Predicting Lateralization Performance at High Frequencies from Auditory-Nerve Spike Timing

by

Anna Alexandra Dreyer
B.S. Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2003

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements of the Degree of

Master of Engineering in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

May 6, 2005 [June 2005]

Author _____
Department of Electrical Engineering and Computer Science
May 6, 2005

Certified by _____
Dr. Bertrand Delgutte
Thesis Supervisor

Certified by _____
Dr. Andrew Oxenham
Thesis Supervisor

Accepted by _____
Arthur C. Smith
Chairman, Department of Committee on Graduate Theses

Predicting Lateralization Performance at High Frequencies from Auditory-Nerve Spike Timing

by

Anna A. Dreyer

Submitted to the

Department of Electrical Engineering and Computer Science

May 6, 2005

In Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science

ABSTRACT

Psychophysical sensitivity to interaural time differences (ITD) in the envelope of high-frequency sinusoidally amplitude-modulated (SAM) tones is generally poorer than that to low-frequency pure tones (PT). ITD sensitivity at high frequencies might be improved using "transposed stimuli" (TS), which seek to produce the same temporal discharge patters in high-frequency neurons as in low-frequency neurons for PT. Here, we study ITD sensitivity for PT, SAM tones and TS using neurophysiology, psychoacoustics and computational models. Phase locking of auditory-nerve fibers in anesthetized cats was characterized using both the synchronization index and autocorrelograms. With both measures, phase locking is stronger for PT than TS, and for TS than for SAM tones. Phase locking to SAM tones and TS degrades with increasing stimulus level, while remaining more stable for PT. ITD discrimination was measured in humans for stimuli presented either in quiet or with band-reject noise intended to restrict listening to a narrow frequency band. Performance improves slightly with increasing stimulus level for all three stimuli both with and without noise. ITD sensitivity to TS is comparable to PT performance only in the absence of noise. To relate psychophysical performance to auditory-nerve activity, we developed a physiologically-based optimal binaural processor model with delay lines and coincidence detectors. In the no-noise condition, model performance is stable with stimulus level, consistent with psychophysics. However, in the band-reject noise condition, model performance for SAM tones and TS degrades with increasing level. These results have implications for the relative roles of peripheral patterns of activity and the binaural processor in accounting for ITD sensitivity at low versus high frequencies.

Thesis Supervisor: Bertrand Delgutte

Title: Associate Professor of Otology and Laryngology and Health Sciences and Technology, Harvard Medical School

Senior Research Scientist, MIT Research Laboratory of Electronics

---

Thesis Supervisor: Andrew Oxenham

Title: Principal Research Scientist, MIT Research Laboratory of Electronics

ACKNOWLEDGEMENTS

I would foremost like to express my most sincere thanks to my two advisors, Dr. Andrew Oxenham and Dr. Bertrand Delgutte. I thank them for their continued and unyielding support and enthusiasm on this project, their consistent availability and dedication to my education and their encouragement to explore deeper issues. They have been and continue to be wonderful mentors. I have greatly enjoyed and have benefited from my interaction and friendship with the students, researchers and staff of the Eaton-Peabody Laboratory of Massachusetts Eye and Ear Infirmary and the Auditory Perception and Cognition Group at MIT's Research Laboratory of Electronics. From the EPL side, I would like to extend special thanks to Ms. Connie Miller for surgical support with the physiological experimentations.

The past two years as a graduate student in the Health Sciences and Technology Speech and Hearing Biosciences and Technology Program have been my most intellectually stimulating and enjoyable years at MIT and for that I also owe thanks to my classmates and other students in the HST SHBT program. The completion of this thesis marks the start of my career as an auditory scientist, to which I look forward a great deal. During this intellectual quest and always, I am most thankful to my greatest sources of unconditional love, support and encouragement: my parents, Elizabeth Chernin and Alexander Dreyer and my fiancé, Richard Hanna.

1. Introduction

Interaural time differences (ITDs) serve two important functions in binaural hearing: localization of sound sources and detection of sounds in noisy environments. ITDs are the dominant cue in localization of sources containing low-frequency components, whereas interaural level differences (ILD) serve as the primary cue for localizing high-frequency sources (e.g., MacPherson and Middlebrooks, 2002). Although listeners are sensitive to ITDs in the envelope of high-frequency sounds, both ITD discrimination performance and the extent of lateralization are poorer for high-frequency sinusoidally amplitude-modulated (SAM) tones than for low-frequency pure-tones. Bernstein (2001) suggested that phase locking at high frequencies (and therefore ITD sensitivity) might be improved by using transposed stimuli formed by modulating a high-frequency carrier by a half-wave rectified sinusoid (van de Par and Kohlrausch, 1997). Transposed stimuli are designed to produce the same temporal discharge patterns in high-characteristic frequency (CF) neurons as those observed in low-CF neurons to pure-tone stimuli. Transposed stimuli might also be useful in cochlear implants, where low-frequency information must be transmitted to high-frequency auditory-nerve fibers (ANFs), because stimulating electrodes can only be inserted into the basal (high-frequency) cochlear region.

While indirect psychophysical evidence suggests that transposed stimuli may be at improve performance in high-CF neurons, there have been no studies of AN responses to this novel class of stimuli. Phase locking may degrade at higher levels where high-CF fibers responding to transposed stimuli begin to saturate. We therefore examined hase locking and lateralization performance of high-frequency and low-frequency stimuli, with special emphasis on level-effects, using physiology, psychophysics and a computational model. By recording

from auditory nerve fibers in cats, we compared responses to pure tones, SAM tones and transposed stimuli. Using human psychophysics, we studied the level dependence of ITD sensitivity for pure tones and transposed stimuli. Lastly, we developed a computational model to relate physiological findings to psychophysics by evaluating whether human psychophysical performance can be predicted from responses at the level of the auditory nerve.

2. Background

Binaural hearing allows humans to localize sound sources in space and to separate sound sources, such as signals in noise. There are two primary spatial cues for determining the azimuth of a sound: interaural time differences (ITDs) and interaural level differences (ILDs). ILDs arise because the head casts an acoustic shadow that attenuates wavelengths shorter than the size of the head. For humans, this acoustic shadow appears for frequencies above 500 Hz and becomes prominent above 2000 Hz. ITDs stem from noticeable time differences between sounds arriving at each ear. Classic experiments by Rayleigh (1906) suggested that ITDs serve as the primary spatial cue for low-frequency pure tones (below approximately 1500 Hz), and as the frequency of the sound increases, localization performance using ITDs deteriorates.

Two main explanations have been proposed for the deterioration in ITD sensitivity with frequency for pure-tone stimuli. The first explanation is purely acoustic, suggesting that sensitivity to ITDs begins to degrade for pure-tones frequencies for which the maximum delay at one ear approaches a phase lag of 180 degrees, making the localization cues ambiguous. For humans, this occurs for frequencies above 1.5 kHz. For animals with smaller heads, the frequency at which ITD cues become ambiguous is lower. However, this explanation is contradicted by evidence that barn owls are able to localize sounds using ITDs in the fine structure at frequencies as high as 7 kHz, but whose head sizes are smaller than those of humans (Knudsen, 1979).

The second explanation suggests that the degradation of ITD sensitivity at high frequencies results from loss of phase locking in the auditory nerve and higher auditory centers. The limit of phase locking in the AN for pure tones is about 5 kHz in cats (Johnson, 1980), more than enough to account for the behavioral limit of ITD sensitivity, about 2 kHz for this species

(Wakeford and Robinson, 1974). ANFs in the barn owl, an animal that can utilize ITDs at high frequencies, phase lock to pure tones of up to 10 kHz (Koppl, 1997). ANFs project to the spherical bushy cells (SBCs) in the anteroventral cochlear nucleus (AVCN), which in turn, provide excitatory input to the medial superior olive (MSO) where ITD processing first occurs (Smith, et al., 1993; Helfert and Aschoff, 1997). The SBCs largely preserve the phase locking limit of the AN (Rhode and Smith, 1986), with the help of specialized giant synapses (Rouiller et al., 1986), fast glutamate receptors (Raman et al., 1994) and low threshold K+ channels (Manis and Marx, 1991, Trussell, 1999). Phase locking in the SBCs is even enhanced somewhat relative to AN below 1 kHz (Joris et al., 1994). Although no systematic data exist on the upper limit of phase locking in the MSO, findings that lateralization (Wakeford and Robinson, 1974) and ITD sensitivity (Yin and Chan, 1990) for pure-tones of 2 kHz are possible in cat suggest that the phase locking limit in the MSO is at least as high. Although the phase locking limit in the AN is higher than the limit of lateralization sensitivity for pure-tones, it is possible that the extra margin is needed in the face of jitter inherent in the transmission through successive synapses. In exploring the neurophysiological basis for ITD sensitivity at the level of the auditory nerve, we make the central assumption that phase locking is necessary for good lateralization performance, with the degradation of phase locking consistent with mammalian degradation in ITD sensitivity to high-frequency pure tones.

The use of ITDs in localization tasks is not restricted to low-frequency stimuli, however. Several studies have shown that, although interaural time differences in the fine structure of high-frequency stimuli cannot be detected, the auditory system can exploit the ITDs present in the *envelope* of high-frequency stimuli to lateralize sound (Leakey et al., 1958; Henning, 1974; McFadden and Pasanen, 1976). However, the perceptual ability to use ITDs in high-frequency

envelope cues seems weaker than that produced by low-frequency fine-structure cues. For instance, the extent of laterality (extent to which image is displaced from the midline) produced by sinusoidal amplitude modulated (SAM) tone is typically less than that measured for low-frequency pure-tones of the same frequency as the modulation (Bernstein and Trahiotis, 1985).

The jitter in synaptic transmission and inner-hair cell membrane time constant account for the main limitations in phase locking in AN for pure-tones. Compared with monaural pathways, the problem of jitter is larger in binaural pathways since the binaural processor needs to decode interaural delays in the presence of jitter in each ear. An additional factor also limits phase locking to the envelope of high-frequency complex sounds. High-frequency stimuli, such as AM tones, have side bands. The distance in frequency between the main band at the carrier frequency and the side bands equals the modulation frequency (f_m). For an AM tone centered at CF, as modulation frequency increases, the two side bands begin to fall outside the passband of the auditory filter, leaving only the high-frequency carrier and reducing or even eliminating phase locking to the envelope. This explanation is consistent with Joris and Yin's (1992) findings of a positive correlation between the upper limitation of modulation sensitivity and CF for fibers with CF's below 10 kHz in the auditory nerve. However, another temporal mechanism, similar to the limitation of phase locking to pure tones, appears to limit phase locking at high modulation rates for higher CF fibers (Joris and Yin, 1992).

ITD sensitivity has been documented in both the MSO and lateral superior olive (LSO) nuclei of the superior olivary complex (SOC) for both pure tones and amplitude modulated stimuli (Goldberg and Brown, 1969; Joris and Yin, 1995; Joris, 1996; Batra et al., 1997). Current view on binaural processing in the low-frequency MSO suggests that MSO cells operate as coincidence detectors, discharging when receiving coincident spikes from excitatory monaural

inputs that contain the temporal information about the stimulus waveform. Support for the coincidence detection mechanism stems from (1) the responses of MSO cells that are strongly modulated by ITD (Yin and Chan, 1990), (2) the invariance of best ITD with respect to pure-tone frequency (Rose et al, 1966; Yin and Kuwada, 1983, Joris et al., 1998) and (3) the predictability of the best interaural phase from response phases to monaural stimulation of each ear (Goldberg and Brown, 1969; Yin and Chan, 1990). However, controversy surrounds the mechanism with which ipsilateral and contralateral inputs are delayed, with axonal delay lines (Smith et al., 1993), and inhibition (Brand et al., 2003) having been suggested. In contrast, cells in the mostly high-frequency LSO operate via the excitation/inhibition mechanism, whereby an ipsilateral stimulus excites and the contralateral stimulus inhibits the LSO cell (Joris and Yin, 1995). The inhibitory input to the LSO is precisely-timed, short-duration, and glycine-mediated (Joris and Yin, 1995; Joris, 1996), allowing this inhibition to mediate ITD sensitivity.

As suggested by Colburn and Esquissaud (1976), the relatively poor resolution of ITDs in high-frequency stimuli is either due to differences in the temporal properties of the spike train after peripheral processing of the signals, differences in how the central nervous system processes low and high-frequency signals or a combination of the two. Van de Par and Kohlrausch (1997) hypothesized that differences in lateralization sensitivity between SAM and pure-tones may be attributed to the difference in "the internal representation after transformation in the inner hair cells" for the two stimuli. As a first order approximation, this transformation can be quantified as half-wave rectification followed by low-pass filtering. The low-pass filtering is attributed to both the inner hair cell membrane time constant and the jitter in the synapses between the hair cell and the auditory nerve. Half-wave rectification arises from the

inner hair cell voltage versus displacement characteristic, as shown in Figure 1, and also from transformation of receptor potential into spikes as there are no "negative spikes".

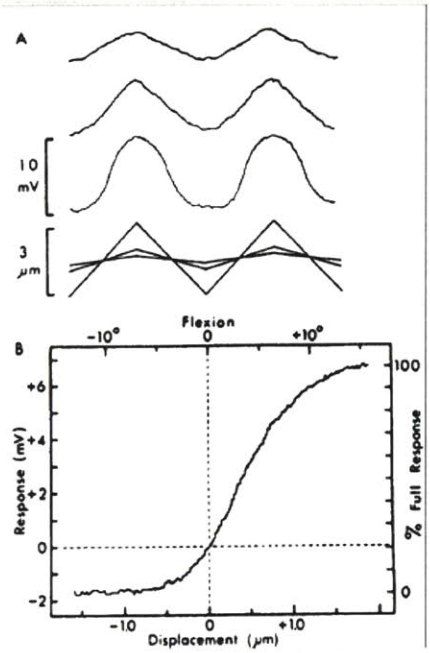

Figure 1: Inner hair cell receptor potential as a function of ciliary bundle displacement (Hudspeth and Corey, 1977). For small ciliary bundle displacements, the response is approximately linear. However, for larger displacements, only the response to positive displacements is linear while the response to negative displacements is compressive.

Hair cell transduction channels are partially open even in the absence of sound, indicated in Figure 1 as the operating point, or point at which there is no ciliary displacement. The IHC transduction input/output function is relatively linear for small positive and negative displacements, but becomes compressive in both directions for larger displacements. However, the operating point is located closer to the compressive region for negative displacements (away from the kinocillium) than for positive displacements (toward the kinocillium), resulting in approximately a half-wave transformation.

Whereas the IHCs approximately transform a pure-tone into a half-wave rectified sine wave, they transform a high-frequency SAM tone to a sinusoid, where the carrier frequency (f_c) of the SAM tone is above the cutoff of the lowpass filter. Therefore, van de Par and Kohlrausch suggested a stimulus, called "transposed stimulus", designed to produce the same auditory discharge patterns in the ANF as a pure-tone stimulus.

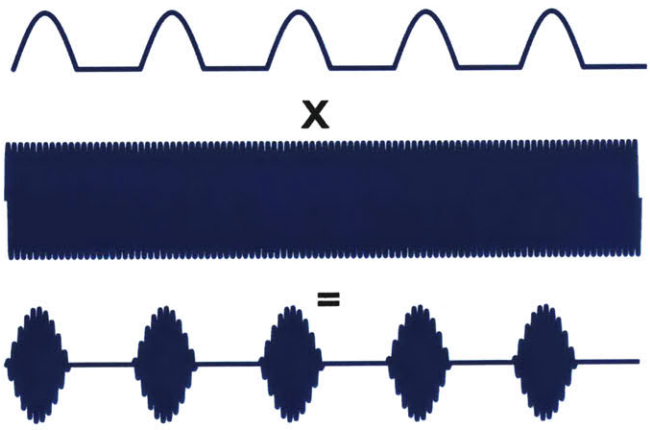

Figure 2: Schematic depiction of creation of the transposed stimulus. Transposed stimuli are created by multiplying a fast sinusoidal carrier by a slow half-wave rectified sinusoid.

As shown in Figure 2, a sinusoidal carrier is multiplied by a half-wave rectified low-frequency tone, producing a high-frequency transposed stimulus with the envelope of the low-frequency half-wave rectified sine wave. After peripheral processing, including low-pass filtering and approximate half-wave rectification by the inner hair cells, the response to transposed stimuli is expected to resemble half-wave rectified sinusoids and theoretically emulates pure-tones, as shown in Figure 3.

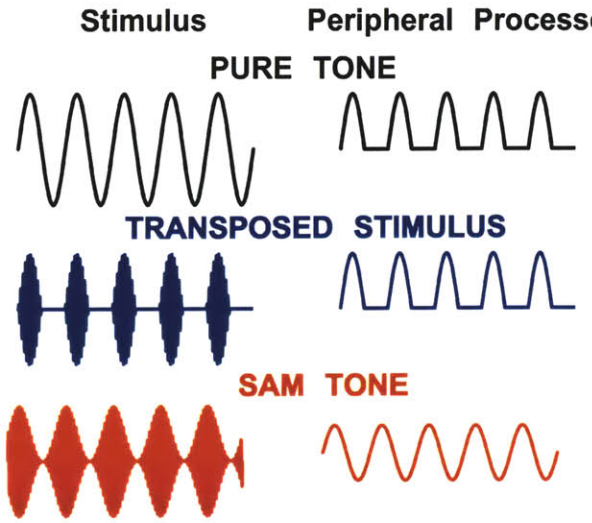

Figure 3: Inputs and outputs of peripheral auditory processor for pure-tone, SAM-tone and transposed-tone stimuli. Adapted from Bernstein, 2001.

Using transposed stimuli, Bernstein (2001) and Bernstein and Trahiotis (2002) showed evidence that poorer sensitivity to ITDs at high frequencies result from differences in the internal stimuli representation produced by the peripheral processor for low modulation rates. They found that high-frequency transposed stimuli produced extents of laterality and ITD sensitivity comparable to those in response to pure tones for modulation frequencies below 150 Hz. Lateralization performance using transposed stimuli markedly improved compared to performance using SAM tones for low modulation frequencies. However, as modulation frequency was increased, lateralization performance in response to transposed stimuli degraded below pure tone performance with a frequency equal to the transposed modulation frequency, suggesting that additional factors beyond peripheral representation also play a role.

Although there are few little physiological data on responses to transposed stimuli, we hypothesize that the coding of transposed stimuli in the auditory nerve will differ from the coding of pure-tones. As shown in Figure 1, the displacement to voltage transfer function of the inner hair cell does not result in perfect half-wave rectification. Consequently, one would expect pure-tones to have small "negative components" (instantaneous rate below spontaneous) after peripheral processing, contrary to the idealized model used for motivating transposed stimuli. Additionally, phase locking to SAM tones has been shown to vary non-monotonically with SPL due to cochlear compressive nonlinearities (Smith, 1988; Joris and Yin, 1992). Given the AM nature of the transposed stimulus, similar level-dependent phase-locking may be expected. Therefore, the first (physiological) portion of this study seeks to compare the phase locking to transposed stimuli and pure-tones, with a particular emphasis on the level dependence of phase locking. Second, a psychophysical study is conducted to test whether the expected non-monotonicity of phase locking with level affects lateralization performance for transposed stimulus even when the listening band is restricted by masking noise to only include neurons near the f_c of the stimulus. In the unrestricted listening condition, fibers with CFs far from f_c could maintain good performance at high levels where fibers near f_c are saturated. The restricted listening condition removes this possibility. Third, we developed a model of the binaural processor to explore the feasibility of predicting psychophysical lateralization performance from auditory nerve spike timing.

3. Phase locking of AN fibers to pure tones, SAM tones and transposed stimuli

3.1. Methods

3.1.1. Animal surgery

Cats with a normal eardrum and healthy middle ear were anesthetized with Dial in urethane (75 mg per kg of body weight) and maintained in an anesthetized state as judged by the toe-pinch reflex. The auditory nerve was exposed by removing the posterior fossa and retracting the cerebellum. The bullae of the middle-ear cavities were opened and the round window was exposed. The cat was administered 10 cc per pound of lactated Ringer's solution, with an additional 10 cc booster per day, to prevent dehydration and dexamethasone (.26 mg per kg of body weight) as a prophylactic against brain edema throughout the experiment. Body temperature was maintained at 37° with a heating blanket.

3.1.2. Acoustic chamber and measurements

Recordings were conducted with the animal placed on a vibration-attenuating table in an electrically shielded, temperature-controlled, sound-proof chamber. A metal electrode at the round window recorded the compound action potential (CAP) in response to click stimuli to assess the condition and stability of the cat's cochlear response. The round-window response was monitored throughout the experiment.

An electrodynamic speaker (Realistic 40-1377) delivered sound into the cat's ear through a closed acoustic assembly, calibrated for accurate delivery of sound-pressure level at the tympanic membrane. Stimuli were delivered with a NIDAC 6052e 16-bit analog I/O board converter with a sampling rate of 50 kHz. The frequency response of the acoustic system was equalized by measuring sound pressure levels at the tympanic membrane as a function of frequency and designing digital inverse filters.

Auditory nerve activity was recorded with glass micropipettes filled with 2M KCl solution. The electrode was inserted into the auditory nerve with the aid of a microscope and then mechanically advanced with a micropositioner (Kopf 650). The microelectrode signal was bandpass filtered and sent to a spike detector which records spike arrival times with an accuracy of 1 μs.

A click stimulus at 55 dB SPL was played while advancing the electrode into the nerve. Upon locating a responsive fiber, a threshold tuning curve was measured with an automatic tracking procedure, (Kiang and Moxon, 1974) and the CF and minimum sound pressure level to excite the unit at the CF, the unit's threshold, were noted. The fiber's rate of response in the absence of sound, the spontaneous rate (SR), was measured over 20 seconds.

3.1.3. Stimulus generation

Low-frequency sinusoids and high-frequency SAM tones and transposed stimuli were synthesized digitally. High-frequency transposed stimuli were generated as described by van de Par and Kohlrausch (1997) and Bernstein and Trahiotis (2002), as shown in Figure 2. Pure-tone frequencies (f) and the modulation frequencies (f_m) of both SAM and transposed stimuli were 60 Hz, 125 Hz, 250 Hz, 500 Hz or 1000 Hz. The f_c of SAM tones and transposed stimuli was usually set at the fiber's CF as shown in Figure 4. Occasionally f_c was set slightly below or above CF.

