges of binaural modeling lies in the design of a system both for the extraction of static interaural differences (i.e., as in localization tasks) and the detection of dynamically varying interaural differences (i.e., as in detection tasks). Most binaural models heavily rely on some form of a crosscorrelation function of the signals from both ears (cf. Albeck and Konishi, 1995), based on the coincidence mechanisms proposed by Jeffress (1948). Some drawbacks of these models can be summarized as follows.

1. Most correlation models do not integrate the processing of static interaural differences and dynamically varying interaural differences, which results in the difficulty to link detection and localization phenomena.
2. Some models are not able to detect static interaural intensity differences at all.
3. Most models cannot explain the influence of masker bandwidth in a binaural detection task.

In this chapter, we present a new binaural model which bases its predictions on a process of contralateral inhibition. The various functional elements of this particular model, as shown in Fig. 1, are discussed below.

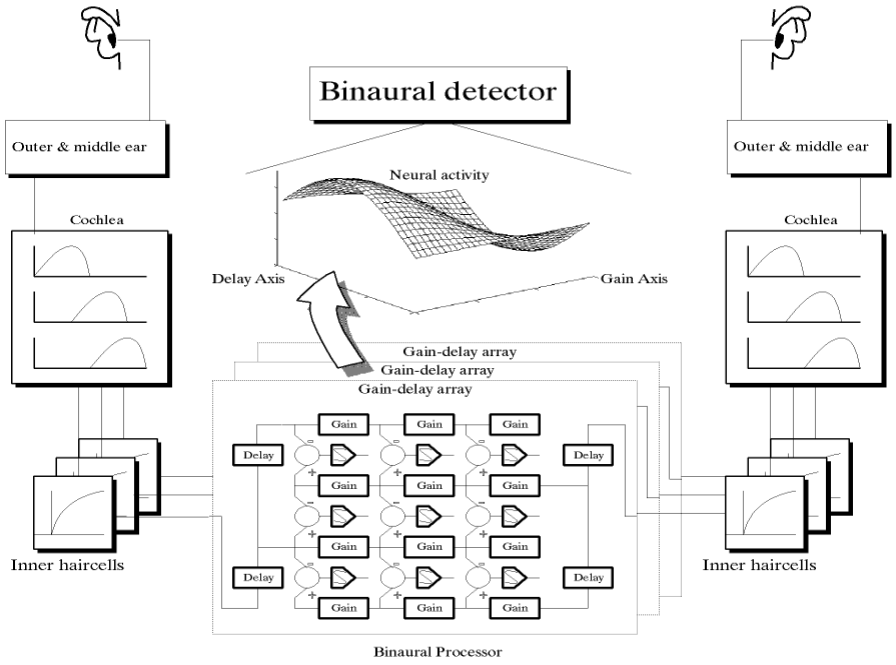


Figure 1 - Structure of the binaural model. Signals presented at both ears enter the outer and middle ear block, followed by the cochlea model, the inner hair-cell model, the binaural processor and finally the binaural detector. The EI-type cells, denoted by circles in the gain-delay array of the binaural processor (see text) record the difference in signals from left and right ear. Finally, a temporal integrator computes the energy of the EI-type cell output. This leads to a neural activity pattern that is analyzed by a binaural detector.

1.1 Outer and middle ear

The first processing block simulates the general spectral shaping of the outer and middle ear, without taking into account the direction specific spectral shaping found in head-related transfer functions. The combined outer and middle-ear response is simulated by a frequency-transfer function with a rolloff of 6 dB/Oct below 1 kHz and -6 dB/Oct above 4 kHz.

1.2 Cochlea and basilar membrane

The second block, the cochlea module, simulates a running spectral analysis of the incoming signals. In the present model, this part is modeled by means of a fourth-order Gammatone filterbank with filters of equivalent rectangular bandwidth. The spacing of the adjacent filters is two per equivalent rectangular bandwidth. The signal levels of this model are expressed in terms of model units

(MU), whereby one MU corresponds to a sound pressure of 20 μPa. In order to introduce an absolute threshold, an interaurally-independent noise is added to the filtered signals from the left and right ear with a level of 10 dB re 1 MU.

