Effect of reverberation on the directional sensitivity of auditory neurons: Peripheral factors

by

Andrew H. Schwartz

B.S. Computer Systems Engineering
Boston University, 2005

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2010

Signature of author:.....
Andrew Schwartz
Department of Electrical Engineering and Computer Science
May 7, 2010

Certified by:.....
Bertrand Delgutte
Senior Research Scientist
Thesis Supervisor

Accepted by:.....
Professor Terry P. Orlando
Chairman, Department Committee for Graduate Studies

Effect of reverberation on the directional sensitivity of auditory neurons: Peripheral factors

by

Andrew H. Schwartz

Submitted to the Department of Electrical Engineering on May 7, 2010 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Electrical Engineering

ABSTRACT

Reverberation poses a challenge for theories of sound localization due to the interaction between the direct sound and the various acoustic reflections. These reflections corrupt binaural cues available to the receiver, resulting in a degradation of directional information available in the acoustic stimulus. Despite this interaction, directionally-sensitive neural responses in the auditory midbrain have been shown to be more robust to reverberation than predicted by a binaural model based on the long- term cross-correlation of the two ear-input signals (Devore et al., 2009, *Neuron* 63(1), pp 123-134).

To determine the extent to which this robustness is central or peripheral in origin, and to quantitatively investigate whether peripheral adaptation contributes to this robustness, we recorded auditory nerve (AN) responses to tokens of noise with varying levels of simulated reverberation. We found many qualitatively similar trends in AN responses as have been previously observed in the midbrain, suggesting a peripheral origin of robust directional coding. In particular, we found that degradation of directional coding in the AN due to reverberation is stronger at high frequencies, and that this degradation is limited near the stimulus onset. We also show that peripheral adaptation plays a positive role in increasing robustness of directional representation in the presence of reverberation.

We also investigated the nature of the frequency dependence observed in the degradation of directional coding due to reverberation. Based on our experimental results and the results of an auditory model, we argue that in addition to being affected by AN fibers' synchrony to stimulus fine structure, the frequency dependence is also produced by temporal structure of the reverberant room response.

Thesis Supervisor: Bertrand Delgutte
Title: Senior Research Scientist

I. Motivation

In ideal anechoic environments, sounds reaching the ears of a listener contain two relatively simple cues that provide information about the sound source's azimuthal location. The first of these is an interaural time difference (ITD) arising from the different lengths of the paths between the sound source and the two ears. The second cue is an interaural level difference (ILD) primarily due to the interaction between the sound field and the listener's head. The presence of echoes and reverberation found in many acoustic environments poses a problem for extracting this information because the superposition of acoustic reflections from various directions on the direct waveform degrades directional information, causing temporal fluctuations in the ITD and ILD and decorrelation of the ear-input signals (Beranek 2005; Shinn-Cunningham & Kawakyu 2003).

In this thesis, we focus on ITD as a localization cue. A common method of extracting ITD is by computing a cross-correlation between the left and right ear acoustic waveforms. In anechoic environments, this approach generally results in a strong peak at the delay corresponding to the ITD. Reverberation can severely attenuate this peak (Shinn-Cunningham 2003), making it much more difficult to determine the location of a sound source. Despite this difficulty, ITD-sensitive neurons in the auditory midbrain show more robust directional responses in the presence of reverberation than predicted by a model based on the average cross-correlation between left and right input signals (Devore et al. 2006). Thus it is apparent that there are neural mechanisms that lead to robust coding of directional information. Here we investigated factors in the auditory periphery that may contribute to this robust directional coding by recording the response of auditory nerve fibers in anesthetized cats to reverberant and anechoic stimuli.

II. Background

Phase-locking of the auditory nerve conveys directional information

The auditory pathway is a remarkable signal processing system capable of extracting much information solely from a two-channel (left/right) acoustic input. The first stage in processing

involves mechanical frequency analysis in the inner ear, followed by transducing the filtered acoustic signals into a series of electrical impulses in the AN. To a rough approximation, we can view the inner ear as a filter bank whose half-wave rectified outputs determines the probability of observing an impulse ("spike") on a given AN fiber at any given time. Each fiber is driven by a single frequency channel of this filter bank, and thus has a characteristic frequency (CF) to which it optimally responds. The mechanisms responsible for transducing the acoustic energy into neural spikes can only follow the acoustic signal up to a few kilohertz. For fibers with CF above this frequency, the probability of observing a spike becomes increasingly dependent on the envelope rather than the fine time structure of the filter bank output. Fibers with low CFs are therefore said to phase-lock to the fine-structure of the auditory stimulus, while higher CF fibers phase-lock only to the stimulus envelope. A more in-depth review of peripheral auditory processing can be found in Geisler 1998.

Due to the differing path lengths taken by a sound to reach either ear, the delay in the sound at one ear relative to the other (up to a few hundred microseconds for a typical head-size of a cat) can be mapped to the azimuthal angle from which the sound originated. As spike patterns on the AN are phase-locked to the auditory stimuli, we expect spikes from one ear to generally follow spikes from the other by this delay with relatively high probability in anechoic conditions. A temporal correlation of the left/right spike trains will therefore have a large peak at this delay. When the fiber's response is phase-locked to the stimulus fine structure, spikes tend to occur at a specific phase of the filtered stimulus. Thus for these fibers the temporal correlation will show additional side-peaks indicative of the stimulus fine-structure. Narrow-band filtering of noise stimuli produces a sinusoid at the filter's center frequency with random frequency and amplitude variations, so for these stimuli we expect temporal correlation peaks to be spaced roughly at periods corresponding to the fiber's CF. When the fiber's response is phase locked only to the stimulus envelope, we expect to see a broad peak in the temporal correlation around the ITD.

The auditory nerve projects to the cochlear nucleus, where some cells known as spherical bushy cells (SBCs) show comparable or better phase-locking than that seen in individual AN fibers

(Smith et al. 1993). Cells in the medial superior olive (MSO) receive input from SBCs from both left and right cochlear nuclei. These MSO cells perform coincidence detection to convert timing cues in the incoming spikes into a rate code (Moushegian et al. 1975; Spitzer & Semple 1995; Yin & Chan 1990). Neural coincidence detection involves a cell that fires a spike only when two (or more) input spikes are received nearly simultaneously. An MSO cell with a contralateral internal delay will fire maximally to inputs with an equal magnitude contralateral-leading ITD; thus MSO cells with differing internal delays can code for an equivalent range of ITDs in the acoustic input. MSO neurons project to the inferior colliculus (IC) where similar ITD sensitivity is found (Aitkin et al. 1985; Kuwada et al. 1987).

Reverberation degrades interaural timing cues

The simple view of extracting directional information from AN spike data is made much more complicated by reverberation. The multiple acoustic reflections from various directions combine with the direct sound, corrupting the directional information and leading to fluctuations in the ITD and ILD cues. These effects increase in severity as the distance between the source and the receiver is increased and the direct-to-reverberant (D/R) energy ratio decreases (Shinn-Cunningham et al. 2005). Figure 1, taken from Shinn-Cunningham & Kawakyu 2003, shows the effect of reverberation on a cross-correlation based model of ITD extraction for two different simulated listening positions within a virtual room. Each column of the nine 2-dimensional plots represents the model output at a given time. These plots demonstrate a predicted shift in correlogram peak with changing source azimuth; more importantly, they show that while the output has a sharp, stable peak for anechoic stimuli, it becomes more variable and less reliable when responding to reverberant stimuli.

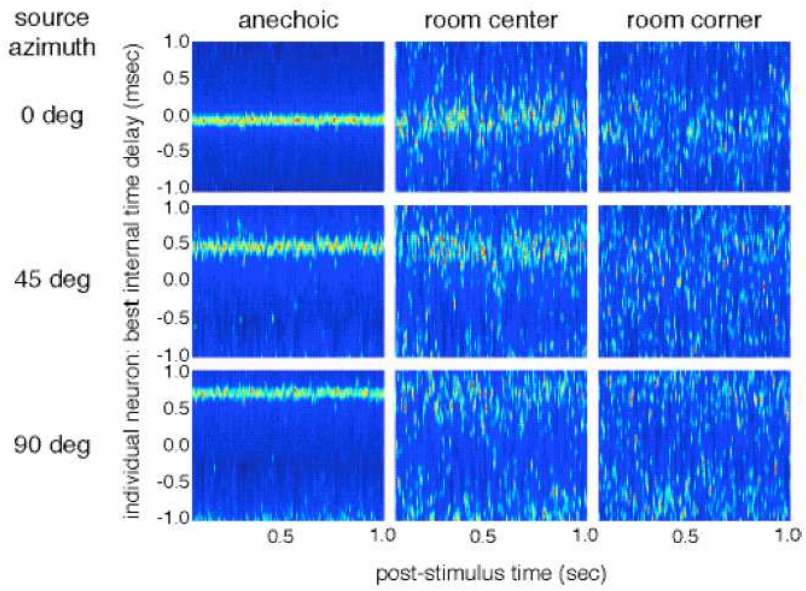

Figure 1 – Time-varying output of a cross-correlation-based neural ITD computation model. Each column represents a different room condition used, where "room center" and "room corner" refer to the simulated position of the listener/receiver. Each row represents a different simulated source azimuth, giving rise to a different ITD in the binaural signal. The model output is corrupted by the presence of reverberation, particularly for the "room corner" condition where early reflections are strong.

Neural coding of directional information is more robust than current models

Previous work in our lab has focused on responses of ITD-sensitive cells in the IC of cats and rabbits to noise bursts with varying simulated azimuths and levels of reverberation (Devore et al. 2009; Devore & Delgutte 2010; Devore et al. 2010). Results confirmed that directional coding was degraded in the presence of reverberation. This degradation was quantified by the ratio of the range of firing rates across azimuths observed in response to the reverberant stimuli to the range observed in response to the anechoic stimuli (hereafter referred to simply as "relative range"). Yet this degradation was less than predicted by a binaural model based on long-term cross-correlation that accurately predicted responses in the anechoic condition (Hancock & Delgutte 2004). This result is interesting in that it implies the existence of a neural processing mechanism not included in the model that allows for this robustness to reverberation. Moreover, directional information was more reliable in the early response of IC neurons than in

the ongoing response, reflecting the fact that reverberation builds up over the time course of the stimulus. Many cells in the auditory pathway exhibit adaptation, where the cell will fire more vigorously near the onset of a stimulus; thus it seems plausible that adaptation may contribute to more robust coding of directional information in reverberation.

Additionally, Devore and colleagues found that the relative range was generally lower (corresponding to more degradation) in high frequency neurons than in low frequency neurons. Specifically, neurons tuned to frequencies below 2 kHz showed significantly larger relative range than neurons tuned to higher frequencies. Yet there were no significant differences within the low and high frequency groups themselves. This result suggested that a dominant factor in this frequency dependence may be the loss of fine-structure sensitivity in neural responses at higher frequencies.