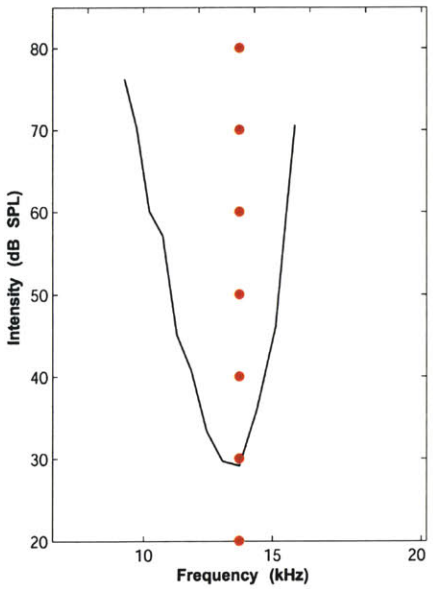

Figure 4: Example tuning curve for an AN fiber shown with overlapped response measurements. The tuning curve for an average high-frequency fiber is shown in black. The frequency of the tip of the tuning curve is the characteristic frequency and the level of the tip is the fiber's threshold. On-CF measurements are shown in red (middle of the tuning curve) and the dots represent the levels at which stimuli are presented.

Fibers of low-CF were utilized to measure responses to pure-tones, and responses to transposed stimuli and SAM tones were recorded from high-CF fibers (CF>3 kHz above the limit of phase locking). Responses were measured as a function of stimulus level from zero to a maximum of 60 dB above threshold in 10 dB steps. Sound pressure levels did not exceed 85 dB SPL so as not to injure the cat's cochlea. Each amplitude modulated stimulus presentation consisted of 20 repetitions per level of a 480-ms of SAM tone, followed by 520 ms of silence, 480 ms of transposed tone at the same modulation frequency and another 520 ms of silence. Pure-tone stimuli were presented as 20 repetitions per level of 480 ms of pure-tone followed by 520 ms of silence. The stimulus presentation times include 20-msec rise and fall times using raised cosine ramps to prevent abrupt transients.

3.1.4. Analysis

Spike discharges do not occur on every cycle of the stimulus waveform, but tend to occur at a particular phase of the stimulus cycle (Javel et al., 1968 for pure tones). To quantify phase locking to the stimulus waveform, period histograms, locked to the modulation period, or, for pure tones, to the tone frequency, were computed from spike arrival times. Period histograms display the distribution of spikes within a stimulus cycle. The synchronization index (SI) was computed from the period histograms to characterize the strength of phase locking. SI is the ratio of the fundamental frequency component of the period histogram to the average firing rate. Also known as vector strength, the SI varies from 0 for a flat histogram with no phase locking, to 1 for an impulsive histogram with perfect phase locking.

To obtain a more general description of temporal behavior, phase locking was also characterized with shuffled autocorrelograms (SACs) (Louage et al., 2004). While the synchronization index is calculated at a predetermined frequency, SACs do not require knowledge of the stimulus, and thus complement the period-histogram based analysis. Temporal analysis using SACs compares the temporal properties of spike trains across different presentations of the stimu lus. Each SAC was computed from the N=20 stimulus presentations at each level. Spike trains from each stimulus presentation were paired, or shuffled, with the N-1 others to create $N*(N-1)/2$ distinct pairs of spike trains, as shown in Figure 5.

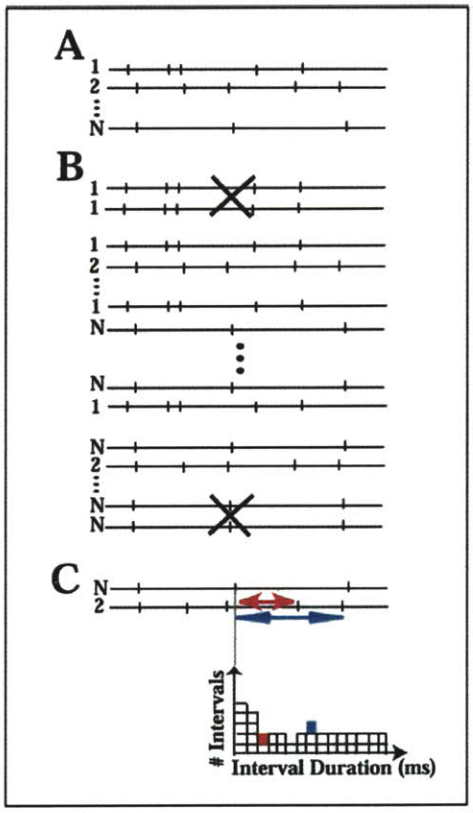

Figure 5: Illustration of shuffled autocorrelogram construction (adapted from Louage et al., 2004). A: Spike trains to several presentations of the stimulus. B: All possible pairs of spike trains. Crosses indicate identical pairs which are not included in SAC computation. C: Measurement of all order intervals and construction of the shuffled autocorrelogram.

For each pair, the forward and backward time intervals between all spikes from the first spike train and the spikes from the second spike train were calculated. All intervals were tallied into a histogram. To eliminate histogram and fiber-dependent information from the SACs, the histograms were normalized by the fiber's average firing rate r_0, histogram bin width b, recording duration D, and number of stimulus presentations N by dividing by $N(N-1)\ r_0^2 bD$. The ordinate axis of the normalized SAC thereby measures the extent to which the fiber temporally codes the stimulus waveform, with a unity indicating the absence of any such coding (Louage et al., 2004). This analysis uses histogram bin widths of 50 μs as suggested by Louage et al. (2004).

Two measures of phase locking are derived from SACs: the maximum peak height of the SAC and the peak half-width (as described in Louage, 2004). The peak height measures the extent to which spikes occur at the same intervals on repeated presentations of the same stimulus relative to the interval expected for a random spike train. Peak half-width measures the temporal precision, or spread, of the arrival times of the spikes around the peak. Unlike the SI, these interval-based measures reflect the increased variability of coincidence detection of spike timings arriving from the two ears and are therefore a particularly suitable measure in this context of using physiological data in a crosscorrelation based model of the optimal binaural processor.

The SI and SAC analysis were applied to all SAM tones, transposed stimuli and pure-tone data. This analysis was then used to characterize how phase locking depends on sound level for each stimulus type and whether the modulation frequency or a fiber's spontaneous rate affect the fiber's ability to phase lock.

4.2. Physiological Results

4.2.1. Comparison of phase locking to pure tones, SAM tones and transposed stimuli

The data includes recordings from 130 auditory nerve fibers in 9 healthy adult cats. Responses to SAM tone, transposed stimulus are collected from 40 fibers in 4 cats and include 70 stimulus presentations of SAM tone and transposed stimulus pairs presented to the same fiber with the same f_m. Because we only had limited pure-tone data from those 4 cats, previously recorded pure-tone response data collected by Tsai and Delgutte (1997) were also used from an additional 5 cats. Period histograms were computed for all responses, and synchrony-level functions were calculated from the histograms. An example of period histograms in response to pure tone, transposed stimulus and SAM tone are shown in the left portion of Figure 6. In this

example, the modulation frequency of the transposed stimulus and SAM tone responses approximately matches the frequency of the pure tone stimulus.

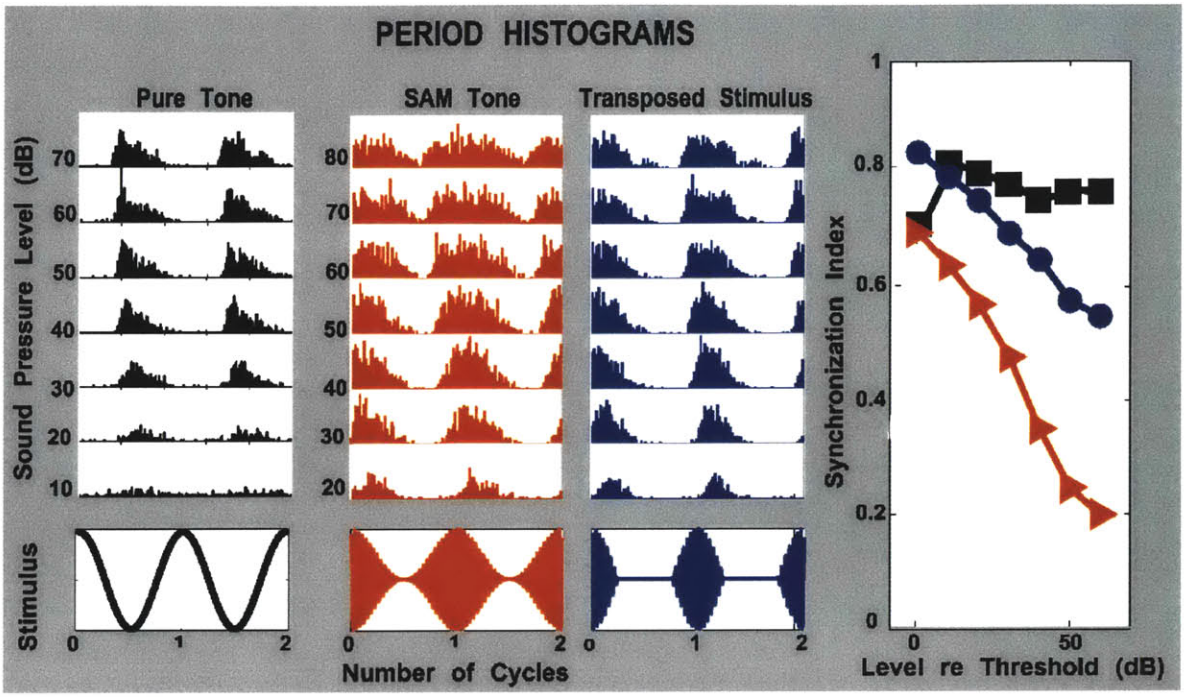

Figure 6: Bottom left: Pure tone with frequency of 354 Hz, SAM tone and transposed stimulus with a modulation frequency of 250 Hz and a carrier frequency of 7000 kHz. Pure tone response recorded from low-frequency fiber with CF=385 Hz, SAM tone and transposed stimuli recorded from high-frequency fiber with fc=CF=7000 Hz. Top left: Period histograms for levels from 10-70 db SPL (pure tone) 20-80 dB SPL (SAM tone and transposed stimulus) in response to each stimulus. Right: Synchronization index calculated from each period histogram for 3 stimuli as a function of level.

At low stimulus levels, responses to both high-frequency stimuli are restricted to approximately half of the stimulus cycle, indicating a high degree of phase locking. As level increases, spike discharges begin to spread to other portions of the stimulus cycle, indicating a decrease in phase locking, although the spread is less severe for transposed stimuli than SAM tones. For each level, responses to transposed stimulus are restricted to a smaller part of the discharge cycle than for SAM tones.

As Figure 6 illustrates, auditory nerve fibers can precisely phase lock to pure tone stimuli even when the stimulus level falls within the saturated portion of the rate-level function (Johnson,

1980). The rate-level function (not shown) for this low-CF fiber saturates at 40 dB SPL, while phase locking remains high for higher stimulus levels. For the low-frequency fiber response to pure tone shown, the fiber's response is restricted to virtually the same portion of the stimulus discharge cycle at threshold (20 dB SPL) and at 60 dB above threshold (80 dB SPL). This fiber phase locks to the stimulus waveform over a larger range than the rate-based dynamic range of this high-SR fiber.

These patterns of response are reflected in the synchrony level functions derived from the period histograms (Figure 6, right panel). For each stimulus level, synchrony is greater for transposed stimuli than for SAM tones. The SI is highest near threshold, and as stimulus levels increase, synchrony to both high-frequency stimuli falls. Synchrony to the pure tone stimulus also reaches a peak near threshold, but synchrony remains fairly stable as levels increase. However, near threshold, fibers show similar values of synchrony to pure tones and transposed stimuli.

Figure 7, shows synchrony-level functions for pure tones, transposed stimuli and SAM tones for the entire data set. Data are shown for all modulation frequencies, and all pure tone frequencies below 1000 Hz, to ensure that phase locking capabilities are maximum. Sound level is expressed in dB re threshold to facilitate comparison between fibers. For the most part, responses are well separated based on the stimulus: low-frequency fibers phase lock with the greatest precision to pure tones, then high-frequency neurons to transposed stimuli, followed by the poorest phase locking to SAM tones by high-frequency neurons. However, near threshold, fibers tend to show equally good phase locking to transposed stimuli and pure tones, indicating that synchrony falls faster with level for transposed stimuli than for pure tones. Synchrony to

SAM tones is poorer near threshold and degrades faster with level than for transposed and pure-tone stimuli.

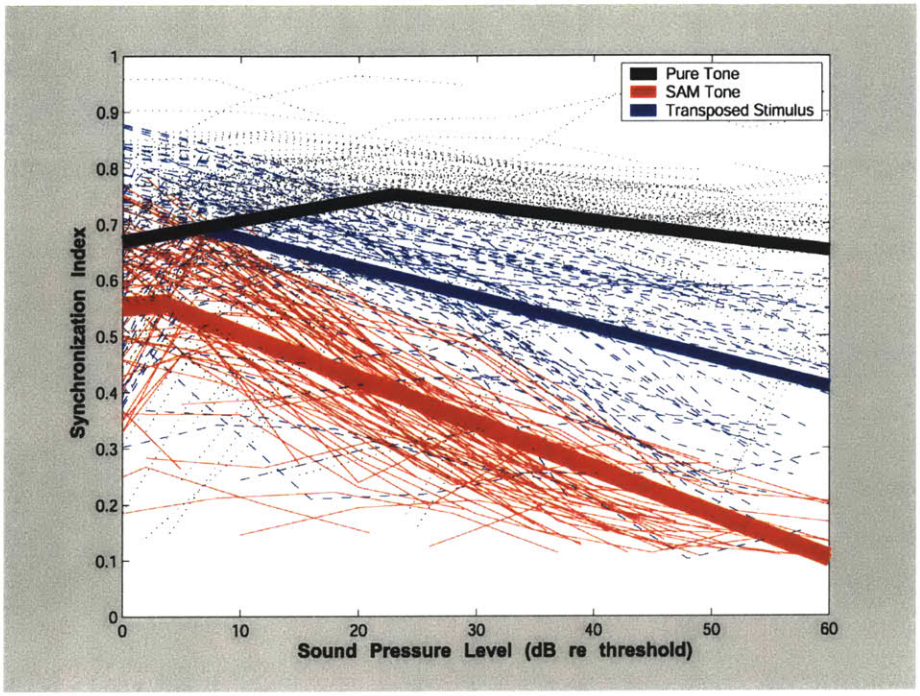

Figure 7: Synchronization Index level functions for fibers in response to pure tones (black dotted), transposed stimuli (blue dashed) and SAM tones (red solid) presented at fiber's CF. Best piece-wise linear fits to mean are shown in bold for each stimulus type. Only responses for pure tones of and below 1000 Hz are included in this data set.

To show general level-dependent trends, we averaged the synchrony-level function for each stimulus and interpolated the average measures at every 1 dB from threshold to 60 dB above threshold, and calculated the best piece-wise linear functions to the mean synchrony-level functions. The best-fit linear functions are also shown in Figure 7 as bold lines. The fitted lines show the same patterns that we observed for the raw data. Piece-wise linear fits to the monotonically-decreasing portions of the SI-level functions predict a slope of -.0027/dB for pure tones (r^2=.99), -.0053/dB for transposed stimuli (r^2=.97), and -.0080/dB for SAM tones (r^2=.96). Analysis of covariance reveals a significant effect of stimulus type (pure tone, SAM tone or transposed stimulus) on the response of the linear-trend of the phase locking responses with level ($p<.001$). The interaction between stimulus type and level above threshold yield statistically different slopes ($p<.001$). Also important to note is not only that phase locking to pure tones falls at least two times slower than that to SAM tones and transposed stimuli, but that phase

locking to transposed stimuli is always greater than phase locking to SAM tones for both the averaged metric and the majority of the data.

Our results suggest that phase locking to transposed stimuli degrades with increasing stimulus level, while phase locking to pure tones remains more stable. In general, transposed stimuli would not be expected to produce the same auditory nerve spike patterns as pure tones even at low frequencies because they do not modulate the instantaneous firing rate around the spontaneous rate as do pure tones. Firing rate is close to SR during the half-wave rectified portion of the transposed stimulus, whereas a pure tone stimulus cause firing with a rate below SR, consistent with Figure 1. Moreover, here we show that cochlear compression affects the envelope differently from the fine structure at high SPLs. In addition, cochlear processing distorts the waveform of a transposed stimulus more than a pure-tone waveform because the energy in transposed stimuli is distributed over several cochlear filters.

Maximum synchrony to the two high-frequency stimuli is expected to vary somewhat with modulation frequency. As modulation frequency increases, the sidebands increasingly fall outside the fiber's tuning curve, decreasing the stimulus power in the tuning curve. We therefore examined the maximum phase locking to each stimulus as a function of modulation frequency. As shown in Figure 7, these maximum values of synchrony occur close to fiber threshold. Figure 8 compares mean pure tone, transposed stimulus and SAM tone maximum synchrony values.

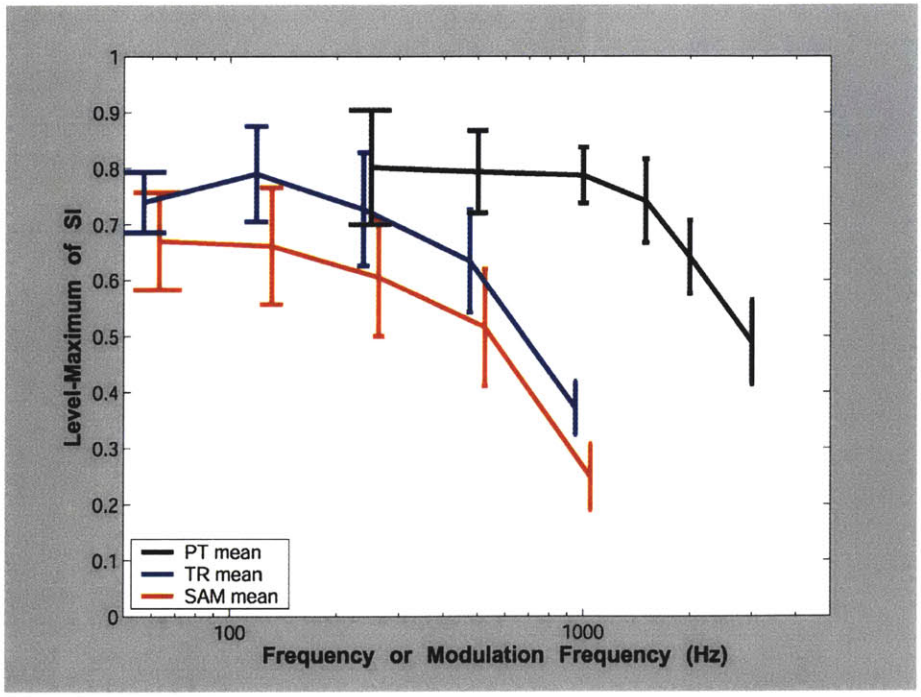

Figure 8: Maximum values of synchronization index as a function of frequency or modulation frequency for pure tones, AM tones and transposed tones. Responses to transposed stimuli are shown in blue and to SAM tones are shown in red with error bars of plus or minus one standard deviation.

For low modulation frequencies (fm=60, 125 and 250 Hz), the level maximum of the SI in transposed stimuli is comparable to that of pure tones below 500 Hz. However, synchrony to transposed-stimuli drops for modulation frequencies of 500 and 1000 Hz. Phase locking to the SAM tones is always lower than phase locking to transposed stimuli, although the rate of decrease of phase locking at high modulation frequencies is similar for both stimuli.

Figure 8 also indicates that maximum phase locking to transposed stimuli begins to degrade at lower modulation frequencies than for comparable pure tone frequencies. Whereas phase locking to pure tones does not start to degrade until 1000 Hz, phase locking to transposed stimuli and SAM tones begins to degrade for modulation frequencies between 250 and 500 Hz. The SAM tone results are comparable to Joris and Yin (1992). Interestingly, Bernstein (2002) finds that lateralization for transposed stimuli is comparable to that of pure tones for modulation frequencies (and pure tone frequencies) below 250 Hz, but begins to degrade for larger

modulation frequencies. As phase locking is essential to interpreting ITD JNDs, these results suggest a possible correlation with psychophysical results at the level of AN.

It is also interesting to observe how synchrony level functions for SAM tones and transposed stimuli compare for the same modulation frequency. Although Figure 8 indicates that the level-maximum of SI decrease by the same amount with fm for transposed stimuli and SAM tones, the level dependence may differ across f_m. Synchrony-level functions for transposed stimuli and SAM tones are shown in Figure 9 for modulation frequencies of 60 Hz and 500 Hz. Data at f_m =500 Hz are shown instead of f_m=1000 Hz because synchrony is severely degraded at the latter modulation frequency.

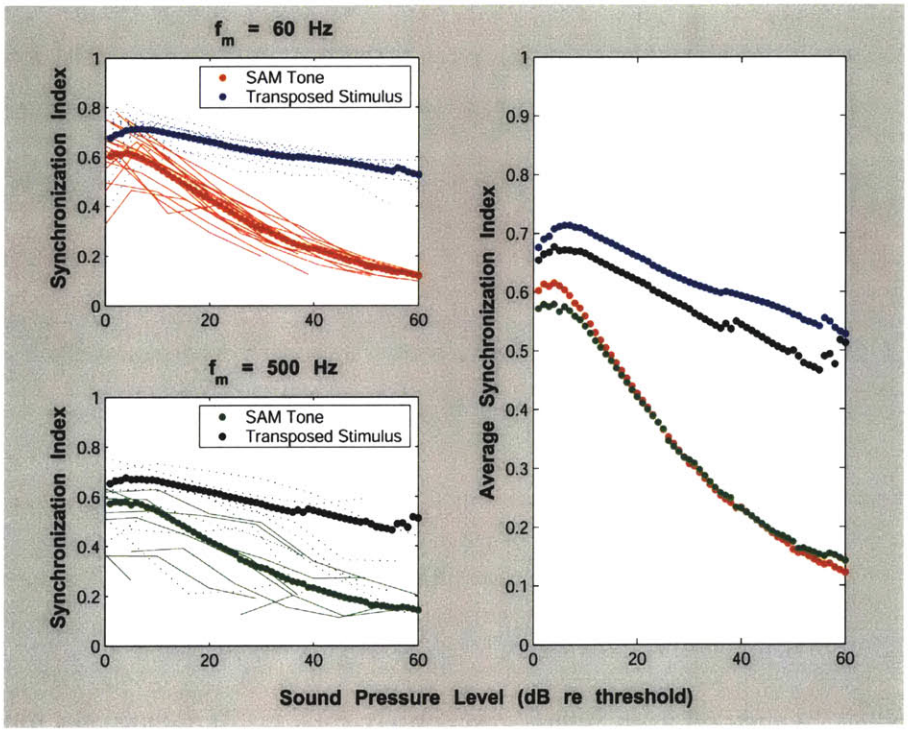

Figure 9: Synchronization index level functions for f_m=60 and f_m=500 Hz. Right panels show responses and averages (bold) to SAM tones (red for f_m =60 Hz and green for f_m =500 Hz) and transposed stimuli (blue for f_m =60 Hz and black for f_m =500 Hz). Left panel compares average SI-level function for SAM tone and transposed stimuli at the two different modulation frequencies.