1.3 Inner hair cells

The third block simulates the effective signal processing of the inner hair cells. Important processing features are:

1. First, there is only response for the positive phases of the temporal waveform. This effect is modeled by a half-wave rectifier.
2. The loss of fine structure is modeled by a 5^{th}-order low-pass filter with a cutoff frequency of 650 Hz.
3. Third, compression. Both physiological and psychophysical findings indicate the presence of nonlinearities in the peripheral auditory system. The model assumes that it is possible to group these nonlinearities together and describe them in terms of a simple time-invariant power law. Therefore, the instantaneous values of the waveforms are raised to the power 0.4.

1.4 Binaural processor

The binaural processor simulates the binaural interaction found in the auditory system. Neurons in the ascending auditory pathway exhibiting binaural interaction are found in the superior olivary complex (SOC) and the inferior colliculus (IC). In both nuclei, two different types of binaural neurons are found: EE-type cells (i.e., coincidence detectors) and EI-type. The EI-type cells are excited by one ear and show *inhibitory* effects when the other ear is stimulated (Excitation-Inhibition).

The activity of EI-type cells has been measured in phsyiological experiments a function of several parameters of the acoustic stimuli (Goldberg and Brown, 1969; Joris and Yin, 1995; Joris, 1996). To account for these neurophysiological results, for each auditory filter, a gain-delay array including EI-type cells is used, as shown in Fig. 1. Each EI-type cell receives the neural activity from the inner hair cell model corresponding to one auditory filter, while each position in the array corresponds to a specific interaural delay and interaural gain. Each cell is characterized by three parameters (filter number, characteristic delay, balance between excitatory and inhibitory influence) and computes a normalized difference-intensity (E') from the incoming signals as described below:

Equation 1

$$E'(\tau,\alpha,n)=\frac{\int\left(e^{-\alpha}l_n(t)-e^{\alpha}r_n(t-\tau)\right)^2dt}{e^{-2\alpha}\int l_n^2dt+e^{2\alpha}\int r_n^2dt}.$$

Here, n represents the Gamma-tone filter number, τ the characteristic delay, α the balance between excitatory and inhibitory influence (here defined as a relative gain), l_n and r_n the signals from Gammatone filter n after being processed by the inner hair cell model of the left and right signal, respectively. To include saturative properties of the EI-type neuron, the following input-output is included:

Equation 2

$$E = E'\exp\{cE'\}.$$

Here, E' is the neural activity as described in Eq. 1-1; E represents the saturated activity and c the saturation coefficient (c=-0.625).
As an example, the neural activity (E) according to Eq. 1-2 as a function of the external interaural time difference for a sinusoid is shown in Fig. 2. The solid line represents the neural activity of the model for α=0.03, τ=250 μs, the circles represent neurophysiological data from Kuwada et al. (1984).

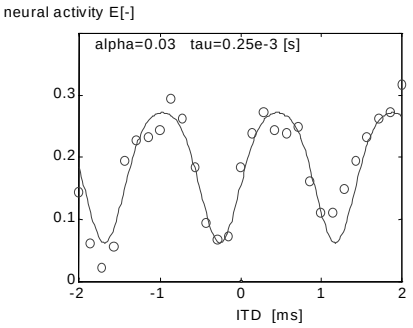


Figure 2. EI-cell activity as a function of the external interaural delay for a sinusoid of 700 Hz center frequency. Solid line: model predictions according to Eq. 4, for α=0.03, τ=250 μs, on-frequency filter. Circles: data adapted from Kuwada et al. (1984), multiplied by a constant factor to convert the spike rate to a normalized neural activity.

