

Linear Processing of Interaural Level Difference Underlies Spatial Tuning in the Nucleus of the Brachium of the Inferior Colliculus

Sean J. Slee¹ and Eric D. Young¹

¹Biomedical Engineering Department, Johns Hopkins University, Baltimore, Maryland 21205

The spatial location of sounds is an important aspect of auditory perception, but the ways in which space is represented are not fully understood. No space map has been found within the primary auditory pathway. However, a space map has been found in the nucleus of the brachium of the inferior colliculus (BIN), which provides a major auditory projection to the superior colliculus. We measured the spectral processing underlying auditory spatial tuning in the BIN of unanesthetized marmoset monkeys. Because neurons in the BIN respond poorly to tones and are broadly tuned, we used a broadband stimulus with random spectral shapes (RSSs) from which both spatial receptive fields and frequency sensitivity can be derived. Responses to virtual space (VS) stimuli, based on the animal's own ear acoustics, were compared with the predictions of a weight-function model of responses to the RSS stimuli. First-order (linear) weight functions had broad spectral tuning (approximately three octaves) and were excitatory in the contralateral ear, inhibitory in the ipsilateral ear, and biased toward high frequencies. Responses to interaural time differences and spectral cues were relatively weak. In cross-validation tests, the first-order RSS model accurately predicted the measured VS tuning curves in the majority of neurons, but was inaccurate in 25% of neurons. In some cases, second-order weighting functions led to significant improvements. Finally, we found a significant correlation between the degree of binaural weight asymmetry and the best azimuth. Overall, the results suggest that linear processing of interaural level difference underlies spatial tuning in the BIN.

Introduction

The formation of auditory objects is essential for navigating complex auditory environments (Shinn-Cunningham, 2008). This perceptual phenomenon is based on the simultaneous processing of multidimensional acoustic cues, including spatial information derived from binaural disparities and the directional filtering properties of the outer ear (Bregman, 1990; Darwin, 2008). Although massive convergence of acoustic features occurs in the central nucleus of the inferior colliculus (ICC) (Oliver et al., 1997; Davis et al., 1999; Delgutte et al., 1999; Chase and Young, 2005; Devore and Delgutte, 2010), a systematic representation of auditory space has not been reported there. Consistent with these results, our previous studies suggest that sound localization cues are not integrated coherently at this level of processing (Slee and Young, 2011). In contrast, the superior colliculus (SC) contains a well-defined spatial map (Knudsen, 1982; Palmer and King, 1982; Middlebrooks and Knudsen, 1984). These results suggest

that a transition to a spatially based stimulus representation takes place between the inferior and superior colliculi.

In barn owls, the external nucleus of the inferior colliculus provides the main auditory input to the SC and is well studied (Knudsen and Konishi, 1978). It combines cues based on interaural time difference (ITD) and interaural level difference (ILD) in a multiplicative fashion to form narrow receptive fields (Knudsen, 1987; Peña and Konishi, 2001). However, results in barn owls may not easily generalize to the mammal. In owls, ITD varies most with azimuth (AZ), whereas ILD varies with elevation (EL) (Moiseff, 1989). Furthermore, these cues overlap substantially in frequency and are present over a large portion of the animal's hearing range. In mammals, the azimuth is encoded by fine structure ITD at low frequencies and ILD (and to some extent, envelope ITD) at high frequencies (Strutt, 1904, 1907; Mills, 1958; Yin et al., 1984). Elevation judgments are based on spectral shape (SS) cues and require broadband sounds (Roffler and Butler, 1968a, b; Huang and May, 1996).

In mammals, the SC receives its auditory input via a topographic projection from the nucleus of the brachium of the inferior colliculus (BIN) (Jiang et al., 1993; King et al., 1998), which in turn receives its input from the ICC (Nodal et al., 2005). Currently there is limited information about the properties of BIN neurons. They are broadly tuned to both frequency and spatial location (Aitkin and Jones, 1992), and show some topographic organization (Schnupp and King, 1997). These results suggest the hypothesis that the BIN plays a primary role in the transition to a spatial representation in the SC.

Received July 18, 2012; revised Jan. 6, 2013; accepted Jan. 11, 2013.

Author contributions: S.J.S. and E.D.Y. designed research; S.J.S. performed research; S.J.S. analyzed data; S.J.S. and E.D.Y. wrote the paper.

This work was supported by NIH Grants DC00115, DC00023, DC05211, and DC012124. We thank Ron Atkinson, Jay Burns, Phyllis Taylor, and Qian Gao for technical assistance; Jenny Estes and Nate Sutuyo for assistance with animal care; and the three anonymous reviewers for their helpful comments on this manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Sean J. Slee Biomedical Engineering Department, Johns Hopkins University, 505 Traylor Research Building, 720 Rutland Avenue, Baltimore, MD 21205. E-mail: sslee1@jhmi.edu.

DOI:10.1523/JNEUROSCI.3437-12.2013

Copyright © 2013 the authors 0270-6474/13/333891-14\$15.00/0

The extent to which the spatial responses of BIN neurons depend on particular localization cues is unknown, as is the extent to which neurons integrate different cues. We measured this by recording from BIN neurons in unanesthetized marmoset monkeys while presenting virtual space (VS) and random spectral shape (RSS) stimuli. The neural responses suggest that the topographical map of spatial tuning in the BIN is primarily formed by linear integration of ILD across frequencies and a topographic variation in the balance between ipsilateral and contralateral inputs.

