JUNE 10 2022

Lower interaural coherence in off-signal bands impairs binaural detection

Bernhard Eurich; Jörg Encke; Stephan D. Ewert; ... et. al

J Acoust Soc Am 151, 3927–3936 (2022)
https://doi.org/10.1121/10.0011673

JASA ARTICLE

Lower interaural coherence in off-signal bands impairs binaural detection

Bernhard Eurich,[a),b)] Jörg Encke,[b)] Stephan D. Ewert,[b)] and Mathias Dietz[b)]
Department für Medizinische Physik und Akustik, Universität Oldenburg, 26111 Oldenburg, Germany

ABSTRACT:
Differences in interaural phase configuration between a target and a masker can lead to substantial binaural unmasking. This effect is decreased for masking noises with an interaural time difference (ITD). Adding a second noise with an opposing ITD in most cases further reduces binaural unmasking. Thus far, modeling of these detection thresholds required both a mechanism for internal ITD compensation and an increased filter bandwidth. An alternative explanation for the reduction is that unmasking is impaired by the lower interaural coherence in off-frequency regions caused by the second masker [Marquardt and McAlpine (2009). J. Acoust. Soc. Am. **126**(6), EL177–EL182]. Based on this hypothesis, the current work proposes a quantitative multi-channel model using monaurally derived peripheral filter bandwidths and an across-channel incoherence interference mechanism. This mechanism differs from wider filters since it has no effect when the masker coherence is constant across frequency bands. Combined with a monaural energy discrimination pathway, the model predicts the differences between a single delayed noise and two opposingly delayed noises as well as four other data sets. It helps resolve the inconsistency that simulating some data requires wide filters while others require narrow filters. https://doi.org/10.1121/10.0011673

(Received 10 November 2021; revised 22 May 2022; accepted 24 May 2022; published online 10 June 2022)

[Editor: Matthew J. Goupell] Pages: 3927–3936

I. INTRODUCTION

The detection of a pure tone in noise is facilitated by differences in the interaural phase between tone and noise (Hirsh, 1948). The improvement in the detection threshold compared to the diotic case is referred to as the binaural masking level difference (BMLD). The maximum BMLD is observed when detecting an antiphasic pure tone target (S_π) in an in-phase noise masker (N_0). Adding an interaural time difference (ITD) to the masker has been observed to reduce the BMLD (Langford and Jeffress, 1964). A particularly simple case is when the noise and the target tone have exactly opposite interaural phase differences. In this case, detection thresholds increase gradually and monotonically with increasing noise ITD (Rabiner *et al.*, 1966). The increase can be simulated accurately by exploiting changes in the cross correlation coefficient of the left and right signals after using a filter with an equivalent rectangular bandwidth (ERB) of 60 to 85 Hz at a center frequency of 500 Hz (Dietz *et al.*, 2021; Rabiner *et al.*, 1966). This bandwidth range resembles the established estimate of the human peripheral filter bandwidth obtained from monaural psychoacoustic experiments at this frequency, which is 79 Hz (Glasberg and Moore, 1990) and referred to as standard filter bandwidth in the following.

Another explanation uses an array of different internal delays, known as delay lines (Bernstein and Trahiotis, 2018, 2020; Stern and Colburn, 1978; van der Heijden and Trahiotis, 1999). Jeffress (1948) suggested that the binaural system has the ability to compensate for the external ITD. The compensation accuracy or efficiency has been assumed to decrease with masker ITD in order to model the decreasing BMLD (Bernstein and Trahiotis, 2017; Stern and Colburn, 1978; van der Heijden and Trahiotis, 1999).

van der Heijden and Trahiotis (1999) generated a new stimulus, which they termed "double-delayed noise" [diamonds in Fig. 1(E)] by adding two noises, one with a positive and one with a negative ITD. We refer to this as opposingly delayed noises (ODN). They found detection thresholds in ODN to be substantially higher than in "regular" delayed noise termed "single-delayed noise" (SDN). Since internal delays can only compensate for the ITD of one noise, ODN limits the usefulness of the putative delay lines. Thus, van der Heijden and Trahiotis (1999) attributed the additional unmasking in SDN, compared to ODN, to the delay lines. Irrespective of the use of internal delays, however, a large part of the threshold differences between the two stimuli is caused by the interaural coherence oscillating as a function of ITD in ODN while monotonically decreasing in SDN [see the coherence $|\gamma|$ pattern in Fig. 1(C)]. More relevant for the role of internal delays are those ITDs that are the multiples of half the period. There, the coherence is the same in SDN and ODN but thresholds differ. For S_0 detection at 500 Hz this is the case at ITD = 1 and 3 ms, for S_π detection at ITD = 2 and 4 ms. So far, only

[a)]Electronic mail: bernhard.eurich@uni-oldenburg.de
[b)]Also at: Department für Medizinische Physik und Akustik and Cluster of Excellence "Hearing4All", Universität Oldenburg, 26111 Oldenburg, Germany.