1.5 Binaural detection mechanism

The binaural detector extracts binaural information from the EI-activity pattern. The analysis of the pattern resulting from a stimulus is done by means of an optimal detector that weights the information for each individual auditory filter. In this model, a reference is formed by the mean excitation pattern for a masker alone. This reference is called a template, and can be obtained easily by averaging the excitation patterns for several maskers alone intervals. In a 3IFC experiment, where the subject (i.e., the model) has to detect the stimulus containing the test signal, the detection mechanism computes the excitation pattern for each stimulus, and compares it to the template, by integrating the difference between excitation pattern and template for each filter. Here, it is assumed that cells with smaller characteristic delays are more frequent than cells with larger characteristic delays.

Therefore, a weighting function p(τ), as described by Colburn (1977) is applied in the integration process:

Equation 3

$$p(\tau) = \begin{cases} 1 & for \quad |\tau| \le 0.15ms \\ \exp\left(\frac{-|\tau| + 0.15}{0.6}\right) & for \quad 0.15 \le |\tau| \le 2.2ms \\ 0.033\exp\left(\frac{-|\tau| + 2.2}{2.3}\right) & for \quad |\tau| > 2.2ms \end{cases}$$

The internal errors (or internal noise) are modeled by adding a Gaussian-distributed random variable with mean zero and fixed variance to all outputs.

2 Simulation Results

2.1 Procedure and stimuli

We compared model predictions with the results from several experimental studies with human observers which were adapted from literature. The procedure used in determining the model predictions was the same as the procedure that was used in the psychoacoustical studies.

2.2 Effect of masker bandwidth

To study the dependence of binaural detection thresholds on masker bandwidth, binaural thresholds were simulated for several binaural masking conditions for different center frequencies, as a function of masker bandwidth. These experiments were adapted from Breebaart et al. (1998a). Fig. 3 shows signal-to-noise ratios of the model (filled blocks) in comparison with experimental data (open symbols). The upper-left panel shows the thresholds for an NoSπ condition at 125-Hz center frequency, while the upper-right panel shows the same condition at 500-Hz center frequency. We see that the model predictions fit the experimental data quite well. For bandwidths beyond 100 Hz, the thresholds increase with 5 dB at 125 Hz and 3 dB at 500-Hz center frequency. For larger bandwidths, the thresholds decrease again. Interestingly, this decrease starts at approximately 2 to 3 times the critical bandwidth.

The middle panels show thresholds for the detection of an in-phase sinusoid added to a time-delayed noise masker (NτSo, blocks, τ equals half the period of the center frequency) and a phase-reverted noise masker (NπSo, triangles) at 125-Hz center frequency (left panel) and 500-Hz center frequency (right panel). The open blocks represent experimental NτSo data from Breebaart et al. (1998a). The

experimental data for the NπSo conditions are not shown, since Breebaart et al. (1998a) showed that the NτSo and NπSo conditions have equal thresholds. We see that the predictions and the experimental data are very similar, and that the NπSo and NτSo conditions show very similar thresholds, perfectly in line with the experimental observations.

The bottom panels show thresholds for the detection of an out-of-phase multiplied noise added to an in-phase sinusoidal masker as a function of the bandwidth of the noise at 500-Hz center frequency. The left panel shows data for

