

A Physiologically Based Model of Interaural Time Difference Discrimination

Kenneth E. Hancock¹ and Bertrand Delgutte^{1,2}

¹Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, and ²Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02143

Interaural time difference (ITD) is a cue to the location of sounds containing low frequencies and is represented in the inferior colliculus (IC) by cells that respond maximally at a particular best delay (BD). Previous studies have demonstrated that single ITD-sensitive cells contain sufficient information in their discharge patterns to account for ITD acuity on the midline (ITD = 0). If ITD discrimination were based on the activity of the most sensitive cell available (“lower envelope hypothesis”), then ITD acuity should be relatively constant as a function of ITD. In response to broadband noise, however, the ITD acuity of human listeners degrades as ITD increases. To account for these results, we hypothesize that pooling of information across neurons is an essential component of ITD discrimination. This report describes a neural pooling model of ITD discrimination based on the response properties of ITD-sensitive cells in the IC of anesthetized cats.

Rate versus ITD curves were fit with a cross-correlation model of ITD sensitivity, and the parameters were used to constrain a population model of ITD discrimination. The model accurately predicts ITD acuity as a function of ITD for broadband noise stimuli when responses are pooled across best frequency (BF). Furthermore, ITD tuning based solely on a system of internal delays is not sufficient to predict ITD acuity in response to 500 Hz tones, suggesting that acuity is likely refined by additional mechanisms. The physiological data confirms evidence from the guinea pig that BD varies systematically with BF, generalizing the observation across species.

Key words: auditory; binaural; hearing; inferior colliculus; localization; psychophysics

Introduction

A fundamental question in neuroscience concerns the number of neurons used by the CNS to make sensory judgments (Parker and Newsome, 1998). Comparisons of neural response properties to psychophysical performance have led to two general models of neural coding. The lower envelope hypothesis applies when the firing pattern of a single neuron accounts for sensory judgments related to a particular stimulus (Parker and Newsome, 1998). For example, vibration detection thresholds for human subjects are quantitatively similar to the lowest response thresholds exhibited by primary somatosensory afferents over a wide range of stimulus frequencies (Mountcastle et al., 1972). In other cases, a neural pooling model better accounts for psychophysical performance. For example, the smallest discriminable change in interocular disparity, on which depth perception is based, increases with the base disparity at which it is measured (Badcock and Schor, 1985). This is inconsistent with the use of the most sensitive neurons in cortical areas V1 and V2 because these neurons are highly sensitive to changes in disparity regardless of the

base disparity to which they are tuned (Poggio and Fischer, 1977). Interocular disparity discrimination is more consistent with a model in which psychophysical judgments are based on the pooled responses of neurons tuned to many different base disparities (Lehky and Sejnowski, 1990).

Neural pooling is also an issue in the context of the discrimination of interaural time difference (ITD), a dominant cue to the location of sounds containing low-frequency components (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002). In response to 500 Hz tones, the just noticeable difference (JND) in ITD is nearly constant at 10 μ sec as a function of ITD (Domnitz and Colburn, 1977). For broadband noise, however, the JND is three to four times larger and systematically increases as ITD increases (Mossop and Culling, 1998).

Previous studies have shown that ITD acuity near the midline is consistent with the responses of single ITD-sensitive inferior colliculus (IC) neurons (Skottun et al., 2001; Shackleton et al., 2003). Because IC neurons are tuned to a wide range of ITDs, application of the lower envelope hypothesis to ITD discrimination predicts uniform acuity for all ITDs. However, this contradicts the observation that ITD JNDs in response to broadband noise systematically increase with ITD (Mossop and Culling, 1998). We hypothesized that a neural pooling model might better account for ITD acuity in general and developed such a model based on the physiological response properties of ITD-sensitive cells in the cat IC. We first developed a version of the cross-correlator model (Colburn, 1973) that can be fit efficiently to

Received March 2, 2004; revised June 22, 2004; accepted June 23, 2004.

This work was supported by National Institute on Deafness and Other Communication Disorders Grants DC00119, DC02258, and DC05295. We thank Leslie Liberman and Connie Miller for surgical support and the two anonymous reviewers for their helpful comments.

Correspondence should be addressed to Kenneth E. Hancock, Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, MA 02114. E-mail: keh@epl.meei.harvard.edu.

DOI:10.1523/JNEUROSCI.0762-04.2004

Copyright © 2004 Society for Neuroscience 0270-6474/04/247110-08\$15.00/0

single-unit rate–ITD curves. The parameters derived from the data were then used to constrain a population model of ITD discrimination. The population model accurately predicts ITD acuity as a function of ITD for broadband noise when responses are first pooled across best frequency (BF). It appears that the auditory system places a premium on the consistency of ITD information across frequency (Stern et al., 1988) at the expense of ITD acuity on the periphery.

Materials and Methods

Physiological methods. Adult cats were anesthetized with intraperitoneal injections of dial-in urethane (diallylbarbituric acid; 75 mg/kg; Sigma, St. Louis, MO) and secured on a stereotaxic apparatus inside an electrically shielded, double-walled sound-attenuating chamber. The inferior colliculi were visualized by opening the posterior fossa and partially aspirating the cerebellum. Single units were isolated using parylene-insulated tungsten microelectrodes (Micro Probe, Potomac, MD).

Acoustic stimuli were generated at a sampling rate of 20 kHz and presented to the animal over calibrated headphones (Realistic 40-1377; RadioShack, Ft. Worth, TX). ITD sensitivity was characterized using frozen broadband noise (bandwidth, 3 kHz) stepped in interaural delay from -2000 (ipsilateral leading) to $+2000$ μsec (contralateral leading) in 200 μsec increments. For some low-frequency units (BF less than ~ 400 Hz), a wider range of ITDs was used (-3000 to $+3000$ μsec) to capture the side lobes of the rate–ITD curve (Fig. 1*b*). Stimulus level was 50 dB sound pressure level, which was typically 15–20 dB above unit threshold.