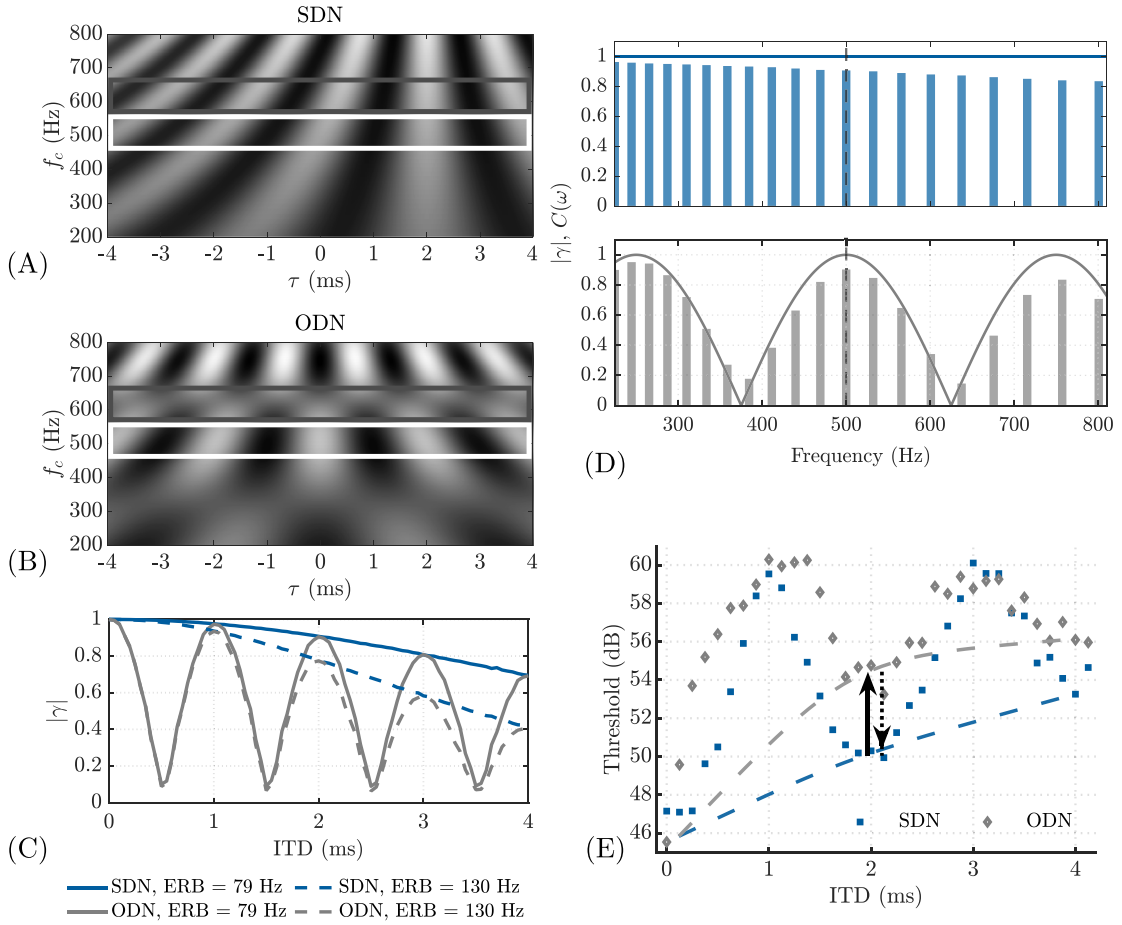

FIG. 1. (Color online) (A) Cross-correlogram of delayed noise (SDN) with ITD = 2 ms. White and black areas represent maxima and minima of the cross correlation functions, respectively. The white box highlights the 500 Hz frequency channel while the gray box highlights a channel centered at 625 Hz. (B) Interaural cross-correlogram as in (A) but for opposingly delayed noises (ODN). (C) Interaural coherence $|\gamma|$ as a function of noise ITD for SDN (blue lines) and ODN (gray lines) for two underlying filter bandwidhts. (D) Continuous lines: Normalized cross-power spectral density (CPSD) at ITD = 2 ms as a function of frequency, $C(\omega)$, as derived in Eq. (A2) and what follows; Bars: Interaural coherence $|\gamma|$ of the signals after peripheral Gammatone filtering. (E) Thresholds of S_π detection in SDN and ODN as a function of ITD from van der Heijden and Trahiotis (1999). The dashed lines symbolize the coherence-decline-induced threshold increase determined by a filter bandwidth of ERB = 79 Hz (lower line) and ERB = 130 Hz (upper line). As denoted by the arrows, the data can be explained in two ways: [(1) dotted downward arrow] The ODN thresholds are determined by the cross correlation function at 500 Hz and a bandwidth $\geq$ 130 Hz. A delay line causes the lower SDN thresholds. [(2) solid upward arrow] The SDN thresholds are determined by the ITD-dependent coherence as derived from an ERB of 79 Hz. Off-frequency incoherence in ODN causes higher ODN thresholds.

the model of van der Heijden and Trahiotis (1999), which is based on delay lines, has precisely accounted for both SDN and ODN detection thresholds. The SDN-ODN detection threshold difference is therefore used as psychoacoustic evidence for several millisecond long delay lines (Stern *et al.*, 2019). The difference was in fact used to derive the length and potency of the delay line system (van der Heijden and Trahiotis, 1999).

However, two problems exist with establishing the psychoacoustically derived delay line length or internal delay distribution function. First, measured delays in binaural neurons of mammals are short compared to the respective period duration (Joris *et al.*, 2006; McAlpine *et al.*, 2001; see also Leibold and Grothe, 2015 for review) and thus too short to fulfil the lengths requirements of delay line models (Marquardt and McAlpine, 2009; Stern *et al.*, 2019; Thompson *et al.*, 2006).

Second, if the delay-line models use their internal delays to account for SDN thresholds while correlation coefficient-based models (Encke and Dietz, 2022; Rabiner *et al.*, 1966) are equally precise for SDN without delay lines, the two model types must differ in some other manner, such as filter bandwidth. van der Heijden and Trahiotis (1999) used ODN thresholds to determine the filter bandwidth. They could best fit their ODN thresholds with filters of various shapes and an ERB of 130 to 180 Hz at 500 Hz center frequency. This is expectedly larger than what models without delay lines, such as Encke and Dietz (2022), required for SDN. The two versions cannot both be correct. Thus, either the SDN-threshold-based filter bandwidth is confounded by not considering delay lines or the ODN-based filter bandwidth fit by van der Heijden and Trahiotis (1999) is confounded by something else. For the latter, Marquardt and McAlpine (2009) offered a possible explanation. They identified the interaural coherence to be lower in certain off-frequency regions in ODN but not in SDN. They argued that the higher detection thresholds in ODN could also originate from some detrimental off-frequency impact related to the

https://doi.org/10.1121/10.0011673

low coherence rather than from a wider filter bandwidth *per se* [upward arrow in Fig. 1(E)]. If this is true, both SDN and ODN thresholds can potentially be predicted using the same standard filter bandwidth. Figures 1(A), 1(B), and 1(D) show that the cross-power spectral density is constant across frequency in SDN but spectrally modulated in ODN (see the Appendix for derivation).

Leaving aside the first physiologic argument, there are two options to account for the SDN-ODN difference, (1) wider filters combined with delay lines [downward arrow in Fig. 1(E)] or (2) filters with standard peripheral bandwidths and a detrimental off-frequency impact [upward arrow in Fig. 1(E)]. However, recent data of SDN thresholds measured for different noise bandwidths can only be accurately simulated with filters falling into the standard peripheral bandwidth category (Bernstein and Trahiotis, 2020; Dietz *et al.*, 2021), causing a logical impasse for the wider-filters assumption even within the psychoacoustic domain and for SDN alone.

The aim of this study is thus to develop a model according to option (2) that accounts for SDN and ODN thresholds at the same time, using a standard filter bandwidth and—consequently—without several millisecond long delay lines. Rather, we suggest an across-frequency incoherence interference mechanism that is inspired by binaural interference (Bernstein and Trahiotis, 1995) and modulation detection interference (Oxenham and Dau, 2001; Yost and Sheft, 1989). With a low coherence qualitatively reflecting strong IPD fluctuations,[1] this can be thought of as an interference of IPD fluctuations across frequency channels. With this mechanism, the same "hardware" causes different detection thresholds for maskers with the same on-frequency coherence but with a lower interaural coherence in off-frequency channels. The here developed incoherence interference will be described in Sec. II and used to predict critical binaural detection data in Sec. III.