Materials and Methods

Animal preparation and recording. All procedures were approved by the Johns Hopkins University Institutional Animal Care and Use Committee and conform to the National Institutes of Health standards. Animal care and recording procedures were similar to those described previously for recording in the auditory cortex and ICC of the awake marmoset monkey (Lu et al., 2001; Nelson et al., 2009). For 1 month before surgery, the marmoset was adapted to sit quietly in the restraint device for the 3–4 h periods used for recording. A sterile surgery was performed under isoflurane anesthesia to mount a post for subsequent head fixation during recordings and to expose a small portion of the skull for access to the IC. The exposed skull was covered by a thin layer of dental acrylic or Charisma composite to protect it and form a recording chamber. During a 2 week recovery period, the animal was treated daily with antibiotics (Baytril, 2.5 mg/kg), and the wound was cleaned and bandaged. Analgesics (buprenorphine, 0.005 mg/kg; Tylenol, 5 mg/kg; and lidocaine, topical 2%) were given to control pain.

For recording, a small (~1 mm) craniotomy was drilled at locations determined by stereotaxic landmarks (e.g., bregma) defined during surgery to target the BIN. We have created atlases that include the BIN from the brain sections of three marmosets. The BIN in the marmoset is located from ~2.5–5 mm caudal of bregma and 3–4 mm lateral to the midline. The exposed recording chamber surrounding the craniotomy was covered with polysiloxane impression material (GC America) between recording sessions. After many penetrations (usually >30), the final track in each craniotomy was labeled as described in Fluorescent track labeling, below, and the craniotomy was filled with a layer of bone wax and dental acrylic before making another craniotomy to provide access to other regions of the same IC. Multiple craniotomies were performed on both hemispheres before the animal was killed and perfused for histological evaluation.

Recordings were made in five marmosets: two females and three males. During daily 3–4 h recording sessions, an epoxy-coated tungsten microelectrode (A-M Systems; impedance between 3 and 12 M Ω) or tetrode (Thomas Recording) was slowly advanced using a hydraulic microdrive (Kopf Instruments) to control its depth relative to the surface of the exposed dura. Using a dorsal or lateral approach, the electrode traversed ~0.9–1.2 cm of brain tissue before reaching the IC. During recording, the animal was awake with its eyes open or had its eyes closed (and was presumably asleep in some cases). Its body was loosely restrained in a comfortable seated position with the head held in the natural orientation. Recordings were conducted in low light levels or in darkness. Stimulus presentation, animal monitoring via infrared video camera, and electrode advancement were controlled from outside a double-walled soundproof booth.

For single electrodes, spikes were isolated using a Schmitt trigger. For tetrodes, the Catamaran clustering program (kindly provided by D. Schwarz and L. Carney) was used to separate single units from the electrode signal (Schwarz et al., 2012). In both cases, single units were defined based on visual inspection of traces and by having fewer than 1% of interspike intervals shorter than 0.75 ms. Of all 92 BIN units studied, 77 met this criterion, whereas the remaining 15 units had a slightly greater percentage of short intervals. These units were included in the sample because their responses were stable, interesting, and not obviously biased by activity from neighboring neurons. Their inclusion did not affect the qualitative nature or the statistical significance of the results.

The BIN was targeted physiologically by driving tracks laterally and anteriorly to the tonotopic map in the ICC and ventrally and posteriorly

to a light-sensitive area, likely the SC (Schnupp and King, 1997). The borders of the SC were roughly located by recording multiunit entrainment to a pulsed (1 or 4 Hz), red LED located in front of the monkey, or ~20° or ~40° to the side (contralateral to the IC under study). Although clear light responses were common, the precise mapping of visual receptive field centers in the SC was difficult, probably because of the uncontrolled eye position of the awake marmoset.

Neurons were isolated using a combination of binaural wideband noise bursts (50 or 200 ms duration), VS stimuli, and pure tones of variable frequency and level. When a neuron was first isolated, diotic pure-tone response maps were measured to determine its frequency selectivity. The tones were presented at 30 dB attenuation (~70 dB SPL) over a six-octave range centered at 5 kHz (16 steps per octave) for a duration of 200 ms with a period of 1 s.

Chronic recording methods did not allow confirmation of the recording site on every track. Instead neurons were classified as “BIN-like” using the following criteria: (1) electrode position relative to the physiologically defined SC and ICC; (2) strong responses to broadband auditory stimulation and weak responses to tones; (3) possible habituation to repeated stimuli; (4) short latencies consistent with previous studies (~5–20 ms); (5) no response or weak responses to spots of light; and (6) position relative to the final track labeled in each craniotomy by coating the electrode with fluorescent dye.

