

Masking with interaurally “double-delayed” stimuli: The range of internal delays in the human brain

Torsten Marquardt and David McAlpine

*UCL Ear Institute, 332 Gray's Inn Road, London WC1X 8EE, United Kingdom
t.marquardt@ucl.ac.uk, d.mcalpine@ucl.ac.uk*

Abstract: Is binaural processing in humans different to that of other mammals? While psychophysical data suggest that the range of internal delays necessary for processing interaural time differences is at least ± 3 ms, physiological data from small mammals indicate a more limited range. This study demonstrates that binaural detection is impeded by reduced interaural coherence in auditory channels remote from the signal frequency, in accordance with the wider critical bandwidths reported for binaural processing. This explains previous psychophysical data without requiring long internal delays. The current psychophysical data support the view that human binaural processing is similar to that of other mammals.

© 2009 Acoustical Society of America

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

Date Received: August 12, 2009 **Date Accepted:** September 29, 2009

1. Introduction

To account for human sensitivity to interaural time differences (ITDs), Jeffress¹ envisaged an array of binaural coincidence detectors, each maximally activated only when the “external” interaural delay was compensated by an equal and opposite “internal” interaural delay (arising from a difference in the axonal conduction delay from each ear). Although physiological investigations confirm the existence of neurons that respond as such binaural coincidence detectors, the range of the compensatory internal delays is open to question. In particular, electrophysiological recordings in a range of mammalian species suggest the existence of a “ π -limit,” with no binaural coincidence detectors tuned to ITDs longer than half a period (relative to the center frequency of the auditory channel under consideration).² This has been difficult to reconcile with psychophysical studies purporting to show evidence of internal delays of at least several times this magnitude. Consequently, the extent to which the π -limit might be extrapolated to the human brain is unclear. van der Heijden and Trahiotis³ addressed the question of the existence of long internal delays in humans, by employing a binaural detection paradigm in which they measured the detection threshold for a diotic tone that was masked by a “double-delayed” noise (DDN). DDN is the sum of two independent binaural noises with ITD of equal magnitude but opposite sign. In contrast, a “single-delayed” noise (SDN) is generated by only one such noise source with either positive or negative ITD—a masker that is routinely employed in measurements of binaural-masking-level-difference (BMLD). Figures 1(a) and 1(b) show cross-correlograms of SDN and DDN stimuli. Since DDN is the sum of two SDN sources, the normalized cross-correlogram for the DDN is the average of those of its two SDN components. This is illustrated in Fig. 1(c) for the frequency channel containing the target tone (500 Hz). The data of van der Heijden and Trahiotis³ [a sub-set of which is reproduced here in Fig. 1(d)] showed that thresholds for SDN maskers were lower than for DDN maskers. Their model simulations suggested that internal delays as large as 3 ms must exist, and are employed under the SDN masker conditions to profit from the deeper modulation in this region of the cross-correlation function [Fig. 1(c)].

The lack of physiological evidence for such large internal-delay magnitudes calls for alternative explanations for van der Heijden and Trahiotis' data.³ For the 500-Hz target em-