Single-neuron model. Each rate–ITD curve was fit with the cross-correlation model of Figure 1*a*. For stimulation with broadband noise, the expected value of the normalized interaural correlation, $\rho(\text{ITD})$, is the cross-correlation of the ipsilateral filter impulse response and the time-delayed contralateral filter impulse response, calculated as follows:

$$\rho(\text{ITD}) = \frac{\int_{-\infty}^{\infty} h_{\text{ipsi}}(t)h_{\text{contra, delayed}}(t - \text{ITD})dt}{\int_{-\infty}^{\infty} h_{\text{ipsi}}(t)h_{\text{contra}}(t)dt} \quad (1)$$

Because the single-neuron model being described will be applied to single IC units, the filters represent the cumulative effects of filtering peripheral to the IC, including the basilar membrane mechanics, and possible convergence of tuned inputs in the brainstem. Fourth-order gammatone filters were used, so that:

$$h_{\text{ipsi}}(t) = \left(\frac{t}{\tau_0}\right)^3 \times \exp\left(-\frac{t}{\tau_0}\right) \times \cos(2\pi \times \text{CF} \times t) \times u(t) \quad (2)$$

where $u(t)$ is the unit step function, CF is the characteristic frequency of the filter, and τ_0 is its decay time constant (Patterson et al., 1988). The quality factor of a gammatone filter is given by $Q = 2\pi \times \text{CF} \times \tau_0$ (Darling, 1991). As a consequence of the narrowband filtering, responses tend to be periodic as a function of ITD. For pure-tone stimuli, the responses are periodic at the stimulus frequency; for broadband noise, they are quasiperiodic at the neural best frequency. Incorporating a characteristic delay (CD) and characteristic phase (CP) on the contralateral side, we use the following:

$$h_{\text{contra, delayed}}(t) = \left(\frac{t - \text{CD}}{\tau_0}\right)^3 \times \exp\left(-\frac{t - \text{CD}}{\tau_0}\right) \times \cos(2\pi \times \text{CF} \times [t - \text{CD}] + \text{CP}) \times u(t) \quad (3)$$

To minimize the number of free parameters, filters on both sides are assumed to have the same CF and time constant τ_0 . The relationship between firing rate (R) and interaural correlation (ρ) in the IC can often

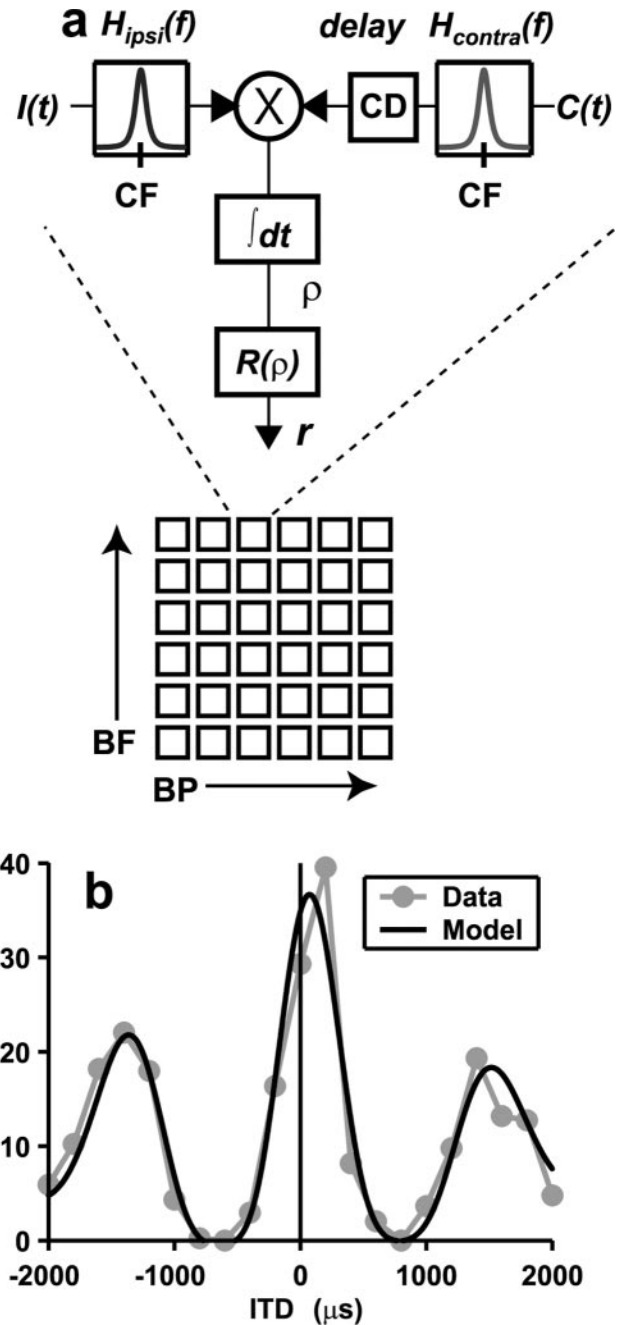


Figure 1. *a*, Top, cross-correlation model for ITD-sensitive neurons. Ipsilateral (I) and contralateral (C) stimuli are passed through narrowband filters representing auditory periphery. Filtered contralateral signal is delayed by a CD and in some cases, phase shifted by a CP . Interaural correlation is converted to a firing rate by the input–output function $R(\rho)$. Bottom, population model is a grid of cross-correlation elements containing independent distributions of BF and BP. *b*, Firing rate as a function of ITD for a typical delay-sensitive unit from cat central nucleus of the IC responding to broadband noise (gray circles). Curve is periodic at BF (673 Hz). Black line is the best model fit to the data.

be categorized as quadratic in shape (Albeck and Konishi, 1995), as follows:

$$R(\rho) = A \times \left(\frac{\rho + 1}{2}\right)^2 + B \quad (4)$$

Together, Equations 1–4 give an implicit expression for the mean firing rate as a function of ITD. For each single unit, a nonlinear algorithm was used to find the parameter set $\{\text{CF}, \tau_0, \text{CD}, \text{CP}, A, B\}$ minimizing the sum

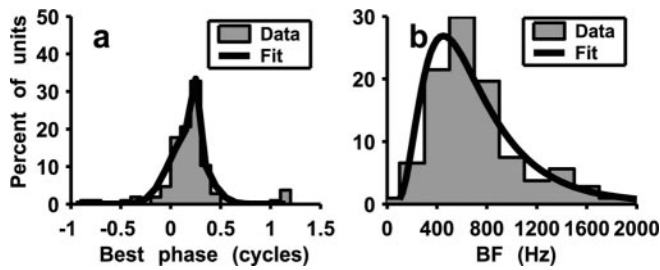


Figure 2. *a*, Gray bars, BF distribution for delay-sensitive units in this study. Black line, Best lognormal fit. *b*, Gray bars, BP distribution. Black line, Best sum-of-Gaussians fit.

of squared errors between the model and the rate–ITD data. An example fit is shown in Figure 1*b*. The estimate of best delay (BD) was taken as the ITD corresponding to the peak of the best fit. The BF of each unit was defined as the best-fitting CF parameter. For clarity, we will use *BF* to refer to the quantity estimated from physiological data and *CF* to refer to the model filter parameter. Statistics summarizing the 107 estimated parameter sets were used to constrain a population model, as described in the following section.