Besides the discussion concerning delay lines in humans and other mammals, the width of filters has caused an unresolved contradiction in the binaural literature that filters need to be narrow to account for some and broad to account for other data [see Verhey and van de Par (2020) for a review]. Generally speaking, detection thresholds in spectrally simpler maskers can be simulated using a standard peripheral filter bandwidth (Breebaart *et al.*, 2001b), whereas more complex maskers appear to be processed by wider filters or alternative across-frequency processes (Kolarik and Culling, 2010). We therefore evaluated our model with data from five different studies in three groups:

(1) van der Heijden and Trahiotis (1999) combined all key aspects required to revisit Marquardt and McAlpine's hypothesis: (a) The SDN thresholds are planned to be determined by the decay of $|\gamma|$ with a 79 Hz-wide Gammatone filter. (b) The ODN thresholds supposedly will, despite the same 79 Hz on-frequency filter, be elevated by the across-channel incoherence interference.
(2) Marquardt and McAlpine (2009) not only presented the above-mentioned hypothesis but also experimental data with a novel type of stimuli. There, SDN and ODN maskers are spectrally surrounded by bands that each have a different, constant IPD. Certain flanking-band IPDs do while others do not cause interaural incoherence at the transitions. Their reported differences impose a challenge for single-channel models that use a constant filter bandwidth.
(3) Sondhi and Guttman (1966), Holube *et al.* (1998), and Kolarik and Culling (2010) reported detection thresholds of an S_π tone centered in an in-phase noise that is spectrally surrounded by antiphasic noise. These simulations are included for an additional discussion about the proposed standard-filter-plus-off-frequency-impact concept, since larger binaural than peripheral bandwidths have previously been derived based on such data.

II. DESCRIPTION OF THE MODEL

Figure 2 shows the processing stages of the proposed model. It is designed as a numerical multi-channel model through all stages, but these were here realized and tailored to predict binaural-detection data with a 500 Hz pure-tone target. The model builds on the analytical single-channel model approach of Encke and Dietz (2022). It furthermore includes an across-frequency incoherence interference mechanism. It consists of a multi-channel binaural processing pathway and a monaural pathway in order to account both for conditions that provide interaural or only energetic cues. In both pathways, multiple tokens of the processed representation of the condition-specific masker only are compared to the representation of signal plus masker. This comparison has been suggested to mimic a subject's strategy of comparing a stimulus to a learned reference template (Bernstein and Trahiotis, 2017; Breebaart *et al.*, 2001a). Based on these comparisons, both pathways deliver a sensitivity index (d'). An optimal combination of the pathways' estimates gives the overall d' estimate of the model (Biberger and Ewert, 2017; Green and Swets, 1966).

A. Peripheral processing

The left and right input signals were processed with a fourth-order Gammatone filterbank that represents basilar-membrane bandpass filtering. The filterbank implementation by Hohmann (2002) was employed with a spacing of five filters per ERB in the range of 67 to 1000 Hz. The grid was defined by centering one filter at 500 Hz. This filter had an ERB of 79 Hz (Glasberg and Moore, 1990) and was indexed with $k = 0$.

To focus on the impact of the spectral masker properties discussed above, the present implementation did not include any other peripheral processing such as low-pass filtering, power-law compression or half-wave rectification. Only Gaussian noise was used as masker and only 500-Hz tones as targets.

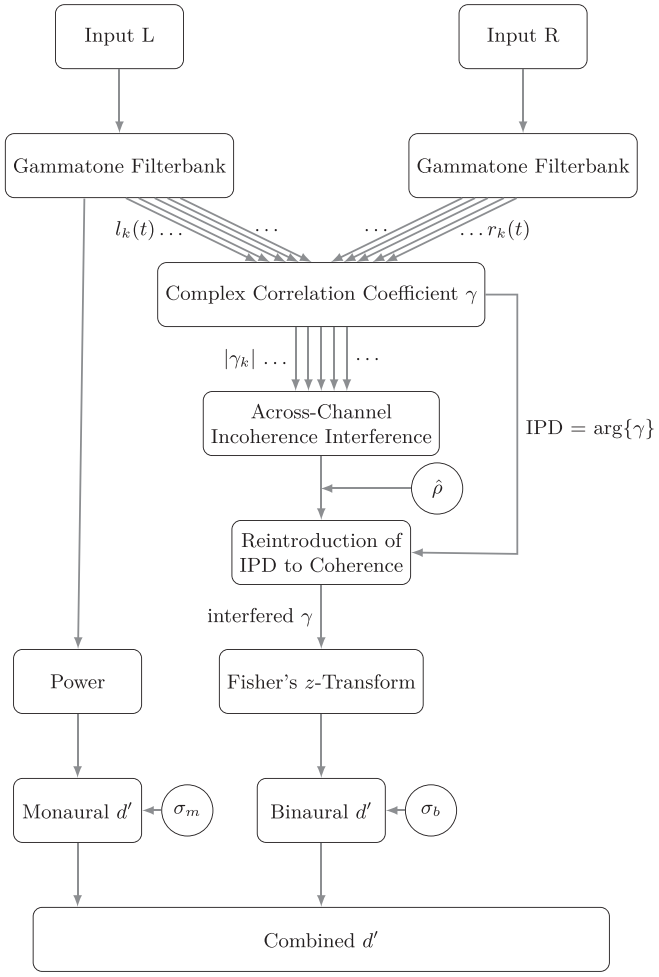

FIG. 2. Processing stages of the proposed model. See main text for details.

B. Binaural pathway

The correlation coefficient $\gamma(\tau = 0) = \gamma$ was derived from the analytical (i.e., complex-valued) left and right signals $l(t)$ and $r(t)$ in the frequency channel k, provided by the Gammatone filterbank:

$$\gamma_k = \frac{\overline{l_k(t)^* r_k(t)}}{\sqrt{\overline{|l_k(t)|^2}\,\overline{|r_k(t)|^2}}}, \tag{1}$$

where $\bar{\bullet}$ denotes the mean over the duration of the signal. This results in one complex correlation coefficient per frequency channel, averaged over the whole stimulus duration. The complex-valued correlation coefficient was used because it conveniently combines information about both the mean IPD as arg{γ} and about the interaural coherence $|\gamma|$. While the Introduction mentioned a mismatch between mammalian physiology and delay line models, it should be noted that the seemingly abstract use of complex-valued correlation is identical to two real-valued correlations with a 90° phase offset. Such two orthogonal correlators exist in the form of the average left- and right hemispheric binaural neuron in mammals (Joris *et al.*, 2006; McAlpine *et al.*, 2001). The physiologic relation of γ is explained in more detail in Encke and Dietz (2022).

As pointed out in the Introduction, the novelty of the present model is the incoherence interference across frequency channels. The term *incoherence interference* was chosen to describe the purely detrimental nature of the interaction. Only channels with lower coherence affect their neighborhood, but not the other way around. This process is implemented as a *restricted* across-channel weighted average of the coherence $|\gamma|_k$: The $|\gamma|_k$ are limited such that they can no more exceed the on-frequency $|\gamma|$, thus referred to as $|\gamma|_{k,\mathrm{lim}}$,

$$|\gamma|_w = \sum_{-m}^{m} w(k)|\gamma|_{k,\mathrm{lim}}. \tag{2}$$

$w(k)$ symbolizes a function that weights the contribution of a channel k to the resulting $|\gamma|_w$. The employed weighting function has an exponential decay described by

$$w(k) = e^{-|k|/(b\sigma_w)}. \tag{3}$$

σ_w represents the decay parameter, normalized by the number of filters per ERB, b. The double-exponential decay shape was chosen by empirical trials. While the exact shape of the window was not crucial, we did not obtain more precise simulations with other shapes.