Fluorescent track labeling. The final track in each craniotomy was labeled by coating the electrode with a fluorescent lipophilic dye (DiCarlo et al., 1996). The electrode was dipped in the dye (Vybrant Cell labeling; Invitrogen), removed, and then allowed to dry for 10 s. This process was repeated 10 times. At the end of this process, the dye on the electrode was usually apparent with the naked eye. Dyes with different emission spectra (DiO, DiI, and DiD) were used in different craniotomies to recover the location of several closely spaced tracks. The electrode track was positioned at a location that contained many BIN-like physiological responses over the course of previous tracks. In addition, at the end of all recordings in an animal, focal electrolytic lesions were made at BIN-like recording sites. The animal was killed, and the brain fixed (0.5% paraformaldehyde), sectioned (50 μ m), and stained (cytochrome oxidase) to localize the fluorescent tracks and lesions relative to the BIN (Fig. 1). The sections were imaged with a 0.5, 10, or 20 \times objective using bright-field and fluorescence microscopy. The fluorescent tracks were isolated using filters B-2E/C, G-2E/C, and ET-Cy5 (Omega Optical) for DiO, DiI, and DiD, respectively.

Acoustic stimuli. Sound stimuli were generated digitally by digital-to-analog (D/A) converters in a Tucker-Davis RP2.1 or a National Instruments 6205 board, and were presented over calibrated headphone drivers (STAX) acoustically coupled to ear inserts (closed system). All stimuli were gated on and off with 10 ms linear ramps. The acoustic system was calibrated *in situ* daily with a probe microphone, and all stimuli (unless otherwise indicated) were corrected for the calibration.

Acoustic measurement of sound localization cues. A pair of head-related transfer functions (HRTFs; one for each ear) contains the ITD, ILD, and SS cues specific to a single position of a sound source. We measured individual HRTFs in each of the five marmoset monkeys at the beginning of the sequence of experiments. The details of this procedure and the results have been described previously (Slee and Young, 2010). Briefly, marmosets were anesthetized with a mixture of ketamine (30 mg/kg) and acepromazine (1 mg/kg) and then secured to a stand in the center of an acoustically isolated chamber. The T60 (time to a level 60 dB down from the peak of the room impulse response) of the chamber was 22 ms (measured in a band from 0–48.8 kHz). The T40 was 7 ms. We also measured T60s and T40s in restricted frequency bands (2 octaves wide) by band-pass filtering the impulse response with a fourth-order Butterworth filter. The band centers (in kilohertz) and T60s (in milliseconds) were (0.25, >83), (0.5, >83), (1, >83), (2, 42), (4, 19), (8, 17), and (16, 16). The band centers (in kilohertz) and T40s (in milliseconds) were (0.25, 36), (0.5, 27), (1, 18), (2, 11), (4, 11), (8, 7), and (16, 7). The 8192-point Golay stimulus limited measurements to below ~83 ms. The energy decay was not exponential. This could be caused by a nonuniform distribution of sound absorption within the room, perhaps from floor space distant from the microphone, which was not covered with acoustic foam.