Robustness to reverberation is critical for real applications

The study of sound localization has a number of applications. For example, hearing-impaired listeners, including users of cochlear implants, suffer from reduced ability to use directional cues to process events in their environment (Arbogast et al. 2005; Loizou et al. 2009; Marrone et al. 2008). A common complaint amongst hearing aid and cochlear implant users is that although they can understand speech reasonably well in quiet, anechoic environments, the task becomes difficult or impossible in many everyday settings. Understanding the origin of the robustness of directional coding to reverberation could lead to processing strategies for hearing aids and cochlear implants that better transmit the necessary information to restore these abilities in hearing-impaired listeners in realistic environments.

Other applications of sound localization studies include any signal processing system that uses auditory information to gain knowledge about its surroundings. Such systems can include automated systems for combat awareness and hostile activity detection, general surveillance and security systems, interactive robotic systems, and automated voice recognition algorithms.

Of course, if any of these applications are to have real-world usability, they need to be robust to the effects of reverberation found in everyday environments. As simple cross-correlation

models fail to account for observed neural performance, a logical step forward is to investigate the origin of this neural robustness to reverberation.

III. Materials & Methods

Preparation

We recorded from single units of the auditory nerve of seven anesthetized cats. The first four of these cats were anesthetized with Dial in urethane (75 mg/kg). Due to drug availability, the remaining cats were anesthetized with urethane and Nembutal (36.65 mg/kg). All cats were given supplementary doses of anesthetic as needed, as well as occasional injections of dexamethasone (0.26 mg/kg) to prevent brain swelling. For each experiment, the posterior portion of the skull was removed, and the cerebellum retracted to expose the AN. A calibrated closed acoustic system was fitted into the cat's left ear canal. All stimuli were digitally filtered to compensate for the transfer characteristic of the acoustic system. A silver electrode was positioned near the exposed round window of the cochlea to record the AN compound action potential to monitor cochlear function during the experiment. Glass micropipettes filled with 2M KCl solution were advanced into the AN using a micro-positioner.

Recording

Spike times, detected by a user-set threshold voltage, were recorded and saved for offline processing. We used click stimuli at around 55 dB SPL to search for units in the AN. To characterize a unit once it was isolated, we first measured its frequency tuning curve (Kiang 1965) to obtain the CF and tone threshold at CF. We then measured the fiber's spontaneous firing rate (SR) recorded over a period of 20 seconds. To ensure normal function in each individual unit we recorded from, we calculated a "z-score" measuring the distance, in units of standard deviations, of a fiber's tone threshold from the mean threshold, for a given CF and SR, from a previously collect data set (Liberman 1978). Units with a z-score exceeding 3 were excluded from all analyses in this thesis.

We also measured a rate-level function (RLF) in response to 200 ms bursts of broadband noise by recording the firing rate in response to randomly varying levels of noise. Noise levels were selected 10 dB apart, and the RLF was linearly interpolated between the measured firing rates. A unit's noise threshold was defined as the level for which the RLF exceeded the SR by three times the standard deviation (the SR measurement was divided into 20 one-second trials to compute standard deviation). AN fiber RLFs lie flat at the fiber's SR for low input levels, rise past some threshold, and level off at a fiber's "saturation rate." We chose noise levels for our main experiment to lie roughly in the middle of the dynamic range over which the fiber's RLF was increasing—typically roughly 10-15 dB above the fiber's noise threshold, but never exceeding 90 dB SPL to avoid damaging the cat's hearing during an experiment. Sound levels ranged from 30 to 90 db SPL with a median of 55 dB SPL.

Stimuli

We used 40 repetitions of an exactly-reproducible 400-ms token of white noise (with 4 ms linear on/off ramps and 1 s inter-stimulus interval) as our stimulus. To simulate reverberation, we generated binaural room impulse responses (BRIRs) using the room-image method (Allen & Berkley 1979) for a room size of 11 x 13 x 3 m and an inter-receiver distance of 12 cm to allow the ITDs to range between ±360 μs, typical of those experienced by cats. We simulated reflections in an empty room (without head interactions), thus the BRIRs contain almost no ILD for any source azimuth. Source distances of 1m and 3m were used to analyze the effects of moderate and severe reverberation. These BRIRs represent the impulse response of the room, for a given source position, as seen by each ear. We convolved a noise source signal with the left / right channels of this filter to obtain simulated reverberant stimuli for arbitrary source azimuth directions. We delivered sound only to one ear of the cat, presenting left and right channel responses separately.

As a first step toward investigating directional coding for more realistic stimuli, we recorded a few AN responses to speech-modulated white noise. A speech utterance ("place the cake next to the table") was recorded and filtered into two bands (above and below 3500 Hz, respectively) using an 8^{th}-order elliptic filter with 1 dB pass-band ripple and 40 dB stop-band attenuation. The

envelope of both signals was extracted by full-wave rectification followed by a 4^{th}-order Chebychev type II low-pass filter with a 100 Hz cutoff frequency. For a few of our AN units, we selected the envelope from the CF-appropriate band and recorded the unit's response to noise modulated by this envelope.

Analysis

Extracting temporal information using Shuffled Correlation

Spike trains in response to anechoic and reverberant noise bursts were analyzed using the shuffled correlation technique (Joris 2003). This technique, illustrated in Figure 2, mimics the output of an array of coincidence detectors operating on the auditory nerve data by computing a discrete correlation (using a fixed coincidence window / bin width) between each possible pair of left/right spike trains, and summing all resulting correlograms.

The resulting shuffled cross-correlogram (SXC) is normalized for average firing rate to the left and right ear stimulus, r_L and r_R respectively, stimulus duration, D, width of the coincidence window, w, and number of trials, N, by dividing by the quantity $N^2 r_L r_R w D$ (Louage et al. 2004). This normalization results in SXC values of 1 representing the number of coincidences expected from two temporally unstructured and uncorrelated spike trains. Values above 1 represent greater-than-chance degrees of coincidence, such as we might expect from phase-locked AN spike trains for certain spike-pair delays. Specifically, for anechoic stimuli we expect to see a peak in the SXC at the stimulus ITD as spikes in response to the left / right BRIR-filtered stimuli will tend to be temporally shifted by this ITD. For low-CF AN fibers that phase-lock to the stimulus fine-structure, we also expect to see multiple side-peaks corresponding to the delays between peaks in the filtered stimulus; i.e., the CF period. For high-CF AN fibers that phase-lock only to the stimulus envelope (which in our noise stimuli is generated by narrowband cochlear filtering of the noise tokens), we expect only a broad peak in the SXC centered around the stimulus ITD. Lastly, we expect reverberation to attenuate these SXC peaks as the acoustic reflections degrade the temporal correlation between left and right-ear signals.

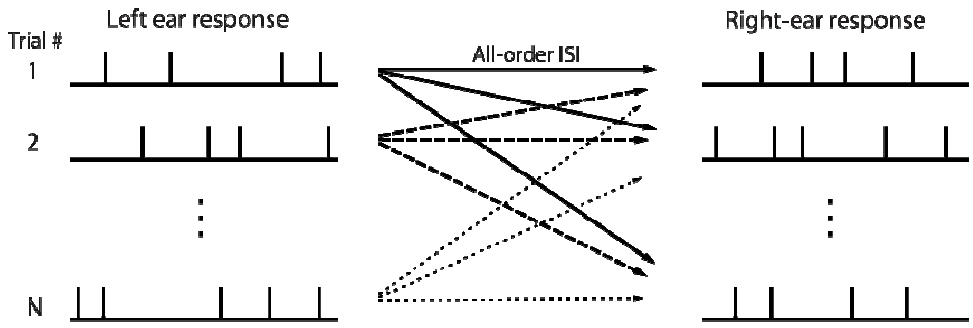

Figure 2 – Computation of the SXC. Delays are counted between all possible pairs of spikes across all possible pairs of left/right spike trains. Results are binned and summed, resulting in a histogram of delay occurrences. Counting delays is mathematically equivalent to computing a correlation on binary spike data as the cross-correlation expression $r[m] = \sum x_1[n] \cdot x_2[n+m]$ *represents exactly the number of spikes in* x_2 *that follow a spike in* x_1 *by m time samples.*

We used a bootstrapping procedure to determine if a given SXC peak was reliable. We sampled trials, with replacement, independently from the 40 left and 40 right stimulus response spike trains. Sampling was performed 2000 times for each set of recordings, and we computed an SXC for each sampling. From these SXCs we generated a distribution of SXC peak locations. We take the area of the bootstrap-generated distribution of peak locations surrounding the actual SXC peak, bounded by the closest local minima in the distribution, divided by the total area in the distribution (2000 peaks) to generate a metric between 0 and 1 representing the "peak reliability index." If an actual SXC peak is reliable, then a high percentage of the peaks generated by this bootstrapping procedure should fall close to this peak, resulting in a peak-reliability near 1. If the peak from the full SXC is simply due to noise, then the bootstrapped peaks will be scattered throughout the range of delays, resulting in low peak reliability index.

To quantify the degree to which a fiber's responses are sensitive to the fine-structure or envelope of the stimulus, we also presented antiphasic noise tokens in which the stimulus polarity is inverted for one ear. If a fiber's response is sensitive to the stimulus fine-structure, the SXC in response to antiphasic noise will undergo a 180° phase-shift relative to the SXC in response to homophasic noise. Alternatively, if a fiber's response properties are dominated by the stimulus envelope, the SXC should be unaffected by this polarity inversion. Thus the

correlation coefficient between the SXCs in response to homophasic vs. antiphasic noise gives us a metric that varies between -1, indicating high dependence on stimulus fine structure, and +1, indicating sensitivity only to the cochlear-induced stimulus envelope (Joris 2003).

Investigating the time-course of reverberation

The interference caused by reverberation is not stationary; specifically, reverberation is minimal near the stimulus onset and builds up over time. To view how temporal information coded in the AN response changes over the time course of a stimulus, we computed a dynamic, time-dependent shuffled correlation (DSXC) for each pair of spike train responses to noise and to speech-like stimuli. The DSXC is computed by binning spike pairs in one dimension by the inter-spike interval and in a second dimension by peristimulus time, resulting in a 2-dimensional display of temporal information in the left/right AN responses. This operation is roughly equivalent to computing the SXC on successive temporal windows of the response spike trains, but includes spike-pairs that cross temporal window edges. We normalized the DSXC like we did for the full-stimulus SXC by dividing each vector, corresponding to a given peristimulus time bin, by $N^2 r_L r_R wb$, where b refers to the width of the peristimulus time bin, and r_L and r_R refer to the average firing rates within that bin.

Modeling the frequency-dependent degradation

Devore et al. 2010 suggested that increasing degradation in directional sensitivity at high frequencies may largely be due to loss of synchrony of neural responses to stimulus fine structure at these frequencies. We investigate this idea more closely using both our AN data and data generated by a simple auditory processing model. Our auditory processing model consists of a gammatone filter bank, followed by half-wave rectification, and finally a low-pass synchrony filter. The impulse response of the gammatone filter bank is given by

$$h(t) = 2\pi b \cdot t^3 \cdot e^{-2\pi b \cdot t} \cdot \cos(2\pi \cdot CF \cdot t) \cdot u(t)$$

where u(t) is the unit step function, CF is the filter's center-frequency / AN unit characteristic frequency, and b is the filter's bandwidth. The synchrony filter was a 4^{th}-order Butterworth low-pass filter with a 1 kHz cutoff frequency.