For a masker coherence close to zero or at the practically irrelevant case of a positive signal-to-noise ratio (SNR), adding a target with an IPD of π relative to the masker can swap the mean IPD from the masker to that of the target. In special cases, the masker alone and masker plus target can have the same coherence but differ in their mean IPD and thus in their correlation. Hence, the interaural coherence $|\gamma|$ is not sufficient as a decision variable. Instead, γ, including both coherence and the mean IPD, is required. Therefore, the original mean IPD is now reintroduced to the coherence after the limitation and interference stage, so that the model can operate on the *complex* correlation coefficient as suggested by Encke and Dietz (2022),

$$\gamma_w = |\gamma|_w e^{\arg\{\gamma_0\}}. \tag{4}$$

Unity-limited measures such as coherence or correlation can be Fisher z (i.e., atanh) transformed for the purpose of variance normalization (Just and Bamler, 1994; McNemar, 1969), as often applied in psychophysics [e.g., Bernstein and Trahiotis, 2017; Lüddemann *et al.*, 2007). As in Encke and Dietz (2022), γ_w is multiplied by a model parameter $\hat{\rho} < 1$ to avoid an infinite sensitivity to deviations from a coherence of unity. This is equivalent to adding uncorrelated noise to the two input signals. The decision variable of the binaural pathway is thus

$$\zeta = z[\hat{\rho}\gamma_w], \tag{5}$$

where $z[\bullet]$ is the Fisher z-transform applied to the modulus of γ_w while leaving the argument unchanged.

https://doi.org/10.1121/10.0011673

In the signal detection stage, the d' is obtained based on the difference between the ensemble averages of the representations of the target signal plus noise, ζ_{N+S}, and the representations of the noise alone, ζ_N,

$$d'_b = \frac{|\zeta_{N+S} - \zeta_N|}{\sigma_b}. \quad (6)$$

The internal noise σ_b defines the sensitivity of the binaural model pathway (Dietz *et al.*, 2021).

C. Monaural pathway

For the monaural pathway, the power P of the on-frequency filter channel was evaluated. It is half the squared mean of the envelope across the whole signal duration (Biberger and Ewert, 2016). The envelope is the modulus of the complex-valued filter output,

$$P = \frac{\overline{|u_0(t)|}^2}{2}. \quad (7)$$

In the stimuli employed in this study, the power is identical in the left and right channels, thus it is sufficient to evaluate only one side.

For a signal-induced power change $\Delta P = P_{N+S} - P_N$, the processing accuracy is limited by a level-dependent internal noise with a Gaussian distribution of amplitudes and a standard deviation of σ_m. Thus, the sensitivity of the monaural pathway is equivalent to

$$d'_m = \frac{\Delta P / P_{\text{avg}}}{\sigma_m}, \quad (8)$$

where P_{avg} represents the average power between P_{N+S} and P_N.

D. Detector

The sensitivity indices of the binaural, d'_b, and monaural pathway, d'_m, were combined assuming two independent information channels (Biberger and Ewert, 2016; Green and Swets, 1966),

$$d'_{b+m} = \sqrt{d'^2_b + d'^2_m}. \quad (9)$$

The d' that corresponds to the experiment-specific detection thresholds was obtained via table-lookup [numerical evaluation in Hacker and Ratcliff (1979)]. This depends on the number of intervals as well as the specific staircase procedure used in the simulated experiments. For each condition, the model was evaluated for a range of target levels. This delivered the psychometric function. The predicted detection threshold was obtained from a straight line fitted to the logarithmic d'. The model parameters were manually adjusted in order to optimize the prediction accuracy. The resulting parameter values are given in Table I.

III. PREDICTIONS OF BINAURAL-DETECTION DATASETS

In all experiments, a 500 Hz S_π or S_0 tone was to be detected in a broadband Gaussian noise masker. Figures 3–5, show the experimental data denoted by symbols and the predictions of the proposed model including incoherence interference (continuous lines), as well as excluding incoherence interference (as dotted lines). Three types of binaural-detection experiments were simulated, as described in detail in the following. Table I summarizes the parameter values used to simulate the experimental conditions. It further lists the non-adjusted coefficient of determination (R^2, interpretable as the proportion of variance in the data explained by the model) and the root mean square error (RMSE) of the simulations both with and without the proposed incoherence interference.

A. van der Heijden and Trahiotis (1999)

In this arguably most central experiment, detection thresholds of an S_0 target tone (Fig. 3, upper panel) as well as of an S_π tone (Fig. 3, lower panel) were measured as a function of the interaural masker ITD in steps of 0.125 ms. The bandwidth of the masker was 900 Hz. As outlined in the Introduction, the ODN consisted of two superimposed noises with opposite ITD. The experiment performed by van der Heijden and Trahiotis (1999) employed a four-interval, two-alternative forced choice task (4I-2AFC, first and fourth intervals always contained only the masker and served as queuing intervals). Their adaptive 2-down 1-up stair case procedure estimated the 70.7% correct-response threshold. This is equivalent to a d' of 0.78 at threshold. Thus, as

TABLE I. Summary of the simulated experiments and predictions. Columns 1–3: Simulated experiment, IPD of the used target signal, independent variable. Columns 4–7: Used model parameters—$\hat{\rho} < 1$: Maximum coherence (internal noise); σ_b: Standard deviation of the internal noise to determine the absolute performance of the binaural pathway; σ_w: Slope parameter of the double-exponential across-channel interaction window (normalized by the number of filters per ERB); σ_m: Standard deviation of the level-dependent internal noise to determine the accuity of the monaural pathway. Columns 8–11: Accuracy of the predictions with and without incoherence interference—Coefficient of determination (R^2); root mean square errors (RMSE) of the predictions.

							With		Without	
Experiment	Signal	Variable	$\hat{\rho}$	σ_b	σ_w	σ_m	R^2	RMSE / dB	R^2	RMSE / dB
van der Heijden and Trahiotis (1999)	π	ITD	0.91	0.20	0.50	0.40	0.94	0.85	0.78	1.45
	0		0.86	0.17	0.65	0.40	0.87	0.86	0.57	1.38
Marquardt and McAlpine (2009)	0	BW	0.89	0.24	0.65	0.40	0.96	0.37	−0.62	1.97
Kolarik and Culling (2010)	0	BW	0.91	0.20	0.50	0.40	0.97	0.67	0.42	3.07