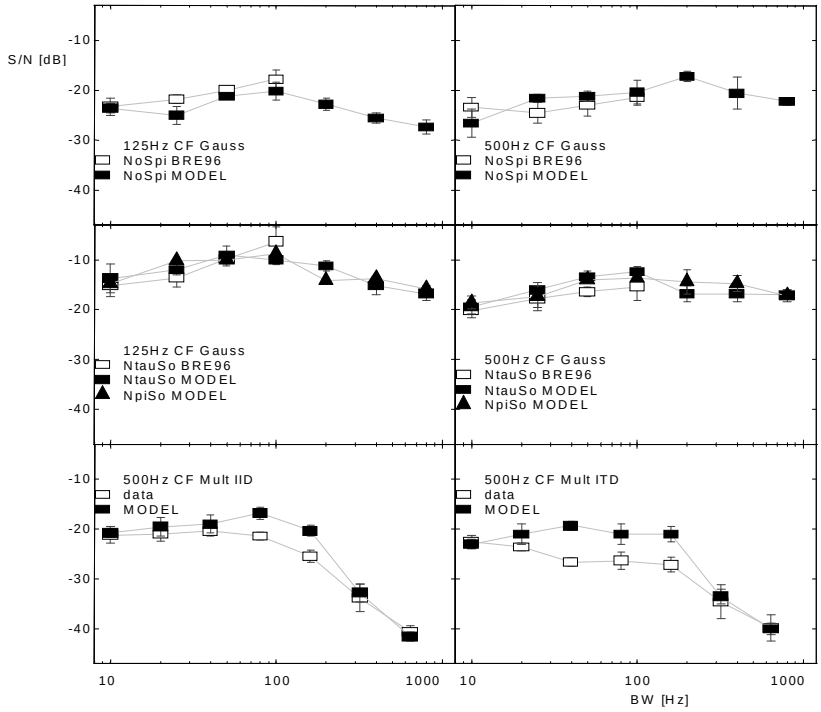


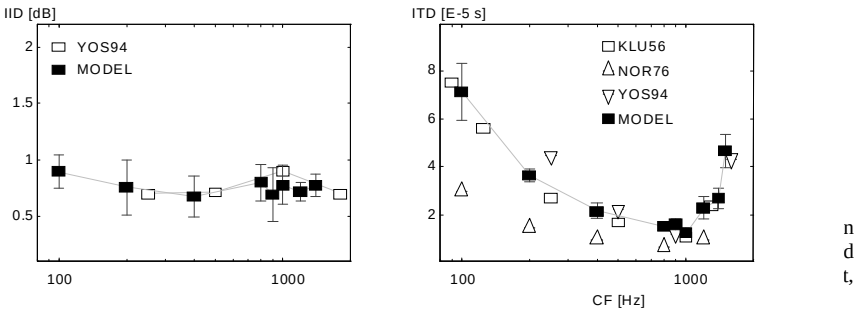
Figure 3 - Bandwidth dependence of the model compared to experimental data. Filled symbols: model predictions. Open symbols: experimental data. Upper panels: NoSπ condition at 125 Hz (left) and 500 Hz (right) center frequency. Middle panels: NπSo and NτSo conditions at 125 Hz (left) and 500 Hz (right) center frequency. Lower panels: NoSπ (sinusoidal masker with out-of-phase multiplied noise test signal for θ=0 (left) and θ=π/2 (right)) at 500 Hz center frequency. Errorbars denote the standard deviation of the mean.

an IID condition (θ=0), while the right panel shows data for an ITD condition (θ=π/2). The open symbols represent the experimental data from Breebaart et al. (1998b), the filled symbols represent the model predictions.

For bandwidths greater than 100 Hz, both experimental and predicted thresholds decrease in a very similar way. However, the ITD condition shows slightly higher predicted thresholds for bandwidths between 40 and 160 Hz.

2.3 Frequency dependence of static interaural differences

The dependence on center frequency of static interaural difference in time (right panel) and intensity (left panel) for sinusoids is shown in Fig. 4. As in the preceding figures, the open symbols represent experimental data adapted from literature, while the filled symbols represent the model predictions. We see that for both the experimental data and the predicted thresholds, the intensity jnd is approximately
constant over the whole frequency range. For ITDs, however, the interaural time difference jnd decreases with center frequency for frequencies below 1 kHz, while for frequencies above 1 kHz, the jnds increase rapidly. The increase in sensitivity with increasing center frequency equals a constant phase jnd, while the decrease in sensitivity above 1 kHz is the result of the decrease of phase locking in the inner hair cell model.



n
d
t,

2.4 Interaural correlation of the masker

To study the dependence on interaural correlation of the masking noise, the detection thresholds of the model were determined for several (negative) interaural correlations of a narrowband masking noise (NρSo condition). Thresholds were measured at a center frequency of 125 Hz and a bandwidth of 50 Hz.