We used four versions of the model; (1) the filter bandwidths were either matched to previously collected physiological data from AN recordings (Carney & Yin 1988), or were constant across frequency, (2) the synchrony filter was either present or bypassed. For equal filter bandwidths, we used the filter bandwidth of the 1 kHz filter for all filters. The filter outputs to left-ear and right-ear stimuli were cross-correlated, and the resulting correlogram was normalized analogously to the SXC normalization so that the resulting correlogram values would be 1 for uncorrelated inputs and greater than 1 for higher degrees of correlation (see appendix A). We compared the output of this model to the results from our physiological experiments to investigate the factors that contribute to the trend of decreasing directional sensitivity at higher frequencies.

IV. Results

Responses to anechoic stimuli

We recorded responses to repeated noise bursts from 88 single units in the AN of seven cats. 22 units were excluded for having threshold z scores above 3.0. Of the 66 remaining units, 6 had low-SR (<1 spike/sec), 23 had mid-SR (between 1 and 20 spikes/sec), and 37 had high-SR (>20 spikes/sec). Figure 3 shows representative, unsmoothed SXCs of the responses to homo-phasic and anit-phasic noises bursts in the anechoic condition. The low-CF fiber demonstrates a clear sensitivity to stimulus polarity, while the high-CF fiber's response appears unchanged. The mid-CF fiber lies in between, demonstrating both phase-sensitivity in its fine structure while having a prominent phase-insensitive envelope component. We quantified the phase/fine-structure sensitivity of a given fiber by the Pearson's correlation coefficient between the SXCs to homo-phasic and anti-phasic noise over the range of delays between -3 ms and +3 ms. The results are plotted in Figure 4, demonstrating a switch from fine-structure sensitivity (correlation coefficient near -1) to envelope-sensitivity (correlation coefficient near +1) around 1-3 kHz. Joris, 2003, found a similar transition to envelope-sensitivity, though perhaps more narrowly focused between 2-3 kHz.

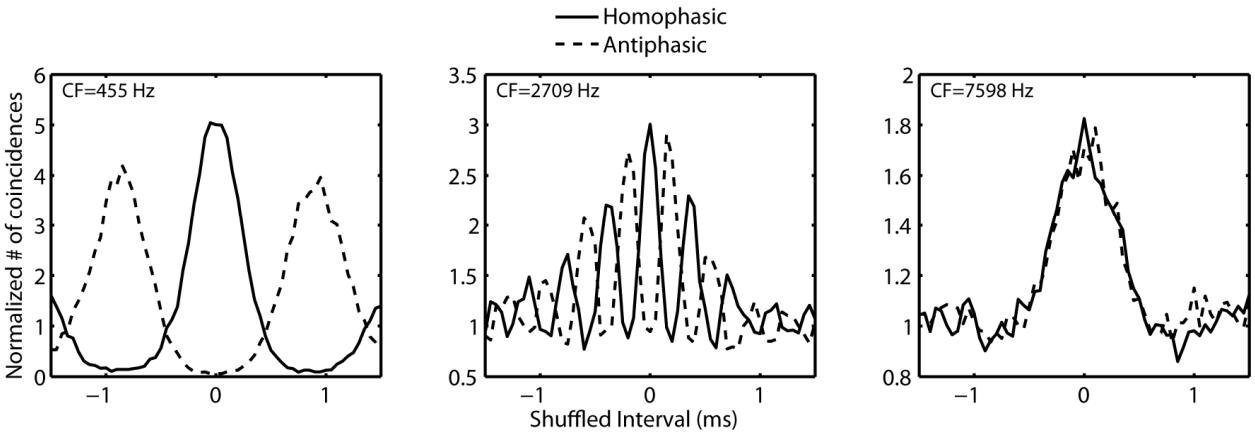

Figure 3 – Representative, unsmoothed SXCs for a low, mid, and high-CF fiber in response to homo-phasic and anti-phasic noise bursts (solid lines and dashed lines, respectively). Stimulus levels for these plots were 12 dB, 18 dB, and 14 dB above noise threshold (see: methods) for the low, mid, and high CF fibers, respectively.

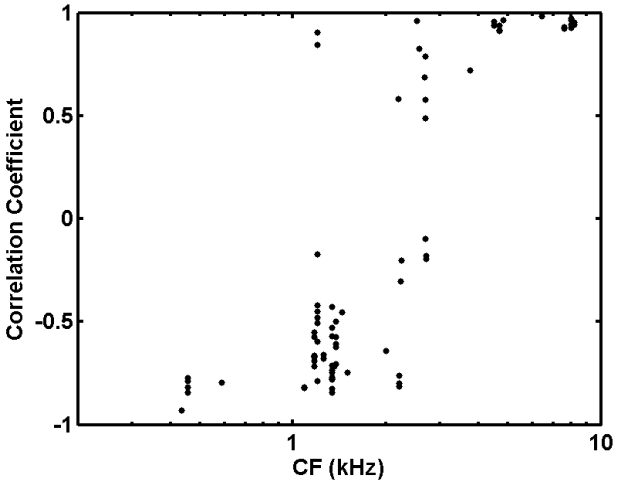

Figure 4 – Correlation coefficient between unsmoothed SXC histograms (between delays of -3 and +3 ms) for anechoic monophasic and antiphasic noise stimuli for all available data (restricted to stimulus levels between 0 and 20 dB above a fiber's noise threshold). We recorded responses to anechoic antiphasic noise bursts in 30 units. Negative coefficients indicate sensitivity to stimulus fine structure, while positive coefficients indicate sensitivity only to the stimulus envelope. These data show a transition from find-structure to envelope sensitivity at roughly 1-3 kHz.

Reverberation degrades SXC peaks

In addition to the anechoic noise bursts, we also played noise bursts with two levels of reverberation, corresponding to 1m and 3m source distances in a virtual room. In general reverberation caused a reduction in SXC peaks. Figure 5 shows three representative SXCs for a low, mid, and high CF fiber, for the three different room conditions for zero-degree azimuth stimuli. These results demonstrate increasing degradation of directional information with increasing reverberation, and also suggest that this degradation is more severe for fibers with higher CFs (in addition to SXC peaks being overall lower at higher CFs).

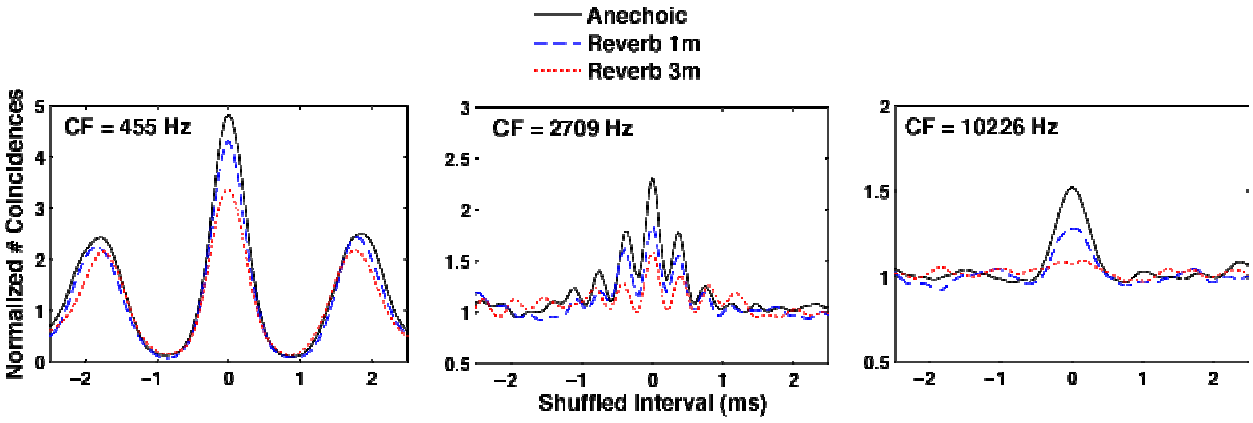

Figure 5 – Representative SXCs computed on responses to 0-ITD stimuli from three fibers of varying CFs. Stimulus levels were 12 dB, 18 dB, and 15 dB above each fiber's noise threshold for the low, mid, and high CF fiber, respectively. The low-CF and mid-CF fiber shows a quasi-periodic structure due to sensitivity to the stimulus fine-structure, with higher-order peaks spaced roughly according to the fiber's CF. The high-CF fiber shows a single, broad peak, indicating synchrony to the stimulus envelope. SXCs shown here were smoothed with a 0.2 ms standard deviation Gaussian window; note that synchrony to the stimulus fine structure dropped substantially even for unsmoothed SXCs at around 2-3 kHz (Figure 4).

Figure 6 shows the SXC peak heights in response to stimuli with 0 degrees simulated azimuth for all units with bootstrapped peak reliability index above 0.5. Peak height decreases both with increasing reverberation and increasing CF. Note in particular that for the 3m reverberation condition, neurons with CF above roughly 5 kHz show SXC peaks close to 1.

To quantify the effect of reverberation, we subtract 1 from the SXC peak height (to represent the height of the peak above the chance value of 1) and, for each fiber, take a ratio of this relative peak height in each reverberant condition to the corresponding peak height in the anechoic condition. Relative peak ratios less than 1 indicate degradation by reverberation. Figure 7 shows these relative peak ratios plotted against CF for all units in our sample that have an anechoic peak reliability index of over 0.5 (for 0 degree stimuli). We use only the anechoic peak reliability index rather than both anechoic and reverberant peak reliability index here as many high-frequency, 3m reverberant data do not have reliable peaks, and by eliminating these data entirely we would bias the data towards units with large peak heights. The degradation increases with both amount of reverberation and CF. Both effects are highly significant (ANOCOVA with reverberation as a factor: $p<0.001$ for both CF and reverberation, with no significant interaction). It is worth noting that the same analysis on only units with CF above 5 kHz still yields a highly significant decrease in peak ratio with CF and with reverberation ($p<0.001$ for both factors). The data above 5 kHz also show a significant interaction ($p<0.001$) between CF and reverb. These results contrast with previous results (Devore et al. 2010) showing a step-like transition in degradation of directional coding in IC neurons with no significant additional degradation beyond 2-3 kHz.

Though we did not collect as many data for a simulated azimuth of 90 degrees, there does not appear to be a large difference in this condition apart from a shift in the peak locations corresponding to the non-zero ITD. Figure 8 shows three representative[1] SXCs for 90 degree azimuth, demonstrating similarity to the results shown in Figure 5 both in degradation by reverberation and lower peaks for higher CF neurons. Figure 9 plots peak heights for 90 degree azimuth against peak heights for 0 degree azimuth for units from which we recorded responses to both azimuths. There is a slight overall reduction in peak heights for reverberant conditions in response to stimuli form 90 degrees compared to the response to stimuli from 0 degrees. This azimuth dependence of the effects of reverberation requires further investigation.