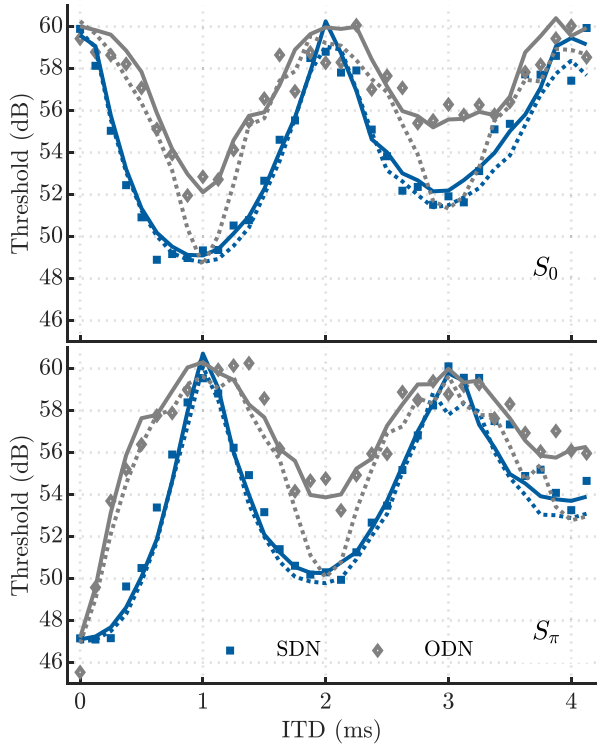

FIG. 3. (Color online) Experimental data from van der Heijden and Trahiotis (1999) (symbols). The continuous lines show the predictions of the presented model including the across-channel incoherence interference. The dashed lines show predictions for ODN excluding interference (single-channel version), equivalent to Encke and Dietz (2022). Upper panel: Detection thresholds with S_0 target; lower panel: S_π target.

described in Sec. II D the model determined the threshold in the form of the signal level producing this d'. The continuous lines in Fig. 3 show the simulations of the presented model, including the across-channel incoherence interference. From visual inspection, the simulations captured all effects from the experimental thresholds and the critical

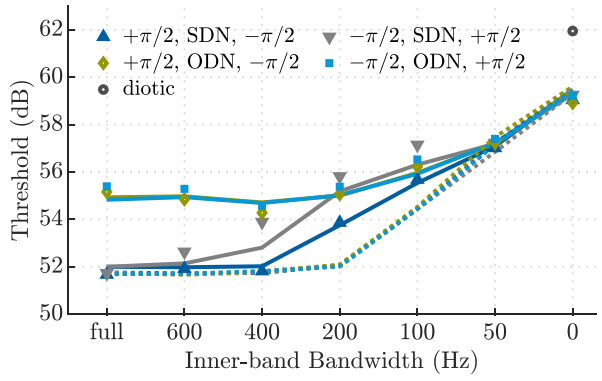

FIG. 4. (Color online) Experimental data from Marquardt and McAlpine (2009) (symbols) and model predictions (lines). Detection thresholds are given as function of the inner-band bandwidth. The inner band contains delayed noise (triangles) or opposingly delayed noises (diamonds and bullets) with a fixed ITD = 1 ms while the flanking bands have a constant IPD of $+\pi/2$ (upward triangle and diamond) and $-\pi/2$, or vice versa (downward triangle and bullet). Continuous and dashed lines again show predictions with and without across-frequency incoherence interference, respectively.

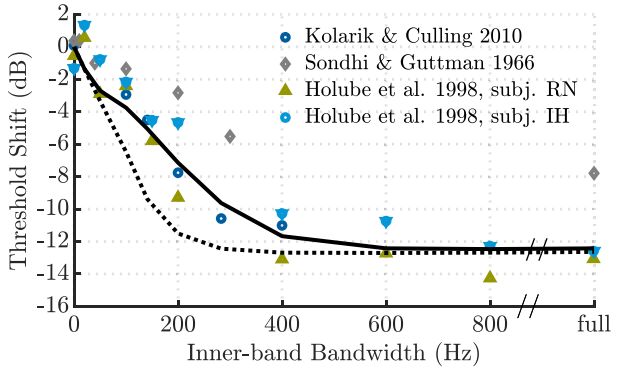

FIG. 5. (Color online) Symbols denote data from binaural detection experiments with the configuration $N_{\pi 0 \pi} S_\pi$ as a function of the inner-band (N_0) bandwidth; continuous and dotted line: model prediction with and without across-incoherence incoherence interference, respectively.

threshold differences between SDN and ODN at all ITDs under both conditions. Specifically, the critical threshold differences of 3.5 dB at an ITD = 1 ms in the S_0 condition and 4 dB at an ITD = 2 ms in the S_π condition are precisely accounted for. This good correspondence is also reflected in the around 90% explained variance under both conditions and RMS errors of less than 1 dB. The dashed lines show simulations excluding the across-channel incoherence interference [single-channel model, cf. Encke and Dietz (2022)] but all other model parameters unchanged. This shows that a large amount of the threshold differences is already explained by differences in the on-frequency coherence. As mentioned in the Introduction: In much the same way as ODN coherence oscillates as a function of analysis frequency [Fig. 1(D)], it also fluctuates as a function of the masker ITD [Fig. 1(C)]. Particularly at ITD = 0.5 ms, ODN is incoherent in the 500-Hz band, whereas SDN is almost fully coherent. This, and not the across-frequency process, causes the difference in the simulated thresholds at this ITD. The across-frequency process only comes into play at those ITDs where the coherence at 500 Hz (on-frequency) is nearly identical in SDN and ODN (upper panel: ITD = 1 and 3 ms; lower panel: ITD = 2 and 4 ms).

B. Marquardt and McAlpine (2009)

The masker of this experiment contained SDN and ODN centered at the frequency of the S_0 target tone with a constant ITD in the inner band. The inner band was spectrally surrounded by bands that each had a constant IPD of $\pi/2$ and $-\pi/2$, or vice versa. Thresholds are given as a function of the inner-band bandwidth. The resulting phase transitions between inner and flanking bands have been hypothesized to impair the detection if they cause a frequency region of low interaural coherence. The lower and upper frequency limits of the composite stimuli are 50 and 950 Hz, respectively. The two-interval-two-alternative-forced choice task with a 3-down 1-up procedure that was used estimated the thresholds to be 79.4% correct. This corresponds to $d' = 1.14$ at the threshold predicted by the model. In Fig. 4, detection thresholds of the S_0 tone are

https://doi.org/10.1121/10.0011673

shown as a function of the bandwidth of the inner band. Again, the model predicted all critical characteristics of the data. These include the 3 dB difference between SDN and ODN at the full inner-band bandwidth (same as ITD = 1 ms in the S_0 condition in van der Heijden and Trahiotis, 1999), the elevated SDN thresholds in the $[-\pi/2, \text{SDN}, +\pi/2]$ compared to the $[+\pi/2, \text{SDN}, -\pi/2]$ condition and the 3 dB BMLD where the inner-band bandwidth is zero. Without the incoherence interference, the predictions cannot be distinguished between the different conditions of the experiment. They deviate more from the mean than the data, resulting in a negative R^2.