Parameterization of the population ITD model. The population model is a two-dimensional grid of ITD-sensitive neural elements with BF represented on one axis and best interaural phase ($= BD \times BF$) on the other. Only one side of the brain is represented to minimize computation; test simulations suggest that the results are not significantly different using a bilateral configuration. Best phase (BP) was used because its distribution is more nearly independent of BF than the distribution of best delay (see Fig. 4). The distribution of best phase (Fig. 2*a*) was parameterized by a weighted sum of two Gaussians, as follows:

$$f(\text{BP}) = \frac{w}{\sigma_1 \sqrt{2\pi}} \exp\left[-\frac{(\text{BP} - \mu_1)^2}{2\sigma_1^2}\right] + \frac{1-w}{\sigma_2 \sqrt{2\pi}} \exp\left[-\frac{(\text{BP} - \mu_2)^2}{2\sigma_2^2}\right] \quad (5)$$

where $w = 0.19$, $\mu_1 = 0.23$, $\sigma_1 = 0.04$, $\mu_2 = 0.16$, and $\sigma_2 = 0.19$. Model BPs were spaced in equal probability steps [i.e., with a constant value for the integral of $f(\text{BP})$ between consecutive points].

To a close approximation, the best delay of the single-neuron model in response to broadband noise is determined by the characteristic delay and the characteristic phase: $BD = CD + CP/CF$. It follows directly that the best phase is as follows: $BP = CD \times CF + CP$ (Fitzpatrick et al., 2000). In the simulations that follow, two versions of the model, corresponding to two extreme conditions, are considered. In the first condition, the best phase is implemented as a pure delay: $CD = BP/BF$; $CP = 0$. In the second condition, it is a pure phase shift: $CP = BP$; $CD = 0$. Either assumption results in the appropriate distribution of best phase (and equivalently best delay) with respect to BF for noise, but each implies a different mechanism and leads to different results for other stimuli, such as pure tones.

Figure 2*b* shows the physiological distribution of BF. The data were fit with a lognormal probability distribution function, as follows:

$$f(\text{BF}) = \frac{1}{\text{BF} \times \sigma \sqrt{2\pi}} \times \exp\left[-\frac{(\ln(\text{BF}) - \mu)^2}{2\sigma^2}\right] \quad (6)$$

where $\mu = 6.5$ and $\sigma = 0.51$. The lognormal distribution was chosen because it captures the asymmetry of the data and because it is defined strictly for positive BF values. Model BFs were spaced in equal probability steps. For each model element, the value of τ_0 was set to $2.3/(2\pi \times CF)$ to give a quality factor equal to the median value, $Q = 2.3$, estimated from the IC single-unit data. This median quality factor is smaller than that estimated from the cat auditory nerve ($Q \approx 3.4$) (Carney, 1993), reflecting slightly broader tuning at more central points in the auditory system. Finally, an identical quadratic input–output function was used for each model element (Eq. 4) with the parameters set to the mean physiological estimates ($A = 31$ spikes/sec; $B = 1$ spike/sec).

Simulating ITD discrimination with the population model. The model

inputs consist of 1-sec-long tokens of noise. At each step in the simulation, two noise tokens are compared, one at a base ITD, itd_0 , and the other at a test ITD, itd_1 . Frozen noise was used throughout the simulation so that the tokens compared at any step differed only in ITD. The model output is the percentage correct discrimination P_C , predicted by signal detection theory for a two-alternative forced-choice procedure. The specific method used was the ideal observer model for multiple Gaussian independent observations (Green and Swets, 1988). For a model element in the i th position on the *CF*-axis and the j th position on the *BP*-axis, a d' value was computed from the firing rate r and its variance σ^2 , as follows:

$$d'_{ij} = \frac{|r_{ij}(itd_1) - r_{ij}(itd_0)|}{\sqrt{\frac{1}{2}[\sigma_{ij}^2(itd_1) + \sigma_{ij}^2(itd_0)]}} \quad (7)$$

Under the assumption that the firing rate variance is proportional to its mean (i.e., $\sigma^2 = k_0 \times r$), we use the following:

$$d'_{ij} = \frac{|r_{ij}(itd_1) - r_{ij}(itd_0)|}{\sqrt{\frac{k_0}{2} \times [r_{ij}(itd_1) + r_{ij}(itd_0)]}} \quad (8)$$

This assumption is consistent with the physiological data of this study (data not shown), as well as that of comparable studies in the visual cortex (Lehky and Sejnowski, 1990). The constant k_0 was determined from the physiological rate–ITD data by computing the mean rate and variance for each ITD used and for all 107 single units. The average variance to mean ratio $k_0 = 0.8$ was used in all model computations. Because the data were collected using frozen noise, the value of k_0 represents the intrinsic response variability of ITD-sensitive neurons in the IC. The cumulative d' for the ideal observer model is as follows:

$$d' = \sqrt{\epsilon \times \sum_i \sum_j (d'_{ij})^2} \quad (9)$$

An efficiency parameter, ϵ , is introduced to ensure that model performance reflects the activity in the whole population and does not reach threshold ($d' \approx 1$) based solely on the activity of the most sensitive model elements (Geisler and Albrecht, 1997). The model population was 15 BFs by 15 BPs in size, and the efficiency was set to $\epsilon = 1/18$. This combination of values produced ITD acuity in response to broadband noise quantitatively similar to that observed psychophysically.

The efficiency parameter has no direct physiological correlate but may encompass several factors. One such factor is stimulus variability. The psychophysical results (Mossop and Culling, 1998) were obtained using random (not frozen) noise and therefore are affected by trial-to-trial variability in the stimulus. Stimulus variability cannot be explicitly modeled simply by generating new noise tokens on each trial because the model is based on the average interaural correlation over 1 sec, which does not change appreciably from trial to trial. An explicit model of stimulus variability requires a thresholding mechanism sensitive to the instantaneous interaural correlation. We avoided this additional level of complexity by assuming that stimulus variability can be approximated by a suboptimal efficiency.

Stimulus variability alone, however, is unlikely to account for the small efficiency value (Shackleton and Palmer, 2004). Other inefficiencies may arise psychophysically from the pooling process or in the decision-making process. Because it is not possible to characterize all sources of inefficiency from the data, the value was empirically adjusted to produce quantitative agreement with the psychophysical data. For this reason, no significance can be attached to the absolute value of the simulated ITD JNDs. Rather, the focus is on trends in JND with changes in the base ITD.