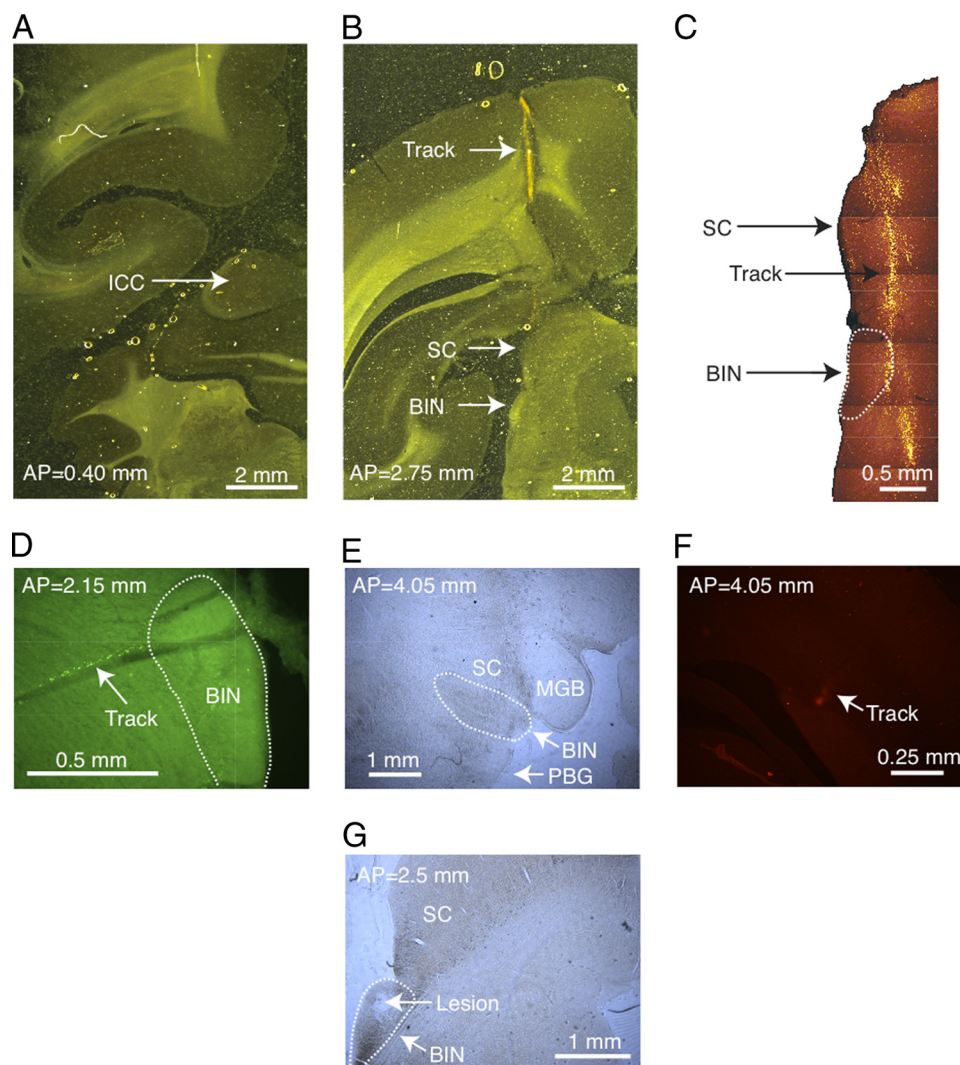


Figure 1. *A*, A brain section stained for cytochrome oxidase from marmoset 12W. The ICC is present in this section, shown by the dark staining indicated by the arrow, but the BIN is not, and there are no tracks. *B*, Another brain section in the same animal located 2.35 mm anterior to the section in *A*. The arrows point to the track mark through cortex, the SC, and the BIN. *C*, A higher magnification of the section shown in *B* viewed with fluorescence microscopy. The picture is a photomontage of several adjacent images. The arrows point to the SC, the BIN (outlined), and the fluorescently labeled electrode track. *D*, Labeled track through the posterior right BIN in a different marmoset (18W) using a lateral electrode approach. *E*, A far anterior section of the right BIN from 18W also showing the SC, medial geniculate body (MGB), and the parabigeminal nucleus (PBG). The track is not labeled at this depth. *F*, The same section as *E* at a position dorsal and lateral that contains a piece of the fluorescently labeled track (arrow). *G*, A posterior section through the left BIN from animal 18W where an electrolytic lesion was made (arrow).

We note that effects of reverberation were greatly reduced in the BIN experiments by windowing the impulse responses used to generate virtual space stimuli presented in closed field. The HRTFs were measured with miniature microphones placed just in front of the tympanic membranes. A broadband Golay stimulus (Zhou et al., 1992) was presented from a sound source located 1 m from the animal. The stimuli were presented at 15° intervals from 150° (ipsilateral to the right ear) to −150° in AZ. ELs ranged from −30 to 90° in 17 steps of 7.5°. In total, a collection of 357 HRTFs were measured in each ear over a period of ~4 h.

Removing spectral shape cues. Marmoset sound localization cues consist of ILDs, ITDs, and SS cues (Slee and Young 2010). In some instances, the SS cues were removed by smoothing the HRTFs in the frequency domain. Each HRTF was initially smoothed with a 17.9 kHz (1501 point) triangular filter from 3 to 40 kHz. The result was smoothed again with a 6 kHz triangular filter from 0.2 to 40 kHz. This filtering algorithm was found to reduce high-frequency notches with minimal effect on frequencies <3 kHz. The result of this signal processing is shown in Figure 7E.

VS stimuli. VS stimuli were constructed using HRTFs measured in both ears of each marmoset (individual HRTFs) to impose sound localization cues on samples of random noise. To present accurate VS stimuli, the transfer function of the acoustical delivery system and the ear canal between it and

the microphone must be removed (Wightman and Kistler, 1989). In all experiments, the face of the microphone used to measure HRTFs and the tip of the probe tube used to calibrate the acoustic delivery system were located in roughly the same position within the ear canal (with an accuracy of ~1–2 mm). For VS stimuli, the acoustic signal was multiplied by the complex gain $H(f)/C(f)$, where $H(f)$ is the HRTF from the desired spatial direction, and $C(f)$ is the daily calibration of the acoustic delivery system (Delgutte et al., 1999). On some days, the acoustic calibration contained large notches above 30 kHz that could not be removed by repositioning the ear inserts. In these cases, the VS stimuli were sharply low-pass filtered at 30 kHz by setting the higher-frequency components to zero.

For each VS stimulus, the calibrated HRTF was interpolated from 8192 to 32,768 points (using resample in Matlab) and multiplied by a random 32,768 point sample of noise in the frequency domain (to impose the VS cues on the noise carrier). The product was then inverse transformed into the time domain and resampled to the sampling rate of the National Instruments D/A board (100 kHz). Finally, the stimulus was converted to analog with the National Instruments D/A converter and presented for a total duration of 100 or 300 ms with a stimulus period of 0.6, 1, or 2 s at ~20 dB above the diotic binaural noise threshold. Note that these stimuli include ITD, ILD, and spectral sound localization cues.

Virtual space receptive fields (VSRFs) were determined using the average firing rate (FR) for VS stimuli located from 150° (ipsilateral to the right ear) to −150° in AZ in steps of 30°, or, in some neurons, from −135° to 135° AZ in steps of 15°. ELs ranged from −30° to 90° in steps of 15°.