---

[1] Unfortunately, we were not able to collect any data from comparably low-CF units as in Figure 5 for 90° azimuth

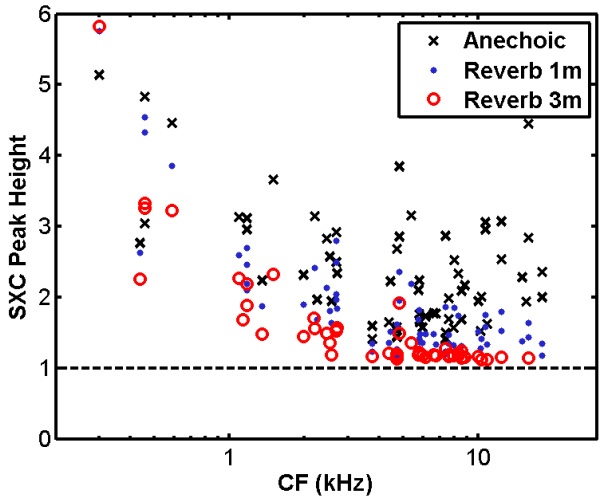

Figure 6 – SXC peaks heights in response to 0-degree stimuli for all units with peak reliability index above 0.5 and threshold z-score below 3, for stimulus levels between 0 and 20 dB relative to the fiber's noise threshold (see: methods). This selection yields 66 units for the anechoic and 1m reverberant conditions, and 50 units for the 3m reverberant condition.

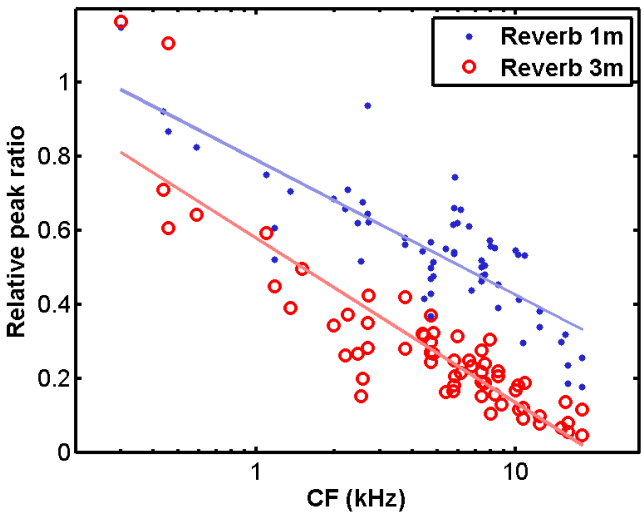

Figure 7 – Relative peak ratios plotted against CF, for 0-degree stimuli, for units with an threshold z-score under 3, an anechoic bootstrapped SXC peak reliability index of over 0.5, and stimulus levels between 0 and 20 dB of the unit's noise threshold (see: methods). Includes 50 units for the 1 m reverberant condition, and 52 units for the 3 m reverberant condition. Solid lines represent linear fits (in log CF)

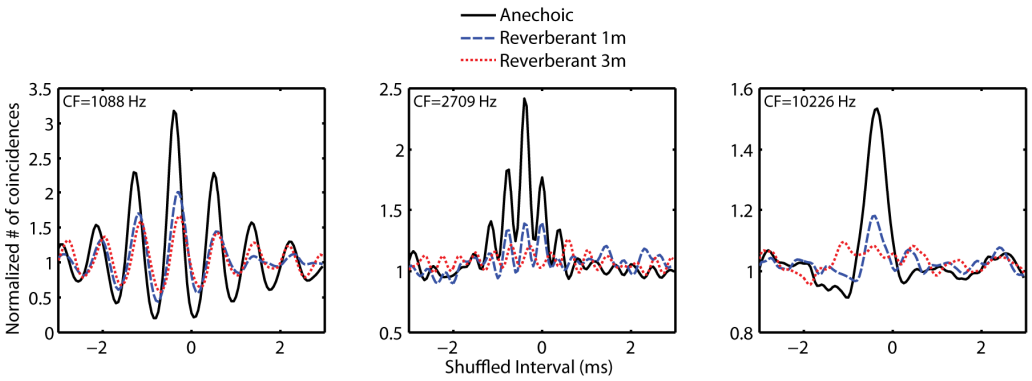

Figure 8 – Representative SXCs for three units to noise bursts with a simulated 90 degree azimuth. Results are similar to those for 9 degree azimuth (e.g. Figure 5) but peaks locations are shifted to the left due to the ITD contained in the stimulus.

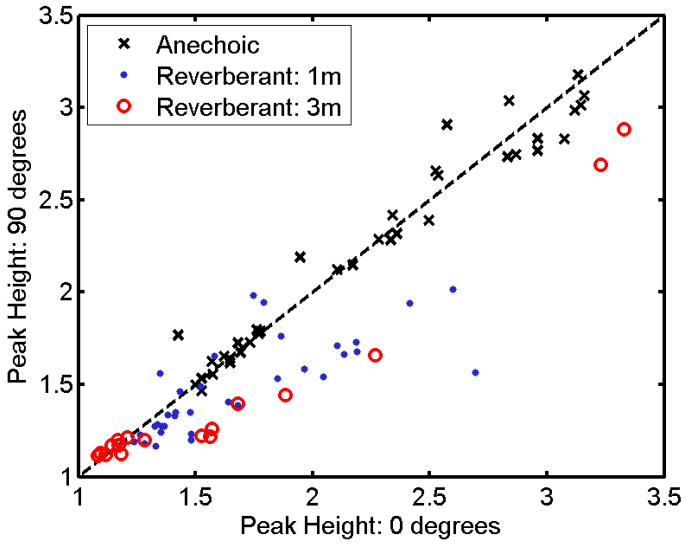

Figure 9 – Peak heights for units in response to stimuli from 90 degrees simulate azimuth plotted against peak heights in response to stimuli from 0 degrees azimuth. There is one additional anechoic data point at approximate coordinates (4.8, 4.6) which is omitted to expand the axes. n = 37 units for anechoic, 34 units for 1m reverberant, and 17 units for 3m reverberant conditions.

Degradation due to reverberation is limited near the stimulus onset

We investigated the time-course of the degradation in SXC peak height due to reverberation. Figure 10 shows representative normalized DSXCs, where each column is individually normalized for the firing rate within that time bin, for 3 representative fibers in anechoic and 3m reverberation conditions and 0-degree azimuth. In the anechoic condition, we see relatively stable peaks at 0 ms. For the low and mid CF fibers, we see side-peaks indicative of sensitivity to stimulus fine structure, while for the high CF fiber we see only a broader, messier peak. There is clear degradation of the timing information in the reverberant condition. This degradation is severe for the high-CF fiber, moderate in the mid-CF fiber, and small in low-CF fiber. In the low-CF and mid-CF fibers, we see not only a reduction in peak height across the duration of the stimulus (more so in the mid-CF fiber), but also greater fluctuations in the peak locations caused by the simulated acoustic reflections. For the high-CF fiber, the degradation is so severe that there is no discernable peak for most of the stimulus. However, in each plot, the response during the first 10-20 ms gives a peak at 0 ms shuffled interval (faint but present, in yellow, for the high CF fiber), suggesting that it may be possible to extract reliable information about stimulus direction even in extreme reverberant environments using a mechanism that emphasizes this early onset response, such as AN adaptation.

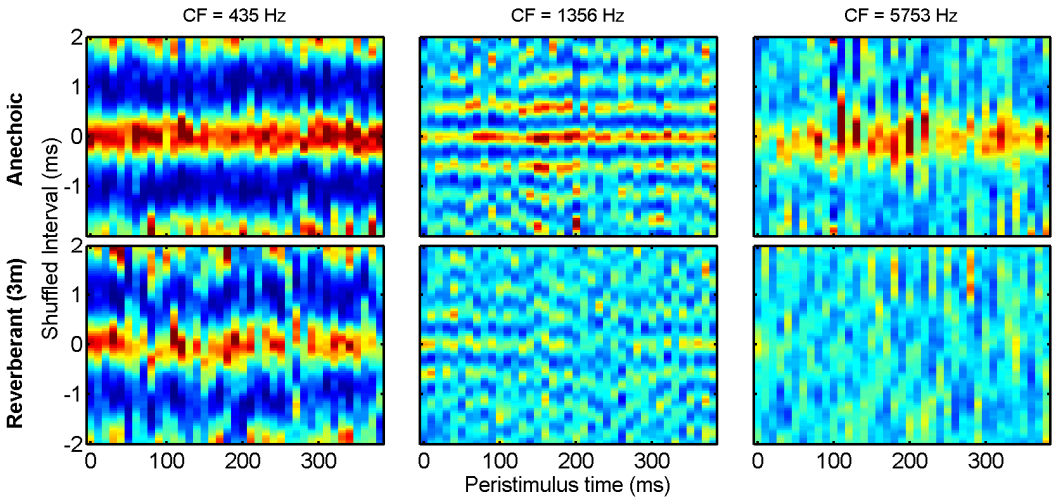

Figure 10 – Normalized DSXCs for three example fibers using 0-degree azimuth stimuli. Spike pairs are binned by inter-spike interval (y axis) and peristimulus time of the right-channel-response spike (x axis). Dark colors represent large values.

To quantify the reliability of timing information near the stimulus onset relative to the ongoing portion of the stimulus, we computed separate SXCs over two windows; the onset (0-50 ms) and the ongoing (51-400 ms) portion of the response. We define an "onset advantage" ratio as the ratio of the SXC peak height (above 1) computed over onset to the peak height (above 1) computed over the ongoing portion of the stimulus. Values above 1 indicate degraded peaks in the ongoing response, indicating an onset advantage. Results are shown against CF in Figure 11. The onset advantage ratio in reverberation increases both with increasing CF and with increasing reverberation, indicating larger onset advantages for the units most affected by reverberation. As expected, the onset advantage is near 1 for all CFs in the anechoic condition. A two-way ANOVA, using CF (in 6 equi-populated bins) and reverberation as factors, shows a significant effect of reverberation ($p<0.001$) but not of CF, with a significant ($p<0.001$) interaction between the two factors. The interaction reflects the CF dependence see in reverberation, but not in the anechoic condition.

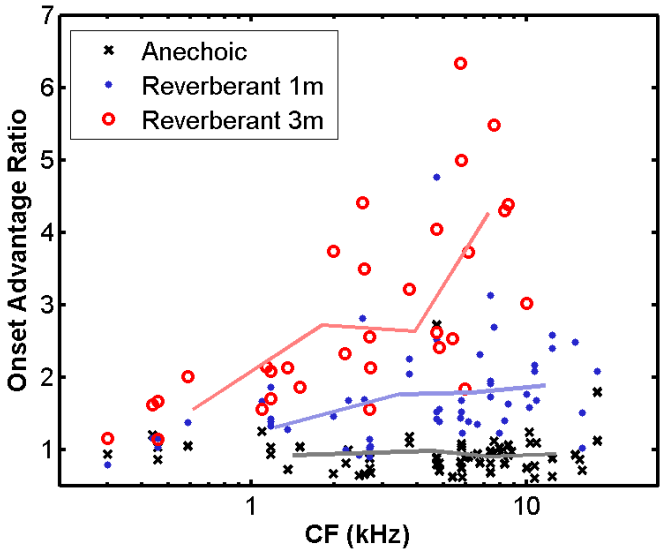

Figure 11 – Ratio of SXC peak height above 1 computed over the ongoing (51-400 ms) portions of the 0-degree stimulus to the peak height computed over the first 50 ms of the same stimulus. Higher ratios indicate a better peak height in the early response compared to the ongoing response. Solid lines plot averages for 4 equi-populated bins for each room condition.