Finally, the percentage correct was computed from d' using the cumulative normal distribution:

$$P_C = \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{d'} e^{-x^2/2} dx - 1 \quad (10)$$

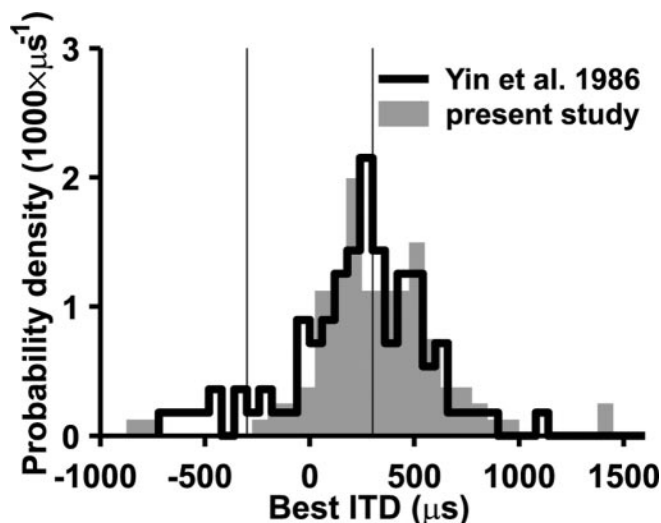


Figure 3. Distributions of best ITD in the anesthetized cat. Gray shading, Present study. Black line, Data from Yin et al. (1986). Vertical lines indicate approximate natural ITD range for the cat. In both studies, best ITD is estimated from rate–ITD curves measured in response to broadband noise.

For each base ITD, the test ITD was adjusted to find the value giving 75% correct.

Results

Distribution of best delay depends on best frequency

Rate–ITD curves were measured for 107 single units in the IC from 13 anesthetized cats. Physiologically relevant parameters were estimated by fitting a cross-correlation model to the data using nonlinear estimation techniques. The model gave an excellent fit to the data, on average accounting for 93% of the variance in the neural responses. As shown in Figure 3, the overall distribution of best delay estimates was essentially the same as the distribution reported previously for the anesthetized cat (Yin et al., 1986).

Figure 4*a* shows that the BD distribution depends on BF, in which the BF for each unit is the CF parameter of the best-fitting single-neuron model. Best delays greatly exceed the natural ITD range ($\sim 300 \mu\text{sec}$; horizontal dashed line) for many low-BF units. For higher BFs, however, the best delays seem to cover this range only partially. The dependence on BF was quantified by dividing the units into four quartiles based on BF and computing the median BD for each quartile (filled squares). The BF ranges for each quartile are <450 , $451\text{--}615$, $616\text{--}855$, and >856 Hz. Figure 4*b* demonstrates that the median BD in the cat IC (squares) and the mean BD in the guinea pig IC (circles) (McAlpine et al., 2001) decrease with BF in a similar manner. Furthermore, the interquartile deviation of BD in the cat and the SD of BD in the guinea pig are similar in magnitude and also decrease with increasing BF (Fig. 4*b*, inset).

As shown in Figure 4*c*, when the BD is instead expressed as a phase relative to BF ($BP = BD \times BF$), the distribution becomes more nearly independent of BF. In the cat IC, the median and interquartile deviation of best phase increase only slightly with BF in a manner quantitatively consistent with the best phase mean and SD in the guinea pig (Fig. 4*d*; squares, cat; circles, guinea pig). Thus, in the ITD processing mechanism of the cat as well as the guinea pig, best phase, rather than best delay, is distributed independently of BF.

In the guinea pig, the maximum slopes of the rate–ITD curves tend to occur within the range of naturally occurring ITDs and

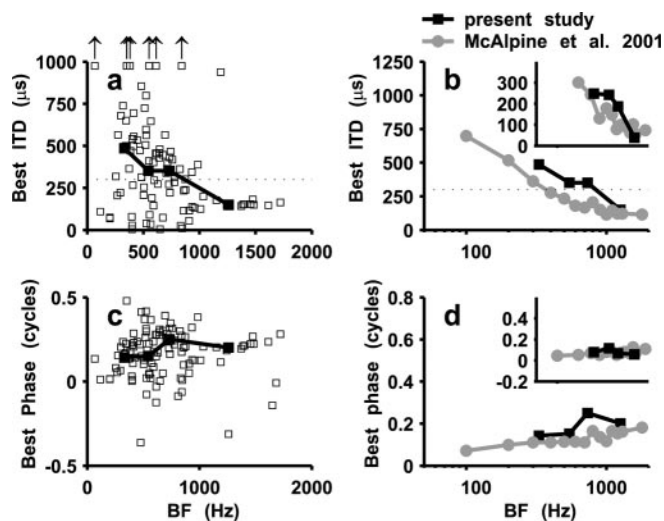


Figure 4. *a*, Scatter plot showing dependence of best delay distribution on BF (open squares). Filled squares, Median best delay in four BF bands. Black line, Approximate maximum naturally occurring ITD for the cat. *b*, Median data from the cat (squares) show a similar trend as the mean BD as a function of BF for the guinea pig IC (circles). In both species, BD distributions are narrower for higher BFs, as quantified by the interquartile deviation (cat, squares) or SD (guinea pig, circles) (inset). *c*, Scatter plot of best phase ($BD \times BF$) versus BF. Median values (filled squares) are shown for the same BF bands used in *a*. *d*, Best phase is more nearly independent of BF in both species. Spread of distribution is also independent of BF (inset).

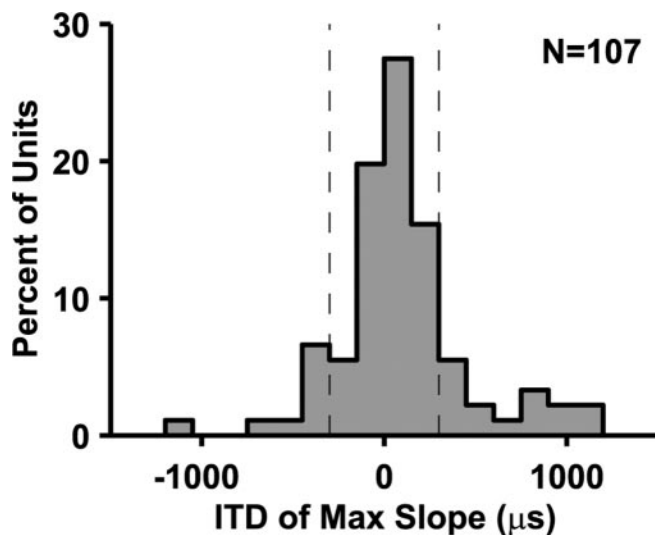


Figure 5. Distribution of the ITD corresponding to maximum slope of rate–ITD curve. Values are clustered around the midline, within the natural ITD range for the cat (dashed lines).

are centered around the midline (McAlpine et al., 2001). This is because, as the peaks shift to higher BDs for lower BFs, the peaks also become wider because the rate–ITD curves are quasiperiodic at BF. The same is true in the cat, as shown in Figure 5. The ITD corresponding to the maximum slope was found for each unit from the best fit of the cross-correlation model. The distribution of these ITDs (gray bars) is centered near $\text{ITD} = 0$ and primarily restricted to the natural ITD range (vertical dashed lines). This is consistent with the notion that a change in stimulus ITD is coded not by a change in the locus of peak activity but by changes in firing rate within the population. The results from the cat, including both this study and that of Joris et al. (2004), extend the results of McAlpine et al. (2001) by demonstrating that these

characteristics are not limited to rodents or to animals with small heads but may be a general property of mammalian ITD coding.