Binaural RSS stimuli. The binaural RSS stimuli were a variant of previous monaural versions (Young and Calhoun, 2005). Each dichotic stimulus consisted of a sum of tones spaced logarithmically at $\frac{1}{64}$ octave. The tones had randomized phase, and the phases were the same in the two ears. For each ear, the stimuli were arranged in quarter-octave bins (16 tones of the same level) that span the marmoset hearing range. The level of the tones in the bin centered at frequency f was fixed at $S(f)$ decibels, drawn from a Gaussian distribution with mean zero and an SD ~ 8.6 dB relative to a reference sound level. A subset of stimuli was presented with $S(f) = 0$ dB at all frequencies (the reference sound level) to calculate the reference spike rate, R_0 , in the model described below.

The spike rates, r_j , were recorded from 422 RSS stimuli with different pseudorandom spectra (or as long as the recording permitted). RSS stimuli were presented at a reference level 20 dB relative to threshold for a duration of 100 or 300 ms with a period of 1 or 2 s. The RSS stimuli were corrected for the acoustic calibration only for a subset of the experiments. This was because our initial algorithm took ~ 25 min to synthesize the RSS stimulus set after the daily calibration was measured. Therefore, we occasionally obtained recordings before this processing was finished. Subsequently, we improved the speed of the algorithm. Because of the construction of the stimuli, calibration has no effect on first-order RSS models, but can influence second-order models. This is so because the sum of the dB stimulus values at each frequency is zero.

The measured spike rates were used to fit the parameters of a quadratic gain function:

$$r = R_0 + \mathbf{w}^T \mathbf{s} + \mathbf{s}^T \mathbf{M} \mathbf{s}, \quad (1)$$

where R_0 is the reference spike rate for all 0 dB stimuli; \mathbf{s} is a vector containing the decibel values of the stimulus at different frequencies in the two ears, i.e., $\mathbf{s} = [S_R(f_1), \dots, S_R(f_N), S_L(f_1), \dots, S_L(f_N)]^T$; \mathbf{w} is the vector of first-order weights [in spikes/(seconds \times decibels)]; and \mathbf{M} is a matrix of second-order weights, m_{jk} [in spikes/(seconds \times decibels squared)], which is the gain for joint energy in the j th and k th frequency bins. Note that the values of m_{jk} measure binaural as well as monaural interactions. The weights were estimated by minimizing the mean square error between the model rates predicted by Equation 1 and the empirical rates r_j (Young and Calhoun, 2005). Because the model depends on the parameters linearly, this is a well-understood optimization problem, which is solved using the method of normal equations. The first-order weights \mathbf{w} reveal the linear spectral tuning of the ipsilateral and contralateral inputs. The second-order weights \mathbf{M} give a measure of sensitivity to intensity and monaural or binaural interaction of two frequencies (within or between ears), as might be produced by pinna-based spectral cues.

Models were fit with and without the second-order terms, as explained in Results.

ITD stimuli. Because the RSS stimuli have fixed ITD cues, ITD sensitivity was measured in some units with calibrated noise stimuli. The ITDs ranged from $\pm 600 \mu\text{s}$ in steps of $50 \mu\text{s}$. The stimuli were presented at 20 dB relative to threshold for 300 ms with a period of 1 s. The stimuli were