We also used the DSXCs investigate the role of adaptation in the auditory nerve in the coding of directional information. Adaptation refers to the fact that AN fibers tend to fire maximally near the onset of a stimulus compared to the ongoing portion of the stimulus. Adaptation may have a key role in creating reverberation-robust directional responses as reverberation is minimal near the stimulus onset. Thus at least for isolated stimuli, the maximal response near this onset provides the most accurate information about the sound source direction.

To assess the effect of adaptation, we computed the SXC in two ways. The first, the "full SXC," as already described above, weighs each spike pair equally. As AN fibers fire more rigorously to the stimulus onset due to adaptation, this method emphasizes this portion of the response in the SXC relative to later portions of the response. The extent of emphasis due to adaptation is demonstrated in Figure 12, which shows, for the same response, a normalized and unnormalized DXC (using different color scale to be able to simultaneous visualize both plots). We can obtain the (unnormalized) full SXC by summing across columns of the unnormalized DSXC. This leads naturally into the second method of computing the SXC, where we take the mean of columns of the normalized DSXC. As each column / time bin of the normalized DSXC is normalized for the firing rate within that time bin, taking the mean of columns in this manner gives us a normalized SXC that does not emphasize the onset response, effectively "normalizing out" the effect of adaptation on the resulting "adaptation-normalized SXC," or ANSXC.

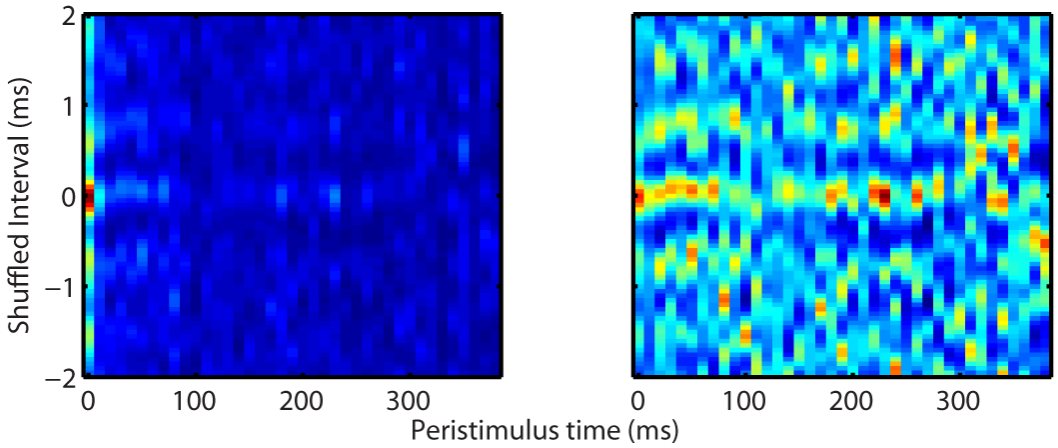

Figure 12 – Unnormalized (left) and normalized (right) DSXC for a 1.1 kHz CF fiber in the 3m reverberant condition. The unnormalized DSXC demonstrates how adaptation can strongly emphasize the response near the stimulus onset (when reverberation is minimal), while the normalized DSXC assigns more equal weighting to the directional information in each time bin The full-stimulus SXC can be obtained by summing across columns of the unnormalized DSXC, while the ANSXC is obtained by taking the average across columns of the normalized DSXC.

We compared the peak height of the ANSXC with the peak height of the full SXC to quantify the effect of AN adaptation on the coding of timing information in the AN. Figure 13 shows full-stimulus SXCs and ANSXCs for anechoic and 3m reverberant conditions for three example units. Peak heights are somewhat reduced for the ANSXCs compared to full-stimulus SXCs. This reduction is most pronounced in the reverberant condition, particular for the mid and high CF fibers. Such a reduction in peak height in part reflects the fact that the neural response is strongest near the stimulus onset, when directional information is more reliable because reverberation is minimal.

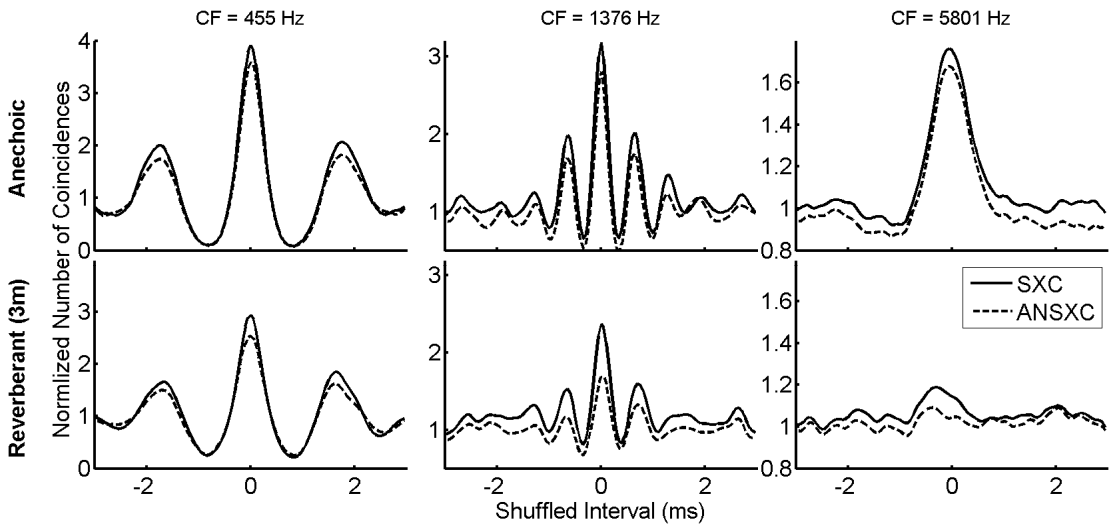

Figure 13 – Full SXCs and ANSXCs from three example fibers in our population. For these fibers, ANSXC peaks are lower than full-stimulus SXC peaks. This effect is more pronounced for reverberant stimuli particularly in the higher CF fibers.

To quantify this peak reduction for all units in our sample, we computed the ratio of the peak height (above 1) of the full-stimulus SXC to the peak height (above 1) of the ANSXC (the "adaptation ratio"). Adaptation ratios above 1 indicate a lower peak height in the ANSXC, indicating that adaptation plays a positive role in the robustness of ITD coding in the AN. Figure 14 shows these adaptation ratios for our entire data set against CF, in all three simulated acoustic conditions. Adaptation ratios in the anechoic condition are near 1, as expected. Adaptation ratios increase with CF and with reverberation, indicating increasingly better extraction of directional information from AN responses using a full-stimulus SXC rather than an ANSXC. As the ANSXC weights all time bins equally while the full-stimulus SXC generally provides more weight to the early response, this result indicates that adaptation may increase robustness of directional coding particularly in heavily reverberant environments. A two-way ANOVA, with CF (using 6 equi-populated bins) and reverberation as factors reveals a significant effect of reverb ($p<0.05$) and a significant interaction between CF and reverb ($p<0.001$), but no main effect of CF. As with the onset advantage ratio, this interaction reflects the CF dependence seen in reverberation but not the anechoic condition.

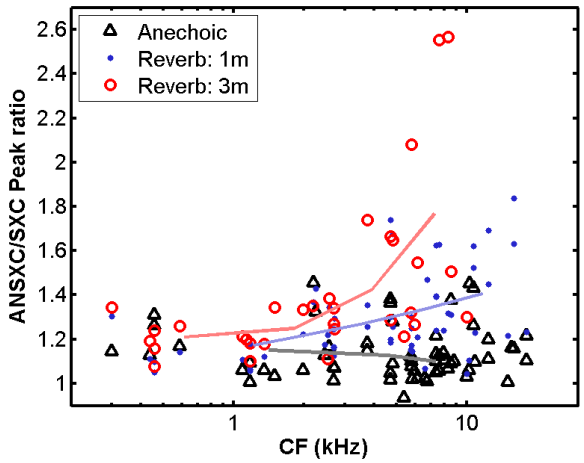

Figure 14 – Ratios of ANSXC peaks to full-stimulus SXC peaks over our entire data set for anechoic (49 units), 1 m reverberant (47 units) and 3m reverberant (27 units). Units with a threshold z-score greater than 3 or bootstrapped peak reliability index less than 0.75 were excluded from this plot (see: methods). Solid lines plot averages for 4 equi-populated bins for each room condition.

Speech onsets contain reliable directional information in reverberation

Thus far we have considered the effect of reverberation for bursts of white noise, finding that adaptation emphasizes reliable onset responses, allowing for more robust coding of directional information. Of course natural stimuli are often more continuous than isolated bursts of white noise. We hypothesize that slow envelope modulations present in many natural stimuli allow the auditory system to take advantage of the same onset dominance that emphasizes directional information when the contribution of reverberation is relatively low. As a first step towards testing this hypothesis we recorded, for a few AN fibers, responses to speech-modulated white noise. An example DSXC for a unit with a CF of 8 kHz is shown in Figure 15, showing both the anechoic and 3m reverberant response. We used a 90 degree simulated azimuth, corresponding to a 360 µs ITD in the direct sound. The anechoic response shows a steady peak at this ITD value (dashed line) at times when the fiber responds to the stimulus. The reverberant response is highly degraded in general, but still demonstrates accurate information near the three main onset peaks (marked by red ellipses). This result supports the plausibility of increased robustness to reverberation by onset emphasis even for natural, more dynamic stimuli.