The predicted (filled symbols) and experimental (open symbols) thresholds are shown in Fig. 5. The open blocks represent data from Breebaart (1996) for a 50-Hz wide masker, while the open triangles represent data from Robinson and Jeffress (1963) for a broadband noise masker. As the correlation is changed from -1 to -.95, thresholds increase by about 3 dB. As the correlation increases further, thresholds increase with increasing correlation, but with a decreasing slope.

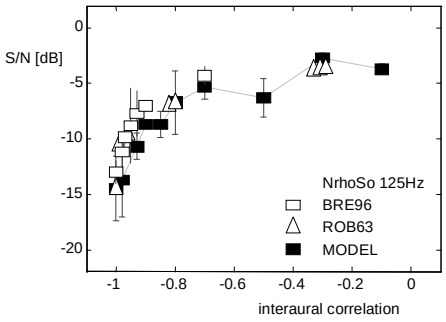


Figure 5 - Detection thresholds as a function of the interaural correlation of the masker at 125Hz center frequency. Errorbars denote the standard deviation of the mean. Filled symbols: model predictions. Legend: BRE96=Breebaart,1996; ROB63 = Robinson and Jeffress, 1963.

2.5 Temporal integration

During the previous simulations, the changes in the internal activity were computed from the whole stimulus, i.e., 400 ms. Data from Grantham & Wightman (1978) and Kollmeier & Gilkey (1990) show that the binaural system has a limited time resolution. To incorporate a limited time resolution in the model, the following condition from Kollmeier and Gilkey (1990) was used.

Here, the first 375 ms of the masker consist of a diotic noise of 500-Hz bandwidth centered at 500 Hz. The next 375 ms (i.e., from 375 to 750 ms) consist of a phase-inverted noise of the same bandwidth and center frequency. Thus, at the temporal center, the masker switches from No to Nπ. The test signal consists of a 20-ms out-of-phase sinusoid, ramped by a 2-ms Hanning window. Furthermore, the total signal consisting of masker plus test signal is windowed temporally by a symmetric exponential window with a variable offset. By varying the temporal position of the signal relative to the phase transition of the masker, the condition shifts from NoSπ to NπSπ, resulting in a change in the detectability with temporal position.

The position of the temporal window resulting in the lowest detection thresholds is not always given by centering the window on the signal. To determine the optimal position of the temporal window for a given signal offset, the window was shifted from -70 to +70 ms away from the temporal center of the test signal. The lowest signal-to-noise ratio achieved was used as a detection threshold for that signal offset.

The simulation results (filled symbols) and data from Kollmeier and Gilkey (1990) are shown in Fig. 6 (left panel). The right panel shows the temporal window with a constant of 25 ms.

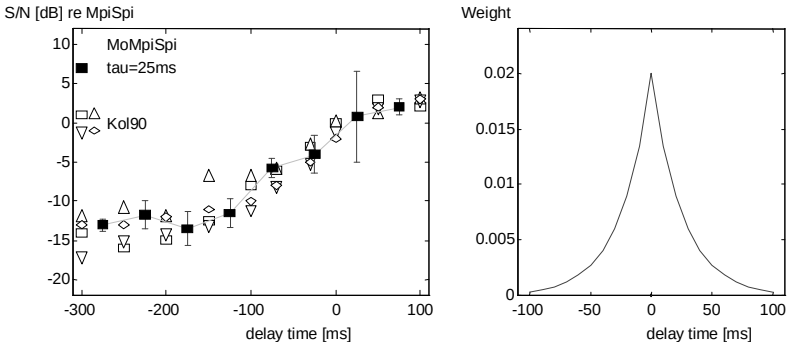


Figure 6 – Signal-to-noise ratios as a function of the delay time of the signal-offset; 0 corresponds to the phase-switch of the masker (left panel) and temporal window (right panel). Note that according to the data from Kollmeier and Gilkey (1990), the thresholds are expressed in comparison to NπSπ thresholds. The errorbars denote the standard deviation of the mean.