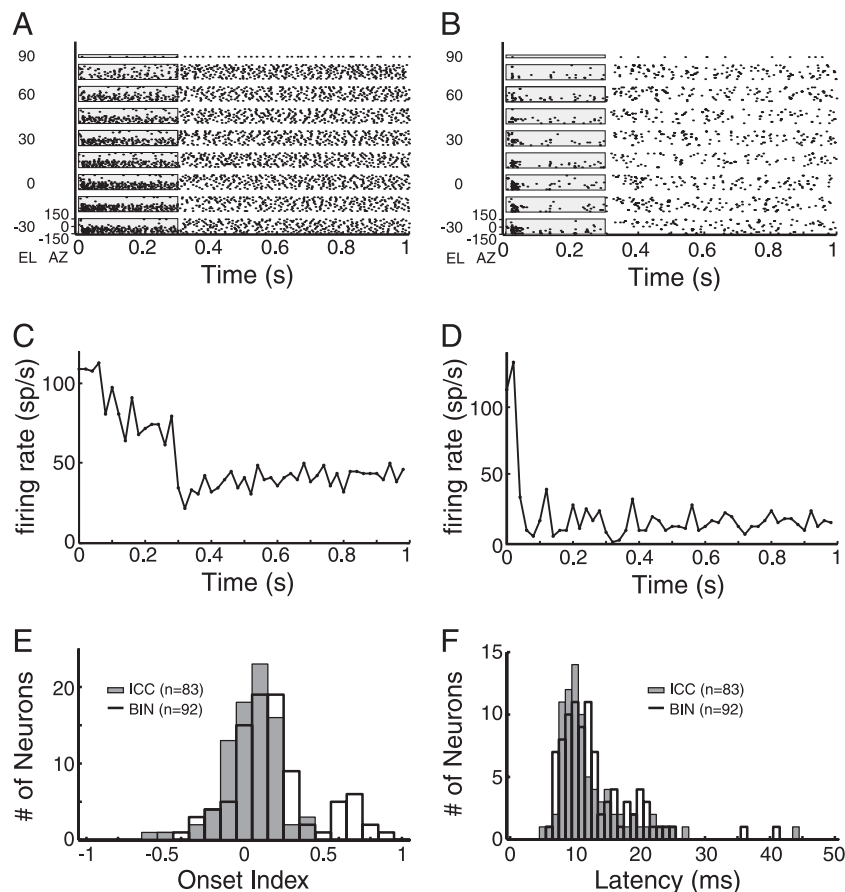


Figure 2. *A*, A spike raster for a neuron with sustained responses to VS stimuli. Each dot represents the time of a single spike. Each row corresponds to a different VS location. Different ELs are separated vertically. The gray box indicates the stimulus time. *B*, Same as *A* for a neuron with onset responses to VS stimuli. *C*, A PSTH for the neuron in *A* averaged across all VS locations. *D*, Same as *C* for the neuron in *B*. *E*, Histogram of onset indices for BIN (black) and ICC (gray) neurons (for definitions, see Materials and Methods). *F*, Histogram of average first-spike latencies for BIN (black) and ICC (gray) neurons (for definitions, see Materials and Methods).

repeated five times. The significance of ITD sensitivity was tested with an ANOVA on the distribution of spike rates.

Spike latency and onset index. The first-spike latency and onset index (Schumacher et al., 2011) were measured in response to VS stimuli. For latency, the spike times (from all positions) were binned at a 1 ms resolution. The first-spike latency was defined as the first time bin where the SD was at least three times greater than the average SD for spontaneous responses. The onset index was defined as follows:

$$O_{\text{index}} = (\text{Onset}_{\text{FR}} - \text{Steady}_{\text{FR}}) / (\text{Onset}_{\text{FR}} + \text{Steady}_{\text{FR}}), \quad (2)$$

where Onset_{FR} is the average firing rate during the first 50 ms of the stimulus, and $\text{Steady}_{\text{FR}}$ is the average firing rate from 50 to 100 ms relative to stimulus onset. The same time windows were used for all neurons independent of the stimulus duration (100 or 300 ms). The O_{index} ranges from 1 for a pure onset response to 0 for a sustained response, and is -1 for a buildup response.

RSS model predictions. RSS models were used to predict the firing rate response to novel RSS stimuli (meaning stimuli not used to derive the model). The model was generated with a pseudorandom sample of 90% of the data and was used to predict the remaining 10% of the data. This was repeated 1000 times so that each rate was predicted ~ 100 times by a different RSS model. The quality of the RSS model prediction was quantified by the fraction of variance (f_v) defined as follows:

$$f_v = 1 - \frac{\sum_j (r_j - \bar{r})^2}{\sum_j (r_j - \bar{r}_j)^2}, \quad (3)$$

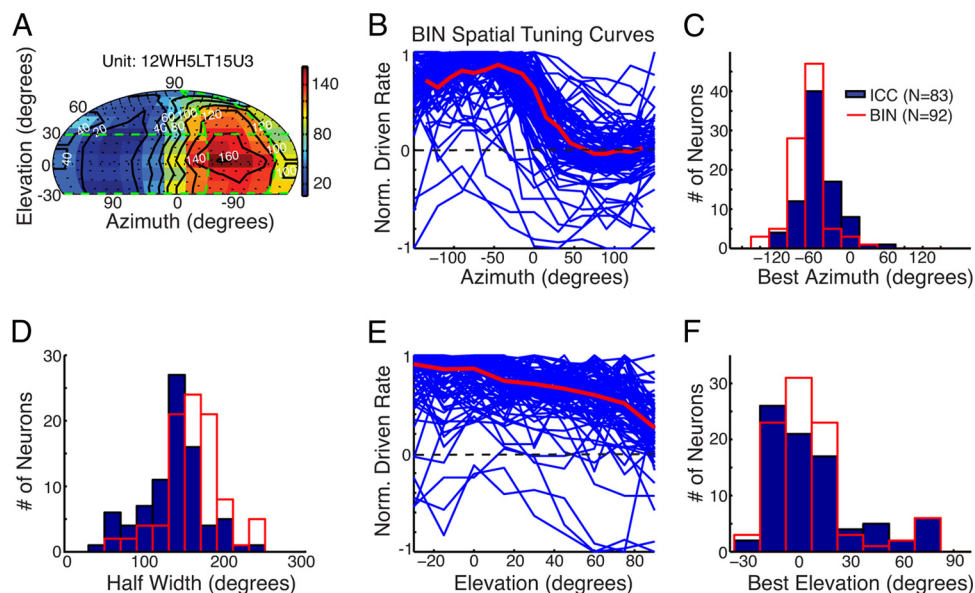


Figure 3. *A*, Example of a VSRF for a BIN neuron using a single pole coordinate system (Slee and Young, 2010). The color bar indicates the average spike rate. The white numbers denote the values of the contours of the constant rate (black lines). The green horizontal and vertical lines indicate the regions for which AZ and EL tuning were computed, respectively. *B*, Azimuthal tuning for all BIN neurons, plotted as the normalized rate (rate minus spontaneous rate divided by maximum rate minus spontaneous rate). The negative values show inhibition. The red line is the median of the data. *C*, The best AZ computed by the geometric center of responses (see Spatial tuning) for 92 BIN neurons (red) and 83 ICC neurons (blue). The negative values are in the contralateral field. *D*, The width of AZ tuning measured halfway between the maximum and minimum rates for 92 BIN neurons (red) and 83 ICC neurons (blue). *E*, *F*, Same as *B* and *C* but for EL.