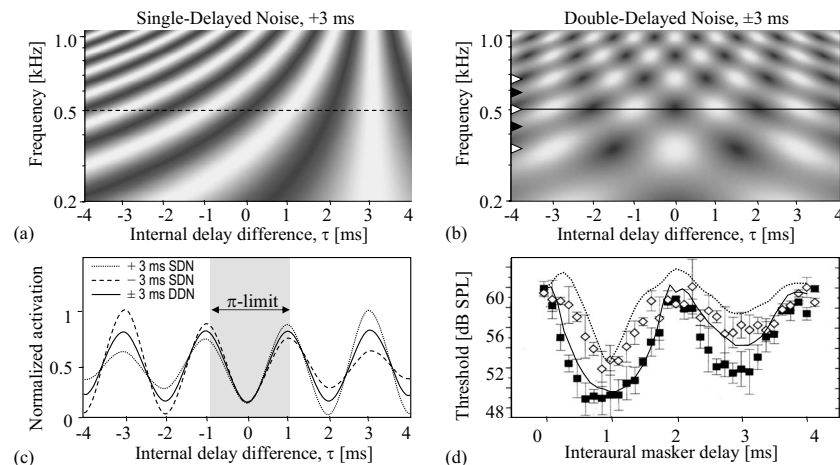


Fig. 1. Rescaled cross-correlograms for SDN (a) and DDN (b) maskers with ITD of 3 ms magnitude. Their cross-correlation values are computed using Eq. (1) (here with $\tau = \pi\{-4 \text{ ms}, \dots, 4 \text{ ms}\}$), and are rescaled to express normalized neural activation: The activity of a neural coincidence detector is one (white), when the signals arriving from the left and right ears are fully correlated, and zero (black), when the arriving signals have a correlation of -1 . The frequency channels of the target tone (500 Hz) are indicated by horizontal lines. Solid triangles and open triangles in (b) mark examples of frequency channels of zero and high IACs, respectively. (c) Normalized activation within the target-tone channel to SDN and DDN. The physiologically observed range of internal delays is shown gray (π -limit). (d) Detection thresholds for a diotic 500-Hz tone with SDN (solid squares) and DDN (open diamonds) maskers obtained by van der Heijden and Trahiotis (Ref. 3). Thresholds predicted by our across-frequency convergence model are plotted as solid (SDN) and dotted (DDN) lines.

ployed, the π -limit is ± 1 ms and the regions of deeper modulation that presumably lowered thresholds for the SDN maskers fall outside this range of internal delays [gray area in Fig. 1(c)]. However, while the π -limited cross-correlation functions to DDN and SDN are almost identical *within* the target-frequency channel, large differences are apparent *across* frequency bands [Figs. 1(a) and 1(b)]. For SDN, the cross-correlogram is strongly modulated over the whole frequency range, indicating high interaural coherence (IAC) (the maximum cross-correlation within a frequency channel) across frequency. For DDN, however, the modulation is only pronounced in certain frequency bands. Frequency bands alternate between high IAC (e.g., at 333, 500, and 667 Hz) and zero IAC (e.g., at 417 and 583 Hz). Here, we demonstrate that these differences in the modulation of off-frequency channels provide an alternative explanation for the observed threshold differences between the two masking conditions; an explanation that does not require long, physiologically unrealistic, internal delays.

2. Psychophysical experiment

To investigate the influence of IAC in off-frequency channels on binaural detection, we measured masked tone thresholds in either DDN or SDN, each flanked by noise bands of different interaural configurations (similar to the paradigm applied by Sondhi and Guttman⁴). The addition of the flanking bands modifies the IAC in channels tuned to frequencies around the transitions between inner and flanking bands. We predict that detection of the target tone will be impaired by these flanking bands in a manner that depends on the IAC at these transitions and the proximity of these transitions to the target-frequency.

The interaural timing configurations of the four masker types employed in our experiment are illustrated in Fig. 2(a). The three frequency bands (one inner band and two flanking bands) had equal power density (45.5 dB sound pressure spectrum level), and spanned the total range 50–950 Hz. The inner band, consisting of either SDN or DDN with an ITD magnitude of 1 ms (0.5 cycles at 500 Hz), was positioned symmetrically around the frequency of the target tone (500 Hz). The upper and lower flanking bands were unilaterally phase delayed such that