Data in Figure 9 suggests that the shape of SI-level functions is affected primarily by the stimulus and secondarily by modulation frequency. The right panel shows that even at 500 Hz, the rate of decrease in synchrony is much larger for SAM tones than for transposed stimuli.

Cochlear filtering has been implicated in the fall of maximum synchrony with modulation frequency seen in Figure 8. As modulation frequency of the stimulus increases, the side bands of SAM tones no longer fall within a fiber's tuning curve and synchrony decreases. If cochlear filtering dominated phase locking, we would expect SI-level functions for transposed stimuli to approach those for SAM tones at high modulation frequencies, since all but two sidebands remain within the cochlear filter bandwidth, making the transposed stimulus look like a SAM tone to the fiber. However, the effect of stimulus type seems to dominate even at high modulation frequencies. Thus, Figure 9 suggests that cochlear filtering plays only a small role in shaping synchrony for SAM tones and transposed stimuli, but that cochlear compression plays a larger role.

4.1.2. Comparison of phase locking using synchronization index and autocorrelation-based methods

The synchronization index is an obvious measure of phase locking to pure tones as it measures the ratio of the first Fourier component (at the frequency of the pure tone) to the average firing rate. However, the synchronization index may not necessarily represent phase locking to the fundamental frequency component in complex stimuli. This section provides support for the use of SI by comparing phase locking as measured by SI and autocorrelation based measures.

Using the same data as in Figure 6, we computed the SACs shown in Figure 10.

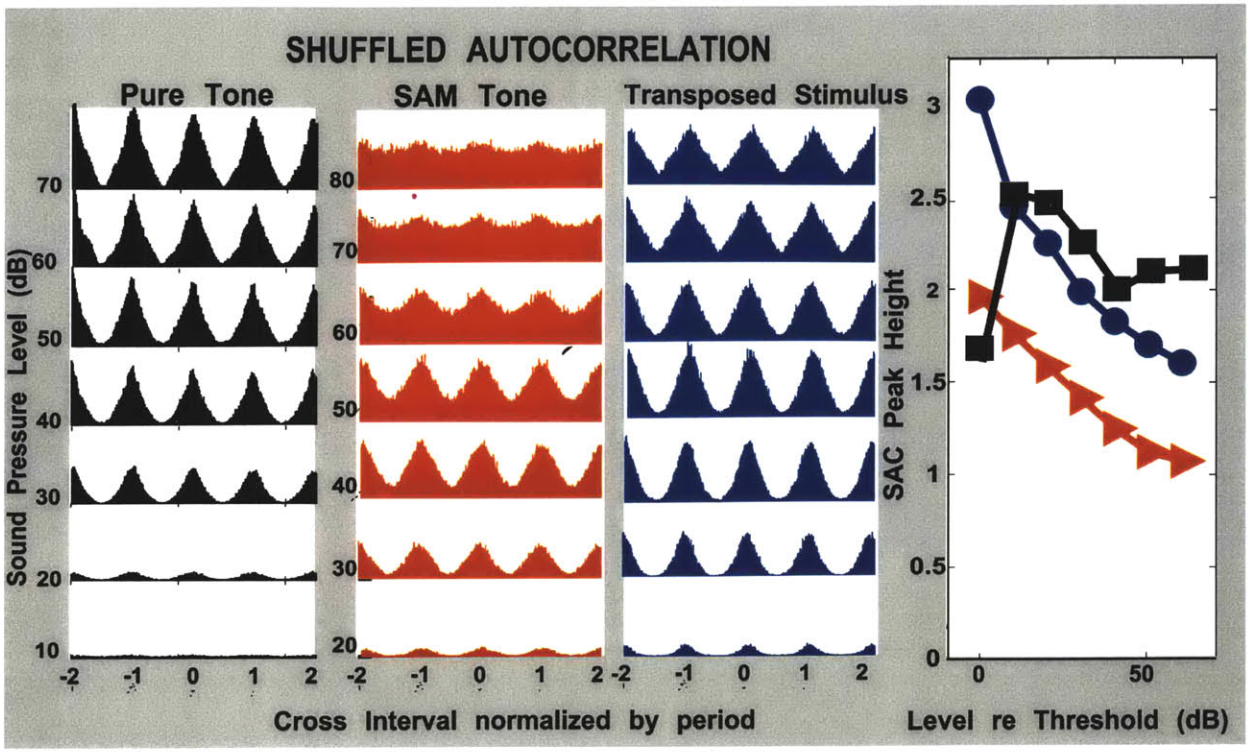

Figure 10: Shuffled autocorrelograms for levels 10-70 dB SPL in response to pure tone, and 20-80 dB SPL in response to SAM tone and transposed stimuli shown in Figure 6. Low-frequency and high-frequency fibers are same as in previous example. Right: Normalized autocorrelation peak height against sound level re. threshold for all three stimuli.

The spread of the SACs for the high-frequency stimuli parallel the spread of period histograms to larger portions of the stimulus cycle as levels increase. To quantify phase locking, we used the peak height metric (Figure 10 right panel) Peak-level functions suggest similar phase locking to pure tones and transposed stimuli near threshold. As stimulus levels increase, peak height for pure tones falls less rapidly than peak height for transposed stimulus. In addition, for all levels, peaks heights calculated from SAM tone responses are always less than those from transposed stimuli. The peak-height level functions show similar patterns to SI-level functions in Figure 6.

Autocorrelation histogram based measures capture the inherent periodicities in the signal and do not necessitate knowing the fundamental frequency or period of the signal. Although Figure 10 showed only the peak-height metric, in general, we inferred phase locking from two autocorrelation based measures as described by Louage et al. (2004): the histogram peak height

and half width. The central peak-height represents the extent to which spikes occur in a certain temporal relationship with respect to each other across stimulus presentations. Since the stimulus is the same for all presentations, discharges in the same temporal relationship tend to correspond to discharges occurring during the same portion of the stimulus discharge cycle. Similarly, the SAC half-width depends on the temporal precision at which spikes occur at certain intervals. For smaller half widths, spikes occur with more temporal precision. Because both measures reflect the periodicity in the response, one would therefore expect the two measures to be empirically related and to share a quasi-inverse relationship. Indeed, Figure 11 shows this inverse relationship of peak height and peak half width. The inverse relationship between half width and half height is clearly seen in Figure 11, with few outliers.

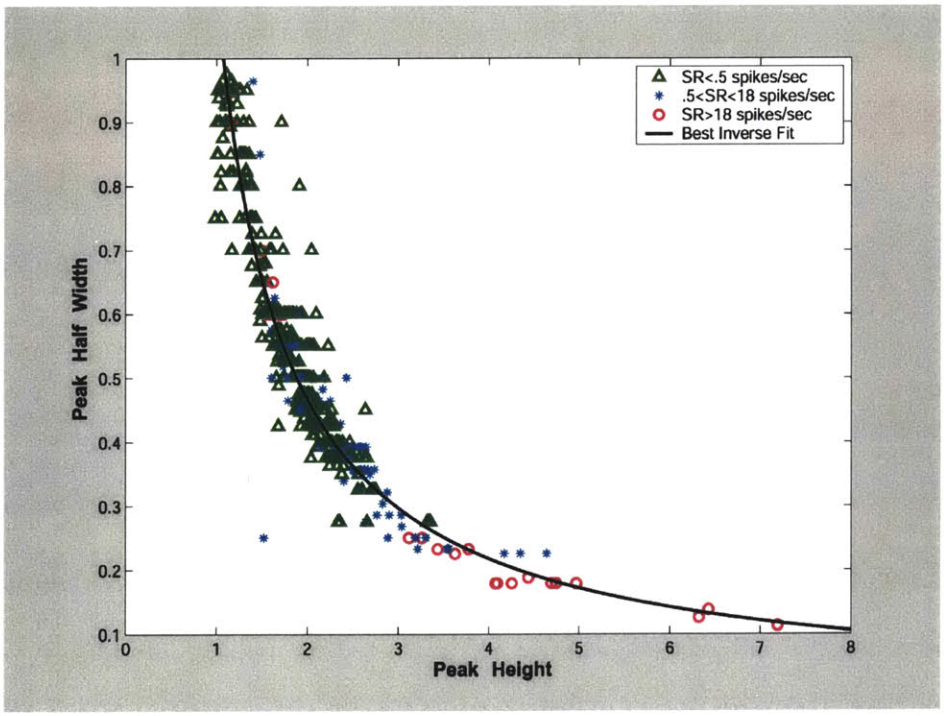

Figure 11: Maximum values of mean autocorrelogram peak width and peak height for all pure-tone data. Mean autocorrelograms were obtained by averaging over common interval multiples. The two phase locking measures are segregated by SR. Magenta circles represent responses of low-SR fibers, cyan stars crosses of medium-SR fibers and green triangles of high-SR fibers as defined by Liberman (1978). Best Inverse function fit is shown in black.

The deviation from the best fit curve most likely arises from the inaccuracies in measuring the peak width. The evidence suggests that one could use either of the two measures derived from SACs to characterize phase locking using interval histogram statistics. For this paper, we have

chosen to characterize phase locking using the peak height measure for two reasons. First, the peak width measure is more difficult to obtain when data are noisy, such as when there are only a few spikes. Second, due to the theoretically unrestricted range of peak heights while the width is limited by zero, the peak height measure could be more sensitive.

Figure 11 shows excellent characterization by an inverse relationship between the peak height and half width. Specifically, the relationship between height and width is well fit by the following function for all pure-tone fibers:

$$HW = \frac{1}{.8149PH - 3.1817} \quad (1),$$

where *HW* represents the peak half width and *PH* represents the peak height. Although the inverse fit describes the height and width relationship well for all SR groups, Figure 11 highlights enhanced phase locking by fibers with lower spontaneous rate. Low-SR fibers phase lock with more precision than medium SR-fibers, that tend to phase lock with more precision than high-SR fibers. This observation is confirmed by analysis of variance that reveals a significant effect of SR on peak height ($p<.001$).

Peak height represents the extent to which spikes occur in the same temporal relationship across stimuli presentations, and unlike the SI, can theoretically reach an infinite value if AN fibers were able to entrain perfectly to the stimulus frequency. SI, on the other hand, would reach an asymptotic value of unity in the entrainment case. Therefore, the degree to which the phase locking capabilities of low-SR fibers is superior to medium and high-SR fibers is shown more clearly using the peak height than the SI as a measure of phase locking. Indeed, some low-SR fibers can reach peak heights of over 7 while medium SR and high-SR fibers have maximum peak heights of less than 5 and less than 4, respectively in this sample. Peak width, similarly to

SI, is normalized to unity, where unity represents a flat SAC, and differences in phase locking ability between fibers of different SRs are not as obviously apparent.

Having selected the peak height measure, we determined to what extent the traditional synchronization index provides similar measures of phase locking. To do so, we used all fibers responding to pure tones, transposed stimuli and SAM tones. For each stimulus, both measures of phase locking are shown as a scatter plot in Figure 12.

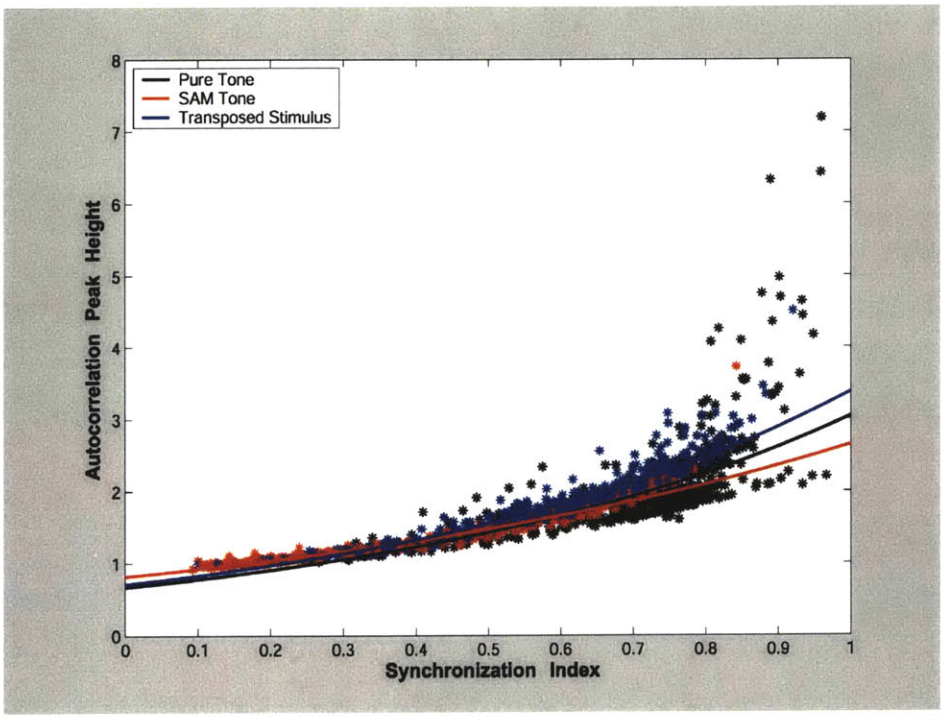

Figure 12: SAC peak height against synchronization indices for all recordings in response to pure tones (black), transposed stimuli (blue) and SAM tones (red). Best fit exponential functions are shown as solid lines for each stimulus. Data from 130 fibers are included.

There is a close relationship between SAC peak height and synchronization index for each stimulus type. Analysis of covariance reveals a significant effect of stimulus type ($p<.001$) in the relationship between peak height and synchrony, with the following descriptions for each stimulus:

$$\text{Pure tone: } PH = e^{1.5153SI-.4019} \textbf{ (2);}$$

$$\text{SAM tone: } PH = e^{1.1723SI-.1970} \textbf{ (3);}$$

$$\text{Transposed stimulus: } PH = e^{1.5588SI-3416} \textbf{ (4),}$$

where *SI* is the synchronization index and *PH* is the autocorrelation peak height. Both the gain and the exponent are significantly different for each stimulus type. The statistically significant effect of stimulus type reflects different period histogram and SAC shapes across stimulus types, as shown in Figure 6 for period histograms and in Figure 10 for SACs. The best-fit curves represented in solid lines describe the data well for low to moderately-high (SI<.85) values of phase locking for each stimulus type. However, for higher synchrony values, data tends to deviate from the best-fit curves. These points are overwhelmingly represented by low-SR fibers, and as we showed in Figure 11, peak heights of low-SR fibers differ significantly from those of medium and high-SR fibers. However, for each stimulus, the relationship between the two metrics is monotonically increasing and a clear relationship does exist. Thus, this analysis suggests that the synchronization index accurately captures phase locking information for transposed stimuli and SAM tones as well as for pure tones.

4. Human lateralization performance

4.1. Methods

4.1.1. Stimulus generation

Detection of ongoing ITDs were measured behaviorally in humans as a function of level in unrestricted and restricted listening conditions using the same stimulus types as in the physiological experiments: low-frequency pure-tones, 100% modulated SAM tones and low-frequency pure-tones transposed to a center frequency of 4 kHz. The pure-tone frequency and the modulation rates of SAM tones and transposed stimuli were always 125 Hz. Bernstein and Trahiotis (2002) showed that ITD sensitivity for transposed stimuli begins to degrade rapidly for rates of modulation above 150 Hz. Therefore, a modulation frequency of 125 Hz was chosen to explore ITD sensitivity to transposed stimuli with level in the region where performance is comparable to ITD sensitivity to pure-tones (Bernstein, 2002). All stimuli were presented at levels between 30 and 80 dB SPL in 10-dB steps. The center frequency of SAM tones and transposed stimuli was fixed above the assumed range of phase locking by individual ANFs.

All stimuli were 500 ms in length, including 100-ms on- and off-ramps to prevent spectral splatter and to deemphasize the salience of the on- and offsets. For the high-frequency stimuli, only the ongoing portion of the temporal envelope was delayed; for the pure-stimuli, only the fine structure of the waveform was delayed. For all types of stimuli, an ongoing ITD was introduced by applying a linear phase shift to the stimulus representation in the frequency domain and transforming the signal to the time domain. The type of delay applied is referred to as *modulator delay* (Bernstein and Trahiotis, 1985) and allows the auditory system to utilize envelope-based cues to extract ITD information. ITD just noticeable thresholds are measured with ITDs ranging from 0 to 1000 μs. Naturally occurring ITDs do not exceed 700 μs for a

typical human head (Moore, 2003). Therefore ITD JNDs above this range are interpreted as an inability to perform the task.

For all presentations of SAM tones and transposed stimuli, uncorrelated diotic noise low-pass filtered from 0 to 400 Hz was added to prevent additional lateralization cues from nonlinear distortion products in the periphery and to ensure that listeners cannot take advantage of any low-frequency information that may be available in the signal. Distortion product levels increase with stimulus level. To account for this increase, the background noise level was adjusted proportionally to the signal level. However, distortion products grow more nonlinearly with tone levels since they are cochlear in origin. Adequate masking of distortion products by background noise was therefore additionally tested by the Vliegen and Oxenham (1999) procedure for the 70 and 80 dB SPL levels to ensure that distortion products were inaudible. The background noise upper frequency cutoff was chosen to mask the frequency region of the first two sets of harmonics (Nuetzel and Hafter, 1976; Henning, 1974, Bernstein and Trahiotis, 1994). Presentation of background noise was not necessary with pure-tones since no sidebands are present in the signal.

Human detection thresholds were measured in two psychophysical conditions: unrestricted listening where only background noise masking distortion products was present, and restricted listening, where both background noise and notched or high-pass noise that restricted off-frequency listening (Patterson, 1976; O'Loughlin and Moore, 1981) were present. Notched noise was used for SAM tones and transposed stimuli and high-pass noise for pure-tone stimuli. The notched noise contained a stop band between $0.9f_c$ and $1.1f_c$ (3600 Hz to 4400 Hz), as shown in Figure 13.

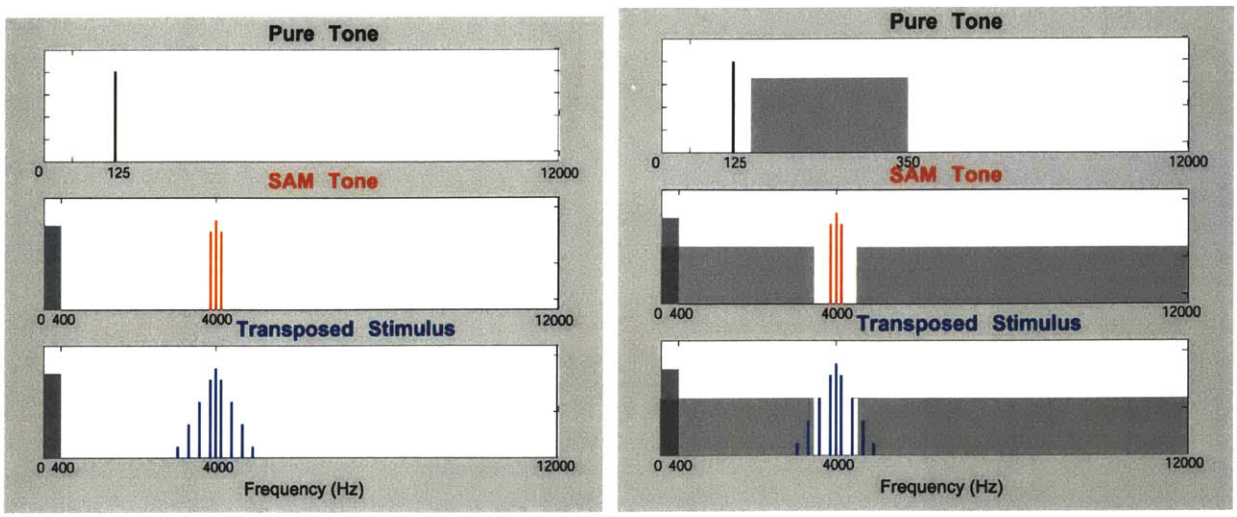

Figure 13: Illustration of three psychophysical stimuli in the two conditions in the frequency domain. The unrestricted listening condition (left) includes only background noise (0 to 400 Hz shown in dark gray), while the restricted listening condition (right) includes high-pass noise (150 Hz to 350 Hz) for pure tone stimuli and notched noise (notch between 3.6 kHz and 4. kHz shown in light gray) for SAM tones and transposed stimuli. Frequency and amplitude scaling are only approximate.

The notch width includes frequencies within +/- 10% of the carrier frequency. As Figure 13 illustrates, the notch noise stop band contains both sidebands of the SAM tone and the first two sets of sidebands of the transposed stimulus, while masking the others. Similarly, off-frequency listening for pure-tone stimuli was restricted by presenting pure-tone stimuli in concurrence with high pass noise from 1.2f (150 Hz) to 350 Hz. The noise levels and spectra were chosen to limit useful information to the CFs around the target frequency (signal-to-noise ratio was too low for neurons with CFs higher than the upper limit of the noise) while keeping the overall intensity of noise from being uncomfortable for subjects. Correlated interaural noise was used for the notched and highpass noise, as listeners could not lateralize transposed stimuli in uncorrelated notched noise within physically realizable ITD limits. Correlated noise should not provide additional lateralization cues using this setup (Nuetzel and Hafter, 1976).

The stimuli are digitally generated and played through a LynxStudio LynxOne soundcard with 24-bit resolution at a 32 kHz sampling rate. The stimuli are presented to the subjects via

Sennheiser HD 580 headphones after being passed though a TDT PA4 programmable attenuator and a TDT HB6 headphone buffer. All testing was conducted in a double-walled, sound-attenuating chamber.

4.1.2. Procedure

Detection thresholds for the three stimuli were measured for each subject in the masking condition using a two-interval, two-alternative forced choice, two-down, one-up, adaptive procedure (Levitt, 1970) using stimuli of the same type and duration as those presented during the lateralization task. The two-down, one-up rule tracks the 70.7% correct point on the listener's psychometric function. Using stimuli of the same duration in detection and lateralization experiments is important because detectability (d') changes with the duration of the signal (Green et al., 1957). Thresholds in quiet were measured to verify normal binaural hearing levels and to ascertain that all stimulus presentations were within the subjects' hearing range. Masked thresholds were measured to obtain the appropriate noise level for each stimulus. The appropriate noise levels were then used in the lateralization experiments, such that the noise levels presented just mask the off-frequency listening region. The masker, at 10, 20, 30 and 40 dB SPL spectrum level, was presented and the signal level was adjusted until the signal was just detectable. A 125-Hz pure-tone was presented in the high pass noise and a 4-kHz pure-tone was presented in the notched noise. Thresholds for transposed stimulus and SAM tones in notched noise are expected to be the same as pure-tone thresholds.