where r_j is the measured response to stimulus j , \hat{r}_j is the predicted response, and \bar{r} is the average rate across all VS stimuli. The fv ranges from 1 for a perfect fit to negative values for poor fits. When $fv = 0$, the model fits as well as a constant at the average firing rate r . fv is used because it is sensitive to both random errors and constant offset errors, whereas the more commonly used Pearson correlation R^2 (among others) is not sensitive to constant offsets; $fv \leq R^2$, but $fv \sim R^2$ for good fits ($fv > 0.5$).

Predicting VSRFs with RSS models. The model in Equation 1 was used to predict VSRFs. The R_0 used in the model was the average firing rate across all VS stimuli. The input stimuli [$S_R(f_j)$ and $S_L(f_j)$] for the binaural RSS model were formed by averaging the HRTF magnitudes in each ear in the same quarter-octave bins used for the RSS stimuli. The model responses were then predicted using Equation 1 (with the first-order terms only) and using three simplifications. (1) For the ILD RSS model, the inputs were the interaural level differences $S_R(f_j) - S_L(f_j)$ between the HRTF magnitudes in each quarter-octave bin. (2) For the contralateral-only RSS model, only the contralateral HRTF magnitudes $S_R(f_j)$ were used as inputs. (3) For the binaural RSS model without SS cues, the SS cues were eliminated by filtering (see above, Removing spectral shape cues), and the result was used to calculate $S_R(f_j)$ and $S_L(f_j)$ as above. All model predictions were quantified by the fraction of variance (Eq. 3).

Nonlinear RSS models. We tested for nonlinear processing in BIN neurons by using the full model in Equation 1, including the second-order terms. (We also tried computing third-order nonlinearities, but the very large number of parameters could not be estimated accurately from our 400-response data set.) We reduced the large number of parameters in the second-order model by systematically adding terms beginning at the frequency of the maximal linear weight (see Fig. 9A). We predicted responses to VS stimuli with the nonlinear model using Equation 1, including a limited range of second-order terms, and evaluated the prediction with fv . Increasing the number of nonlinear weights at first increased the fv because the model had a greater number of free parameters. However, RSS models with too many second-order weights were overfit, and the fv actually decreased. The nonlinear model with the number of nonlinear weights that maximized fv for VS stimuli was used. Notice that this is a cross-validation method, because the model parameters are estimated using the responses to RSS stimuli, but the model is tested using its ability to predict responses to the VS stimuli.

Results

Targeting the BIN

We targeted single-neuron recordings to the BIN using physiological criteria. Briefly, BIN-like recording sites were defined by their broad frequency tuning and lack of a clear tonotopic progression as well as their location relative to a light-sensitive region located dorsally (likely the SC) and a tonotopic region located caudally/medially (likely the ICC). Upon completion of neural recordings, the general targeting accuracy was confirmed with electrolytic lesions and by recovering tracks labeled by coating electrodes with fluorescent dyes (see Materials and Methods). A schematic of the BIN in the ferret has been published previously (King et al., 1998). Figure 1A shows a section of the left hemisphere from one of the marmosets stained with cytochrome oxidase and viewed with bright-field microscopy. The section was taken at a position 0.4 mm anterior to the posterior border of the ICC, so it contains the ICC and no BIN. The ICC is prominent (Fig. 1A, arrow), as indicated by the dark cytochrome staining. The cerebellum is also present ventral to the IC. When imaged using fluorescent microscopy, this section did not show evidence of a labeled track.

A bright-field image of a section from the same animal taken 2.75 mm anterior to the posterior border (AP) of the ICC is shown in Figure 1B. The dark cytochrome staining of the ICC is not seen, however the SC and BIN are apparent (arrows). A track through the cortex is visible dorsal to these structures. A photomontage of higher-resolution fluorescent images of this section is shown in Figure 1C. This figure shows a successfully labeled track that passes through the SC and the dorsal-medial part of the BIN (Fig. 1C, arrows; BIN is outlined with a dotted line). This track was labeled with an electrode coated with DiO (peak emission, 501 nm) and was positioned at approximately the same angle where 29 neurons with BIN-like physiological responses were recorded from this craniotomy. In the same animal, a track (data not shown) labeled with DiD (peak emission, 665 nm) was recovered.