Simulations of ITD discrimination for broadband noise

The population model of Figure 1*a* is an implementation of an ITD processor consistent with the physiological results of Figure 4. The model was constrained using the parameters obtained by fitting the cross-correlation model to the rate–ITD data for noise stimuli and tested by predicting human psychophysical ITD discrimination performance for both tones and noise.

Figure 6 shows simulations of ITD discrimination using broadband noise as the stimulus. Human performance is characterized by a greater than twofold increase in JND between the midline and ITD = 600 μ sec (Mossop and Culling, 1998). The model, however, predicts a nearly constant JND for all ITDs (triangles). This behavior is a direct result of the symmetry of the model ITD curves with respect to best ITD, as illustrated in Figure 6*b*. An individual model element is maximally sensitive to changes in ITD along the rising slope of the central peak but is equally sensitive to ITD changes along the falling slope. The ITD acuity of the model is thus improved by the existence of falling slopes that lie within the physiological ITD range. Such slopes arise from all high-CF neurons regardless of best ITD and also from the small minority of low-CF neurons having best ITDs near the midline.

This argument suggests that a mechanism that suppresses the falling slopes while preserving the rising slopes will lead to decreased sensitivity away from the midline. One such mechanism is simply to average rate responses across BF at each best phase. The midline slopes of the central peaks tend to align across BF, but the large variation in falling slope location with BF causes these slopes to misalign across BF (Fig. 6*c*). Therefore, the result of pooling across BF is to preserve the large rising slope near the midline while reducing the magnitude of the falling slope, as shown by the heavy black line in Figure 6*c*. The effect on ITD discrimination was computed in the model by averaging firing rates of neural elements across BF. The firing rate of a model element in the *i*th position on the CF-axis and the *j*th position on the BP-axis was computed as follows:

$$r_{i,j} = \frac{1}{N_{CF}} \sum_{k=1}^{N_{CF}} r_{k,j} \quad (11)$$

where N_{CF} (= 15) is the length of the CF-axis. The ideal observer model was then applied to the averaged firing rates, retaining the assumption that the variance of the firing rate is proportional to the mean. No change was made to the proportionality constant k_0 because any such change can be offset by an appropriate change in the efficiency parameter to maintain quantitative consistency with psychophysical data. The circles in Figure 6*a* demonstrate that this configuration of the model accurately predicts ITD JNDs for human listeners in response to broadband noise.

Simulations of ITD discrimination for pure tones

The model was further tested by applying it to ITD discrimination for 500 Hz tones (Fig. 7). Across-BF integration was used

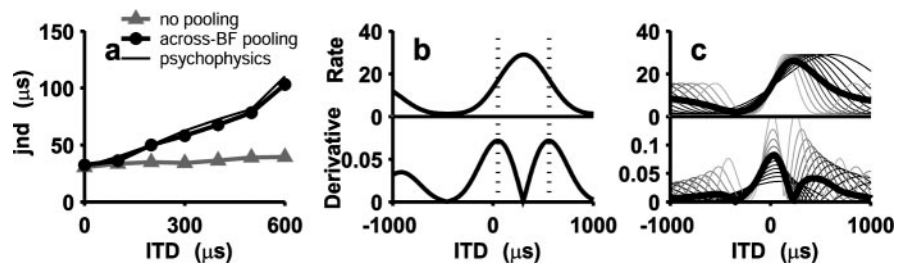


Figure 6. *a*, Population model predictions of JNDs as a function of ITD for broadband noise stimulation, with across-BF integration (circles) and without (triangles). Mean human performance from Mossop and Culling (1998) shown by black line. *b*, Top, Rate versus ITD for a single model element with $BF = 629$ Hz and $BP = 0.19$ cycles. Bottom, Derivative of rate curve with respect to ITD. Model element has a maximum sensitivity to ITD changes on both slopes of the central peak (dashed vertical lines). *c*, Top, Model rate–ITD curves for all BFs and $BP = 0.19$. Average across BF has a shallower slope away from the midline and no side lobes (bold black line). Bottom, Corresponding derivatives.

because doing so provides the most accurate predictions for broadband noise (circles, repeated from Fig. 6*a*). Two configurations of the model were tested differing only in the implementation of the best phase assigned to each model neuron (Fig. 1*a*). In the “pure delay model,” the mechanism takes the form of a pure delay (i.e., $BP = CD/CF$), whereas in the “pure phase shift model,” it takes the form of a pure phase shift ($BP = CP$). Both configurations reproduce the observed dependence of BD on BF for broadband noise. The pure delay model is intuitively appealing because of the long-standing idea that spike conduction delays underlie ITD tuning. The notion of a pure phase shift model requires additional comment. It is not intended to suggest that internal delays are identically zero; the existing evidence clearly indicates otherwise. Rather, the pure phase shift model is a more abstract construction whose properties overcome certain limitations of the pure delay model. Possible physiological interpretations of the pure phase shift model will be discussed below.

In response to 500 Hz tones, human listeners exhibit JNDs that are nearly constant as a function of ITD and ~ 10 μ sec in magnitude (bottom black line) (Domnitz and Colburn, 1977). In contrast, the pure delay model (upward triangles) predicts that JNDs increase slightly with ITD up to ~ 200 μ sec and decrease for higher ITDs, attaining their overall minimum value at ITD = 600 μ sec, contrary to the psychophysical data. Furthermore, the pre-

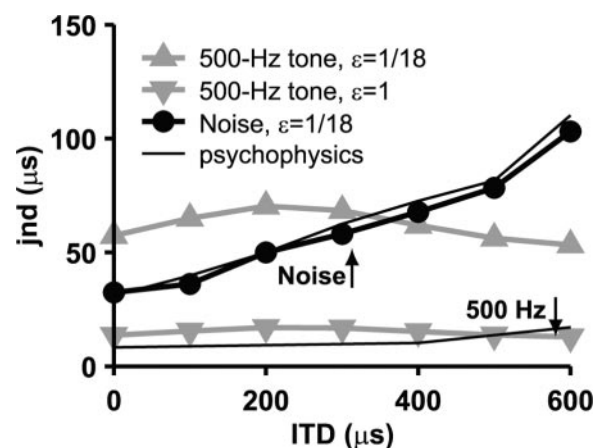


Figure 7. Predictions of ITD acuity for 500 Hz tones (upward triangles) from pure delay model. Average human data from Domnitz and Colburn (1977) (bottom black line) is much smaller in magnitude than model predictions. Increasing model efficiency to $\epsilon = 1$ (downward triangles) brings predictions in line with psychophysical data. Broadband noise results are repeated from Figure 6*a* (black circles).