C. Experiments on the operating bandwidth in binaural detection

Several studies investigated the operating bandwidth in binaural detection using maskers that contain two flanking bands that differ in their interaural configuration from the inner band (Holube *et al.*, 1998; Kolarik and Culling, 2010; Sondhi and Guttman, 1966). The masking noise is diotic (N_0) in the inner band and antiphasic (N_π) in the flanking bands. Detection thresholds of an S_π target tone were again measured as a function of the inner-band bandwidth. Results are expressed as the difference between thresholds in the flanked condition and the threshold without an inner band, i.e., $N_\pi S_\pi$. In Fig. 5, the circles mark the threshold differences reported by Kolarik and Culling (2010) (centered condition), which represent averages across their three participants. The triangles show individual thresholds of the two participants in the study by Holube *et al.* (1998) (rectangular condition). The gray diamonds show the data from Sondhi and Guttman (1966). Our model predictions were oriented on the 2-down 1-up 2I-2AFC paradigm employed in Kolarik and Culling (2010), equivalent to $d' = 0.78$ at threshold. The black continuous line shows the model predictions with the same parameter settings (see Table I) as used to predict the S_π detection thresholds in van der Heijden and Trahiotis (1999) [see our predictions in Fig. 3(B)]. The dotted black line shows model predictions without the across-channel incoherence interference, so that detection was purely determined by the ERB = 79 Hz Gammatone filter centered at 500 Hz. Despite the large deviations between and within experiments, the model predictions involving the incoherence interference captured the shape of the decreasing thresholds with increasing inner-band bandwidth.

IV. DISCUSSION

In this study, those binaural detection thresholds that previously have underpinned the psychoacoustic necessity of several millisecond long delay-lines were simulated involving across-frequency incoherence interference and monaurally derived peripheral filter bandwidths.

As long as the masker coherence is fairly constant across frequency bands, experiments on binaural detection can be explained purely on the basis of the coherence $|\gamma|$ defined by a 79 Hz wide Gammatone filter at $f_c = 500$ Hz (Encke and Dietz, 2022; Rabiner *et al.*, 1966). This includes fully coherent broadband noise maskers (Hirsh, 1948; van de Par and Kohlrausch, 1999), mixtures of correlated and uncorrelated noise (Bernstein and Trahiotis, 2014; Pollack and Trittipoe, 1959; Robinson and Jeffress, 1963), and experiments where the interaural coherence of the masker is reduced by an ITD (Bernstein and Trahiotis, 2020; Langford and Jeffress, 1964; Rabiner *et al.*, 1966). However, the on-frequency coherence does not account for thresholds obtained with maskers where these properties change substantially across filter bands. Specifically, the single-channel model version as proposed in Encke and Dietz (2022) is neither able to predict all of the threshold differences between SDN and ODN nor experiments like Marquardt and McAlpine (2009) and Kolarik and Culling (2010) that involve IPD transitions in the masker spectrum (see dashed lines in Figs. 3–5, as well as the corresponding R^2 and RMSE given in Table I).

Marquardt and McAlpine (2009) hypothesized across-channel processing in the binaural system to explain the reduced binaural benefit under such conditions. Here, we extended the analytical model by Encke and Dietz (2022) to a multi-channel numerical signal-processing model with incoherence interference. The proposed model differs from approaches assuming wider binaural filters [e.g., Kolarik and Culling (2010) and van der Heijden and Trahiotis (1999)], as, for example, wider filters reduce the interaural coherence of SDN, whereas incoherence interference does not reduce it. For stimuli with spectrally constant coherence and masker-target phase relations, like SDN and all conditions simulated by Encke and Dietz (2022), the incoherence interference has no effect and the model operates on the standard filter bandwidths of its peripheral filterbank. Modeling an interference process, our approach also differs from the symbolic model suggested by Marquardt and McAlpine (2009), which sums interaural cues after binaural interaction. Their implementation is also different from wider filters but still causes a stronger damping of binaural sensitivity with increasing masker ITD, which is not seen in the data.

The proposed concept of a detrimental incoherence interference is comparable to modulation detection interference, as shown and discussed by, e.g., Yost and Sheft (1989) and Oxenham and Dau (2001). Similar to the proposed across-channel incoherence interference, this is modeled by modulation patterns interacting across channels, while energetic spectral masking properties are spectrally limited by peripheral filters (Dau *et al.*, 2013; Piechowiak *et al.*, 2007). Furthermore, a similar process is thought to underlie binaural interference as observed by, e.g., Bernstein and Trahiotis (1995), Best *et al.* (2007), and McFadden and Pasanen (1976).

The dataset of van der Heijden and Trahiotis (1999) contains both SDN and ODN and is therefore the critical challenge for binaural detection models.[2] Both van der Heijden and Trahiotis' and our model simulate the data very

accurately. Therefore, the discussion focuses on consequences and plausibility of the two different concepts.

The bandwidth of the signals immediately prior to binaural interaction dictates the temporal coherence and thus the decline of BMLD with increasing noise ITD in the absence of internal ITD compensation (Dietz *et al.*, 2021; Langford and Jeffress, 1964; Rabiner *et al.*, 1966; van der Heijden and Trahiotis, 1999). To date, two of the arguably most comprehensive datasets of dichotic tone-in-noise detection, van der Heijden and Trahiotis (1999) and Bernstein and Trahiotis (2020), have self-reported mutually exclusive requirements for the filter bandwidth (ERB = 130···180 Hz vs ERB ≤ 100 Hz at 500 Hz).

A variety of studies aim to estimate the bandwidth at the binaural input stage by means of dichotic tone-in-noise detection, but no consistent picture emerges. There is, for example, a difference in estimated bandwidth between band-widening and notched-noise BMLD data, and between stimuli with different flanking bands [e.g., Kolarik and Culling (2010)]. Particularly this stimulus-type dependence of the "apparent bandwidth" challenges the assumption that all stimuli are processed by the same filters. To us, the most reasonable "unifying" explanation is that filter properties arise from the basilar membrane and also the binaural system can make full use of this spectral resolution. The observation that there is less spectral resolution in some cases is then best explained by an across-frequency process for certain stimulus features—but in contrast to wider filters it is not affecting all features. The proposed incoherence interference may be this missing across-frequency process. At least it appears to reduce or even eliminate inconsistencies in estimating the bandwidth from various binaural detection experiments.

Another mechanism that has been proposed in the context of bandwidth estimation is an optimal combination of target detectability across frequency channels. Masking patterns in dichotic band-widening experiments have a knee-point at larger bandwidths than their diotic counterparts (Bourbon and Jeffress, 1965; van de Par and Kohlrausch, 1999). van de Par and Kohlrausch (1999) hypothesized that in narrowband maskers, the similar SNR across frequency channels can be exploited to reduce masking. A model that includes such a mechanism (Breebaart *et al.*, 2001b) accounts for the band-widening masking pattern using standard filter bandwidths [i.e., bandwidths as proposed by Glasberg and Moore (1990)]. It also correctly predicts that the knee-point is only shifted to a higher bandwidth if the masker is fully or almost fully correlated (van der Heijden and Trahiotis, 1998).

Most recent binaural models, such as Bernstein and Trahiotis (2017) and Encke and Dietz (2022) already assume a bandwidth as narrow as the peripheral bandwidth. This is also in line with direct measurements of the bandwidth in ITD-sensitive inferior colliculus neurons in cats by Mc Laughlin *et al.* (2008). For delayed noise, as used by van der Heijden and Trahiotis (1999), they found that damping of the cross correlation function corresponds to the peripheral bandwidth at the respective center frequency.