Maskers for SAM tones and transposed stimuli were presented at levels 20 dB below that required to mask the signal, so that off-frequency listening was prevented but the masking noise did not interfere with the on-frequency thresholds (e.g., O'Loughlin and Moore, 1981). Findings of Hershkowitz and Durlach (1969) and Zwislocki and Feldman (1956) indicate that ITD JNDs

do not change significantly when the signal level increases more than 20 dB above threshold. Furthermore, as the level difference between signal and noise was kept constant, the underestimation of performance due to the noise was constant across all levels. High-pass masker levels for the pure-tones were set so that the tones were 10 dB above their threshold in broadband noise. Removing the masking noise from 0 to 400 Hz lowers thresholds by approximately 10 dB, such that the tones are also approximately 20 dB above their masked threshold in the high-pass noise (Glasberg and Moore, 1990). The high-pass filter cut-off was chosen such that the level of the noise plus the masker does not exceed 85 dB SPL for all listeners.

Four normal-hearing subjects served as listeners. Each subject underwent audiological evaluation to ascertain normal hearing pre and post experiment. Subjects' diotic thresholds in quiet and each of the two noise conditions were measured in the same order for all listeners. In the lateralization tasks, stimuli were presented via a two-alternative, forced-choice, two-down, one-up algorithm (Levitt, 1971) with listeners asked to decide whether the right ear leads in the first or second presentation. Thresholds were obtained in each listening condition in random order and each condition was revisited three times. The initial ITD was 500 μs and initial step size was a factor of 2, which was reduced to 1.4 after two reversals and to 1.18 after another two reversals. For each run, testing continued until four reversals at the smallest step size occurred and the threshold was taken as the mean of these four reversals. If the standard deviation of the logarithm of the last eight reversals exceeded .3, the run was repeated. Final thresholds were averaged for each listener and stimulus condition over the three adaptive runs. Each subject began with a training phase with feedback until they have reached asymptotic performance.

Training for each subject lasted from 4 to 6 hours. Feedback was also provided during the data collection phase. Subjects were paid an hourly wage for their services.

4.1.3. Analysis

ITD JND thresholds were obtained for each listener as a function of level for the three stimuli with and without notched noise, yielding six conditions at each of 6 levels. Statistical analysis was conducted to:

- determine whether lateralization discrimination performance improves or deteriorates with level;
- compare lateralization performance for pure tones and transposed stimuli in the restricted listening condition;
- compare lateralization discrimination for pure-tones and transposed stimuli in the unrestricted listening condition;
- compare lateralization performance in the two listening conditions;
- characterize the shapes of lateralization performance curves with level for the six conditions;

4.2. Psychophysical Results

The prominent fall of phase locking with increased stimulus level for transposed stimuli and SAM tones led us to question whether a similar degradation of performance would be seen psychophysically and whether we could predict the psychophysical results from auditory nerve spike timing using an optimal binaural processor model. This sections describes the measured lateralization performance using pure tones, transposed stimuli and SAM tones as discussed in the methods section.

Psychophysical data were collected with and without notched noise, in unrestricted and restricted-listening conditions, as described in the methods section. The unrestricted listening condition should incorporate the performance of all active neurons. On the contrary, the listening band restricted condition is assumed to restrict lateralization information to a small range of frequencies around the carrier frequency of the high-frequency stimuli (3600 to 4400 Hz).

In the unrestricted listening condition, shown in Figure 14, listeners perform equally well (differences statistically insignificant using ANOVA analysis) lateralizing transposed stimuli and pure tones ($f_m = f = 125$ Hz), with performance using SAM tones worse by about 100 μs at each level. Additionally, listeners performed better when stimulus level was increased[1].

[1] Pure tone performance at the lowest stimulus level is not shown because one listener was not able to complete the task.

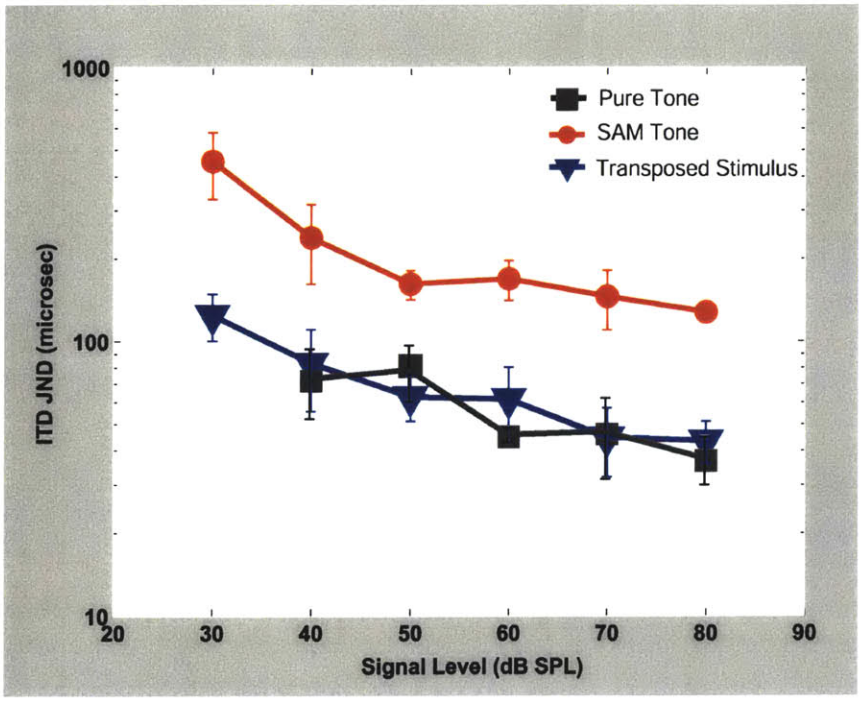

Figure 14: Interaural temporal just noticeable differences measured in an adaptive listening task for pure tones (black), transposed stimuli (blue) and SAM tones (red) with off-frequency listening. Stimuli were pure tones at 125 Hz, SAM tone and transposed stimulus with f_m=125 Hz and f_c=4000 Hz. Standard Errors are shown.

These observations of comparable performance in lateralizing pure tones and transposed stimuli are consistent with the study of Bernstein and Trahiotis (2002) who found that for very low modulation frequencies, extents of lateralization and lateralization sensitivity using transposed stimuli are as good as, if not better than, those using pure tones. Specifically, Bernstein and Trahiotis (2002) measured ITDs for transposed stimuli with a 4-kHz carrier and 128-Hz modulation frequency and pure tones of the same frequency at 75 dB SPL as 76 and 69 μs, respectively, comparable with the measurements above. However, this study shows that the result holds across a wide range of levels.

In the second experiment, we restricted the listening band to a small region surrounding the carrier frequency with notched noise (Figure 15). With our choice of notch width, 2 sets of sidebands of the transposed stimulus are contained within the notch.

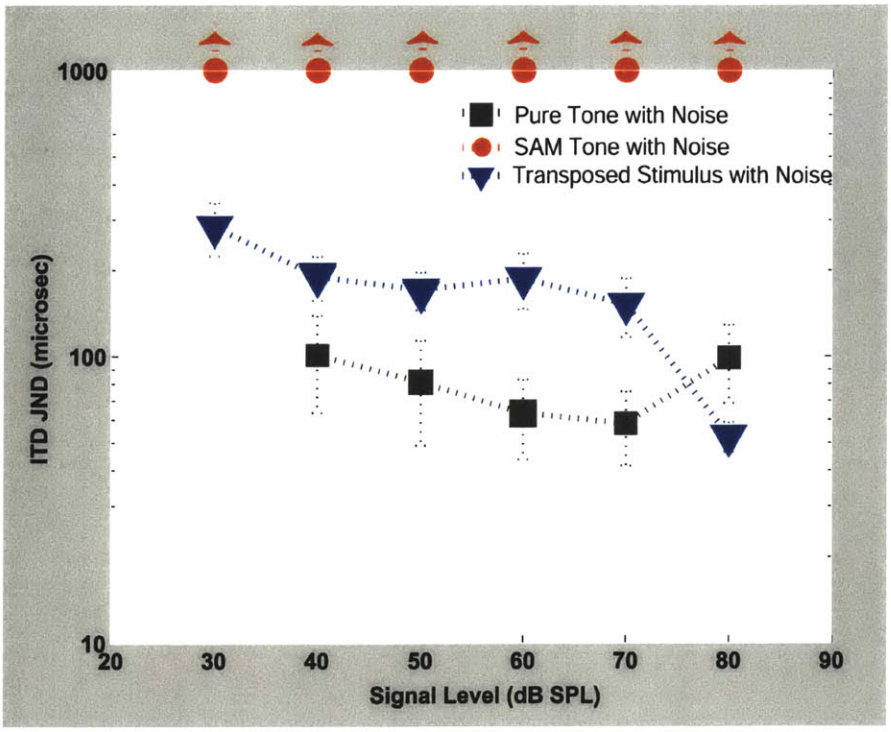

Figure 15: Interaural temporal intensity just noticeable differences measured in an adaptive listening task for pure tones (black), transposed stimuli (blue) and SAM tones (red) without off-frequency listening. Stimuli were pure tones at 125 Hz, SAM tone and transposed stimulus with f_m=125 Hz and f_c=4000 Hz. Standard Errors are shown. Off-frequency was restricted by notched noise for high-frequency stimuli (from 0-3600 Hz and from 4400-12000 Hz) and by high-pass noise (150-350 Hz) for pure-tone stimuli. Upward errors represent SAM tones as ITDs exceeded 1000 us.

In the restricted condition, listeners do performance decreases slightly with pure tone and substantially with transposed tone stimulus, compared to the unrestricted condition. A repeated-measures analysis of variance revealed a statistically significant decrease in performance for transposed stimuli ($F(3)=46.4$, $p=.006$) but not for pure tones in the two listening conditions. Specifically, measured thresholds for transposed stimuli fell well below 100 μs for levels greater than 30 dB SPL, while transposed stimulus thresholds in this condition all exceed 100 μs, except for the 80 dB SPL measurement. SAM tone lateralization performance is not shown as listeners were not able to complete the task. These results suggest that "loss" of neurons, and not sidebands (as no sidebands are masked for SAM tones) add information important for lateralizing the stimulus comparably to pure tones in the unrestricted listening condition.

All but the 80 dB SPL points suggest that even in the restricted band listening condition, ITD sensitivity does not degrade with increased stimulus levels. The result for transposed stimuli raises the possibility that the distortion products at the modulation frequency may

become audible above the background noise. We used a technique described by Vliegen and Oxenham (1999) to determine the level of distortion products at f_m and $2f_m$. The technique involved measuring two masked thresholds, one for a pure tone (with random phase from trial to trial) with f=125 Hz in the presence of a transposed stimulus masker (f_c=4000 Hz, f_m=125 Hz), and one for the same pure tone in the presence of another sinusoidal masker (also at f=125 Hz). The former establishes the masking produced by the combination tone at the pure tone frequency and the later uses this estimate the find the effective level of the combination tone. Consequently, we measured the background noise level needed to mask these measurements of combination tone levels at f_m and $2f_m$ for two listeners and found that the necessary masker levels fall below the level of background noise we used at 70 and 80 dB SPL conditions. The results of this additional test suggest that even at the highest levels used, the improvement in ITD for transposed stimulus cannot be attributed to distortion products becoming audible. The 80-dB SPL pure tone point suggests that ITD sensitivity degrades at very high stimulus levels in the restricted listening condition.

In summary, the major psychophysical findings are comparable ITD discrimination performance for transposed stimuli and pure tones, and slightly decreasing thresholds with increasing levels in the unrestricted listening condition. In the restricted band listening condition, listeners performed worse in lateralizing transposed stimuli relative to pure tones, but performance remained roughly constant, or even improved, as sound pressure level was increased.

5. Physiologically-based model of optimal binaural processor

5.1. Methods: Model overview

This section describes a model for predicting psychophysical lateralization performance described in section 4 based on neurophysiological data presented in section 3. At the core of the model is Colburn's (1973) strategy for predicting binaural interaction based on auditory-nerve firing data. Colburn models a suboptimal binaural processor with some internal noise, which limits the precision at the stage of the auditory nerve, and is consistent with psychophysical testing of the fundamental capabilities of the auditory system. The central processor is suboptimal in that it assumes same-CF inputs to the binaural processor from the two sides and coincidence detection mechanisms that ignore long intervals between spikes from the two sides. Although suboptimal overall, his model of the central binaural processor is optimal given these two physiologically plausible constraints. Colburn's work is a direct extension of Siebert (1968) on monaural phenomena indicating that psychophysical performance may be dominated by peripheral limitations. Here, the Colburn central auditory processor model of binaural detection was used to model predict lateralization performance based on AN spike timing information and to compare the predicted and psychophysical performance. The peripheral component of the model was modified to fit our physiological data on phase locking.

The Colburn model makes several assumptions. First, it assumes stochastic behavior of auditory nerve fibers, the statistical independence of the processes at the two ears and the statistical independence of different fibers given a stimulus waveform. The model also assumes the AN firing pattern are non-homogeneous Poisson processes. The first three assumptions are either widely used or supported by neurophysiological data. The assumption of a Poisson process is supported by the exponential decay of interval histograms and the independence of

successive intervals for spontaneous activity (Kiang et al, 1965). However it must be empirically verified for high-frequency periodic stimuli. Appendix 2 shows empirical evidence that the auditory nerve firing patterns are consistent with the Poisson assumption.

The present model of lateralization discrimination consists of two components: a peripheral component and a central binaural processor (Figure 16). The peripheral component is a cascade of three stages: (1) a bank of linear bandpass filters representing the tuning characteristics of the array of auditory nerve fibers, (2) a rule for predicting the level dependence of single ANF rate and synchrony to each stimulus, and (3) an expression for Poison-based instantaneous firing rate from physiological average rate and synchrony. The binaural component derives ITD discrimination based on an array of coincidence detectors with a distribution of best interaural time delays. Predicted responses from one binaural pair of ANFs constitute the "single-channel' model and responses are integrated across CF to form the "multiple-channel" model (Figure 16). Table 1 details the model variable names and their descriptions. Variables are also shown in Figure 16.

Variable Name	Description
$P_{in}(CF)$	Input stimulus power spectrum
$P_{out}(CF)$	Output stimulus power spectrum after cochlear filtering
$R_m(CF)$	Mean firing rate with respect to SR group
$SI_m(CF)$	Mean SI with respect to SR group
$r_m(CF)$	Instantaneous firing rate defined by Exponential Rate Model
$Q_{i,m}(CF)$	Performance of single pair of fibers at an internal delay

Table 1: Model variable definitions and descriptions. Variables indexed on *l* incorporate low, middle and high SR responses, and variables indexed on *m* incorporate the distribution of internal delays.

We constructed a model of the *human*, and not cat, peripheral processor so as to compare the modeled outputs to human psychophysical performance. In constructing a human model, the

cochlear filters and the cochlear frequency map are based on psychophysical data and models derived from humans (Glasberg and Moore, 1990, Greenwood, 1990). However, synchrony-level and rate-level functions based on cat data are used that we and Liberman (1978) have measured. It is assumed that physiological measurements in individual human auditory nerve fibers would be similar to those of cat, although the CF distribution of fibers may differ. Although filter bandwidths and CF distribution of auditory nerve fibers have been measured in cats (Shera et al., 2002; Liberman, 1978), it is known that these measurements differ from those of humans and would not produce model results comparable to human psychophysics (Shera et al., 2002; Greenwood, 1990).

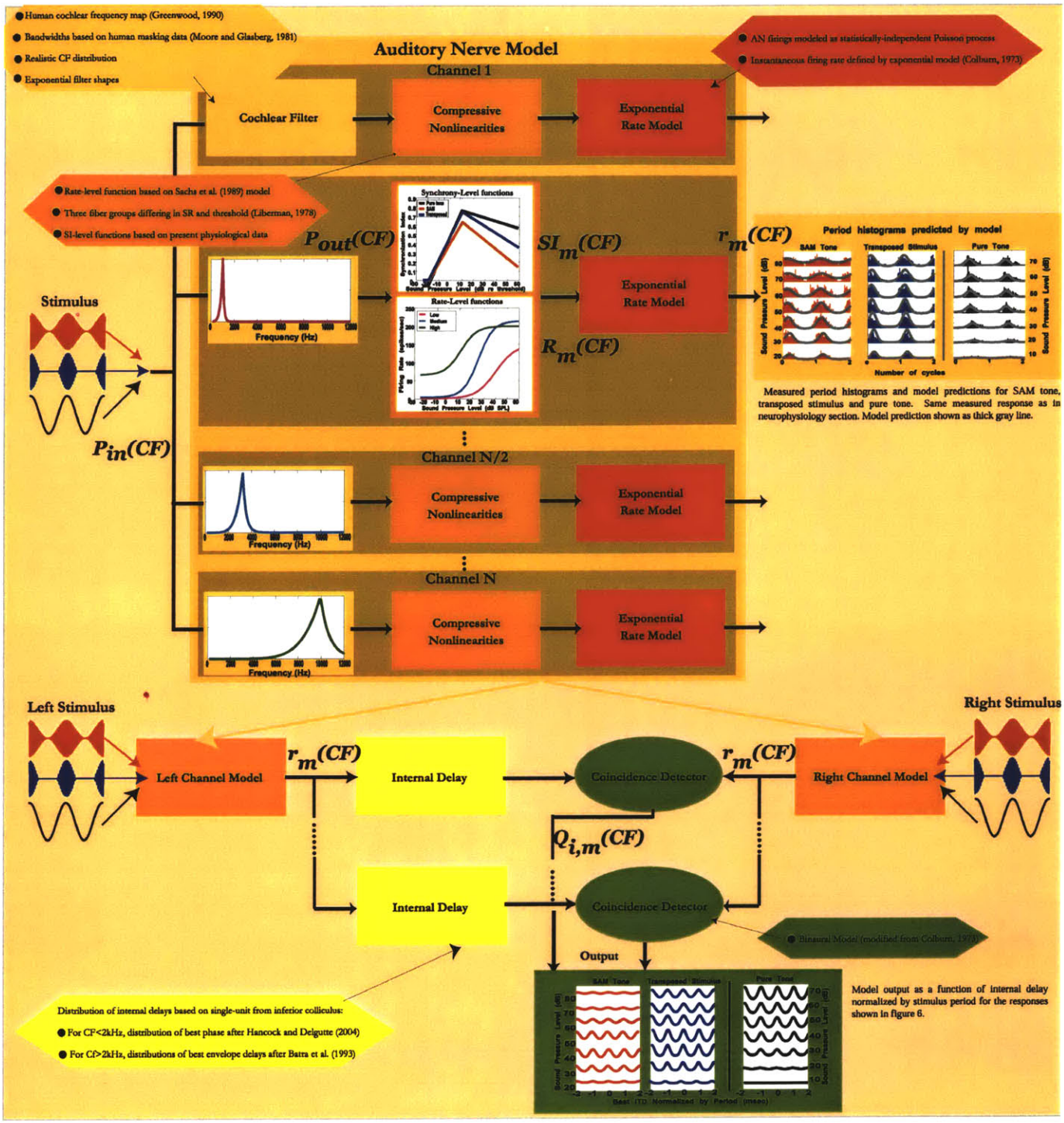

Figure 16: Model Overview. We modeled the binaural processor as a series of coincidence detectors and internal delays such that the calculated ITD JND best represents the delay between the right and left stimulus to the auditory periphery. The periphery model is shown expanded, composed on cochlear filters with bandwidths fit to Glasberg and Moore (1990) ERBs, physiologically measured compressive rate-level and synchrony-level nonlinearities and the Colburn exponential rate model. The fit of the peripheral model is shown overlaying neurophysiological data from Figure 6 as well as model output.

The modeled performance was compared to psychophysical performance in two conditions: unrestricted and restricted listening condition. In both cases, the stimulus parameter, duration, modulation frequency, carrier frequency and level are the same as in psychophysics. We modeled the unrestricted listening condition by incorporating the responses of all responsive AN fibers. Specifically, we included responses of all fibers with CF from 400 Hz to 20 kHz for the high-frequency stimuli and from 20 Hz to 20 kHz for the low-frequency (f = 125 Hz) pure-tone stimulus. We simulated the band-restricted listening condition by only including AN fibers within the noise rejection band (fibers with CF from 3.6 kHz to 4.4 kHz for the high-frequency stimuli and 20 Hz to 150 Hz for the low-frequency pure tone). By implementing this condition with this restricted CF range, we assume (1) that fibers outside the included CF range do not respond to the stimulus (i.e. they are completely masked) and (2) that responses of included fibers are not affected by the masking noise. The notched noise might produce suppression and compression in the noise-rejection band, and we acknowledge the possibility that a more realistic model incorporating suppression and compression might differ in performance. We also assume (3) that auditory filters behave linearly in the response area of interest and (4) that the binaural processor can be modeled as a cross-correlation mechanism. We will return to these points in the discussion.

5.2. Model description

5.2.1. Model Part I: Peripheral Processor

5.2.1.1. Model of cochlear filtering

The cochlear mechanical response can be modeled to a first approximation as a band of linear bandpass filters. The first module implements this bank of auditory bandpass filters and

outputs the effective sound pressure level at each CF by integrating the root-mean-square (RMS) amplitude in each channel. Although this module can use arbitrary sound inputs, we focused on the pure-tone, SAM tone and transposed-tone stimuli as they were presented in the neurophysiological and psychophysical experiments.

The auditory filters in the model are the simple triangular filters (in double logarithmic coordinates) used in the Colburn (1973) model. The filter frequency response $H_m(f)$ is defined by:

$$H_m(f) = \left(\frac{f}{CF}\right)^{a(CF)}, \text{ for } f \leq CF \qquad (5)$$

$$= \left(\frac{CF}{f}\right)^{2a(CF)}, \text{ for } f > CF,$$

where $a(CF)$ is defined as:

such that $24.7\left(\frac{4.37CF}{1000}+1\right) = \frac{6a(CF)}{(2a(CF)+1)(4a(CF)-1)}$, and an analytical solution for $a(CF)$ is derived.

The factor $a(CF)$ specifies the bandwidth of the individual filter. In our implementation, equivalent rectangular bandwidths (ERBs) were matched to the Glasberg and Moore (1990) measured ERBs.