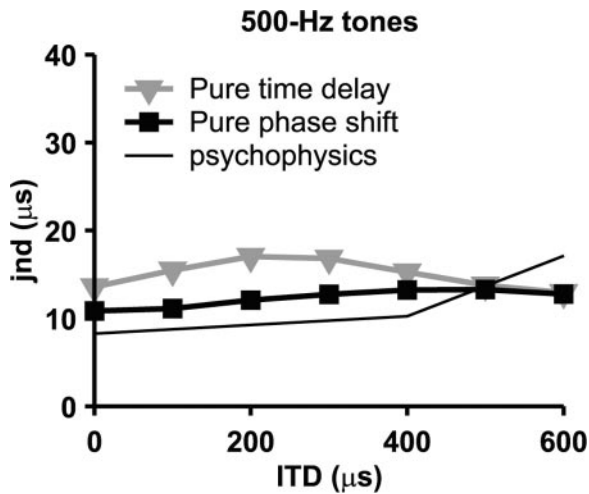


Figure 8. Predictions of ITD acuity for 500 Hz tones with $\epsilon = 1$ (triangles). Curve is non-monotonic and is minimal off the midline. Pure phase shift model (squares) more accurately predicts the shape of the discrimination curves for tones. Black lines, Average human data.

dicted JNDs are $>50 \mu\text{sec}$, more than five times larger than the values measured experimentally.

The ITD discrimination curve for 500 Hz tones can be brought in line with psychophysical observations by increasing the efficiency value ϵ from $1/18$ to 1 (downward triangles). As discussed in Materials and Methods, this parameter may reflect any number of inefficiencies in the auditory system not explicitly incorporated into the model. The discrepancy in acuity between tones and noise is a problem that afflicts not just our model but is a property of the psychophysical data for which there is no satisfactory explanation in the literature of which we are aware. One possibility is that the difference lies in the temporal details of the responses to each stimulus. ITD discrimination may be less efficient for noise because, unlike the model, the biological system cannot average over long times and is subject to short-term fluctuations in estimates of interaural correlation. A quantitative treatment of such short-term variability requires a specifically designed study and is not practical with the data at hand. As stated above, however, the value of the efficiency parameter affects only the absolute value of the simulated JNDs, whereas this study is primarily concerned with trends in the data with changes in ITD.

As shown in Figure 7, increasing the value of ϵ also flattens the ITD discrimination curve for 500 Hz tones. However, when the JND-axis is expanded, the nonmonotonic character of the response is still apparent (Fig. 8). The underlying cause for this behavior is illustrated in Figure 9, which shows rate–ITD curves for all of the model elements at a single best phase ($BP = 0.19$, the median physiological value). For clarity, only one best phase is represented; elements at all best phases contribute to the data shown in Figure 8. In the pure delay model, the best delay varies inversely with BF ($BD = BP/BF$). Hence, the peaks of the rate–ITD curves shift toward zero with increasing BF. For broadband noise, the curves are quasiperiodic at BF so that the peaks narrow as they shift (Fig. 9a), causing the midline slopes to align across BF. For tonal stimuli, however, the curves are periodic at the stimulus frequency, and hence the peaks just shift without becoming narrower. Consequently, the midline slopes do not align (Fig. 9b), leading to a shallower slope after pooling across BF.

For the pure phase shift model, the peripheral filter outputs in response to noise are quasiperiodic at BF, and the central peaks

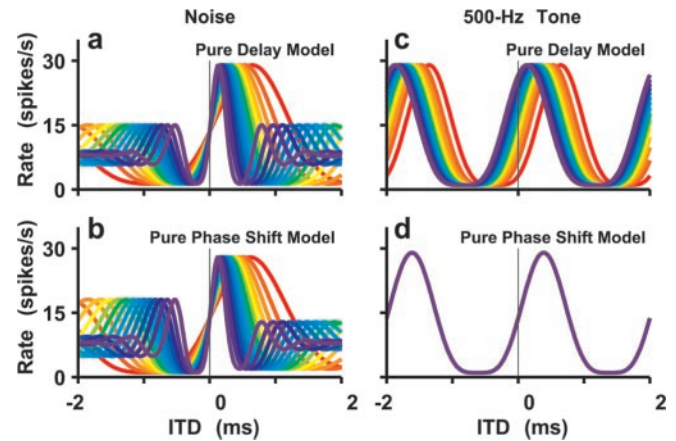


Figure 9. Comparison of pure delay and pure phase shift models. In response to broadband noise, rate–ITD curves across BF for $BP = 0.19$ align on the rising slopes of the central peaks for both the pure delay model (a) and the pure phase shift model (b). c, For the pure delay model using 500 Hz tones, however, neither peaks nor slopes align across BF. d, Rate–ITD curves for 500 Hz tones exactly align across BF for $BP = 0.19$ using a pure phase shift.

shift with BF exactly as in the case of the pure delay model (Fig. 9c). The pure phase shift causes an asymmetry in the magnitudes of the side lobes. This asymmetry does not affect the ITD discrimination curve because the side lobes are smaller in magnitude than the central peaks, and most occur outside the range of base ITDs tested. In response to tones, the rate–ITD curves are periodic at the stimulus frequency, and, because the best delay is determined by the characteristic phase, the best delay also depends on the stimulus frequency ($BD = BP/f$). This means that, for tonal stimuli, the rate–ITD curves are identical regardless of BF and the midline slopes are perfectly aligned (Fig. 9d).

The pure phase shift model thus provides alignment of the rate–ITD curve central slopes in response to both tones and noise. As a result, the predicted ITD JNDs for 500 Hz tones are minimal on the midline and increase only slightly as the base ITD increases (Fig. 8, squares), consistent with the psychophysical data. The results of Figures 7–9 suggest that an ITD processing mechanism that incorporates an internal phase shift better accounts for some aspects of psychophysical ITD discrimination than does an internal time delay alone.