With the present implementation, the binaural pathway parameters ($\hat{\rho}$, σ_b, σ_w) had to be adjusted slightly between conditions with S_π targets and conditions with S_0 targets (see Table I). This is due to the binaural system's sensitivity depending on the baseline IPD (Hirsh, 1948). An angular compression of the decision variable space $\{\zeta\}$ at large IPDs is a possible model extension. Delay-line models can account for this dependence with a corresponding $p(\tau)$ function, which defines the sensitivity of the model as a function of its internal delay. However, they then incorrectly predict better unmasking with $N_{ITD}S_0$ compared to $N_\pi S_0$ when ITD = $T/2$ (Breebaart *et al.*, 1999). Simulating the data of Marquardt and McAlpine (2009) required slightly different parameter values because their listeners obtained different thresholds compared to van der Heijden and Trahiotis (1999) for identical stimuli. This may be due to the different number of presented intervals. Identical model parameters were used for the S_π conditions of van der Heijden and Trahiotis (1999) and Kolarik and Culling (2010).

V. CONCLUSION

Interaural incoherence interference enables the presented binaural model to simulate detection thresholds both for maskers with a spectrally constant and with a spectrally modulated coherence. Employing auditory filters with monaurally estimated bandwidth (Glasberg and Moore, 1990), it predicts the reduced unmasking in opposingly delayed noises (van der Heijden and Trahiotis, 1999) compared to regular delayed noise. The concept can help to resolve the inconsistency that binaural models require filter bandwidths as estimated monaurally for most data sets (Bernstein and Trahiotis, 2017, 2020), but at least 1.6 times wider filters for broadband opposingly delayed noises (van der Heijden and Trahiotis, 1999) and other spectrally complex maskers (Verhey and van de Par, 2020). The main consequence of using a standard filter bandwidth is that the decline of the binaural benefit with masker ITD can be simulated without internal ITD compensation, as first suggested by Langford and Jeffress (1964).

ACKNOWLEDGMENTS

We thank Steven van de Par for fruitful discussions and comments on the manuscript as well as Matthew J. Goupell and two anonymous reviewers for their very helpful comments. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 352015383–SFB 1330 B4.

APPENDIX: DERIVATION OF CROSS-POWER SPECTRAL DENSITY IN OPPOSINGLY DELAYED NOISE

In ODN, two two-channel signals $u(t) = [u(t)\, u(t + ITD)]$ and $z(t) = [z(t)\, z(t - ITD)]$ with opposite ITDs, ITD and −ITD, are summed. The cross-power spectral density (CPSD) functions are

https://doi.org/10.1121/10.0011673

$$S_{UU}(\omega) = 0.5e^{iITD\omega},$$
$$S_{ZZ}(\omega) = 0.5e^{-iITD\omega}. \quad \text{(A1)}$$

The power spectral density is 0.5 1/Hz each, so that the ODN has the same energy as the SDN. Summation of the time signals is equivalent to a summation of their CPSD functions, which leads to

$$S_{UZ} = S_{UU}(\omega) + S_{ZZ}(\omega) = \cos(\omega ITD). \quad \text{(A2)}$$

This resulting cosine pattern is determined by the sum of the CPSDs' phases adding up or canceling each other at different frequencies. This normalized CPSD $C(\omega)$ represents the coherent energy of the signals as a function of frequency (Gardner, 1992),

$$C(\omega) = \frac{|S_{UZ}(\omega)|}{\sqrt{S_{UU}(\omega)S_{ZZ}(\omega)}} = |\cos(\omega ITD)|. \quad \text{(A3)}$$

If $|\gamma(\tau)|$ is based on an ensemble average, then $C(\omega) = \mathcal{F}\{|\gamma(\tau)|\}$, with $\mathcal{F}\{\bullet\}$ the Fourier transform. As a continuous function of ω it gives a coherence for any frequency ω representing an infinitesimally small bandwidth, illustrated as continuous lines in Fig. 1(D). The coherence for peripherally filtered, i.e., finite-bandwidth signals is an average of the frequencies' normalized CPSDs $C(\omega)$. The coherence decreases with increasing ITD and increasing bandwidth, as illustrated by the bars in Fig. 1(D). Two superimposed noises with ITD = ±2 ms are in phase at 500 Hz. At 625 Hz, however, they have IPDs of $\pi/2$ and $-\pi/2$, respectively. The coherence between left and right signals at 625 Hz is therefore zero.

[1]In contrast to measures of IPD fluctuations, such as the variance of the instantaneous IPD (Dietz *et al.*, 2021), both the interaural coherence $|\gamma|$ and the interaural cross correlation function inherently weight the instantaneous IPD with the amplitudes. This is instrumental to quantitatively account for the masking of different noises with different statistics, such as low-noise noise or multiplicative noise. We therefore expect the present model to also account for detection thresholds obtained with such maskers.

[2]The most comprehensive simulation of dichotic tone in noise detection thresholds using a cross correlation-based model is by Bernstein and Trahiotis (2017). It is not expected to simulate the ODN detection thresholds of van der Heijden and Trahiotis (1999) with a good accuracy, because an ERB of at least 130 Hz is necessary. Other ODN stimuli, used experimentally by Bernstein and Trahiotis (2015), were included in the model test battery by Bernstein and Trahiotis (2017). Those ODN stimuli, however, differed in several ways from the former. First, the target frequency is 250 Hz, compared to 500 Hz in van der Heijden and Trahiotis (1999) and in all other studies here simulated. Second, instead of fixing the target tone to S_0 or S_π, the target is delayed by the same amount as one of the two noises, i.e., $(N_0)_{\pm ITD}(S_\pi)_{ITD}$. Such an approach is useful for SDN, as it ensures a constant π difference between the IPDs of the noise and of the tone. For ODN, however, the IPD of the second noise relative to the tone is offset from π by $2 \times$ ITD. This type of stimulus therefore causes an even more complex ITD-dependence of threshold, which offers no advantage over the ODN from van der Heijden and Trahiotis (1999) for filter estimation. With both definitions, corresponding SDN and ODN stimuli can be generated only if the ITD is an integer or a half-integer multiple of the target period (i.e., ITD = $n/2f$, $n \in \mathbb{N}$). In Bernstein and Trahiotis (2015), Fig. 1(a), these are the two data points at ITD = 2 and 4 ms. SDN and ODN thresholds are, however, very similar at those points. Third, the masker bandwidth is 50 Hz. For such a masker bandwidth smaller than the peripheral filter width, neither van der Heijden and Trahiotis (1999) nor our model would predict a considerable threshold difference between SDN and ODN at an ITD of 2 and 4 ms, since there are no off-frequency regions of considerably lower coherence.

Bernstein, L. R., and Trahiotis, C. (**1995**). "Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms," J. Acoust. Soc. Am. **98**(1), 155–163.

Bernstein, L. R., and Trahiotis, C. (**2014**). "Accounting for binaural detection as a function of masker interaural correlation: Effects of center frequency and bandwidth," J. Acoust. Soc. Am. **136**(6), 3211–3220.

Bernstein, L. R., and Trahiotis, C. (**2015**). "Converging measures of binaural detection yield estimates of precision of coding of interaural temporal disparities," J. Acoust. Soc. Am. **138**(5), EL474–EL479.