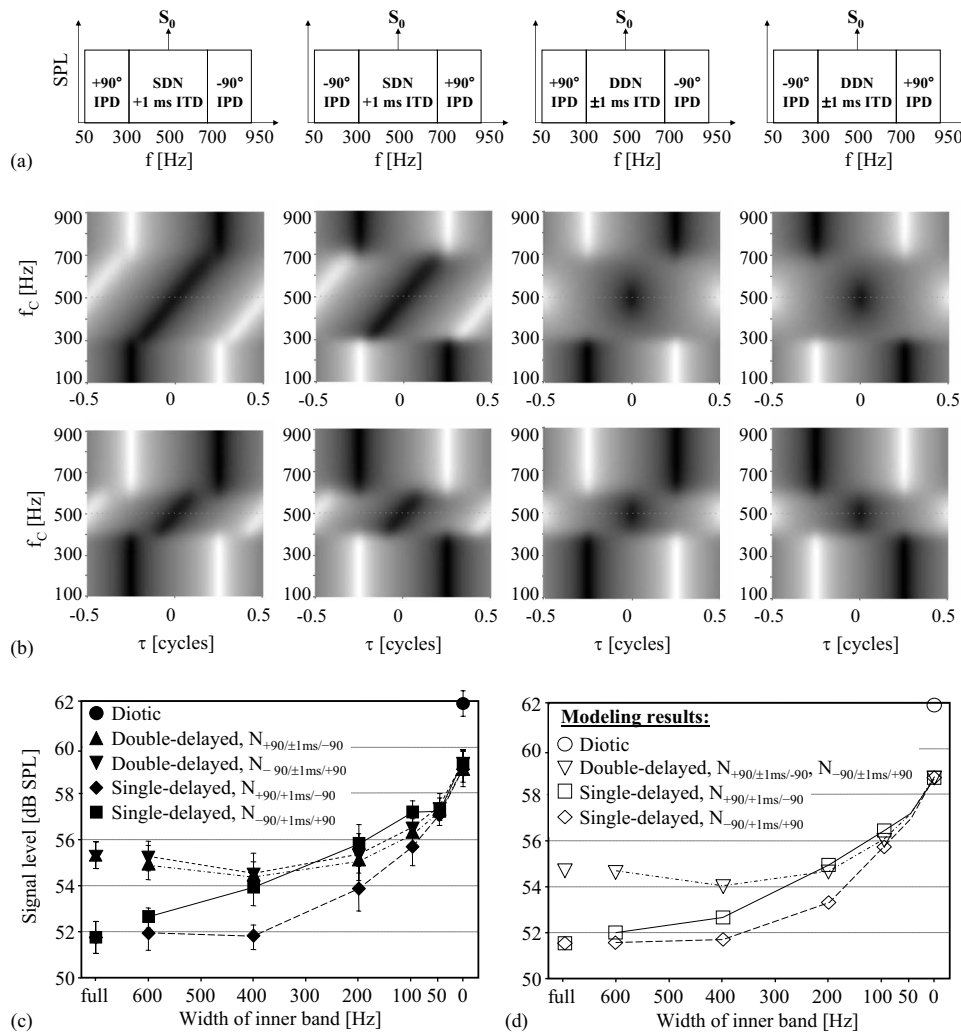


Fig. 2. (a) Long-term average spectra of the four masker configurations: $N_{+90^\circ/+1\text{ ms}/-90^\circ}$, $N_{-90^\circ/+1\text{ ms}/+90^\circ}$, $N_{+90^\circ/\pm 1\text{ ms}/-90^\circ}$, and $N_{-90^\circ/\pm 1\text{ ms}/+90^\circ}$ (left to right column). Labels inside the sub-bands denote their interaural timing configurations. Below each spectrum, in panel (b), are rescaled cross-correlograms of each of these maskers with 400-Hz (upper row) and 200-Hz (lower row) inner bandwidths. In accordance with the π -limit, the range of internal-delay differences (τ) is limited to ± 0.5 cycles and expressed in phase. The rescaled cross-correlation values are computed using Eq. (1) (for rescaling, see caption of Fig. 1). (c) Average detection threshold as a function of the width of the inner band for the four masker types (see legend). Error bars indicate across-listeners standard errors. (d) Thresholds predicted by our across-frequency convergence model.

each band had an interaural *phase* delay (IPD) of 90° , but of opposite sign to the other. This produced a total of four masker configurations: $N_{+90^\circ/+1\text{ ms}/-90^\circ}$, $N_{-90^\circ/+1\text{ ms}/+90^\circ}$, $N_{+90^\circ/\pm 1\text{ ms}/-90^\circ}$, and $N_{-90^\circ/\pm 1\text{ ms}/+90^\circ}$. The masker noises were generated in the frequency domain from spectral components with 1-Hz-spacing and Rayleigh-distributed amplitude. The phases of the spectral components of the left ear stimulus were uniformly distributed, and the right ear phases were assigned relative to the left ear phases to produce the spectrum of ITDs illustrated in Fig. 2(a). For maskers containing DDN, two independent SDNs of half power were added: one with +1 ms ITD and one with -1 ms ITD. After inverse Fourier transformation, the 1-s long masker was truncated to 300 ms (including 10-ms cosine ramps).