2.6 Masker-level dependence

In Fig. 7, the signal level at threshold is shown for a 500-Hz out-of-phase sinusoid added to an in-phase 500-Hz-wide noise as a function of the spectral level of the noise.

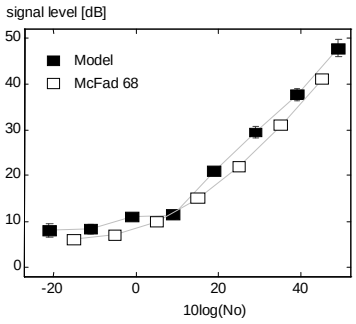


Figure 7 - signal level at threshold for a 500-Hz out-of-phase sinusoid added to a 500-Hz wide masker as a function of the spectral energy density of the noise. Open symbols: data adapted from literature (McFadden, 1968). Filled symbols: model predictions.

2.7 Discussion of the bandwidth dependence of the model

In the model, three effects play an important role in the dependence of detection thresholds on the bandwidth of the masker (see Fig. 3):

1. **The peripheral filtering**. For narrowband maskers, the masker power in the on-frequency filter is (almost) equal to the total power of the masker. With increasing masker bandwidth, the masker power in the on-frequency filter decreases due to the bandpass characteristic of the filter if the total masker power is kept constant. For masker bandwidths that extend the bandwidth of the peripheral filter, a -3 dB/Oct decrease of noise energy is expected, resulting in lower signal-to-noise ratios. This effect plays a minor role for masker bandwidths smaller than the ERB.
2. **The change of information across filters**. For a narrowband diotic masker, the output variable of the on-frequency filter is (almost) equal to the output variable of the adjacent filters. Therefore, the model effectively computes the mean of these output variables and uses it as a decision variable. Since the errors induced by the internal noise are independent in each filter, the model reduces the internal error, resulting in low detection thresholds. With increasing masker bandwidth, the off-frequency filters contain more and more noise power and a relatively decreasing amount of the signal to be detected. Therefore, the output variable of the off-frequency filters contains less information about the presence of an out-of-phase signal, resulting in less cancellation of the internal noise. Therefore, the thresholds increase with increasing masker bandwidth, even for bandwidths up to 3 times the ERB. This explains the phenomenon that for an NoSπ condition, the effective auditory filter bandwidth is wider than the ERB.
3. **Variance due to decorrelation**. The output of an EI-type cell can be written in terms of the interaural correlation at a delay equal to the internal delay of the EI-type cell (Breebaart et al., 1997). Since the EI-activity is computed as a funtion of the internal delay, the change in activity of the EI-type cell due to the addition of the test-signal can be written in terms of a change of the interaural correlation function. With increasing masker bandwidth, the subsequent periods of the masker become less similar (i.e., damping of the cross-correlation function). Furthermore, the variance of the normalized cross-correlation function increases with increasing masker bandwidth. This increase of the variance results in higher detection thresholds with increasing masker bandwidth. The damping of the cross-correlation function is limited by the peripheral filtering. Therefore, this effect only plays a role for bandwidths smaller than the ERB.

3 Conclusions

The presented model is able to predict the dependence of binaural thresholds on several stimulus parameters. The tested conditions show a good fit for changes in bandwidth or correlation of the masker, signal durations, static and dynamic cues, level and frequency dependencies.

The combination of information across auditory filters seems to work fine if the separate information is summed according to the described optimal criterion. This approach also accounts for the wider effective auditory filters for an NoSπ condition. However, not all phenomena concerning binaural detection are covered. For example, the phenomenon of binaural interference (Bernstein and Trahiotis, 1995) cannot be accounted for by this model nor do thresholds increase as a result of distortion products at the basilar membrane (Van der Heijden et al., 1997). These effects need a more sophisticated model of the binaural detection mechanism and the basilar membrane, respectively.