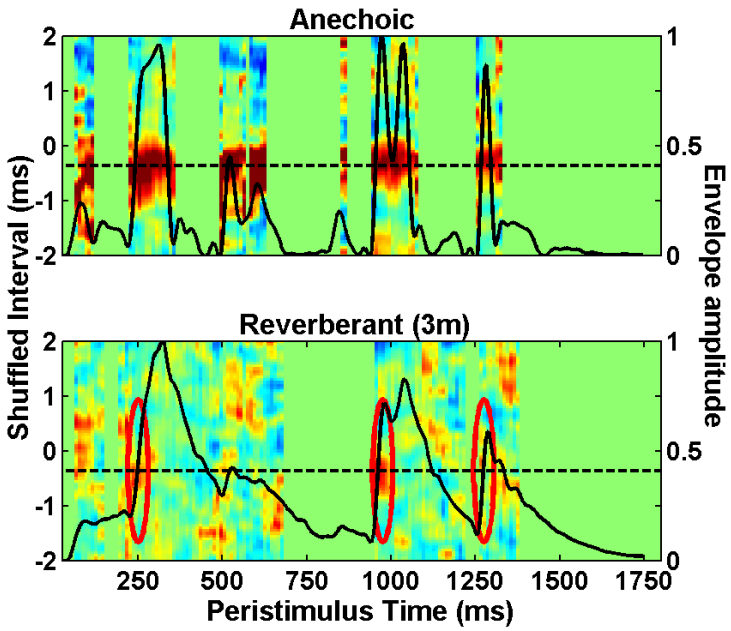

Figure 15 – Example DSXC for speech-modulated noise, using a simulated azimuth of 90 degrees. The unit had a CF of 8 kHz. The solid line indicates the stimulus envelope. The dashed line indicates the direct-sound ITD of 360 µs. The anechoic DSXC shows a steady peak at this ITD value. The reverberant DSXC is largely corrupted, but shows accurate information near the three main stimulus onsets marked by red ellipses. DSXC columns where spiking activity fell below a threshold (175 spike pairs for anechoic, 600 spike pairs for reverberant) were replaced with vectors of ones for visual clarity.

Model Results

To investigate the factors that contribute to the observed pattern of CF-dependent degradation of directional coding in reverberation, we constructed a simple peripheral auditory processing and cross-correlation model. The auditory processing model consisted of a gammatone filter bank, followed by half-wave rectification, and finally a 1000-Hz low-pass filter. The model responses to the same left-ear and right-ear stimuli as used in the experiment were cross-correlated and normalized analogously to the SXC operation above. We computed correlogram peaks and reverberant-to-anechoic peak ratios for each frequency and plotted the results against filter center frequency. Figure 16 shows three example normalized correlograms, paralleling the example SXCs shown in Figure 5

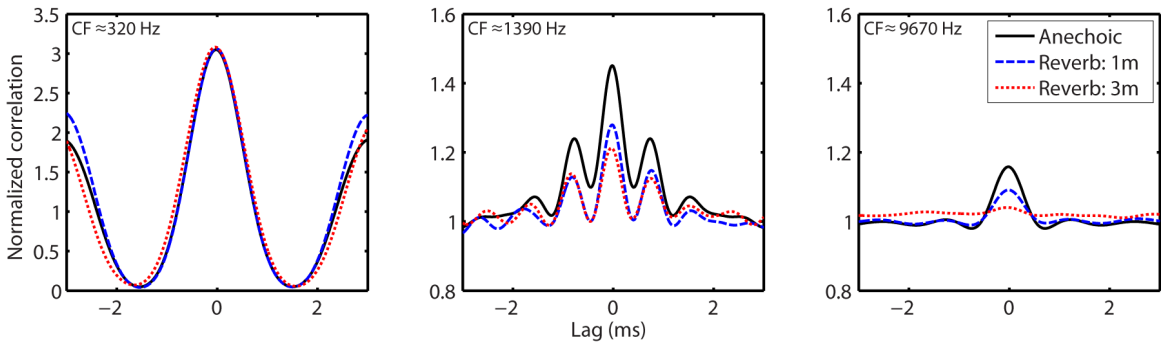

Figure 16 – Example normalized correlograms computed on model output for three filter center frequencies

Qualitatively, the model produces similar results as our AN experiments. Peak heights are reduced for higher frequencies and for increasing reverberation. The model does, however, miss some aspects of auditory nerve function. This is readily apparent in the plot of correlogram peaks vs. CF, shown in Figure 17, where peak heights are generally lower than in the actual data, and fall to 1 for all room conditions much more quickly above the synchrony filter's cutoff frequency of 1 kHz (compare with Figure 6). However, the peak ratio, shown in Figure 18 (solid lines), follows the data more closely (compare with Figure 7). Thus while the model poorly predicts absolute sensitivity to directional information measured by the peak height, we use it here as a rough method of investigating the *relative effects* of reverberation as revealed by the peak ratios. In particular we are interested in the frequency-dependent peak ratio decline in reverberation, whose overall trend is more reasonably captured by the model.

To test for the factors underlying this frequency dependence, we used four versions of the model for each reverberation level. Based on previous results in IC (Devore et al. 2010), we hypothesized there may be an effect of loss of synchrony on the frequency dependence of the peak ratios, so we either left in place or bypassed the 1000 Hz low-pass synchrony filter. Additionally, we hypothesized there may be an effect of increasing filter bandwidth with CF, so we tested the model using data-matched filter bandwidths as well as constant bandwidths across frequency. Three of these model versions' outputs are plotted in Figure 18; the condition bypassing the low-pass synchrony filter and using constant filter bandwidths was omitted for visual clarity.

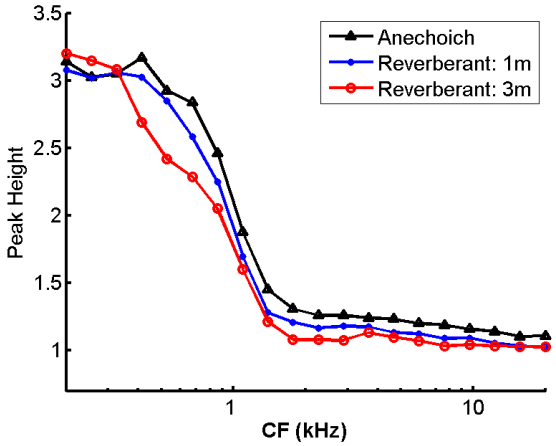

Figure 17 – Normalized correlogram peaks from the simple auditory model with the same input stimuli used in our physiology experiments, plotted against filter center frequency.

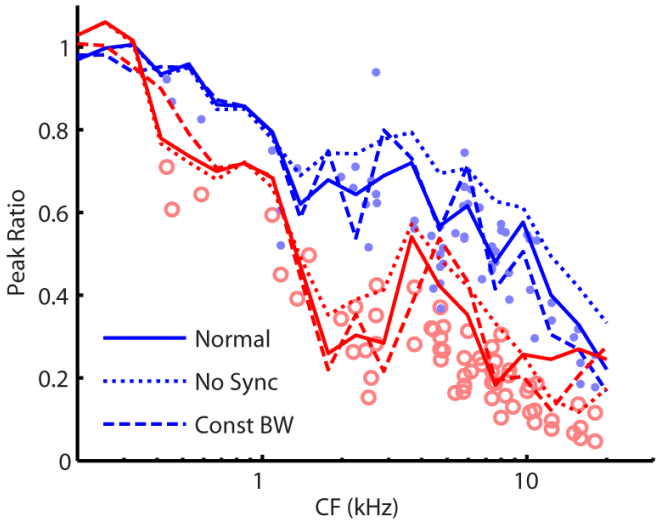

Figure 18 – Peak ratios vs. CF from the model using room-image BRIRs. Three model versions were used; normal, low-pass synchrony filter bypassed, and constant filter bandwidths. The fourth condition (constant bandwidth, low-pass synchrony filter bypassed) condition is not plotted for visual clarity as it did not qualitatively differ from the condition with increasing bandwidths and bypassed synchrony filter. Blue / upper traces represent the 1m reverberation condition; red / lower traces represent the 3m reverberant condition. Scatter plot data (blue dots and red circles for 1m and 3m reverberation, respectively) are peak ratios from our experimental data (same as in Figure 7) for qualitative comparison.

Interestingly, we found surprisingly little effect of either model manipulation. Bypassing the synchrony filter does appear to slightly raise peak ratios for frequencies above the filter's cutoff (dotted lines in Figure 18 are generally above solid and long-dashed lines for frequencies above 1 kHz), but a large frequency-dependent decline is still visible. Filter bandwidths does not appear to make an appreciable difference in these plots. Thus while loss of synchrony does seem to play some role in peak ratios, neither loss of synchrony nor increasing filter bandwidth appear to dominate the frequency-dependent pattern of results seen in our data, suggesting that some property of the BRIR itself, rather than properties of auditory processing, may be sufficient to account for most of this overall frequency dependence.

Given this result, we generated new BRIRs using white noise with an exponentially decaying envelope, and again bypassed the low-pass synchrony filter. We tested both interaurally uncorrelated noise and partially correlated noise, with an average IACC matching that in the room-image BRIRs (excluding the first 50 ms to eliminate the direct response and a few of the stronger early reflections). For the exponential envelope, we used a time constant of 150 ms, which roughly matched the decaying tail of our room-image BRIRs. These exponential BRIRs were scaled and added to a direct-sound impulse to match the D/R energy ratios in our 1 m and 3 m room-image BRIRs.

We ran the model 100 times under each condition, generating different noise impulse responses and using a unique noise token as input for each run. We plot the results for the D/R ratio matching the 3m reverberation condition in Figure 19. Using constant filter bandwidths there is no decline in correlogram peak ratio with frequency. Contrasted against the results in Figure 18, this result highlights that some structure inherent in the reverberant impulse response, rather than simply loss of synchrony to stimulus fine-structure at higher frequencies, is responsible for some of the frequency-dependent decline in peak ratios in our model. Using CF-dependent filter bandwidths imparted a downward-sloping trend, more notable for uncorrelated noise than for partially correlated noise. For this reason we focus on constant filter bandwidths when using these simple, random BRIRs, comparing against the clear frequency-dependent trend seen regardless filter bandwidth dependence using the room-image BRIRs.

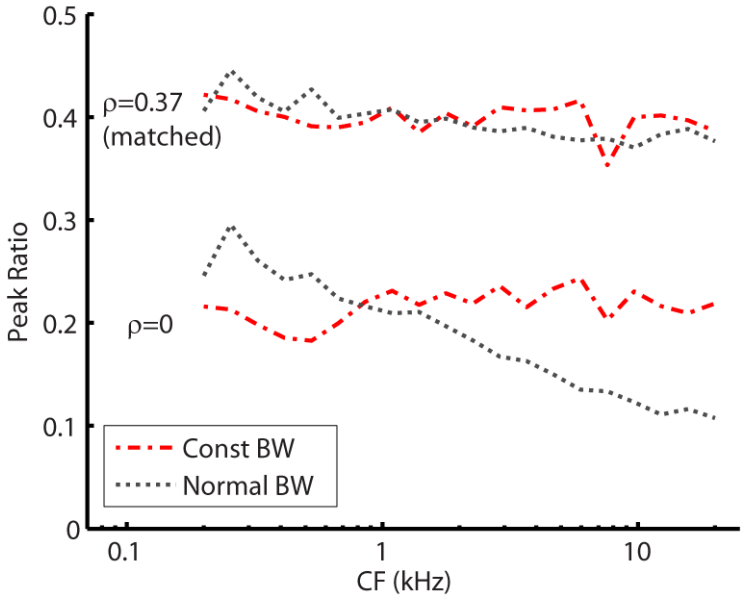

Figure 19 – Means, over 100 runs, of cross-correlation model peak ratios (3m reverberant : anechoic) using exponentially-decaying white noise BRIRs with a time constant of 150 m, and an IACC either matched to the room-image BRIRs or equal to zero, and no synchrony filter. Red, dash-dotted lines represent results with constant filter bandwidths, and grey, dotted lines represent results with CF-dependent bandwidths.