To find the effective power, $P_{out}(CF)$, after filtering by $H_m(f,CF)$ of the stimulus power spectrum $P_{in}(f)$, we calculated the dot product across frequency of the stimulus and the magnitude squared of the filter function, as follows:

$$P_{out}(CF) = \sum_k |H(f_k, CF)|^2 P_{in}(f_k) \quad (6).$$

This effective level vector, $P_{out}(CF)$ is the level of input seen by the range of active fibers. Example effective level functions as a function of CF are shown in Figure 17 for pure tones,

SAM tones and transposed stimuli with fm=125 Hz and levels ranging from 0 to 70 dB SPL. These functions are the effective spread of excitation induced by each stimulus along the basilar membrane.

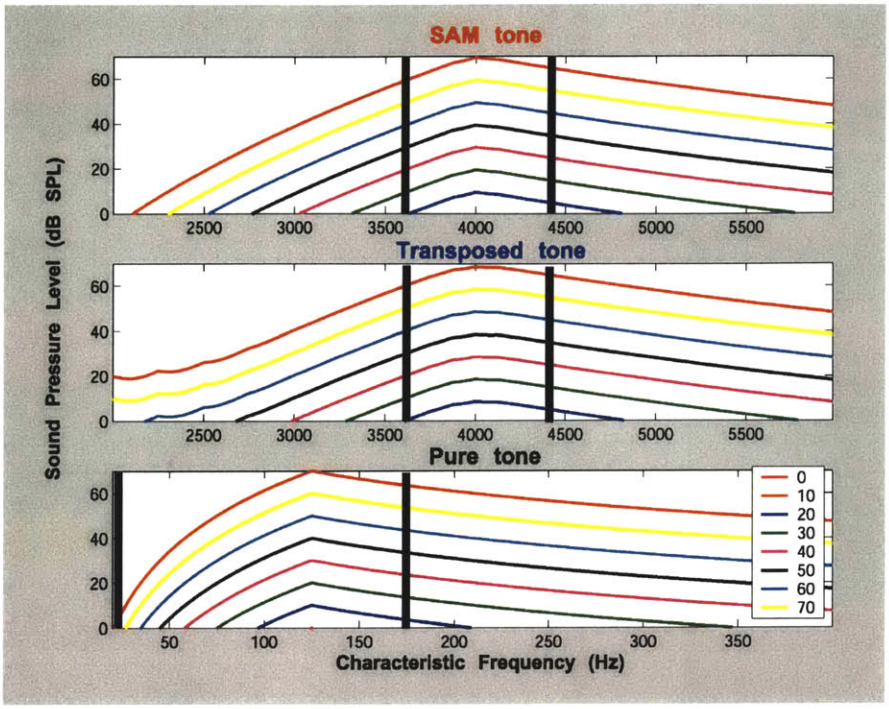

Figure 17: Excitation patterns $P_{out}(CF)$ for SAM tone (top), transposed stimuli (middle) and pure tones (bottom) filtered by the filter $Hm(f,CF)$. All stimuli have frequency or modulation frequency =125 Hz. Stimulus sound pressure levels range from 0 to 70 dB SPL in increments of 10 dB. The listening band here is shown from 2.6 kHz to 5.4 kHz. Solid vertical lines indicate the width of the stop-band of the notched noise used in physiology.

Each curve corresponds to an input level 10 dB larger than the previous curve. This proportional increase in response with stimulus level occurs because the filters are linear. Also note that the high-frequency partials of the SAM tone and transposed stimuli are not resolved for a modulation frequency of 125 Hz, as evidenced by the lack of peaks in the effective level every 125 Hz away from the f_c of 4000 Hz. However, lower-frequency partials (f<2500) are somewhat resolved because cochlear filter widths narrow with characteristic frequency.

5.2.1.2. Model of Cochlear Nonlinearities

The next module in the model incorporates the cochlear non-linearities observed in auditory nerve rate-level (e.g. Liberman, 1978; Delgutte, 1987; Sachs et al. 1989) and synchrony-level function (Johnson, 1980; Joris and Yin, 1990). Numerous studies have

measured auditory nerve responses to pure tones with varying level and have observed a 20-40 dB range in which rate grows monotonically with level between threshold and saturation. For the purposes of this model, we assume that the rate level functions in response to pure-tone, SAM tone and transposed tone stimuli do not differ significantly. We base this assumption on our observations (not shown) that the dynamic range of the fibers in response to these high-frequency stimuli is consistent with the dynamic range in response to pure tones. Here, we modeled rate-level functions in response to the three stimuli based on the work of Sachs and Abbas (1974) incorporating a realistic distribution of SR and thresholds, as found in cat. We assume that the relative distributions of SR and thresholds in cat and human are similar, and that they are the same across CF.

We used the Sachs and Abbas (1974) computational model of rate-level functions for single fibers. The compressive nonlinearity fits the following form:

$$R(CF) = R_D(P_{out}) + R_{SP} = \frac{R_{\max}}{1+\exp^{N(\theta_{50}-L_{out})}} + R_{SP} \quad (7),$$

In this model, $L_{out} = 10\log_{10} P_{out}$ is the effective level as defined in the previous section; R_{max} is the driven saturation rate of the fiber; θ_{50} is the level at which the fiber's firing rate reaches 50% of the maximum firing rate (re threshold), and N determines the steepness of the rate-level function.

We chose the model parameters to fit with physiologically-based rate-level functions for low, medium and high spontaneous rate fibers. Our model spontaneous firing rate, R_{sp}, and maximum firing rate, R_{max}, are consistent with Liberman's (1978) measurements of SR and maximum firing rate of low, medium and high-SR fibers in response to pure tones. Dynamic

range for low, medium and high-SR fibers was selected as 20, 30 and 40 dB, respectively (Liberman, 1978). The following parameters were used in our model for each SR group:

	Low SR	Medium SR	High SR
R_{sp} *(spikes/s)*	.025	4	65
R_{max} *(spikes/s)*	148	214	200
θ_{50} *(dB SPL)*	20	15	10
N	.12	.16	.15

Table 2: Model values for the spontaneous rate R_{sp}, maximum driven rate R_{max}, level of 50% firing rate re threshold θ_{50}, and steepness value N.

The rate level functions for the 3 SR groups are shown in Figure 18.

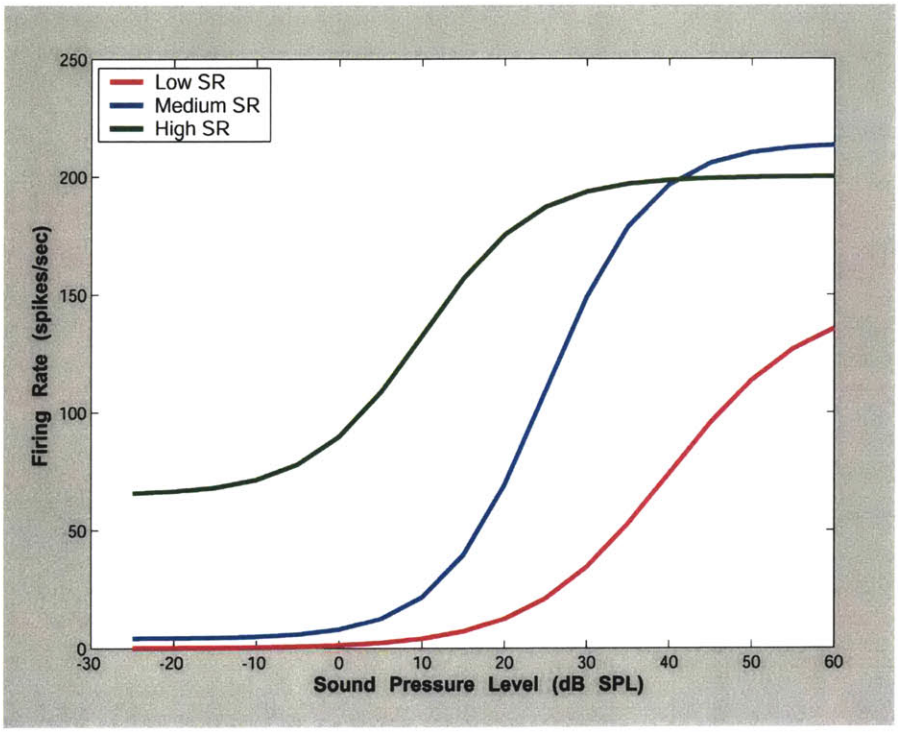

Figure 18: Modeled rate level functions for low, medium and high-SR fibers. Model is based on Sachs and Abbas (1974) model relating sound pressure at the eardrum to discharge rate in auditory nerve.

The compressive nonlinearity of synchrony-level function was also included in this model. Unlike the traditional Colburn model, here we based our SI functions on our neurophysiological recordings consistent with each stimulus type. We are thus able to capture

timing information inherent in the coding of each stimulus in a mathematical model without using actual spike trains.

We pooled all of our data grouped by stimulus type to characterize the synchrony-level functions. Our synchrony-level functions are described by a piece-wise linear function: an upward sloping line from 20 dB below the threshold of the fiber to the level maximum, and a downward sloping line above the level maximum. This linear model fits the data well, as shown by the synchrony-level function below.

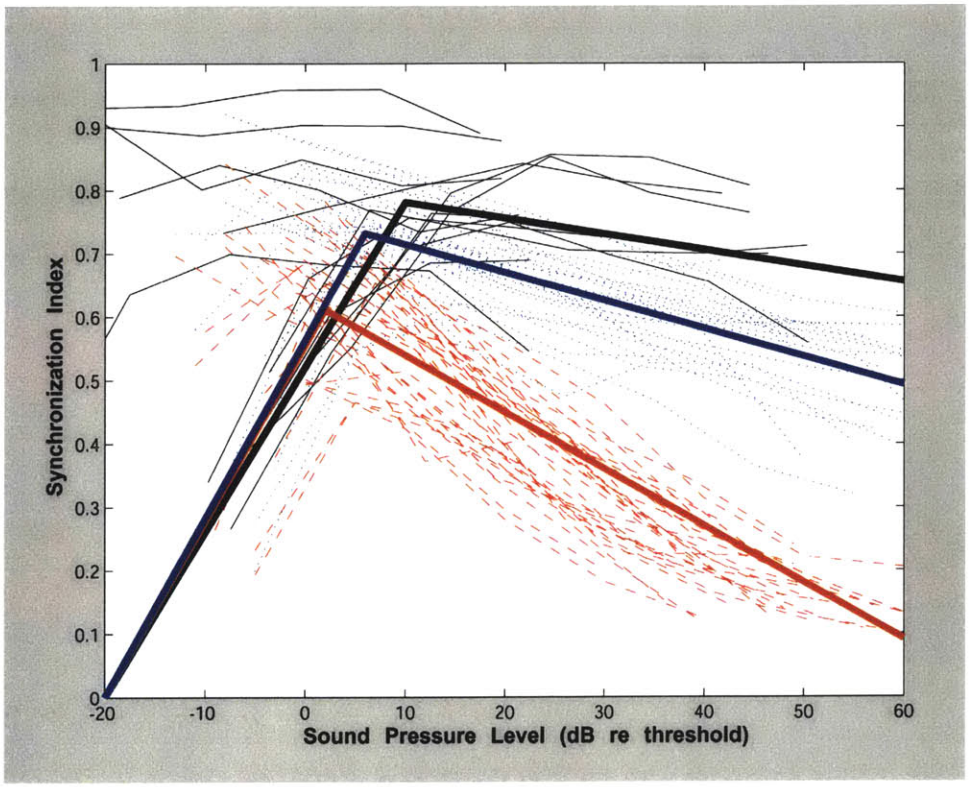

Figure 19: Synchrony-level functions for on-CF presentations of SAM tones (dashed lines) and transposed stimuli (dotted lines) with a modulation frequency, or frequency of 125 Hz. Pure tone (solid lines) data shown for f<400 Hz with CF close to f. Dashed lines depict data and solid lines depict best fit linear model of data.

We developed the model to describe auditory-nerve responses to modulation frequencies of 125 Hz for comparison to psychophysical experiments. Ideally, our model for synchrony-level functions, $SI_m(CF)$ in model, would be based on 125 Hz pure tones. However, we had very few data at such low frequencies. We therefore used all pure-tone data with frequencies below

400 Hz to generate the model synchrony-level functions. We assume that SI for lower-frequency pure tones will be similar over that frequency range (Johnson, 1980).

Linear fits are estimated from data pooled across fibers. The upward sloping portion of the synchrony-level functions are described by the mean value and mean level of the maximum synchronization index across individual fiber synchrony-level functions for each stimulus type. For simplicity, the model assumes a synchronization index of 0 at 20 dB below threshold. Although the falloff to zero synchrony in 20 dB is somewhat arbitrary, this choice would primarily affect performance at low sounds pressure levels where psychophysical performance was not measured. The more important downward sloping portion of the synchrony-level function is fully determined by the mean value and mean level of the SI maximum and slope of degradation of SI with level. This slope is estimated as the mean of the best-fit line of to individual fiber SI-level functions.

After deriving analytical expressions for the rate and synchrony-level function, these analytical expressions are then applied to the effective-level functions to derive the rate and synchrony distributions across the cochlea for each stimulus type. Model synchrony-level functions (Figure 19) are adjusted along the coordinate axis for the fiber threshold in each SR group (Figure 18). Rate and synchronization-index values are then used to generate the instantaneous firing rate of the Poisson process, as described in section 5.2.1.2.

The rate and synchrony excitation patterns across the tonotopic array of fibers reflect the nonlinearities of the rate and synchrony-level functions. Figure 20 shows this variation for a SAM tone, transposed tone and pure tone with stimulus frequency or modulation frequency of 125 Hz and for levels from 0 to 70 dB SPL. These functions are shown for high-SR fibers only.

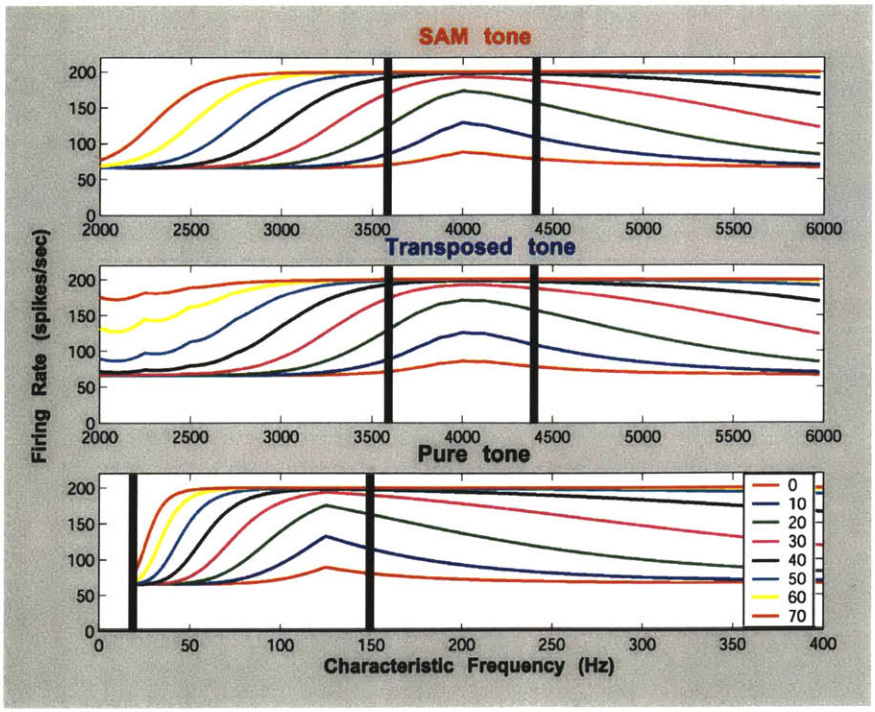

Figure 20: Firing rate variation along basilar membrane for SAM tone (top), transposed stimulus (middle) and pure tone (bottom) with modulation frequency or frequency of 125 Hz and at sound pressure levels from 0 to 70 dB SPL. Curves are obtained by mapping the rate level function to the effective level functions shown in Figure 17. Curves are shown for responses of high-SR fibers only. Solid vertical lines indicate the width of the stop-band of the notched noise used in physiology.

Figure 20 shows that low-level stimuli elicit rate variation with CF proportional to the effective level of the stimulus because these stimuli fall along the monotonically-increasing portion of the rate-level curve. However, the rate excitation patterns begin to saturate as portions of the effective level functions begin to fall above the dynamic range of the fibers. This saturating nonlinearity appears as a plateau for fibers with CF close to f_c, where effective stimulus levels are largest. As stimulus level increases, the plateau region spreads further away from f_c as a larger range of tonotopic locations begin to fall inside the saturating portion of the rate level function.

Adding low and medium-SR fibers reduces the extent of saturation in the rate-excitation pattern in two ways. First, since lower spontaneous-rate fibers have higher thresholds (Liberman, 1978), their rate-level functions are effectively shifted to higher stimulus levels and higher stimulus levels are needed to reach the saturating region. Second, low-SR fibers have a larger dynamic range than high-SR fibers. A larger dynamic range extends the range of levels for

which the effective level function will fall inside the linear region of the rate level function. Figure 21 compares rate-CF profiles for low, medium and high-SR fibers for a stimulus level of 30 dB. In this example, while high-SR fibers responses close to f_c are nearly saturated, responses of medium SR fibers are in the monotonically-increasing region of the rate-level function, and the responses of low-SR fibers are very close to threshold. These enlargements of the dynamic range will be crucial in improving performance at higher levels.

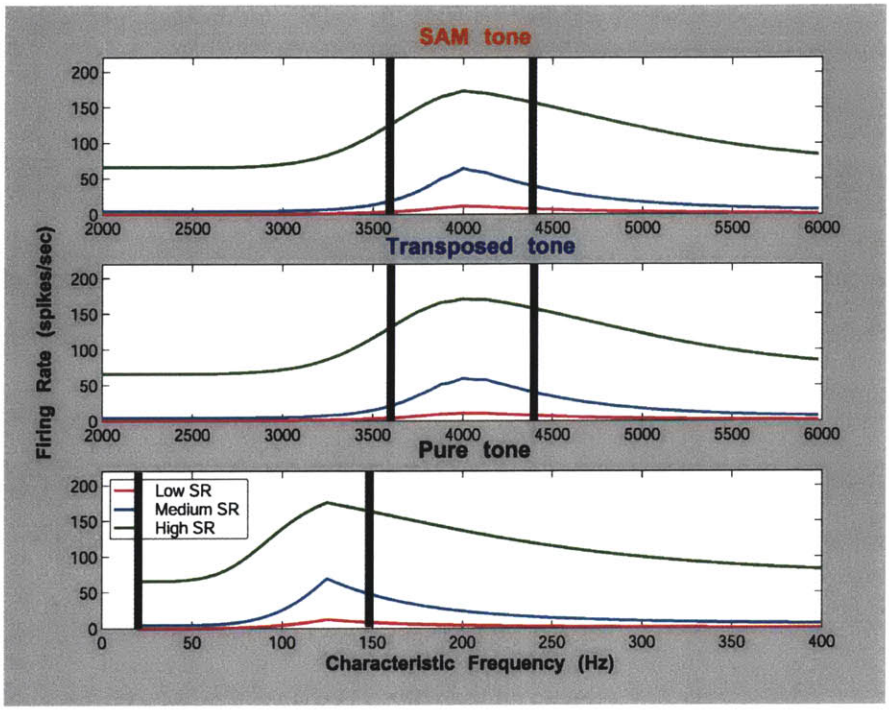

Figure 21: Firing rate variation along the basilar membrane in response to a SAM tone (top), transposed stimulus (middle) and pure tone (bottom) of modulation frequency or frequency of 125 Hz and carrier frequency (SAM and transposed tone) of 4000 Hz. Responses to stimuli at 30 dB SPL for low (magenta), medium (light blue) and high (green) SR fibers. Solid vertical lines indicate the width of the stop-band of the notched noise used in physiology.

Synchronization index excitation patterns also exhibit a form of saturation. However, the shape of the saturating nonlinearity for synchrony excitation spread differs from that of rate excitation spread, so that SI actually decreases at high SPL.

Figure 22 shows the pattern of synchrony across CF for the same stimulus conditions as in Figure 20.

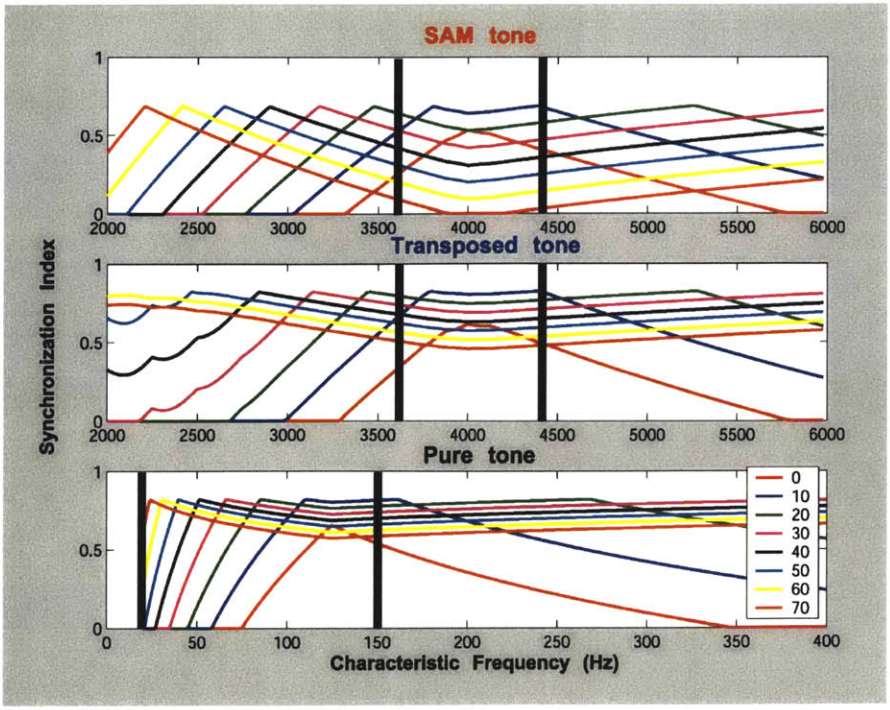

Figure 22: Pattern of synchrony for levels 0 to 70 dB SPL of SAM tone (top), transposed tone (middle) and pure tone (bottom). Graphs were obtained by mapping analytical expressions for synchronization index level functions onto effective level functions (Figure 17). Shown are functions for modulation frequency and frequency^2 of stimuli at 125 Hz and for high-SR fibers only. Solid vertical lines indicate the width of the stop-band of the notched noise used in physiology.