Discussion

This study was based on the rate responses to broadband noise as a function of ITD for 107 units in the IC of anesthetized cats. The median best delay was found to be negatively correlated with BF, whereas median best phase was more nearly independent of BF. This is consistent with recent findings from the cat (Joris et al., 2004) and extends the generality of previous results from the guinea pig IC (McAlpine et al., 2001) and gerbil medial superior olive (MSO) (Brand et al., 2002). The data were used to parameterize a population model of ITD coding in which individual model elements cross-correlate the signals from the two ears after applying to one side either a characteristic delay (pure delay model) or a characteristic phase (pure phase shift model). For the case of broadband noise, both models align the central peaks of the rate–ITD curves across BF. Both accurately predict the ITD dependence of human ITD discrimination when neural responses are first integrated across BF. For stimulation with 500 Hz tones, however, the pure delay model incorrectly predicts that maximum ITD acuity occurs off, rather than on, the midline. This inaccuracy is corrected by the pure phase shift model.

Distribution of best delays

Figure 4 demonstrates that the BD distribution changes systematically with BF. Specifically, higher BFs are characterized by smaller BD medians and smaller BD ranges. Median BD falls approximately as $1/\text{BF}$ so that the best phase is nearly constant with BF (Fig. 4*d*). We cannot rule out that the posterior approach we used was biased against finding units with high BFs and long BDs. Using a dorsal approach, however, Joris et al. (2004) observed a similar relationship between BF and BD. It remains possible that areas of the IC poorly sampled by either approach have a different dependence of BD on BF.

Demonstration of this relationship in the cat is significant because the head size of the cat, in terms of ITD, is approximately twice that of the guinea pig (300 vs 150 μsec) and three times that of the gerbil ($\sim 120 \mu\text{sec}$). Thus, it is unlikely that the findings in the guinea pig IC and gerbil MSO are anomalies related to species or head size and more likely that the results represent a fundamental principle of organization underlying mammalian ITD coding.

Simulated ITD discrimination for broadband noise stimuli

The results of Figure 6 suggest that ITD processing involves significant pooling of information across neurons. Previous studies have demonstrated that the information in the firing patterns of single delay-sensitive neurons in the IC is sufficient to account for human ITD acuity near the midline (Skottun et al., 2001; Shackleton et al., 2003). A model that discriminates ITD on the basis of only the most sensitive neurons, however, predicts constant JNDs as a function of ITD. This is because, regardless of how best delays are distributed, for any ITD that can be produced by naturally occurring sounds, there is certain to be at least one neuron highly sensitive to changes around that ITD. Thus, although single-neuron ITD curves are consistent with ITD acuity on the midline, pooling of information across many neurons is required to account for the degradation of acuity away from the midline. The results of Figure 6 illustrate the importance of neural pooling to ITD coding and suggest that, even for the relatively simple task of discriminating two ITDs, the system is unable to isolate the acuity at the most sensitive internal delay.

Integration across BF

Integration across BF was proposed previously in a model of ITD coding that accounts for the lateralization of certain dichotic stimuli (Shackleton et al., 1992). In that study, integration was described as performing a “straightness weighting” to emphasize ITDs that are consistent across BF (Stern et al., 1988). Integration across BF was used here to reduce the off-midline slopes of the rate–ITD curves, causing ITD acuity to decay with increasing distance from the midline (Fig. 6). Although our data were not measured with sufficient resolution to quantitatively compare slopes of ITD curves, in the guinea pig IC, 80% of the noise delay functions have steeper midline slopes than off-midline slopes (McAlpine et al., 2001). One possibility is that this difference in slopes results from convergence across BF in the MSO projection to the IC, which would be consistent with the mechanism used in this model (McAlpine et al., 1998). It is also possible that inhibitory inputs, either intrinsic to the IC or arising from the dorsal nucleus of the lateral lemniscus, underlie the relative steepness of the midline slope. Whatever the mechanism, the modeling results suggest that processing beyond the cross-correlation model is essential to the formation of the neural code for ITD and, furthermore, that the auditory system seems to place a premium on the consistency of ITD information across frequency at the expense

of ITD acuity on the periphery. The advantage gained by pooling across BF to emphasize consistent ITD cues may be robustness of the ITD code to such degradations as reverberation or the presence of multiple sources.

Simulated ITD discrimination for tones

Human ITD discrimination of 500 Hz tones differs from that of broadband noise in that JNDs are smaller in magnitude and nearly constant with ITD. The generality of the model was examined by predicting ITD discrimination for both stimulus conditions. The pure time delay model predicts JNDs that vary non-monotonically with ITD (Fig. 7), whereas the pure phase shift model produces JNDs that remain nearly constant (Fig. 8). As illustrated in Figure 9, the difference arises because, with a pure phase shift, the midline slopes of the rate–ITD curves align across BF regardless of stimulus, whereas with a pure time delay, the slopes align only for broadband noise.

The results of Figures 7–9 suggest that ITD tuning based on internal phase shifts better predicts observed ITD acuity in response to tones than does tuning based on systematically arranged internal delay lines. The key to the success of the pure phase shift model is the consistent alignment of rate–ITD slopes across BF. Any mechanism producing comparable alignment would also be expected to account for ITD acuity. For example, the central peaks of rate–ITD curves in the rabbit IC appear to remain relatively constant in width as the stimulus frequency changes (Fitzpatrick and Kuwada, 2001). As a result, the slopes align to a greater degree than predicted by the cross-correlation model. The mechanism underlying this behavior is not fully understood but likely involves inhibitory processes and the convergence of delay-sensitive inputs in the IC (McAlpine et al., 1998; Fitzpatrick and Kuwada, 2001).

Another possible mechanism is stereausis, an interaural difference in cochlear traveling wave delay that results from a mismatch in CF between the ears (Shamma et al., 1989; Joris et al., 2004). We performed simulations (data not shown) in which ITD tuning was implemented by systematically mismatching CFs, but the ITD discrimination curves were similar to those of the pure delay model. This is consistent with the theoretical results of Bonham and Lewis (1999), which demonstrate that the phase versus frequency relationship for the stereausis model is not constant but deviates only slightly from a linear relationship.

The conclusion to be drawn from our study is that ITD tuning by a system of internal delays alone is not sufficient to account fully for ITD discrimination; an additional mechanism must be present.