Bernstein, L. R., and Trahiotis, C. (**2017**). "An interaural-correlation-based approach that accounts for a wide variety of binaural detection data," J. Acoust. Soc. Am. **141**(2), 1150–1160.

Bernstein, L. R., and Trahiotis, C. (**2018**). "Effects of interaural delay, center frequency, and no more than 'slight' hearing loss on precision of binaural processing: Empirical data and quantitative modeling," J. Acoust. Soc. Am. **144**(1), 292–307.

Bernstein, L. R., and Trahiotis, C. (**2020**). "Binaural detection as a joint function of masker bandwidth, masker interaural correlation, and interaural time delay: Empirical data and modeling," J. Acoust. Soc. Am. **148**(6), 3481–3488.

Best, V., Gallun, F. J., Carlile, S., and Shinn-Cunningham, B. G. (**2007**). "Binaural interference and auditory grouping," J. Acoust. Soc. Am. **121**(2), 1070–1076.

Biberger, T., and Ewert, S. D. (**2016**). "Envelope and intensity based prediction of psychoacoustic masking and speech intelligibility," J. Acoust. Soc. Am. **140**(2), 1023–1038.

Biberger, T., and Ewert, S. D. (**2017**). "The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking," J. Acoust. Soc. Am. **142**(2), 1098–1111.

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

Breebaart, J., 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**(2), 979–992.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001a**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**(2), 1074–1088.

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**(2), 1105–1117.

Dau, T., Piechowiak, T., and Ewert, S. D. (**2013**). "Modeling within- and across-channel processes in comodulation masking release," J. Acoust. Soc. Am. **133**(1), 350–364.

Dietz, M., Encke, J., Bracklo, K. I., and Ewert, S. D. (**2021**). "Tone detection thresholds in interaurally delayed noise of different bandwidths," Acta Acust. **5**, 60.

Encke, J., and Dietz, M. (**2022**). "A hemispheric two-channel code accounts for binaural unmasking in humans," arXiv:2111.04637 [cs, eess].

Gardner, W. A. (**1992**). "A unifying view of coherence in signal processing," Sign. Process. **29**(2), 113–140.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**(1-2), 103–138.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos Hills, CA).

Hacker, M. J., and Ratcliff, R. (**1979**). "A revisted table of d′ for M-alternative forced choice," Percept. Psychophys. **26**(2), 168–170.

Hirsh, I. J. (**1948**). "The influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**(4), 536–544.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a Gammatone filterbank," Acta Acust. Acust. **88**(3), 433–442, avaialble at https://www.ingentaconnect.com/content/dav/aaua/2002/00000088/00000003/art00015.

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**(4), 2412–2425.

Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**(1), 35–39.

Joris, P. X., de Sande, B. V., Louage, D. H., and van der Heijden, M. (**2006**). "Binaural and cochlear disparities," Proc. Natl. Acad. Sci. U.S.A. **103**(34), 12917–12922.
Just, D., and Bamler, R. (**1994**). "Phase statistics of interferograms with applications to synthetic aperture radar," Appl. Opt. **33**(20), 4361–4368.
Kolarik, A. J., and Culling, J. F. (**2010**). "Measurement of the binaural auditory filter using a detection task," J. Acoust. Soc. Am. **127**(5), 3009–3017.
Langford, T. L., and Jeffress, L. A. (**1964**). "Effect of noise crosscorrelation on binaural signal detection," J. Acoust. Soc. Am. **36**(8), 1455–1458.
Leibold, C., and Grothe, B. (**2015**). "Sound localization with microsecond precision in mammals: What is it we do not understand?," e-Neuroforum **6**(1), 3–10.
Lüddemann, H., Riedel, H., and Kollmeier, B. (**2007**). "Logarithmic scaling of interaural cross correlation: A model based on evidence from psychophysics and EEG," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, Berlin), pp. 379–388.
Marquardt, T., and McAlpine, D. (**2009**). "Masking with interaurally 'double-delayed' stimuli: The range of internal delays in the human brain," J. Acoust. Soc. Am. **126**(6), EL177–EL182.
McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). "A neural code for low-frequency sound localization in mammals," Nat. Neurosci. **4**(4), 396–401.
McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization at high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**(3), 634–639.
Mc Laughlin, M., Chabwine, J. N., van der Heijden, M., and Joris, P. X. (**2008**). "Comparison of bandwidths in the inferior colliculus and the auditory nerve. II: Measurement using a temporally manipulated stimulus," J. Neurophysiol. **100**(4), 2312–2327.
McNemar, Q. (**1969**). *Psychological Statistics*, 4th ed. (Wiley, New York).
Oxenham, A. J., and Dau, T. (**2001**). "Modulation detection interference: Effects of concurrent and sequential streaming," J. Acoust. Soc. Am. **110**(1), 402–408.
Piechowiak, T., Ewert, S. D., and Dau, T. (**2007**). "Modeling comodulation masking release using an equalization-cancellation mechanism," J. Acoust. Soc. Am. **121**(4), 2111–2126.
Pollack, I., and Trittipoe, W. J. (**1959**). "Binaural listening and interaural noise cross correlation," J. Acoust. Soc. Am. **31**(9), 1250–1252.
Rabiner, L. R., Laurence, C. L., and Durlach, N. I. (**1966**). "Further results on binaural unmasking and the EC model," J. Acoust. Soc. Am. **40**(1), 62–70.
Robinson, D. E., and Jeffress, L. A. (**1963**). "Effect of varying the interaural noise correlation on the detectability of tonal signals," J. Acoust. Soc. Am. **35**(12), 1947–1952.
Sondhi, M. M., and Guttman, N. (**1966**). "Width of the spectrum effective in the binaural release of masking," J. Acoust. Soc. Am. **40**(3), 600–606.
Stern, R. M., 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**(1), 127–140.
Stern, R. M., Colburn, H. S., Bernstein, L. R., and Trahiotis, C. (**2019**). "The fMRI Data of Thompson *et al*. (2006) do not constrain how the human midbrain represents interaural time delay," JARO **20**(4), 305–311.
Thompson, S. K., von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T. D., and McAlpine, D. (**2006**). "Representation of interaural time delay in the human auditory midbrain," Nat. Neurosci. **9**(9), 1096–1098.
van de Par, S., and Kohlrausch, A. (**1999**). "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," J. Acoust. Soc. Am. **106**(4), 1940–1947.
van der Heijden, M., and Trahiotis, C. (**1998**). "Binaural detection as a function of interaural correlation and bandwidth of masking noise: Implications for estimates of spectral resolution," J. Acoust. Soc. Am. **103**(3), 1609–1614.
van der Heijden, M., and Trahiotis, C. (**1999**). "Masking with interaurally delayed stimuli: The use of 'internal' delays in binaural detection," J. Acoust. Soc. Am. **105**(1), 388–399.
Verhey, J. L., and van de Par, S. (**2020**). "Binaural frequency selectivity in humans," Eur. J. Neurosci. **51**(5), 1179–1190.
Yost, W. A., and Sheft, S. (**1989**). "Across-critical-band processing of amplitude-modulated tones," J. Acoust. Soc. Am. **85**(2), 848–857.