The shape of the synchrony spread reflects the interaction of the synchrony-level nonlinearity and effective level. For low levels, the synchronization index is a linearly increasing function of level. Consequently, the synchrony spread functions resemble scaled version of low-level BM excitation patterns. As the stimulus level is increased, the excitation level for fibers close to the stimulus carrier frequency begin to fall on the downward-sloping portion of the synchrony level curve. The synchrony pattern becomes bimodal, because the synchrony is lowest for fibers with CF near f_c where the effective levels of excitation are highest. The rate of decrease depends directly on the slope of the synchrony level function. For pure tones, the rate of decrease is so shallow that the synchronization index never falls very much with level. However, the rate of decrease of the fall of SI is so steep in response to SAM tones that the synchrony for fibers with CF close to f_c falls to 0 at a high stimulus levels. The response to transposed stimuli are intermediate between those of pure tones and SAM tones.

Figure 22 only shows the pattern of synchrony across CF only for high-SR fibers. Although the compressive nonlinearity for synchrony is assumed to be the same across SR groups, fibers with lower SR are able to respond linearly up to higher stimulus levels because they have higher thresholds. While earlier data (Figure 11) showed small differences in maximum SI between the three SR groups, the model does not reflect those differences because they are not likely to change model performance. An example of the synchrony CF functions for fibers with different SR is shown in Figure 23 for a stimulus level of 30 dB SPL.

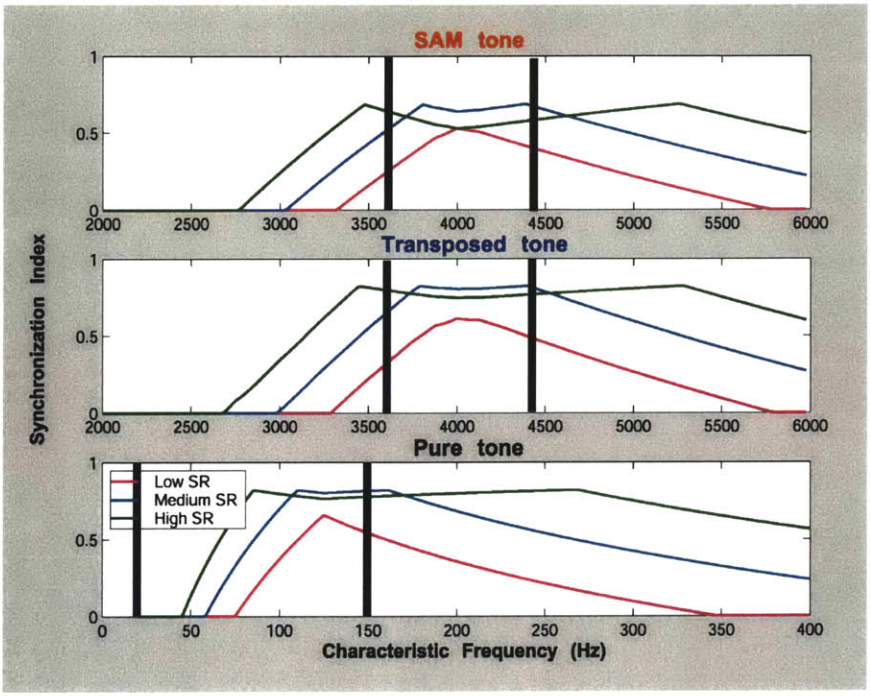

Figure 23: Synchrony-CF functions in response to a SAM tone (top), transposed tone (middle) and pure tone (bottom) with a frequency or modulation frequency of 125 Hz and a characteristic frequency (SAM tone and transposed tone only) of 4 kHz. The level of each stimulus is 30 dB SPL. The magenta represents the synchrony of low-SR fibers, the light blue medium-SR fibers and the green high-SR fibers. Solid vertical lines indicate the width of the stop-band of the notched noise used in physiology.

While synchrony patterns of high-SR fibers are already bimodal for stimulus levels of 30 dB SPL, responses of low-SR fibers are still unimodal. It should also be pointed out that because the excitation level across CF in response to a transposed stimulus at a given level is somewhat lower than for SAM tones due to cochlear frequency selectivity, it is quite possible that a low level of stimulus will produce a bimodal synchrony-CF distribution for SAM tones but not transposed stimuli, as seen in this example for medium-SR fibers.

5.2.1.2.Exponential Rate Model

Inherent in Colburn's binaural processor model is an exponential rate model of auditory nerve firing originally proposed by Siebert and Gray (1964). This imbedded model assumes that period histograms of spike times can be modeled as an exponential function of sine waves. The exponential functions (Siebert, 1970) are specified by the relation:

$$r(t) = a_m \exp(g_m \cos(2\pi ft)) \quad \textbf{(8)}$$

where a_m is related to the average firing rate and synchronization index, and g_m is monotonically related to the synchronization index.

To fit this exponential rate model to the physiological data, the Colburn processor needs only the mean firing-rate and synchrony values. The model actually used two derived parameters a_m and g_m that depend on both the firing rate and the synchrony-level function. The synchronization index alone determines g_m by the empirical equation:

$$\frac{I_1(g_m)}{I_0(g_m)} = SI \quad \textbf{(9)}$$

where $I_0(.)$ is the zeroeth-order modified Bessel function of the first kind, and SI is the value of the synchronization index. The second parameter a_m is derived from both the firing rate and a function of the synchronization index as:

$$a_m = \frac{R}{I_0(g_m)} \quad \textbf{(10)}$$

where R is the average firing rate of a fiber responding to the same stimulus and level as above.

Although there is evidence that the Exponential Rate Model fits responses to pure tone stimuli (Colburn, 1973), no such evidence exists for our high-frequency transposed and SAM stimuli. This section shows empirical evidence that period histograms of AN fibers responding to SAM tones and transposed stimuli can also be fit by the Exponential Rate Model. We have fit

this model to the data shown in Figure 6 for pure-tone, SAM-tone and transposed-tone stimuli (Figure 24).

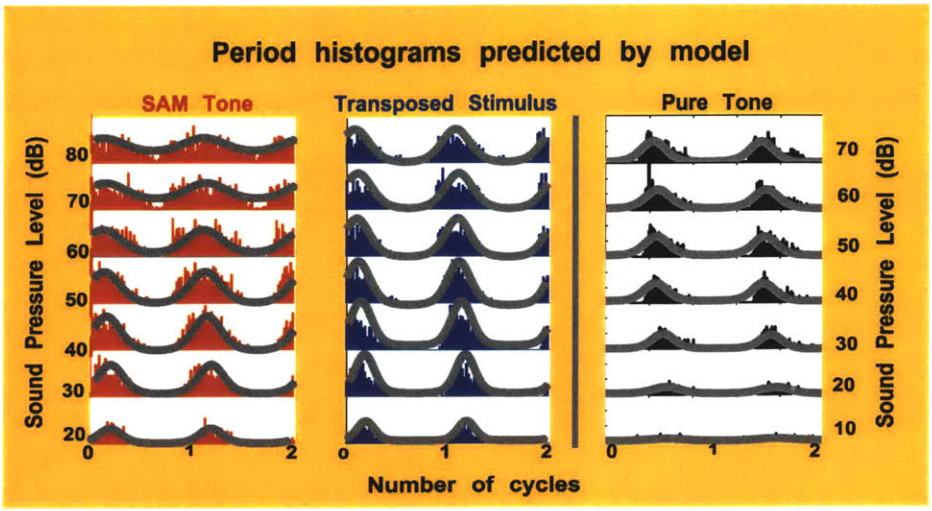

Figure 24: Period histograms and fit of exponential intensity functions for same stimuli as shown in Figure 6. Gray lines shows the fit of the exponential intensity function.

In all three cases, and especially when fibers are not saturated, the exponential intensity functions represent the data well. The discrepancy in all three cases arises because the exponential rate functions are symmetric, while the period histograms are not. AN fibers tend to respond strongly at the onset of each cycle, with waning responses after the stimulus has been on. Overall, the model does capture the data well in this example and throughout the data sample (not shown). The small differences between the model and the data become even less important after processing by the central binaural processor because this processing removes the asymmetry.

5.2.2. Model Part II: Central Binaural Processor

This section describes the use of the Colburn model of the binaural processor to obtain a prediction of system level lateralization performance. The structure of the binaural processor includes a series of delay lines and coincidence detectors. The inputs to the models of the binaural cells are carrying timing information from the auditory nerve from two fibers (one on each side) with the same CF_m. The binaural model cells modeled behave like coincidence detectors in that they respond maximally when the input spikes from two sides arrive within a

coincidence window, and are therefore very sensitive to inter-arrival times. One input from the fiber pair is delayed by an internal delay, and each pair is subject to every delay in the internal delay distribution. The binaural cells respond maximally when the internal delay matches the interaural delay in the stimulus. Given auditory spike timing information (firing rate and synchrony), the binaural model calculates the incremental time difference, or ITD JND, detected based on an assumed detectability metric *d'*, an assumed distribution of internal delays and a coincidence detector window.

5.2.2.1. Internal Delay Distributions

We modeled lateralization performance using both physiologically-based and uniform distributions. The uniform distribution of internal delays range from -2 to +2 ms, and is the same for all CF regions. This distribution extends over ½ cycle of the 125-Hz stimuli, and therefore always includes the best neuron for the stimuli. Since same internal-delay distributions are used for low-frequency and high-frequency stimuli, modeled lateralization performance will not be affected by differences in delay distribution across stimuli.

However, physiological data suggests that longer internal delays are much less prevalent than shorter delays and that the delay distribution depends on CF at low frequencies (Batra et al. 1993; Hancock and Delgutte, 2004, McAlpine et al., 2001). To incorporate this known physiology, we have also modeled lateralization performance with a physiologically-based distribution of best delays. The existing data on best delays in the IC shows different distributions for low-frequency pure-tone stimuli and high-frequency amplitude-modulated stimuli. Best delays at low frequencies are based on a model of delays in the inferior colliculus by Hancock and Delgutte (2004), who described the IPD distribution as the weighted sum of two Gaussians (Figure 25) as:

$$f_{LOW}(BP) = \frac{w}{\sigma_1\sqrt{2\pi}}\exp\left[\frac{(BP-\mu_1)^2}{2\sigma_1^2}\right] + \frac{1-w}{\sigma_2\sqrt{2\pi}}\exp\left[\frac{(BP-\mu_2)^2}{2\sigma_2^2}\right] \quad (11)$$

where w=.19, μ_1=.23, σ_1=.04, μ_2=.16, and σ_2=.19 and best interaural phase, BP, is equal to the product of best delay and the characteristic frequency of each unit. Because the best phase distribution is constant across CF, the best delay distribution scales inversely proportional to CF.

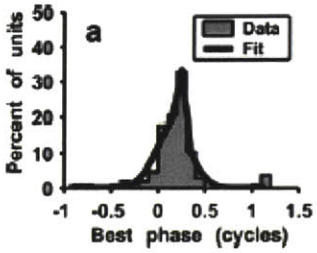

Figure 25: Distribution of best phase for delay-sensitive units. Grey bars represent data and black line is best lognormal fit. Figure from Hancock and Delgutte (2004)

In contrast, best delays from neurons responding to high-frequency amplitude modulated stimuli (CF above 2000 Hz) are distributed according to the distribution shown below regardless of characteristic frequency (Batra, et al., 1993):

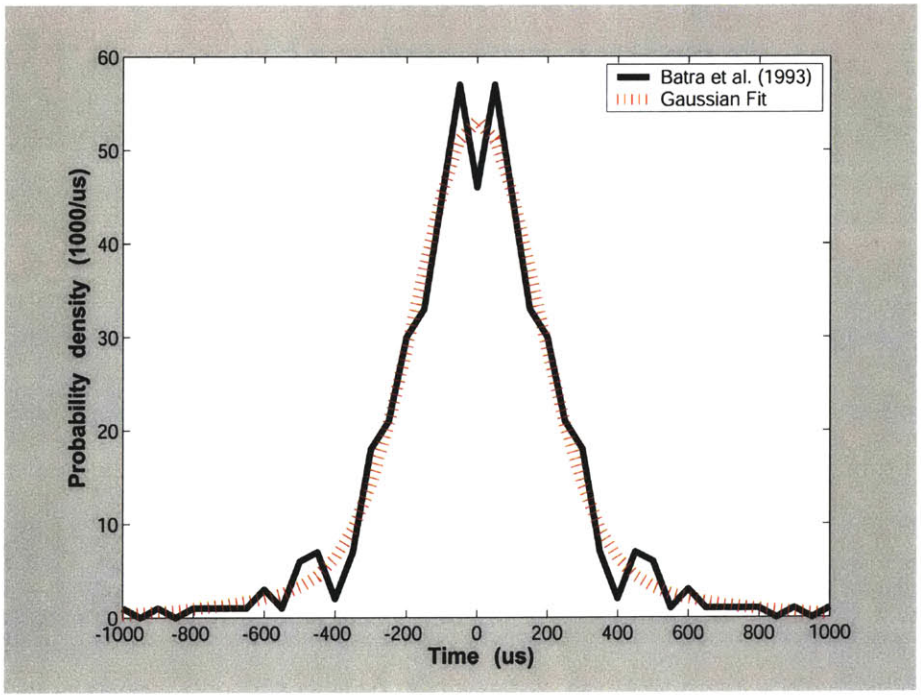

Figure 26: Distribution of best delays in the medial division of the ventral nucleus of the lateral lemniscus for neurons responding to binaural stimulation with CF>2000 Hz (from Batra et. al, 1993) (solid lines) and best fit of weighted sum of two Gaussians (dashed line).

Similarly to Hancock and Delgutte (2004), we parameterized this distribution as the sum of two Gaussians:

$$f_{HIGH}(t) = \frac{w}{\sigma_1\sqrt{2\pi}}\exp\left[\frac{(t-\mu_1)^2}{2\sigma_1^2}\right] + \frac{1-w}{\sigma_2\sqrt{2\pi}}\exp\left[\frac{(t-\mu_2)^2}{2\sigma_2^2}\right] \quad (12)$$

where w=.79, μ_1=0, σ_1=170, μ_2=0, and σ_2=430 and best interaural delay, t, in microseconds. The low-frequency distribution was used when calculating the lateralization performance for pure-tone stimuli and the high-frequency distribution was used to calculate the lateralization performance to SAM tone and transposed stimuli.

5.2.2.2. Lateralization Performance

Using the output of the peripheral models (one for each ear), the binaural processor then calculates the performance of each pair of fibers, $Q_{i,m}$ with internal delay τ_i and CF=CF_m. For the Exponential Rate Model, the performance of a pair of fibers can be calculated as follows:

$$Q_{i,m} = \Delta^2 4\pi^2 f_{\mathrm{mod}}^2 a_m^2 T \frac{\left(\int_{-\infty}^{\infty} c(x) 2g_m \sin\pi f_{\mathrm{mod}}(\tau_s - \tau_i - x) I_1\left[2g_m \cos 2\pi f_{\mathrm{mod}}(\tau_s - \tau_i - x)\right]dx\right)^2}{\int_{-\infty}^{\infty} c^2(x) I_0\left[2g_m \cos\pi f_{\mathrm{mod}}(\tau_s - \tau_i - x)\right]dx}, \quad (13)$$

where

Δ is the incremental time difference in the ITD to be detected;

T is the duration of the stimulus (500 ms);

τ_s is the base interaural delay (ITD) and 0 ms for our experiments;

τ_i is the internal delay;

f_{mod} is the modulation frequency (125 Hz);

$c(x)$ is the coincidence window, rectangular in shape, with a width of 50 μs.

Given the psychophysical experimentation conditions and neurophysiologically based assumptions, several simplifications can be made to the above formula. Psychophysical experiments measured the incremental time difference with the interaural delay at the midline (τ_s=0). The correlation window, $c(x)$, over which spike times are compared is modeled as

rectangular with a very short duration of 50 μs. Therefore, *c(x)* can be approximated as an impulse, integrating to its width ($\int c(x)dx$ =50 μs from Colburn, 1973). In addition, values of τ_m are derived from the internal delay distributions, discussed previously.

Total system performance, Q_{TOTAL}, is a function of both the internal delay distribution and the integral of performance of each pair of fibers at CF_m. We first combine performance $Q_m(\tau_i)$ across the distribution of internal delays, $f(\tau_i)$, as:

$$Q_m = \int Q_{i,m}(\tau_i) f(\tau_i) d\tau_i \quad \textbf{(14).}$$

If the performance score reaches a certain threshold, the processor declares that the ITD inherent in the stimulus is discernable. Next, we combined performance across CF as:

$$Q_{TOTAL} = \sum_m Q_m \quad \textbf{(15),}$$

with additivity justified by statistically independent Poisson processes. We sum performance for fibers in each SR group in physiologically-based proportions (Liberman, 1978). Next, we calculated the discrimitability *d'* as the square root of total performance,

$$d' = \sqrt{Q_{TOTAL}} \quad \textbf{(16).}$$

This quantity is proportional to Δ*JND*. To derive a measure of model performance equivalent to psychophysically measured ITD JNDs, we calculated Δ*JND*, the incremental time difference in the ITD to be detected so that *d'* is equal to 1, such that:

$$\Delta JND = \frac{d'_{critical}}{\sqrt{Q_{TOTAL}}} = \frac{1}{\sqrt{Q_{TOTAL}}} \quad \textbf{(17).}$$

This assumption is valid so long that the JND is small enough.

To model the psychophysical task, we calculated Δ*JND* for each of the stimulus levels used in the psychophysical experiment. However, it is well documented in the literature that

spike timings of individual fibers provide more than the necessary amount of information needed to produce the physically realizable range of lateralization performance. Given the uncertainties regarding the number of physiological delays and their distributions, we equated the JND for the model and the psychophysical data at 50 dB SPL for the pure tone and use the scaling factor to scale the model performance for all other stimuli. We selected this point because predicted and observed performance is relatively stable in this region.

In modeling the unrestricted and restricted psychophysical conditions, we assume that the auditory system makes use of information in the cochlear regions not masked by noise. We therefore modeled the unrestricted listening condition as responses of fibers with CF from 400 Hz to 20,000 kHz for transposed stimulus and SAM tone and 20 Hz to 20 kHz for the pure-tone stimulus, incorporating the bounds of the human cochlear-frequency map (Greenwood, 1990). We modeled the restricted listening condition with fiber responses from 3.6 kHz to 4.4 kHz for the transposed stimulus and SAM tone and 20 to 150 Hz for the pure tone stimulus.

5.3. Comparison of modeled with psychophysical performance

We modeled psychophysical performance using a physiologically based and a non-physiologically based distribution of best delays. Figure 27 shows a comparison of the observed and predicted performance for the unrestricted listening condition using the physiologically based distribution. In this condition, the model captures fairly well the stability of lateralization performance with sound pressure level, although listeners show slight improvement in performance at high levels that the model does not capture. However, the model predicts worse than observed performance for transposed stimuli and SAM tones relative to pure tones.

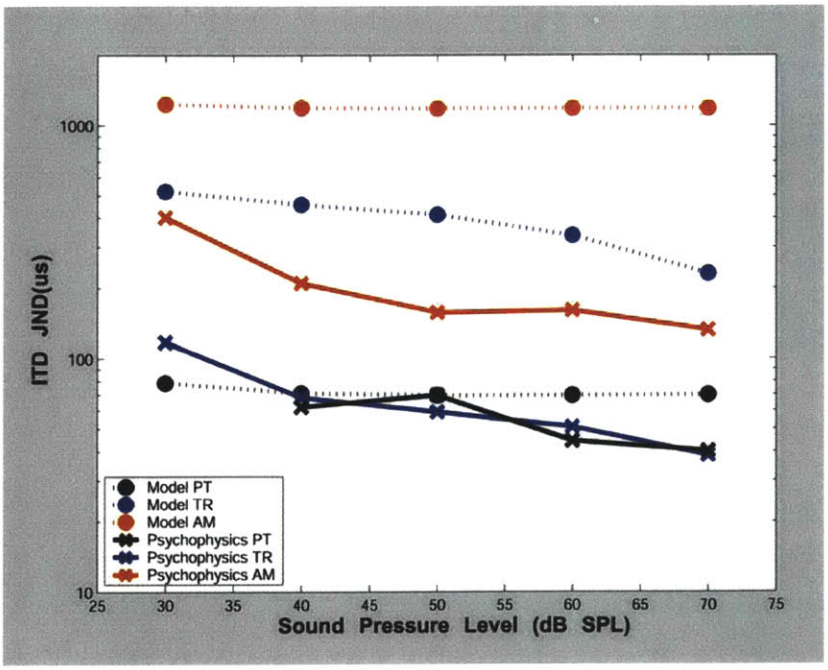

Figure 27: Comparison of psychophysical and modeled performance in the all-frequency listening condition. For the model, this condition was simulated by including fibers with a broad range of CFs (.4 kHz<CF<20 kHz for high-frequency stimuli and 20 Hz<CF<20 kHz for pure tones). Physiologically-based delay distribution is used.

Figure 28 shows the comparison of actual and predicted performance in the listening restricted band condition.

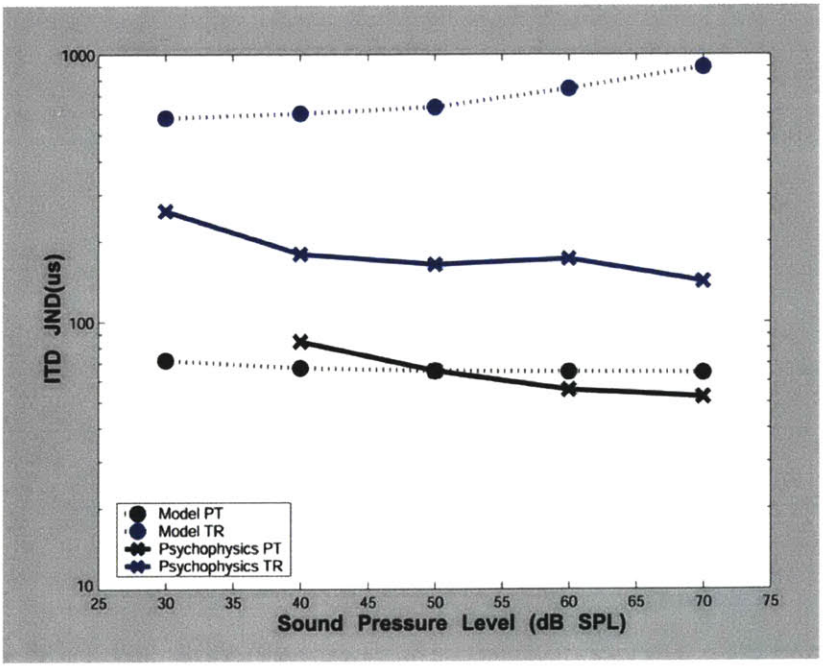

Figure 28: Comparison of psychophysical and model performance in the noise-restricted listening condition. For the model, this condition was simulated by only including fibers within a restricted range of CFs around the carrier frequency (3.6 kHz<CF<4.4 kHz for high-frequency stimuli and 20<CF<150 Hz for pure tones). Physiologically-based delay distribution is used.