To gain insight into how BRIR-related structure may govern this type of frequency-dependent behavior, we generated another set of BRIRs using a series of 5000 impulses[2] distributed randomly throughout the duration of the room-image BRIRs (about 500 ms). Impulses in the right response were randomly jittered from impulses in the left response using a zero-mean uniform distribution[3] of varying width, to roughly capture the fact that the maximum ITD of any given reflection is limited by interaural distance. Like with the noise BRIRs, a decaying exponential envelope (using a time constant of 150 ms) was imposed on the resulting responses. The model's synchrony filter was bypassed, and equal-bandwidth auditory filters were used. Figure 20 shows peak ratios for the D/R energy ratio matching the 3m reverberant

[2] 500 and 50,000 impulses were also used and did not make a qualitatively significant difference.
[3] Gaussian and triangular distributions were also used and did not make any significant qualitative difference.

condition, with jitter width as a parameter. Peak ratios are increasingly low-pass with increasing jitter, reflecting that the left and right signals are strongly correlated for interaural separations roughly less than the wavelength corresponding to the CF of interest. As the jitter distribution is widened, the fall-off in peak ratio falls to lower and lower frequencies. The 2 ms condition is not particularly physiologically relevant, but helps demonstrate the dependence of results on the imposed temporal structure in the BRIR. These results demonstrate how BRIR-inherent structure might govern some frequency-dependent trend in SXC peak ratios.

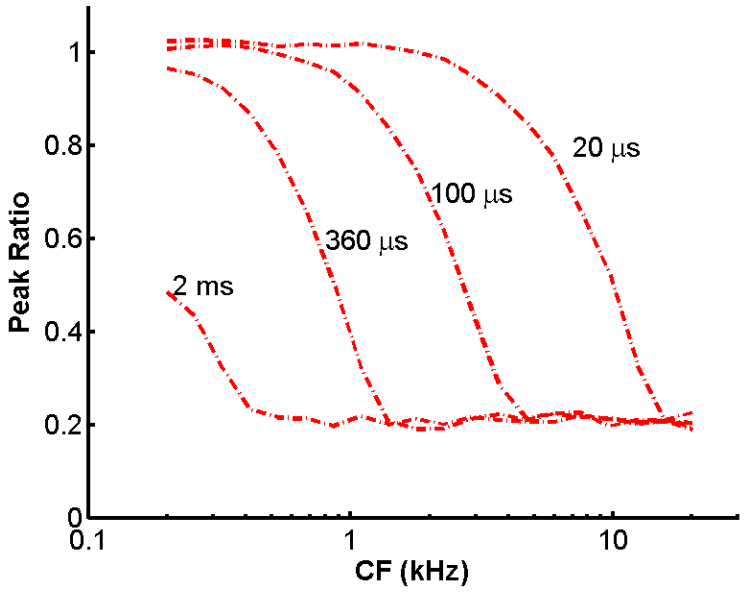

Figure 20 – Mean peak ratios, over 100 runs, from the cross-correlation model using randomly spaced impulse filters (3m D/R energy ratio) using jitter amount as a parameter. Impulses in the right response were jittered randomly from impulses in the left response by a rectangular pdf with zero mean and width as marked. The model's low-pass synchrony filter was bypassed, and constant filter bandwidths were used.

V. Summary and Discussion

Peripheral mechanisms contribute to robust coding of directional information

Previous work (Devore et al. 2009) demonstrated that responses in the IC of cats shows more robust coding of directional information than predicted by a binaural cross-correlation model (Hancock & Delgutte 2004). Neural responses near the onset of a stimulus were more reliable than responses to the ongoing portion of the stimulus. Similar recordings in the IC of the awake rabbit (Devore & Delgutte 2010) showed that degradation of IC response range in reverberation was more severe for higher-CF neurons. We found some similarities between the AN and the IC data. The frequency-dependent increase in degradation with CF due to reverberation observed in the IC was also present qualitatively in AN data. Similarly, robust directional coding near stimulus onsets that may help explain this robustness was demonstrated in the AN data. This suggests that at least part of the robustness to reverberation observed in IC recordings is present in the monaural inputs to the binaural system, rather than being a property of binaural processing.

Reverberation affects ongoing stimuli more than onsets

To investigate the time-course of degradation of timing cues in reverberation, and to address the role of AN adaptation, we extended the shuffled correlation analysis introduced by Joris (2003) to a time-dependent analysis yielding the correlation as a function of peristimulus time. Figure 10 and Figure 15 show the value of this analysis by demonstrating the time-varying structure in the available timing cues in the presence of reverberation. For noise bursts, timing cues in the AN response were most reliable near the stimulus onset, providing a possible means of accurately localizing the sound even in reverberant environments. This result appears to extend to sharp rises in the stimulus envelope of speech.

We quantified the increased reliability of onset timing information in two ways. The first, shown in Figure 11, parallels previous work in our lab by Devore and colleagues, who quantified the effect of reverberation as the “relative range,” defined as the ratio of the range of firing rates to stimuli of varying azimuths in reverberation to the range of firing rates to anechoic stimuli. For

almost all AN fibers, the directional information in reverberation was more reliable in the first 50 ms than in the remaining portion of the stimulus. This result parallels those seen in the IC, shown in Figure 21, where the relative range in the first 50 ms tended to be larger than the relative range over the remaining portion of the stimulus. This similarity is not unexpected as the onset robustness reflects the buildup of reverberant energy in the stimulus.

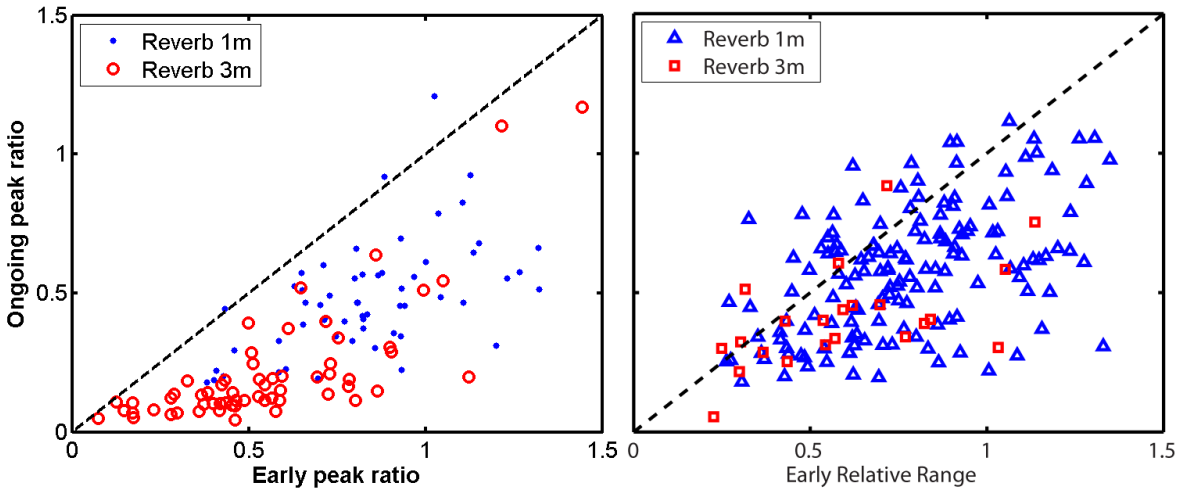

Figure 21 – AN response reverberant-to-anechoic peak ratio (left) and IC response relative ratio (right) computed over the early (0-50 ms) vs. ongoing (51-400 ms) portion of the stimulus. Points tend to fall below the diagonal (dashed line), indicating that reverberation results in greater degradation of directional information in the ongoing response of both cell types compared to the early response, reflecting the buildup of reverberant energy in the acoustic stimulus. Due to differences in the computation of peak ratio vs. relative range, a direct quantitative comparison is not particularly informative.

Peak heights from the ANSXC, which removes the effect of adaptation on the SXC, were generally smaller than peak heights computed from the full-stimulus SXC, indicating that peripheral adaptation contributes to robust directional coding by emphasizing the response near the stimulus onset. These results suggest that some of the robustness to reverberation seen in directional responses in the IC may originate in the auditory periphery rather than in the brainstem.

Frequency-dependent degradation is not be solely due to loss of synchrony

Recordings from IC units in awake rabbits (Devore et al. 2010), similarly to our AN recordings, show a transition from fine-structure to envelope ITD sensitivity with increasing CF. These experiments used the same BRIRs, recording the range of responses to the same noise token of varying simulated azimuths. The effect of reverberation, quantified by the relative range of firing rates of IC units in reverberant and anechoic environments, was stronger for higher-CF units, similar to the data presented in this thesis. Devore and colleagues showed that this degradation followed a step-like pattern with a cutoff near 2 kHz rather than a gradual decline. This result seemed to indicate that the CF-dependence of degradation may not be an effect of frequency per se, but instead may be due to reverberation more severely affecting the envelope timing cues than the fine-structure cues.

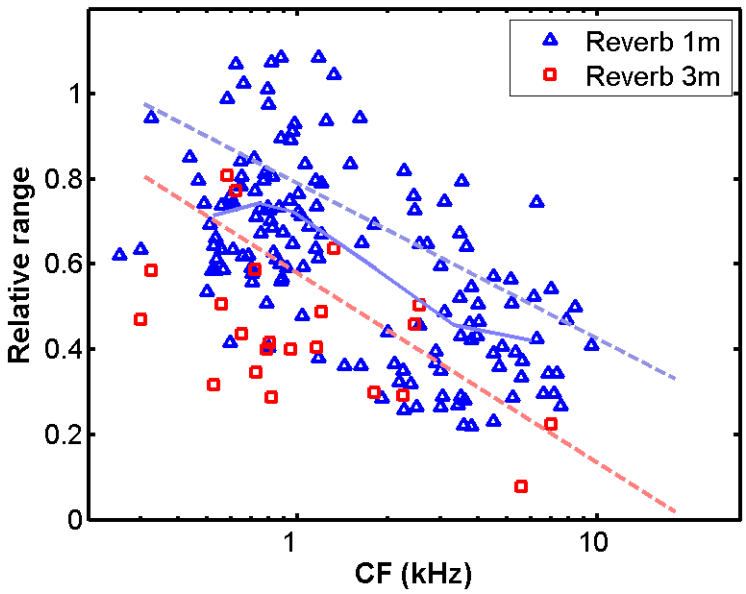

Figure 22 – Relative range in the IC of awake rabbits (from Devore & Delgutte 2010), showing a step-like degradation in directional coding for the 1 m reverberant condition. Relatively few data were collected for the 3 m condition, precluding a thorough analysis. Solid line shows the means of equi-populated bins. Dashed lines represent linear fits to our AN peak ratio data (Figure 7) for qualitative comparison. The overall trend of declining directional sensitivity with increasing frequency in reverberant conditions is consistent with our results, but the specific pattern of degradation differs in that our data do not show a step-like decline.