As in the study by Sondhi and Guttman,⁴ pure-tone thresholds were measured as a function of the width of the inner masker band (0, 50, 100, 200, 400, 600, and 900 Hz). A two-interval, two-alternative forced-choice task was employed with an inter-stimulus interval of 300 ms. For each stimulus interval, a new masker noise was generated. A diotic target tone (500 Hz, 280-ms long including 10-ms cosine ramps) was computed in the time domain and added (temporally centered) to the masker in one of the two intervals (chosen randomly on each trial with equal probability). This trial sequence was then converted to electric analog signals by a 16-bit soundcard (22 050 Hz sampling rate), and presented acoustically via earphones (Beyerdynamic DT 48A with supra-aural cushions). The subject was directed to identify the interval that contained the target tone. Correct-answer feedback was given after each response. The target-tone level, initially set well above the expected thresholds, was varied adaptively using a “3down-1up” procedure to estimate the 79.4% correct threshold. The initial step size of the adaptive track was 4 dB, and was reduced to 2 dB and then to 1 dB following two reversals at each of the former step sizes. A track was terminated after 12 reversals using 1-dB steps, and the threshold for that track calculated as the average target-tone level at the last ten reversals.

Four subjects (aged 20–33; trained until threshold stabilized) were tested under all masker conditions plus one diotic condition. Each subject was tested four times, and each time all conditions were tested in a different random order. The median formed the subject’s threshold. Figure 2(c) shows the average thresholds across all subjects. For all four masker types, the thresholds increase as the flanking bands close in on the target-frequency. Under the $N_{-90^\circ/+1 \text{ ms}/+90^\circ}$ masker conditions, the threshold clearly increased when the flanking bands were still 200 Hz from the target (i.e., with inner bandwidth of 400 Hz). In contrast, for the DDN and the $N_{+90^\circ/+1 \text{ ms}/-90^\circ}$ conditions, thresholds did not increase until the flanking bands were approximately 100 Hz from the target. Although SDN maskers with large inner bandwidths (>400 Hz) produced lower thresholds than the DDN maskers, this advantage was not so clear for maskers with lower inner bandwidths, especially with the SDN masker, $N_{-90^\circ/+1 \text{ ms}/+90^\circ}$.

The data can be understood by examining the cross-correlograms in Fig. 2(b). Within frequency channels, where spectral components from neighboring bands with different interaural timings merge, the IAC and consequently the modulation in activation (along the τ -dimension) are reduced. In the case of the $N_{+90^\circ/+1 \text{ ms}/-90^\circ}$ masker, the IPD difference between bands at the transition remained small such that, for an inner bandwidth of 400 Hz, IAC in the transition channels was hardly reduced and detection threshold remained low. In contrast, for the $N_{-90^\circ/+1 \text{ ms}/+90^\circ}$ configuration with 400-Hz inner bandwidth, the difference in IPD was almost 180° , leading to near zero IAC at both transitions, and presumably causing the substantial threshold increase. Thus, reduced IAC appears detrimental to binaural detection, even outside the target’s frequency channel (which has a critical bandwidth of just 78 Hz).⁵

Of particular interest are the DDN maskers $N_{+90^\circ/\pm 1 \text{ ms}/-90^\circ}$ and $N_{-90^\circ/\pm 1 \text{ ms}/+90^\circ}$. Due to the symmetry of DDN around zero IPD, these two masker types produced almost identical results. DDN has inherently low IAC at certain frequencies outside the target channel [Fig. 1(b)]. As long as the flanking noise did not enter the frequency channels of high IAC in the proximity of the target, its addition had little effect on the already elevated thresholds produced by the DDN masker alone. Thus, we suggest that the elevated thresholds observed under DDN masker conditions, in both our data and those of van der Heijden and Trahiotis,³ are the result of the off-frequency bands of low IAC inherent in the DDN masker.