Modeled performance differs from observed performance in two ways. First, the model predicts worse than actual transposed stimulus performance by almost an order of magnitude. Second, the model predicts a degradation in performance at high sound pressure

levels for transposed stimuli that is not seen psychophysically. SAM tones are not included in this comparison since subjects could not perform the psychophysical task within physically realizable limits of ITDs. Model results were consistent in that predicted JNDs were extremely large (not shown).

The distribution of internal delays in the binaural model is based on measurements of best ITD distributions in the IC neurons (Hancock and Delgutte, 2004; Batra et al., 1993). These distributions differ for low-CF and high-CF neurons. Consequently, Figure 27and Figure 28 show results computed with different internal delay distributions for modeling lateralization performance for pure tones and the high-frequency stimuli (transposed stimuli and SAM tones). Additional simulations with a uniform distribution of +/- 2 ms for both high-frequency and low-frequency neurons suggest that the distribution of internal delays, particularly differences in the distributions at low and high CFs, can significantly affect relative model performance for the low and high-frequency stimuli. These simulations are shown in Figure 29 for the unrestricted listening condition and in Figure 30 for the restricted listening condition.

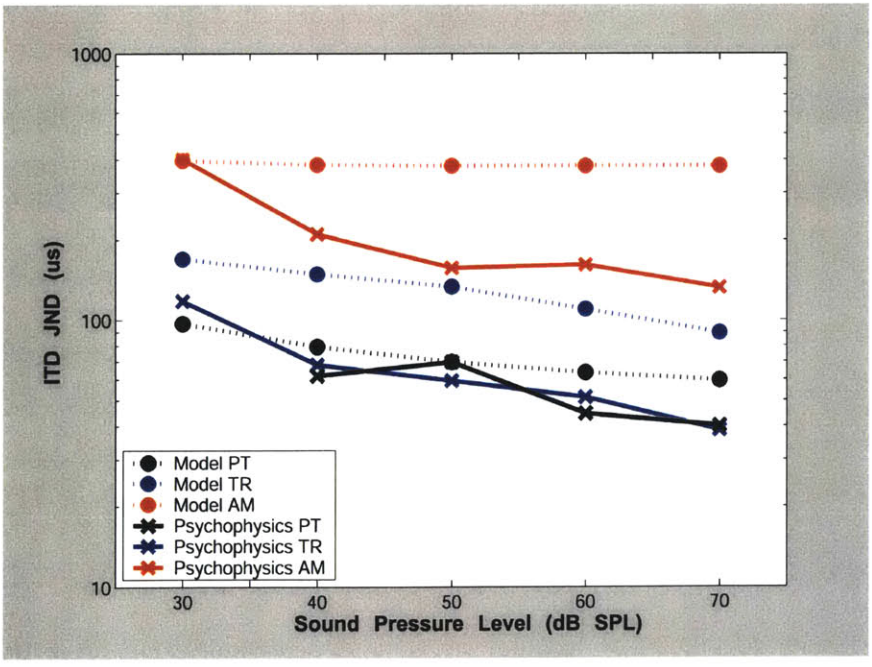

Figure 29: Same as Figure 27 but with identical uniform distribution (+/- 2ms) for high-CF and low-CF fibers.

In the unrestricted listening condition (Figure 29),

modeled lateralization performance for transposed stimuli is more similar to pure tone performance than with physiologically-based distributions, although the performance for transposed stimuli exceeds pure tone by nearly a factor of 2. This finding is also seen in Figure 30 for the restricted listening condition.

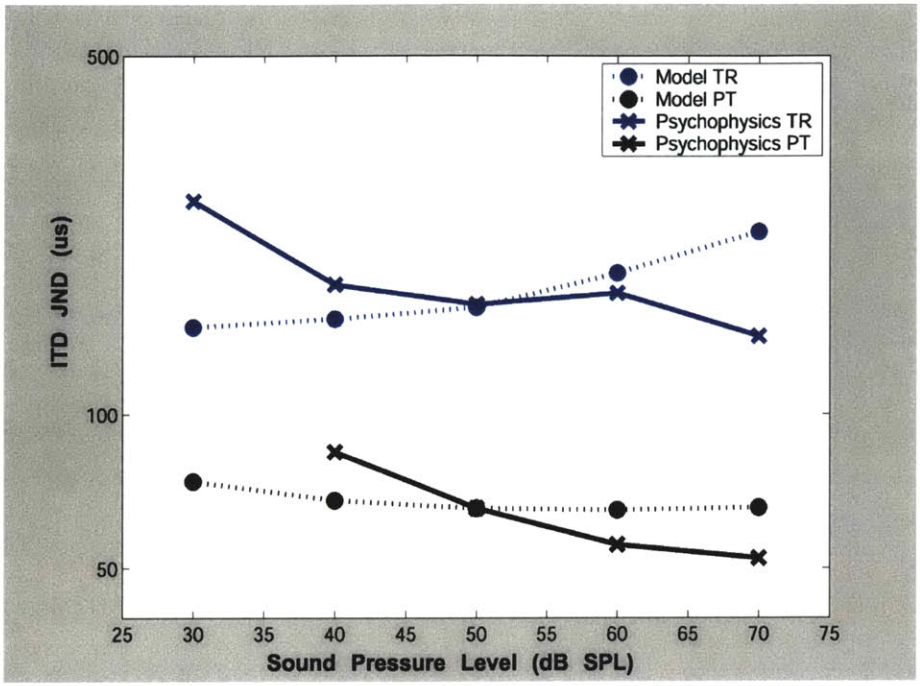

Figure 30: Same as Figure 28 but with identical uniform distribution (+/- 2 ms) for high-CF and low-CF fibers.

Although the model using the uniform internal delay distribution in the unrestricted condition predicts worse than measured performance for transposed stimulus and SAM tones, the discrepancy between model and measurement is not as great as when using the physiologically-based delay distributions. In fact, when the listening band is restricted, the model with uniform delay distributions correctly predicts the transposed performance relative to pure tone performance (Figure 30). This finding indicates that the discrepancy between modeled transposed and pure-tone performance arises partly from the physiological distribution and partly from the faster rate of phase locking with level for transposed stimuli. However, in the restricted listening condition using the uniform distribution, modeled performance for transposed stimuli

also degrades with level as with the physiologically based distribution. This degradation in performance is not seen psychophysically, even when the listening band is restricted. The consequences of this finding are discussed in the next section.

6. Summary and Discussion

The neural mechanisms underlying lateralization performance at high-frequencies have been approached from neurophysiological, psychophysical and modeling perspectives. Our physiological results suggests similar levels of phase locking to pure tones and transposed stimuli only for low stimulus levels close to fiber threshold. Phase locking to transposed stimuli and SAM tones falls faster than for pure tones, although phase locking to transposed stimuli always exceeds phase locking to SAM tones. On the contrary, no degradation in performance with level was seen psychophysically, even in when the listening band was restricted by masking noise. In fact, subjects typically showed a small improvement in lateralization performance with level. In the broadband listening condition, subjects lateralized transposed stimuli and pure tones equally well, with SAM tone lateralization performance worse than performance for pure tones and transposed stimuli. When the listening band was restricted, however, lateralization performance for transposed stimuli degraded relative to pure tone performance and SAM tones could not be lateralized at all. Using our temporal spike patterns from the auditory nerve and an ideal coincidence detector model of the binaural processor, the stability of lateralization performance with level could only be predicted in the unrestricted listening condition, but not the restricted listening condition. Although the model was able to predict the relative performance with pure tones and transposed stimuli in the restricted listening condition, modeled performance for transposed stimuli degraded at high stimulus levels, unlike psychophysical measurements.

6.1. Peripheral Model issues

6.1.1. Off-CF responses and linear filters

Synchrony level function for transposed stimuli and SAM tones were only presented when the carrier frequency of the stimulus matched the characteristic frequency of the fiber, and the model was based on these on-CF responses. Although cochlear filtering is generally non-linear, our use of linear filters in the model is justified in with on-CF responses. The rate-level and SI-level function used reflect the effects of the cochlear amplifier (Sachs and Abbas, 1974), because they are directly derived from AN responses at CF. However, in the unrestricted listening condition, fibers with CF above and below the f_c of the stimulus were also contributing to the overall lateralization performance. The off-CF rate-level and SI-level responses may differ from the on-CF curves used in the model because the cochlear amplifier only operates near CF. In a few cases, we measured SI-level functions above and below CF for SAM tones and transposed stimuli and observed similar level dependence with some rather subtle differences. In order to keep the model simple, we only use rate-level and SI-level functions from on-CF recordings. A more realistic model would include both on-CF and off-CF responses, by more explicitly incorporating the cochlear amplifier.

6.1.2. Compressive and Suppressive effects of Notched Noise

In the restricted listening condition, the model did not include any possible suppression or compression effects resulting from the notched noise used psychophysically. Noise can have two effects on AN rate-level and synchrony-level functions: suppression, manifested as a shift in response toward higher stimulus levels, and compression, resulting from an increase in spontaneous rate and a decrease in maximum firing rate (Palmer and Evans, 1982). Suppressive effects are expected to dominate at low noise levels, and act to improve performance by allowing

discrimination of intensities for which the fiber would otherwise be saturated. Compressive effects dominate at higher noise levels, and degrade performance since fibers have a smaller range of responses, and therefore less sensitivity. Tsai and Delgutte (1997) have found that rate-level and synchrony-level responses of AN fibers to pure tones experience dynamic-range shifts (suppression) in the presence of notched noise, while rate-level functions were also subject to compression. They found higher shifts and compression thresholds for low and medium SR fibers, effectively extending non-saturated fiber response towards higher stimulus levels. Winslow and Sachs (1985) show that the activation of the olivocochlear bundle (most likely inhibited by the anesthesia in our preparation) can partly counteract rate compression at high noise levels. Another factor is that neurons at the edge of the noise band may partly respond to the tone and therefore contribute to performance, although we do not include these in our model. Since compression and suppression are nonlinear, and not enough data exists to capture their effect in the context of this model, we use a simpler model. However, we acknowledge that a more comprehensive model would include the compressive and suppressive effects of the notched noise in predicting lateralization performance in the restricted listening condition.

6.1.3. Rate-Synchrony interactions

The interaction between rate-level functions and the level at which the SI maximum occurs also influences the dependence of modeled performance on level. It is possible that changing the location of the level maximum in the SI-level function by just a few dB such the maximum would occur directly in the linear region of rate-level function would tend to improve performance. However, these variations in SI-rate interactions would primarily affect modeled performance at low-stimulus levels, where psychophysical performance was not measured.

6.2. Binaural Model Issues

6.2.1. Crosscorrelation (MSO) vs. Excitation/Inhibition mechanism (LSO)

The model uses a coincidence detection mechanism to model the binaural processor for all stimulus types. This mechanism is consistent (Goldberg and Brown, 1969; Smith, 1995; Grothe and Sanes, 1994; Rose et al, 1966; Yin and Kuwada, 1983) with the low-frequency MSO (Yin and Chan, 1990) where ITD processing for low-frequency pure tones is likely to occur. On the other hand, ITD sensitivity at high frequencies may be processed in the high-frequency LSO nucleus (Joris and Yin, 1995). The LSO uses a different excitation/inhibition (IE) mechanism, such that LSO cells are excited by inputs to from the ipsilateral ear and inhibited by inputs from the contralateral ear (Joris and Yin, 1995). However, our model is an optimal binaural processor that quantifies the ITD information available in AN regardless of the underlying mechanisms for extracting this information. In addition, the LSO circuit is less developed than the MSO circuit in humans (Adams, personal communication) so that an MSO role for processing of high-frequency ITDs cannot be ruled out in humans. Thus, our choice of a model based on coincidence detection seems justified.

6.2.2. Delay Distribution and Species Differences

The physiologically-based distributions of internal delay in the model differed for high-CF and low-CF fibers, consistent with available data from the IC. Consequently, pure tone responses from low-frequency neurons were subject to a different delay distribution from those of high-frequency neuron responses to SAM tones and transposed stimuli. The low-frequency distribution of best delays is bimodal, while the high-frequency distribution is unimodal. Moreover, the low-frequency distribution scales inversely with CF, while the high-frequency distribution is constant. When a uniform distribution of internal delays was substituted for the

physiological distributions, the discrepancy between modeled and measured lateralization performance decreased, indicating that differences in internal delay distributions between high and low CF fibers contributed to differences in performance between pure tones and transposed stimuli (and SAM tones). Species differences and anesthesia could account for some, but most likely not all, differences in low-frequency and high-frequency distributions as the low-frequency distribution was measured in anesthetized cats while high-frequency distributions were from awake rabbits. Our model is intended to be a human model, and a human delay distribution would be preferred to those in cats and rabbits. Unfortunately, there is no direct information about human distributions. However, since delay distributions do not appear to scale with head size (McAlpine et al., 2001; Joris et al., 2004; Hancock and Delgutte, 2004), the human distributions may not greatly differ from those used in our model.

When modeled using the same uniform distribution of fibers for low and high-CF fibers, measured and modeled performance was more similar than with the physiological distributions. However, even the model with uniform distributions predicted worse than measured performance for transposed stimuli for all levels in the unrestricted listening condition and for high levels in the restricted listening condition. These findings suggest that the worse-than-measured performance using the physiologically-based distribution of best ITDs is not entirely a result of using different delay distributions for low and high-CF fibers, but also a result of degrading phase locking with level to the high-frequency stimuli.

6.3. Solutions for Dynamic Range Problem

Model results suggest that phase locking in the auditory nerve can account for the robust psychophysical performance with level for pure tones and transposed stimuli in the unrestricted listening condition, but cannot account for stability of performance in the restricted listening

condition because phase locking to transposed stimuli degrades with increasing level for fibers tuned to the carrier frequency. The stable performance of the model in the unrestricted listening condition relies on the skirts of the excitation patterns where there are unsaturated neurons with good phase locking. However, the stability with level of psychophysical performance in the restricted listening condition seems to rule out the idea that this mechanism is the only one contributing to good performance at high SPLs. We are thus faced with a "dynamic range problem" in ITD sensitivity to the envelope of high-frequency sounds similar to the classic dynamic range problem in the coding of intensity. The problem arises at high stimulus levels where AN fibers are saturated but psychophysical performance does not degrade. Some neuron types in the cochlear nucleus (e.g. Onset chopper (O_c) and pauser buildup (P/B) neurons)have a wider dynamic range than auditory nerve fibers for amplitude modulation encoding. Such neurons may help maintain good performance for high-frequency stimuli at higher sound pressure levels (Rhode and Greenberg, 1994). The effect of notched noise on AM encoding needs to be examined in theses neurons to evaluate whether responses are robust in this condition. While these neurons are attractive to solve the dynamic range problem, they either project to the IC or form local circuits in the CN, and therefore do not play a role in binaural processing in the LSO and MSO. Another possible mechanism for solving the dynamic range problem includes disproportionate weighting of low-SR responses given their large dynamic range and greater response shifts from compressive effects of notched noise (Tsai and Delgutte, 1997; Delgutte, 1987). However, low-SR fibers are relatively few in the AN (Liberman, 1978) and there is no strong evidence for such a selective weighting mechanism centrally. The neuronal mechanisms underlying stable perceptual sensitivity to ITD in transposed stimuli with a restricted listening band remains an open question.

Appendix I: Detectability thresholds

This section describes measurements of detectability thresholds used in the psychophysical experiments. The first set of data supplements the monaural screening hearing test. This data is used to (1) ascertain normal binaural levels of hearing (2) ascertain that stimuli levels presented are above listeners' detection thresholds in quiet. Pure tone detection thresholds were measured at 125 Hz, corresponding to pure-tone presentations, and at 4000 Hz corresponding to high-frequency stimuli presentations in the lateralization study. Data shown in Figure 31 some variability in detection thresholds for the 125 Hz pure tone across subjects, with S2 and S3 performing at normal levels, S4 performing some what better and S1 somewhat worse than normal.

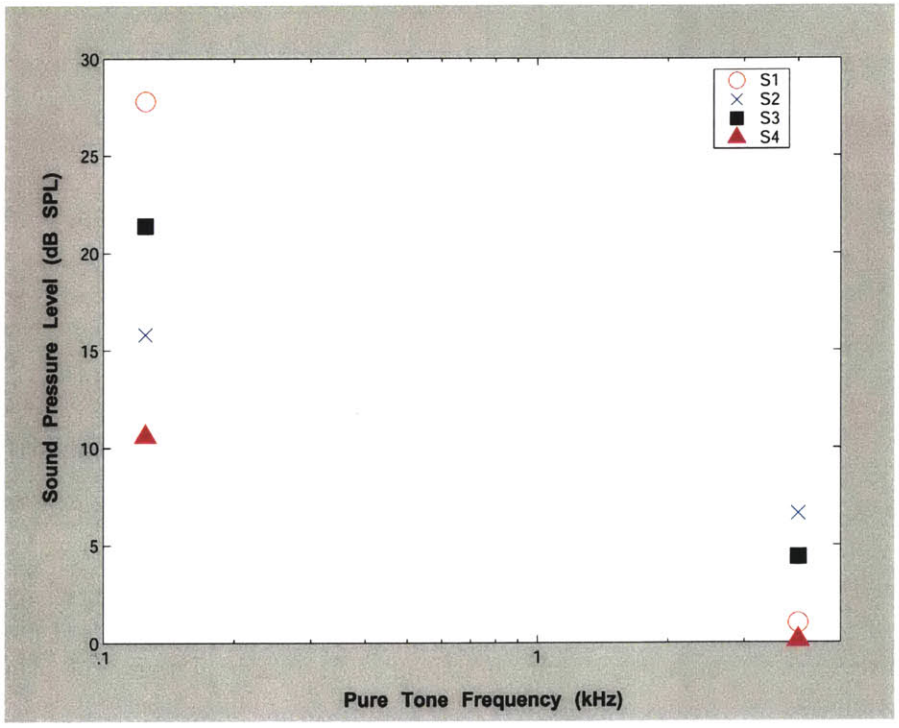

Figure 31: Pure tone thresholds in quiet for a pure tone of frequency 125 Hz and 4000 Hz. Data shown is averaged pure tone diotic threshold in quiet for each of the four listeners.

This elevated threshold for subject S1 contribute to his inability to perform the lateralization task with pure tones at signal levels of 30 dB SPL. The subject's monaural thresholds were elevated by about 10 dB in the low-frequency range in both ears. However, these differences are not

crucial to the use of this subject in the experiment since monaural thresholds are the same in both ears and that some (up to several dB) of the increase over normal levels could be accounted for by the leakage of the speakers at low frequencies. Thresholds for transposed stimuli and SAM tones would be expected to be comparable to pure tone thresholds at 4 kHz. Based on the measured thresholds, all listeners should be able to perform lateralization task on transposed stimuli and AM tone at the lowest level of 30 dB SPL.

The second set of threshold measurements determines the noise level in restricted listening condition. The level of noise needed to just mask pure tones and the high-frequency stimuli was measured with high pass filtered noise between 150 Hz and 12000 Hz in the presence of a 125 Hz pure tone and with notched noise between 0 and 3600 Hz and 4400 and 12000 Hz in the presence of a 4 kHz pure tone. We measured pure tone thresholds in the presence of broadband noise with both pure tone and noise of the same duration as those used in the lateralization experiments. Pure tone threshold measurements are shown in Figure 32.

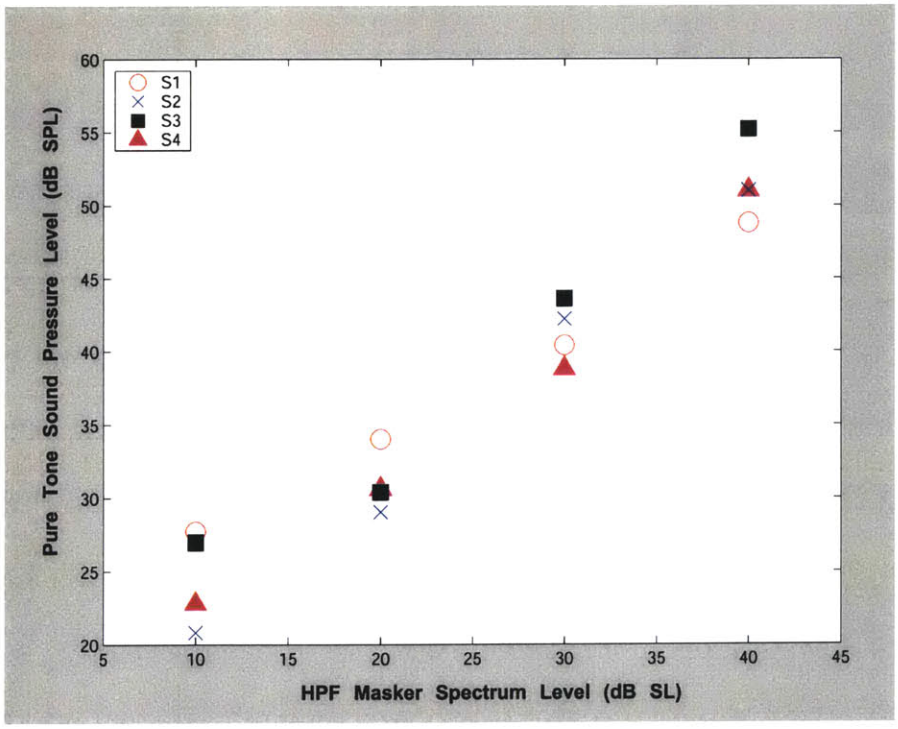

Figure 32: Broadband noise masker thresholds for detection of pure tone at 125 Hz averaged across two trials for each listener. Data is shown for the same subjects that completed the lateralization task.. Notched noise used had content between 150 and 12000 Hz. Masker and signal durations and ramps are same as those specified for the lateralization experiment in METHODS.

The data shows a fairly linear relationship between the noise and pure tone signal levels. Because the trends and levels of broadband masker are fairly similar across subjects, this data is averaged across all subjects for the lateralization study. Pure tone masker levels in the restricted band listening condition are presented 20 dB below the average of the values shown in Figure 32. It is expected that high pass filtered thresholds would be on the order of 10 dB above those shown because listeners would obtain a significant amount of information about the tone in the frequencies between 0 and 150 Hz. The upper frequency cutoff of the pure-tone masker was decreased in the lateralization study since noise intensity at high stimulus levels would have been uncomfortably loud. However, the signal-to-noise ratio is maintained such that responses of fibers beyond the noise in the pure-tone condition are sufficiently masked by the noise.