Summarizing, a system based on contralateral inhibition can account for a wide range of effects found in binaural detection tasks. One of the main advantages of an EI-type cell based system in comparison with an EE-type (i.e., correlation) based system is the incorporation of static interaural intensity differences.

4 Bibliography

1. Albeck, Y., Konishi, M. (1995). Responses of neurons in the auditory pathway of the barn owl to partially correlated binaural signals. *J. Neurophysiol*. 74, 1689-1700.
2. Bernstein, L.R., Trahiotis, C. (1995). Binaural spectral interference in detection and discrimination paradigms. In *Advances in Hearing Research: proceedings of the 10th international symposium on hearing*. Eds. G.A. manley, G.M. Klump, C. Koppl, H. Fastl, H. Oekinghaus.
3. Breebaart, J. (1996). *Modeling the influence of masker bandkwidth on BMLDs*. Institute for Perception Research (IPO), Eindhoven, The Netherlands. Report 1109.
4. Breebaart, J. (1997). *Binaural signal detection: experiments, modeling and applications. Applications for sound source coding*. Institute for Perception Research (IPO), Eindhoven, The Netherlands. Report 1155.
5. Breebaart, J., Van de Par, S., Kohlrausch, A. (1998a). Binaural signal detection with phase-shifted and time-delayed noise maskers. *J. Acoust. Soc. Am*. 103, 2079-2083.

6. Breebaart, J., Van de Par, S., Kohlrausch, A. (1998b). The contribution of static and dynamically varying ITDs and IIDs on binaural detection. *J. Acoust. Soc. Am.* Submitted for publication.
7. Glasberg, B.R., Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*. 47, 103-138.
8. Goldberg, J.M., Brown, P.B. (1969). Response of binaural neurons of dog superior olivary complex: an anatomical and electrophysiological study. *J. Neurophysiol*. 31, 613-636.
9. Grantham, D.W., Wightman, F.L. (1978). Detectability of varying interaural temporal differences. *J. Acoust. Soc. Am*. 63, 511-523.
10. Jeffress, L.A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology*. 41, 35-39.
11. Joris, P.X. (1996). Envelope coding in the lateral superior live. II. Characteristic delays and comparison with responses in the medial superior olive. *J. Neurophysiol*. 76, 2137-2156.
12. Joris, P.X., Yin, T.C.T. (1995). Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences. *J. Neurophysiol*. 73, 1043-1062.
13. Klumpp, R.G., Eady, H.R. (1956). Some measurements of interaural time difference thresholds. *J. Acoust. Soc. Am*. 28, 859-860.
14. Kollmeier, B., Gilkey, R.H. (1990). Binaural forward and backward masking: evidence for sluggishness in binaural detection. *J. Acoust. Soc. Am*. 87, 1709-1719.
15. Kuwada, Y., Yin, T.C.T., Syka, J., Buunen, J.F., Wickesberg, R.E. (1984). Binaural interaction in low-frequency neurons in he inferior colliculus of the cat. IV. Comparison of monaural and binaural response properties. *J. Neurophysiology*. 51, 1306-1325.
16. McFadden, D. (1968). Masking level differences determined with and without interaural disparities in masker intensity. *J. Acoust. Soc. Am*. 44, 212-223.
17. Nordmark, J.O. (1976). Binaural time discrimination. *J. Acoust. Soc. Am*. 60, 870-880.
18. Oxenham, A.J., Moore, B.C.J. (1994). Modeling the additivity of nonsimultaneous masking. *Hearing Research*. 80, 105-118.
19. Van der Heijden, M., Trahiotis, C., Kohlrausch, A., Van de Par, S. (1997). Binaural signal detection with spectrally non-overlapping signals and maskers: Evidence for masking by aural distortion products. Submitted for J. Acoust. Soc. Am.
20. Yost, W.A. (1994). *Fundamentals of hearing. An introduction*. Academic Press, London.