In contrast, our AN data do not show a similar step-like transition in direction coding robustness. Peak ratios gradually declined with CF over the entire range, even at higher CFs where auditory neurons are not sensitive to stimulus fine-structure. This result suggests that some factor beyond loss of synchrony may be involved in the observed frequency-dependent degradation of directional sensitivity. Using a simple auditory processing and cross-correlation model, we showed that while the loss of synchrony can contribute to some decline in peak ratios at frequencies above the synchrony cutoff, much of the decline with increasing CF is still present even when the model's low-pass synchrony filter is bypassed. This model falls short in predicting peak heights, missing factors in auditory processing such as compression, which is likely to affect overall peak heights. Still, the model appears to capture the relative effects of reverberation, and this result suggests that, in addition to the likely role of loss of synchrony, some properties of the BRIR itself might be responsible for the frequency-dependent effects of reverberation. Changing the BRIRs to partially-correlated white noise with decaying envelopes rather than room-image BRIRs, we observed that the model loses this frequency dependence, strengthening the suggestion of a role of BRIR-inherent properties.

Forcing temporal proximity of reflections received by the left and right ears produced a clear frequency-dependent pattern in peak ratios of the model output, providing insight into how constraints on BRIR structure might give rise to such patterns. Real room BRIRs clearly have much more detailed structure not captured by this jittered-impulse model. For example, reflections are generally sparse at first and increasingly dense in the later response. Additionally, for a given room, the ITD of a reflection is not independent of the ITD of at least some temporally proximate reflections, especially for earlier reflections. Much more realistic temporal structure is captured by room-image BRIRs, and using these more-realistic BRIRs gives us a more complicated frequency dependence than shown in Figure 20. While we are careful in generalizing any conclusions due to the simplicity of the auditory processing model and because the model does poorly in matching peak heights of the actual data, these results demonstrate why we might expect BRIR-inherent structure, in addition to auditory processing, to be partly responsible for the frequency-dependence of directional coding in reverberation.

Due to the different metrics used (relative range vs. peak ratio), a direct comparison between the data collected from the IC and the data presented in this thesis is not possible. However, the difference between the step-like degradation with CF in the IC compared to the more gradual decline observed in the AN suggests some difference in the robustness of directional coding in reverberation between the two locations in the auditory pathway. This difference may reflect further loss of envelope-related timing cues upstream of the AN in the auditory pathway, or may suggest an additional mechanism that partially compensates for reverberation but only when fine structure cues are available in the periphery. Further insight into these possibilities would require a more thorough simulation of the processing mechanisms that lie between the AN and the IC.

Envelope modulations in speech may allow more robust localization in reverberation

We looked at AN responses to speech-modulated noise to determine if the natural modulations in the speech envelope might provide multiple onsets that would allow for robust coding of directional information in reverberation. We found that even in relatively severe reverberation, DSXCs showed reliable peaks near major rises in stimulus amplitude. Speech, and many other natural sounds, contain many such onsets that might allow for robust determination of source direction in reverberation. Stop consonants such as /p/, /k/, and /t/ are good examples, containing sudden bursts of noise usually preceded by tens to hundreds of milliseconds of silence during which reverberation to previous syllables can die down. In addition, vocalization energy moving in and out of a given frequency band (such as by a formant transition) may provide enough envelope modulation to allow for robust direction coding emphasized by peripheral adaptation. Further work should be conducted to more fully characterize the role of adaptation in such situations.

Many current hearing aids employ some type of compression algorithm that reduces the dynamic range of incoming auditory signals. This serves a useful purpose to deal with loudness recruitment in people with sensorineural hearing loss (e.g. Moore et al. 1992), yet our results suggest that this type of acoustic processing may diminish users' abilities to robustly encode directional information as dynamic compression can amplify otherwise soft reverberant tails

from preceding stimuli/syllables. This may cause a listener's AN to be continuously driven without allowing recovery from adaptation, thus diminishing the adaptation-driven emphasis to stimulus onsets. While a detailed analysis of hearing aid compression algorithms is beyond the scope of this thesis, the present results warrant further research into the effects of such processing algorithms in reverberant environments.

VII. Acknowledgements

We would like to thank Connie Miller for performing the surgical procedures and Kenneth Hancock for providing support and software. We also thank Sasha Devore for invaluable assistance in the experiment design and for providing data from the IC. We would also like to thank Grace Wang and Bo Wen for assistance in experiments, and Jay Desloge for providing the room image software. This work was supported by Training Grant NIDCD T32 DC00038 and grants RO1 DC002258 and P30 DC005209 from the National Institute of Health.

Appendix A

Given two positive discrete stochastic inputs X_L and X_R of length N, where X_L represents the left-ear signal and X_R represents a shifted version of the right-ear signal, we wish to find a normalization factor α such that the normalized correlation function $R = \alpha \cdot \sum_{n=1}^{N} X_L[n] \cdot X_R[n]$ (where the factor of 1/N associated with the discrete correlation function is absorbed into α) has an expected value of 1 when the inputs X_L and X_R are uncorrelated. We will assume that X_L and X_R are wide-sense stationary over their support, and will ignore the bias in correlation estimates for increasing lag as it is small for lag values much less than the length of support. Since X_L and X_R are uncorrelated, $E[X_L X_R] = E[X_L]E[X_R]$, where $E[\cdot]$ denotes the expected value operation. We now compute the expected value of R given uncorrelated inputs:

$$E[R] = E\left[\alpha \cdot \sum_{n=0}^{N-1} X_L[\mathrm{n}] \cdot X_R[\mathrm{n}]\right] = \alpha \cdot \sum_{n=0}^{N-1} E[X_L \cdot X_R]$$

$$= \alpha \cdot \sum_{n=0}^{N-1} E[X_L] \cdot E[X_R]$$

$$= \alpha \cdot N \cdot E[X_L] \cdot E[X_R]$$

We use $\frac{1}{N} \cdot \sum X_L$ and $\frac{1}{N} \cdot \sum X_R$ to estimate $E[X_L]$ and $E[X_R]$, respectively, where n represents the number of samples in each vector. This gives us

$$E[R] = \alpha \cdot \frac{1}{N} \cdot \sum X_L \cdot \sum X_R$$

Thus to normalize this cross-correlation so that E[R] = 1 (given uncorrelated X_L and X_R), we use the normalization factor

$$\alpha = \frac{N}{\sum X_L \cdot \sum X_R}$$

So we normalize the correlation by the length of the input vectors divided by the product of the sums of each vector, resulting in a metric that is 1 for uncorrelated inputs. Note that this metric is also 1 when both signals are constant DC values, as is approximately the case in summed responses of auditory nerve fibers to pure tone at high frequencies.

References

Aitkin L.M., Pettigrew J.D., Calford M.B., Phillips S.C., Wise L.Z. (**1985**). Representation of stimulus azimuth by low-frequency neurons in inferior colliculus of the cat. *J Neurophysiol*, **53**(1), 43-59.

Allen J.B., Berkley D.A. (**1979**). Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.*, **65**(4), 943–950.

Arbogast T.L., Mason C.R., Kidd J. (**2005**). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.*, **117**(4), 2169-2180.

Beranek L.L. (**2005**). Concert halls and opera houses: music, acoustics, and architecture. *J. Acoust. Soc. Am.*, **117**, 987.

Carney L.H., Yin T.C. (**1988**). Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model. *J Neurophysiol*, **60**(5), 1653-1677.

Devore S., Ihlefeld A., Shinn-Cunningham B.G., Delgutte B.: (**2006**). Neural and behavioral sensitivities to azimuth degrade with distance in reverberant environments. *International Symposium on Hearing*.

Devore S., Delgutte B. (**2010**). Effects of reverberation on the directional sensitivity of auditory neurons across the tonotopic axis: Influences of ITD and ILD. *J. Neurosci.*, **In Press**,

Devore S., Ihlefeld A., Hancock K., Shinn-Cunningham B., Delgutte B. (**2009**). Accurate Sound Localization in Reverberant Environments Is Mediated by Robust Encoding of Spatial Cues in the Auditory Midbrain. *Neuron*, **62**(1), 123-134.

Devore S., Schwartz A.H., Delgutte B. (**2010**). Effect of reverberation on directional sensitivity of auditory neurons: Central and peripheral factors. *Advances in Auditory Research: Physiology, Psychophysics and Models*,

Geisler C.D. (**1998**). From sound to synapse: physiology of the mammalian ear. *Oxford University Press, USA*.

Hancock K.E., Delgutte B. (**2004**). A Physiologically Based Model of Interaural Time Difference Discrimination. *J. Neurosci.*, **24**(32), 7110-7117.

Joris P.X. (**2003**). Interaural Time Sensitivity Dominated by Cochlea-Induced Envelope Patterns. *J. Neurosci.*, **23**(15), 6345-6350.

Kiang N.Y. (**1965**). Stimulus coding in the auditory nerve and cochlear nucleus. *Acta Oto-Laryngologica*, **59**(2-6), 186–200.

Kuwada S., Stanford T.R., Batra R. (**1987**). Interaural phase-sensitive units in the inferior colliculus of the unanesthetized rabbit: effects of changing frequency. *J. Neurophysiol.*, **57**(5), 1338-1360.

Liberman M.C. (**1978**). Auditory-nerve response from cats raised in a low-noise chamber. *J. Acoust. Soc. Am.*, **63**(2), 442-455.

Loizou P.C., Hu Y., Litovsky R., Yu G., Peters R., Lake J., Roland P. (**2009**). Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *J. Acoust. Soc. Am.*, **125**, 372.

Louage D.H.G., van der Heijden M., Joris P.X. (**2004**). Temporal properties of responses to broadband noise in the auditory nerve. *J. Neurophysiol.*, **91**(5), 2051-2065.

Marrone N., Mason C.R., Kidd J. (**2008**). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *J. Acoust. Soc. Am.*, **124**(5), 3064-3075.

Moore B., Johnson J., Clark T., Pluinage V. (**1992**). Evaluation of a dual-channel full dynamic-range compression system for people with sensorineural hearing-loss. *Ear and Hearing*, **13**(5), 349-370.

Moushegian G., Rupert A.L., Gidda J.S. (**1975**). Functional characteristics of superior olivary neurons to binaural stimuli. *J. Neurophysiol.*, **38**(5), 1037-1048.

Shinn-Cunningham B.: (**2003**). Acoustics and perception of sound in everyday environments. *Proc of 3rd International Workshop on Spatial Media*.

Shinn-Cunningham B., Kawakyu K.: (**2003**). Neural representation of source direction in reverberant space. *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*

Shinn-Cunningham B.G., Kopco N., Martin T.J. (**2005**). Localizing nearby sound sources in a classroom: Binaural room impulse responses. *J. Acoust. Soc. Am.*, **117**(5), 3100-3115.

Smith P.H., Joris P.X., 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. Comparative Neurology*, **331**(2), 245-260.

Spitzer M.W., Semple M.N. (**1995**). Neurons sensitive to interaural phase disparity in gerbil superior olive: diverse monaural and temporal response properties. *J. Neurophysiol.*, **73**(4), 1668-1690.

Yin T.C., Chan J.C. (**1990**). Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.*, **64**(2), 465-488.