3. The frequency convergence model

To demonstrate how IAC in the off-frequency channels can quantitatively account for both our data and those of van der Heijden and Trahiotis,³ we introduce a simple phenomenological model based on binaural cross-correlation *followed* by across-frequency convergence. The activation of the binaural coincidence detector neurons is modeled by a rescaled cross-correlogram using 81 frequency channels ($f_c = \{100, 110, \dots, 900 \text{ Hz}\}$) with 201 internal-delay channels per frequency channel. The range of internal delays was restricted to the π -limit as suggested by physiological data ($\tau = \{-0.5/f_c, \dots, 0.5/f_c\}$).

$$CC(f_c, \tau) = \sum_{f=1 \text{ Hz}}^{2 \text{ kHz}} H_{st}^2(f) H_{AF}^2(f_c, f) \left\{ \eta \frac{1}{2} \cos[\text{IPD}_{st}(f) - 2\pi\tau f] + \frac{1}{2} \right\}. \quad (1)$$

The computation of the cross-correlation function within a frequency channel f_c in Eq. (1) is based on the sum of sinusoidal cross-correlation functions of individual spectral components, which are power-weighted by the square of the cochlear filter shape (H_{AF}). (The cross-correlation function is the inverse Fourier transform of the cross-power spectrum.) Although white noise spectra, and consequently cross-correlograms, vary from one stimulus sample to the next, the long-term average of their magnitude spectra is “flat.” Therefore, we defined the masker magnitude spectrum in our model as $H_{st}(50\text{--}950 \text{ Hz})=1$, so that the activity of a neural coincidence detector, $CC(f_c, \tau)$, constitutes a long-term average, and is normalized such that its activity is one, when the signals arriving from the left and right ears are fully correlated, and zero, when the arriving signals have a correlation of -1 (for $\eta=1$). The parameter η modeled the effect of internal noise in the (monaural) auditory periphery, which reduces IAC and, consequently, the modulation of the cross-correlation function (for $\eta < 1$). H_{AF} had the shape of a fourth-order gammatone filter with unit power transmission, and equivalent rectangular bandwidth (ERB) as estimated monaurally.⁵ $\text{IPD}_{st}(f)$ describes the frequency dependence of the stimulus IPD. Input spectra were specified with 1-Hz frequency resolution.

Rescaled cross-correlograms were calculated separately for the signal alone (CC_{signal}) and the noise alone (CC_{noise}). Element-wise division of CC_{signal} by CC_{noise} produces the local signal-to-noise ratio (SNR) at each coincidence-detecting neuron. In each frequency channel f_c , the neuron with the maximum SNR was then identified [$\tau_{\text{max}}(f_c)$]. The activations of these neurons are then summed across frequency channels in a weighted manner, before the global SNR was calculated [Eq. (2)]. This ensured that also activations beyond the target-frequency channel influenced signal detection (“across-frequency convergence”).

$$\text{SNR}_{\text{global}} = \frac{\sum_{f_c} \omega(f_c) CC_{\text{signal}}[f_c, \tau_{\text{max}}(f_c)]}{\sum_{f_c} \omega(f_c) CC_{\text{noise}}[f_c, \tau_{\text{max}}(f_c)]}. \quad (2)$$

The $\text{SNR}_{\text{global}}$ [Eq. (2)], calculated for a diotic stimulus condition, normalized by the $\text{SNR}_{\text{global}}$ calculated for the diotic stimulus, gives the BMLD estimate of our model. (Because the model predicts only BMLD, note that simulated masked thresholds, as plotted in the figures, are the BMLD relative to the experimental masked threshold for the diotic condition.)

The impact of IAC in off-frequency channels on the BMLD can be easily understood when considering that signal activation in these channels is almost negligible. Therefore, $\tau_{\text{max}}(f_c)$ is, here, essentially the internal delay producing the lowest masker activation. Because the cross-correlation function in frequency channels with lower IAC is less modulated, the activation minimum at $\tau_{\text{max}}(f_c)$ and, consequently, the noise contribution to the global SNR from such channels are larger compared to an off-frequency frequency channel with a higher IAC.

The only free parameters of our model are the spectral weighting function $\omega(f_c)$ and the internal noise parameter η , which were manually adjusted to produce the closest agreement between the model and our data. Figure 2(d) shows the threshold functions predicted by our model. They were obtained using $\eta=0.93$, and a two-box weighting comprised of a low-weight convergence over a range larger than ± 3 ERB [$\omega(220 \text{ Hz} < f_c < 780 \text{ Hz})=0.02$], and a higher-weight convergence within approximately ± 1 ERB of the target channel [$\omega(410 \text{ Hz} < f_c < 590 \text{ Hz})=0.3$, $\omega(500 \text{ Hz})=1$]. We felt that this simple two-box description illustrates best the dual character of the across-frequency convergence. It reproduces the features of our psychophysical results surprisingly well. Attempts to use more sophisticated weighting functions did not yield significantly better fits.