Neuronal pooling and the lower envelope hypothesis

These results have demonstrated the importance of response pooling in predicting ITD acuity across the full range of physiological ITDs. In apparent contrast, other studies have shown that binaural unmasking based on ITD coding can be consistent with the lower envelope hypothesis. Lane et al. (2004) described a neural correlate of spatial release from masking (SRM), which occurs by spatially separating a signal from a masker. For a variety of signal and masker locations, psychophysical masked thresholds were accurately predicted by the lowest masked threshold among ITD-sensitive neurons in the IC. However, because both signal and masker were broadband, one would expect similar neural masked thresholds at all BFs, and thus either model, lower envelope or neural pooling, might actually account for the psychophysical data. When SRM is measured using narrowband signals, psychophysical thresholds increase relative to the broad-

band condition, suggesting that pooling across frequency is indeed important for this task (Kopco et al., 2003).

Conclusion

The combination of experimental physiology and modeling described here has provided new insights into the processing of ITDs in the central auditory system. The dependence of the best delay distribution on BF demonstrated here in the cat IC generalizes similar results from the guinea pig across species and head size. Simulations of ITD acuity for broadband noise suggest that the ITD code involves processing beyond the simple cross-correlation model and that pooling of information across neurons is an essential feature of the code. Finally, simulations of ITD discrimination for tonal stimuli suggest that ITD tuning is refined by a mechanism beyond a system of internal propagation delays.

References

- 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.
- Badcock R, Schor C (1985) Depth-increment detection function for individual spatial channels. *J Opt Soc Am A* 2:1211–1216.
- Bonham BH, Lewis ER (1999) Localization by interaural time difference (ITD): effects of interaural frequency mismatch. *J Acoust Soc Am* 106:281–290.
- Brand A, Behrend O, Marquardt T, McAlpine D, Grothe B (2002) Precise inhibition is essential for microsecond interaural time difference coding. *Nature* 417:543–547.
- Carney LH (1993) A model for the responses of low-frequency auditory nerve fibers in cat. *J Acoust Soc Am* 93:401–417.
- Colburn HS (1973) Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination. *J Acoust Soc Am* 54:1458–1470.
- Darling AM (1991) Properties and implementation of the gammatone filter: a tutorial. In: *Speech hearing and language*, pp 43–61. London: University College.
- Domnitz RH, Colburn HS (1977) Lateral position and interaural discrimination. *J Acoust Soc Am* 61:1586–1598.
- Fitzpatrick DC, Kuwada S (2001) Tuning to interaural time differences across frequency. *J Neurosci* 21:4844–4851.
- Fitzpatrick DC, Kuwada S, Batra R (2000) Neural sensitivity to interaural time differences: beyond the Jeffress model. *J Neurosci* 20:1605–1615.
- Geisler WS, Albrecht DG (1997) Visual cortex neurons in monkeys and cats: detection, discrimination, and identification. *Vis Neurosci* 14:897–919.
- Green DM, Swets JA (1988) *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula.
- Joris PX, van der Heijden M, Louage D, Van de Sande B, Van Kerckhoven C (2004) Dependence of binaural and cochlear “best delays” on characteristic frequency. In: *Auditory signal processing: physiology, psychoacoustics, and models* (Pressnitzer D, de Cheveigne A, McAdams S, Collet L, eds), pp 396–402. New York: Springer.
- Kopco N, Lane CC, Shinn-Cunningham BG (2003) Spatial unmasking of chirp trains in a simulated anechoic environment: behavioral results and model predictions. *Assoc Res Otolaryngol Abstr* 26:541.
- Lane CC, Kopco N, Delgutte B, Shinn-Cunningham BG, Colburn HS (2004) A cat’s cocktail party: psychophysical, neurophysiological, and computational studies of spatial release from masking. In: *Auditory signal processing: physiology, psychoacoustics, and models* (Pressnitzer D, de Cheveigne A, McAdams S, Collet L, eds), pp 341–347. New York: Springer.
- Lehky SR, Sejnowski TJ (1990) Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *J Neurosci* 10:2281–2299.
- Macpherson EA, Middlebrooks JC (2002) Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *J Acoust Soc Am* 111:2219–2236.
- McAlpine D, Jiang D, Shackleton TM, Palmer AR (1998) Convergent input from brainstem coincidence detectors onto delay-sensitive neurons in the inferior colliculus. *J Neurosci* 18:6026–6039.
- McAlpine D, Jiang D, Palmer AR (2001) A neural code for low-frequency sound localization in mammals. *Nat Neurosci* 4:396–401.
- Mossop JE, Culling JF (1998) Lateralization of large interaural delays. *J Acoust Soc Am* 104:1574–1579.
- Mountcastle VB, LaMotte RH, Carli G (1972) Detection thresholds for stimuli in humans and monkeys: comparison with threshold events in mechanoreceptive afferent nerve fibers innervating the monkey hand. *J Neurophysiol* 35:122–136.
- Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci* 21:227–277.
- Patterson R, Nimmo-Smith I, Holdsworth J, Rice P (1988) Implementing a gammatone filter bank. SVOS final report: the auditory filter bank. Medical Research Council Applied Psychology Unit contract report 2341.
- Poggio GF, Fischer B (1977) Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J Neurophysiol* 40:1392–1405.
- Shackleton TM, Palmer AR (2004) Sensitivity to changes in interaural time difference and interaural correlation in the inferior colliculus. In: *Auditory signal processing: physiology, psychoacoustics, and models* (Pressnitzer D, de Cheveigne A, McAdams S, Collet L, eds), pp 313–319. New York: Springer.
- Shackleton TM, Meddis R, Hewitt MJ (1992) Across frequency integration in a model of lateralization. *J Acoust Soc Am* 91:2276–2279.
- Shackleton TM, Skottun BC, Arnott RH, Palmer AR (2003) Interaural time difference discrimination thresholds for single neurons in the inferior colliculus of guinea pigs. *J Neurosci* 23:716–724.
- Shamma SA, Shen NM, Gopalaswamy P (1989) Stereausis: binaural processing without neural delays. *J Acoust Soc Am* 86:989–1006.
- Skottun BC, Shackleton TM, Arnott RH, Palmer AR (2001) The ability of inferior colliculus neurons to signal differences in interaural delay. *Proc Natl Acad Sci USA* 98:14050–14054.
- Stern RM, Zeiberg AS, Trahiotis C (1988) Lateralization of complex binaural stimuli: a weighted-image model. *J Acoust Soc Am* 84:156–165.
- Wightman FL, Kistler DJ (1992) The dominant role of low-frequency interaural time differences in sound localization. *J Acoust Soc Am* 91:1648–1661.
- Yin TC, Chan JC, Irvine DR (1986) Effects of interaural time delays of noise stimuli on low-frequency cells in the cat’s inferior colliculus. I. Responses to wideband noise. *J Neurophysiol* 55:280–300.