Figure 33 shows just detectible signal levels (with pure tone at f=4000 Hz) in notched noise that were used to noise masker levels for the high-frequency stimuli in the restricted listening condition.

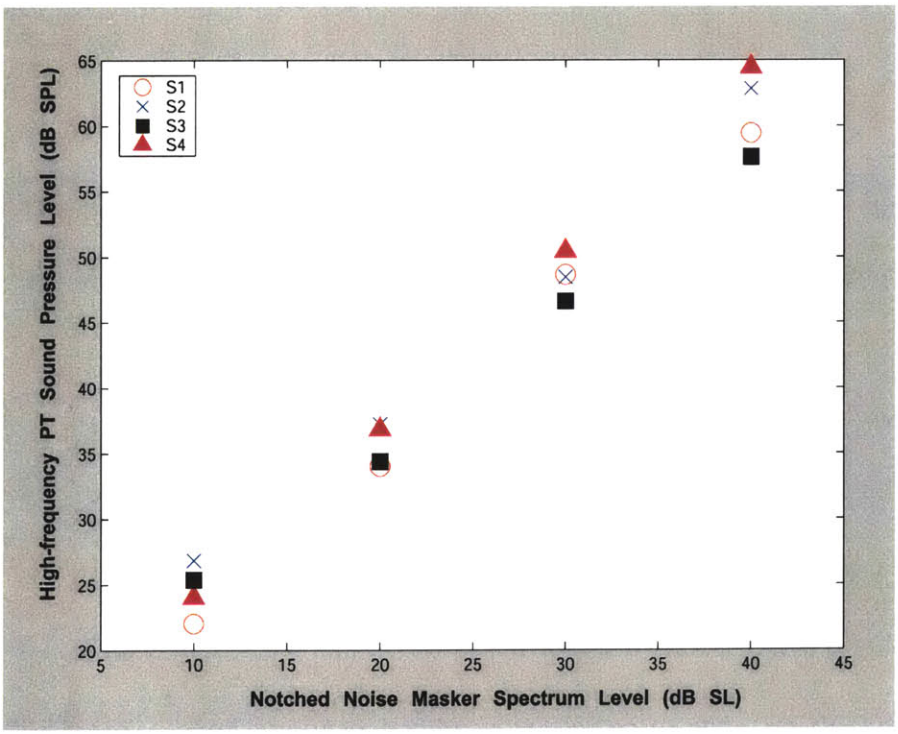

Figure 33: High-frequency pure tone level (frequency = 4000 Hz) that is just masked by notched noise. Each data point shown is averaged over 2 repetitions of the experiment, with no single difference between experiments per subject greater than 3 dB SPL. Notched noise used had content between 0 and 3600 Hz and 4400 and 12000 Hz. Masker and signal durations and ramps are same as those specified for the lateralization experiment in METHODS.

Each subject has a different, but fairly linear, relationship between the just audible signal level and masker spectrum level. That the slope of the best fit line for each subject is greater than unity is consistent with high-side suppression, where the signal level grows faster than the masker level (Ruggero et al, 1992).

Appendix 2: Empirical evidence for Poisson process fit for high-frequency stimuli

If the inputs from the two ears are statistically independent Poisson processes with the same intensity, the cross-correlation is equivalent to the SAC. Therefore, the fit of the model output to the SACs is more important than the fit of the intensity functions examined in Figure 24. Figure 34 compares the fit of binaural processor to the SACs shown in Figure 10.

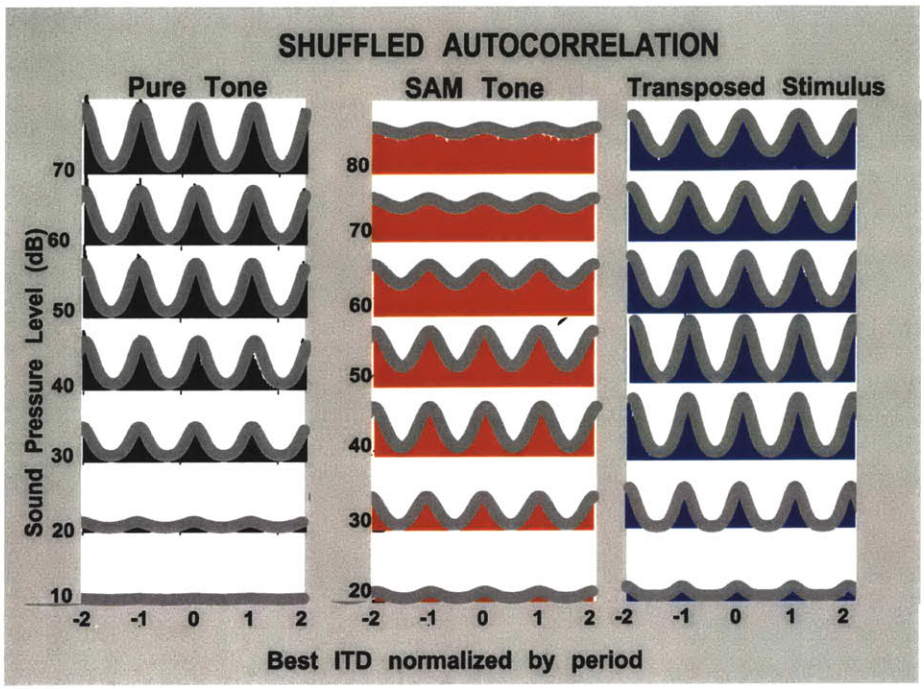

Figure 34: Shuffled autocorrelation histograms and fits by autocorrelation of exponential intensity functions. SAM tone results shown in left panels and transposed stimuli shown in right panel for same fiber as in Figure 24. Model fits are shown in red.

The model fit to the SACs is excellent. Whereas the intensity function fits suffered due to their inherent symmetry, the SAC is inherently symmetric. The model fit shown for the example fiber parallels the fit of the model for all the fibers in the data ensemble (not shown).

Next, we examined whether the autocorrelation functions are comsistent with Poisson processes for the high-frequency stimuli. Poisson processes are characterized by equality of the mean and variance of the process statistics. For each pair of stimulus presentations, the mean and variance were empirically derived from crosscorrelation (Figure 35).

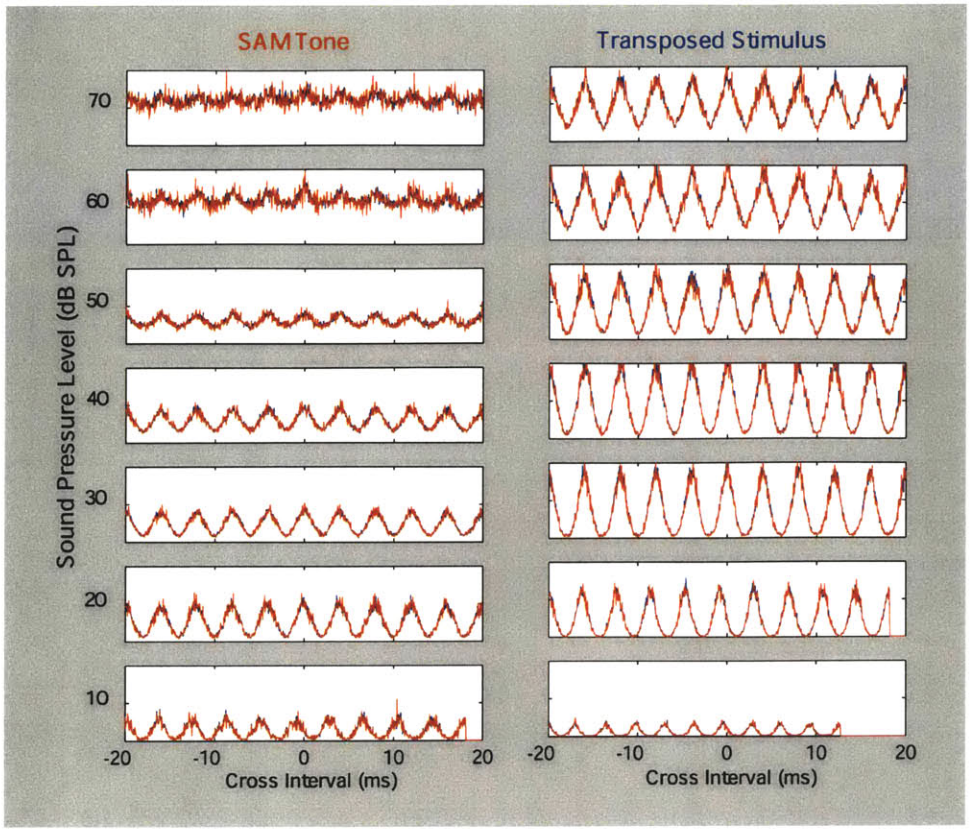

Figure 35: Mean and variance across shuffled stimulus trial number. Means are shown in blue and variance in red. Equality of the mean and variance implies a Poisson distribution. Left: SAM tone. Right: Transposed stimulus.

Figure 35 shows that the mean and the variance for the SAC across all pairs of trials are nearly identical. This result gives strong evidence that the autocorrelation of spikes can be modeled as a Poisson process for these high-frequency stimuli, and these observations hold for the ensemble of fibers sampled. The Poisson nature of the process satisfies this assumption in the Colburn optimal processor model.

References

Batra, R, Kuwada, S. and Stanford T.R. (1993). "High-frequency neurons in the inferior colliculus that are sensitive to interaural delays of amplitude-modulated tones: evidence for dual binaural influences." *J. Neurophysiol.* 70: 64-80.

Beckius, G.E., Batra, R., Oliver, D.L. (1999). "Axons from anteroventral cochlear nucleus that terminate in medial superior olive of cat: observations related to delay lines." *J. Neurosci.* 19: 3146-3161.

Bernstein, L.R., and Trahiotis, C. (1985). "Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities," *J. Acoust. Soc. Am.* 77: 1868-1880.

Bernstein, L.R., and Trahiotis, C. (1994). "Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise," *J. Acoust. Soc. Am.* 95: 3561-3567.

Bernstein, L.R. (2001). "Auditory processing of interaural timing information: new insights," *J. Neuroscience Research* 66: 1035-1046.

Bernstein, L.R., and Trahiotis, C. (2002). "Transposed stimuli improve sensitivity to envelope-based interaural timing information for stimuli having center frequencies of up to 10 kHz," *J. Acoust. Soc. Am.* 111: 2466-2466.

Bernstein, L.R., and Trahiotis, C. (2003). "Enhancing interaural-delay-based extents of laterality at high frequencies by using 'transposed stimuli,'" *J. Acoust. Soc. Am.* 113: 3335-3347.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (2002). "Precise inhibition is essential for microsecond interaural time difference coding," *Nature.* 417:543-547.

Carr, C.E., and Konishi, M. (1988). "Axonal delay lines for time measurement in the owl's brainstem," *Proc. Natl. Acad Sci. USA.* 85(21): 8311–8315.

Carr, C.E., and Konishi, M. (1990). "A circuit for detection of interaural time differences in the brain stem of the barn owl," *J. Neurosci.* 10(10): 3227-3246.

Colburn, H.S. (1973). "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination," *J. Acoust. Soc. Am.* 54: 1458-1470.

Colburn, H.S., and Esquissaud, P. (1976). "An auditory-nerve model for interaural time discrimination of high-frequency complex stimuli," *J. Acoust. Soc. Am. Suppl.* 1 59: S23.

Delgutte, B. (1987). "Peripheral auditory processing of speech information: implications from a physiological study of intensity discrimination," *In the Psychophysics of Speech Perception*, Schouten, M.E.H, ed. Boston: Martinus Nijhoff Publishers, 333-353

Delgutte, B. (1988). "Physiological mechanisms of masking," In H. Duifhuis, J.W. Horst, and H.P. Wit (Eds.), *Basic Issues in Hearing*. London: Academic , 204-214.

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

Goldberg, J.M., and Brown, P.B. (1969). Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J. Neurophysiol.* 32: 613-636.

Goldstein, J.L. (1990). "Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering," *Hear. Res.* 49:39-60

Green, D.M. (1958). "Detection of multiple component signals in noise," *J. Acoust. Soc. Am.* 30: 904-911.

Green, D.M., Birdsali, T.G, and Tanner, W.P. Jr. (1957). "Signal detection as a function of signal intensity and duration," *J. Acoust. Soc. Am.* 29: 523-531.

Greenwood, D. (1990). "Critical bandwidth and consonance in relation to cochlear frequency-position coordinates," *Hear Res.* 54: 164-208.

Grothe, B., and Sanes, D.H. (1994). "Synaptic inhibition influences the temporal coding properties of medial superior olivary neurons: an in vitro study," *J. Neurosci.* 14: 1701-1709.

Hancock K., and Delgutte, B. (2004). "A physiologically based model of interaural time difference discrimination," *J. Neurosci.* 24: 7110-7117.

Helfert, R. H. and Aschoff, A. (1997). "Superior olivary complex and nuclei of the lateral lemniscus." In *The Central Auditory System,* ed. Ehret, G. & Romand, R., pp. 193-258. Oxford University Press, New York.

Henning, G.B. (1974). "Detectability of interaural delay in high-frequency complex waveforms," *J. Acoust. Soc. Am.* 55: 84-90.

Hershkowitz, R.M., and Durlach, N.I. (1969). "Interaural time and amplitude JNDs for a 500-Hz tone," *J. Acoust. Soc. Am.* 46: 1464-1467.

Hudspeth, A.J., and Corey, D.P. (1977). "Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli," *Proc. Natl. Acad. Sci. USA*, 74: 2407-2411.

Javel, E., McGree, J.A., Horst, W., and Furley, G.R. (1968). "Temporal mechanisms in auditory stimulus coding." *In* G.M. Edelman, W.E. Gall, and W.M. Cowan (Eds.), *Auditory Function: Neurological Bases of Hearing* (pp. 515-558). Wiley: New York.

Jeffres, L.A. (1948). "A place theory of sound localization," *J. Comp. Physiol. Psychol.* 41: 35-39.

Johnson, D. (1980). "The relationship between spike rate and synchrony in responses of auditory nerve fibers to single tones," *J. Acoust. Soc. Am.* 68: 1115-1122.

Joris, P.X. (1996). "Envelope coding in the lateral superior olive. II. Characteristic delays and comparison with responses in medial superior olive," *J. Neurophysiol.* 76: 2137-2156.

Joris, P.X., Carney, L.H., Smith, P.H., and Yin, T.C.T. (1994). "Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency," *J. Neurophys.* 71: 1022-1036.

Joris, P.X., Heijden, M., Louage, D., van de Sande B., and van Kerckhoven, C. (2004) "Dependence of binaural and cochlear 'best delays' on characteristic frequency," *In Auditory signal processing: Physiology, psychoacoustics, and models*, ed. Pressnitzer, D., de Cheveigne, A., McAdams, S., and Collet, I. New York: Springer-Verlag.

Joris, P.X., Smith, P.H., and Yin, T.C.T. (1998). "Coincidence detection in the auditory system: 50 years after Jeffres," *Neuron* 21: 1235-1238.

Joris, P.X., and Yin T.C.T. (1992). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. 91: 215-232.

Joris, P.X., and Yin T.C.T. (1995). "Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences," *J. Neurophysiol. 73: 1043-1062*

Levitt, H. (1970). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* 49: 467-477.

Leakey, D.M,, Sayers, B.McA,. and Cherry, C. (1958). "Binaural fusions of low- and high-frequency sounds," *J. Acoust. Soc. Am.* 30: 222(L).

Liberman, M.C. (1978). "Auditory-nerve response from cats raised in a low-noise chamber," *J. Acoust. Soc. Am.* 63: 442-455.

Liberman, M.C. (1982). "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," *J. Acoust. Soc. Am.* 72: 1441-1449.

Louage, D.H.G., van der Heijden, M., and P.X. Joris. (2004). "Temporal Properties of Responses to Broadband Noise in the Auditory Nerve." *J. Neurophysiology.* 91: 2051-2065.

Kiang, N.Y., and Moxon, E.C. (1974). "Tails of tuning curves of auditory-nerve fibers," *J Acoust Soc Am.* 55, 620-30.

Kiang, N.Y.S., Wantanabe, T., Thomas, E.C., and Clark, L.F. (1965). "Discharge patterns of single fibers in the cat's auditory nerve" (MIT, Cambridge), Research Monograph No. 35.

Knudsen, E.I., and Konishi, M. (1979). "Mechanisms of sound localization in the barn own (Tyto alba). *J. Comp. Phys. A.* 133:13-21.

Koppl, C. (1997). "Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, Tyto alba," *J. Neurosci.* 17: 3312-3321.

MacPhereson, E.A., and Middlebrooks, J.C. (2002). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *J. Acoust. Soc. Am.* 111: 2219-2236.

Manis, P.B., and Marx, S.O. (1991). "Outward currents in isolated ventral cochlear nucleus neurons," *J. Neurosci.* 11: 2865-2880.

McAlpine, D., Jiang, D. and Palmer, A.R. (2001). "A neural code for low-frequency sound localization in mammals," *Nat. Neurosci.* 4: 396-401

McFadden, D., and Pasanen, E.G. (1976). "Lateralization at high frequencies based on interaural time differences," *J. Acoust. Soc. Am.* 59: 634-639.

Moore, B.C.J. (2003). *An Introduction to the Psychology of Hearing*. Ed. 5. London: Academic Press. *J. Acoust. Soc. Am.* 70: 1003-1014.

Moore, B.C.J., and Glasberg, B.R. (1990). "Effects of signal-to-noise ratio on the frequency discrimination of complex tones with overlapping or nonoverlapping harmonics," *J. Acoust. Soc. Am.* 89: 2858-2865.

Nuetzel, J. M. and Hafter, E.R. (1976). "Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration," *J. Acoust. Soc. Am.* 60: 1339-1345.

O'Loughlin, B.J., and Moore, B.C.J. (1981). "Off-frequency listening: Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking," *J. Acoust. Soc. Am.* 69: 1119-1125.

Overholt, E.M., Rubel, E.W., and Hyson, R.L. (1992). A circuit for coding interaural time differences in the chick brainstem." *J. Neurosci.* 12: 1698-1708.

Palmer, A.R., and Evans, E.F. (1982). "Intensity coding in the auditory periphery of the cat: Responses of cochlear nerve and cochlear nucleus neurons to signals in the presence of bandstop masking noise," *Hear. Res.* 7: 305-323.

Patterson, R.D., (1976). "Auditory Filter Shapes Derived with Noise Stimuli," *J. Acoust. Soc. Am.* 59: 640-654.

Raman, I.M., Zhang, S., and Trussell, L.O., (1994). "Pathway-specific variants of AMPA receptors and their contribution to neuronal signaling," *J. Neurosci.* 14(8): 4998-5010.

Rayleigh, L. (1907). "On our perception of sound direction," *Phil. Mag.* (Ser. 6) 13:214-232.

Rhode, W.S., and Greenberg, S. (1994). "Encoding of amplitude modulation in the cochlear nucleus of the cat," *J. Neurophysiology.* 71: 1797-1825.

Rhode, W.S., and Recio, A. (1971). "Study of mechanical motions in the basal region of the chinchilla cochlea," *J. Acoust. Soc. Am.* 107: 3317-3332.

Rhode, W.S., and Smith, P.H. (1986). "Encoding timing and intensity in the ventral cochlear nucleus of the cat," *J. Neurophys.* 56: 261-286.

Rose, J.E., Gross, N.B., Geisler, C.D., and Hind, J.E. (1966). "Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source," *J. Neurophysiol.* 29: 288-314.

Rouiller, E.M., Cronin-Schreiber, R., Fekete, D.M., Ryugo, D.K, (1986). "The central projections of intracellularly labeled auditory nerve fibers in cats: an analysis of terminal morphology," *J. Comp. Neurol.*: 249: 261-278.

Ruggero, M.A., Robles L., and Rich, N.C. (1992). "Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression." *J. Neurophysiol.* 68: 1087-1099.

Sachs, M.B., Winslow, R.L., Sokolowski B.H.A. (1989). "A computational model for rate-level functions from cat auditory-nerve fibers," *Hear Res.* 41: 61-70.

Shera, C.A., Guinan, J.J., and Oxenham, A.J. (2002). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements." *Proc Natl Acad Sci U S A.* 99(5): 3318–3323.

Siebert, W.M. (1970). "Frequency Discrimination in the auditory system: place or periodicity mechanisms?" *Proc of IEEE.* 58:723-730

Smith R.L. (1988). "Encoding of Sound Intensity by Auditory neurons," in *Auditory Function: Neurobiological Basis for Hearing,* edited by G.M. Edelman, W.E. Gall, and W.M. Cowan (The Neurosciences Institute, Wiley, New York), pp. 243-273.

Smith, P.H., Joris, P.X., and Yin, T.C.T. (1993). "Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: Evidence for delay lines to the medial superior olive," *J. Comp. Neurol.*, 331: 245-260.

Trussell, L.O. (1999). "Synaptic mechanisms for coding timing in auditory neurons," *Annu. Rev. Physiol.* 61: 477-496.

Tsai, E.J., and Delgutte, B. (1997). "Neural mechanisms underlying intensity discrimination: responses of auditory-nerve fibers to pure tones in band-reject noise," *Assoc. Res. Otolaryngol. Abstr.* 20:154.

Van de Par, S. and Kohlrausch, A. (1997). "A new approach to comparing binaural masking level differences at low and high frequencies," *J. Acoust. Soc. Am.* 101: 1671-1680.

Wakeford, O.S., and Robinson, D.E. (1974). "Lateralization of tonal stimuli by the cat," *J. Acoust. Soc. Am.* 55: 649-652.

Winslow, R.L., Sachs, M.B. (1988). Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise and with stimulation of the crossed olivocochlear bundle," Hear. Res. 35: 165-190.

Yin T.C.T., and Kuwada, S. (1983). "Binaural interaction in low-frequency neurons in inferior colliculus of the cat. III. Effects of changing frequency," *J. Neurophysiol.* 50: 1020-1042.

Yin T.C.T., and Chan, J.C.K. (1990), "Interaural time sensitivity in medial superior olive of cat," *J. Neurophysiol.* 64: 465-488.

Young, S.R., and Rubel, E.W. (1983). "Frequency-specific projections of individual neurons in chick brainstem auditory nuclei." *J. Neurosci.* 3: 1373-1378.

Zhou, Y., Carney, L.H., and Colburn, H.S. (2005). "A Model for Interaural Time Difference sensitivity in the Medial Superior Olive: Interaction of excitatory and inhibitory synaptic inputs, channel dynamics, and cellular morphology," *J. Neurosci.* 25(12): 3046-3058.

Zwislocki, J., and Feldman, R.S. (1956). "Just noticeable differences in dichotic phase, "*J. Acoust. Soc. Am.* 28: 860-864.