Our model (with unaltered parameters) also provides a reasonable account of the data of van der Heijden and Trahiotis³ [Fig. 1(d)], in that it captures the general shape of the func-

tions, e.g., the wider troughs with SDN maskers and, most importantly, the higher thresholds with DDN maskers. It falls short, however, in two aspects: First, our model predicts slightly stronger damping of the oscillations than the psychophysical data of van der Heijden and Trahiotis,³ although we ameliorated this side-effect of across-frequency convergence largely by introducing it after, rather than before, the stage of binaural cross-correlation. Note that the latter would be equivalent to a simple widening of the 500-Hz target channel. Such a π -limited model cannot reproduce the BMLD difference between SDN and DDN at 1 and 3 ms ITDs, because the best internal delay τ_{\max} is then zero, where SDN and DDN produce identical cross-correlation values [see Fig. 1(c)]. Second, our binaural model does not reduce the impact of off-frequency channels near diotic conditions (around 0, 2, and 4 ms masker ITDs), when the critical bandwidth is expected to decrease to that found for monaural listening.⁵ Since our purely binaural model does not incorporate a monaural detector path, those simulated thresholds are somewhat elevated.

The model of van der Heijden and Trahiotis,³ on the other hand, cannot account for our data since it only considers the target-frequency channel, having a 3-dB-bandwidth of just 90 Hz. The long internal delays proposed in their model will always provide an advantage under SDN masker conditions (as long as the flanking noise is outside the target-frequency channel). Our experiment, however, showed similar thresholds for SDN and DDN masker conditions ($N_{-90^\circ/+1 \text{ ms}/+90^\circ}$ and $N_{-90^\circ/\pm 1 \text{ ms}/+90^\circ}$), even when the inner bands were as wide as 400 Hz.

4. Conclusion

Our findings suggest that binaural detection is influenced by IAC outside the target-frequency channel, which is in line with the phenomenon of wider binaural, compared to monaural, critical bandwidths.⁴ Convergence of binaural neurons across different frequency bands has been observed physiologically,⁶ although the questions remain as to how exactly this is implemented, and what function it might serve. Additionally, our findings provide an explanation for the observed difference in thresholds for DDN and SDN maskers that does not require long internal delays. Together with a recent functional imaging study,⁷ our data further support the view that the human binaural system is similar to that of other mammals.

Acknowledgments

Data were collected during a Bogue Visiting Fellowship to TM at Dalhousie University (NS, Canada), hosted by Dennis Phillips and Susan Boehnke. We thank her and Isabel Dean for critical reading of the manuscript, and John Agapiou for many helpful suggestions.

References and links

- ¹L. A. Jeffress, "A place theory of sound localization," *J. Comp. Physiol. Psychol.* **41**, 35–49 (1948).
- ²P. X. Joris and T. C. T. Yin, "A matter of time: Interaural delays in binaural processing," *Trends Neurosci.* **30**, 70–78 (2007).
- ³M. van der Heijden and C. Trahiotis, "Masking with interaurally delayed stimuli: The use of 'internal' delays in binaural detection," *J. Acoust. Soc. Am.* **105**, 388–399 (1999).
- ⁴M. M. Sondhi and N. Guttman, "Width of the spectrum effective in the binaural release of masking," *J. Acoust. Soc. Am.* **40**, 600–606 (1966).
- ⁵B. G. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138 (1990).
- ⁶D. McAlpine, D. Jiang, T. M. Shackleton, and A. R. Palmer, "Convergent input from brainstem coincidence detectors onto delay-sensitive neurons in the inferior colliculus," *J. Neurosci.* **18**, 6026–6039 (1998).
- ⁷S. K. Thompson, K. von Kriegstein, A. Deane-Pratt, T. Marquardt, R. Deichmann, T. D. Griffiths, and D. McAlpine, "Representation of interaural time delay in the human auditory midbrain," *Nat. Neurosci.* **9**, 1096–1098 (2006).