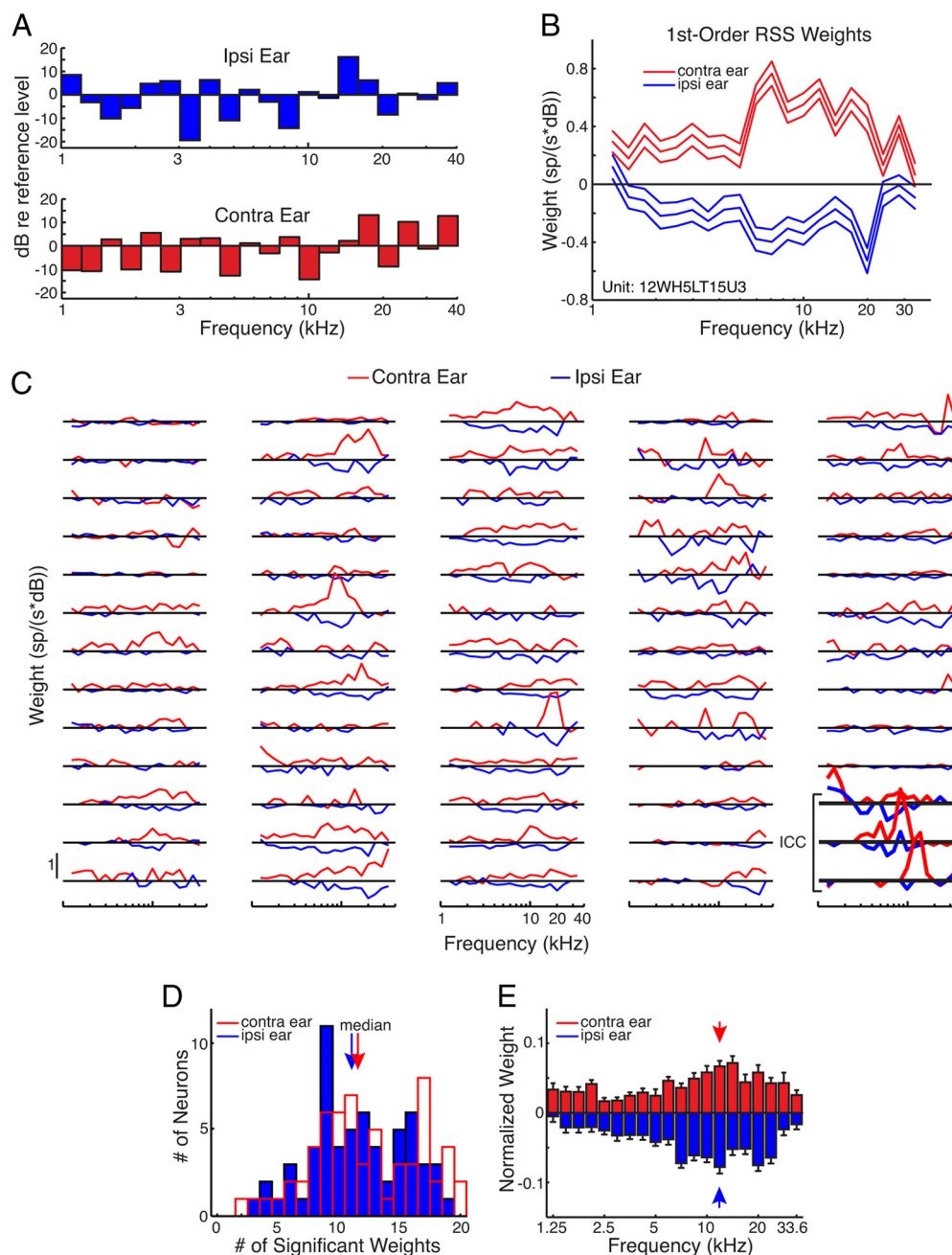


Figure 4. *A*, Example of a binaural RSS stimulus. The height of each bar indicates the spectral level (relative to a reference 20 dB above threshold) in each quarter-octave frequency bin. Note that the levels are independent between frequency bins and the two ears. *B*, An example of a first-order binaural RSS weighting function. The weight in each bin is plotted as a function of frequency. The center line is the mean weight estimate, and the two outer lines are ± 1 SD away. *C*, Binaural RSS weighting functions for 62 BIN neurons and 3 ICC neurons (bottom right, bold). Insignificant weights are set to zero. *D*, Histogram of the number of significant ipsilateral (blue) and contralateral (red) weights in the population of BIN neurons. *E*, RSS weight function for the ipsilateral (blue) and contralateral (red) ear averaged across all units. The weight function in each ear/unit was first normalized by the sum of the weight magnitudes across frequencies. The error bars indicate one SD, and the arrows indicate the median.

ered from a craniotomy located ~ 1 mm anterior and lateral to the one in Figure 1*B*. The second track corresponded to recordings from 10 BIN-like neurons. The location of these tracks is in agreement with BIN-like physiological responses. The locations of the labeled tracks as well as track marks through cortex suggest that recordings were made from roughly the caudal two thirds of the BIN in this animal.

In another marmoset (18W), we made neural recordings from BIN-like sites in four craniotomies using a lateral approach. Figure 1*D* shows a labeled track through the caudal end of the right BIN (AP, 2.15 mm). Figure 1*E* shows a section through the ros-

tral end of the right BIN (AP, 4.05 mm). The track in this section did not label as deeply as the BIN, but did show labeling dorsal and medial to the BIN, suggesting that this track passed through the BIN (Fig. 1*F*). Finally, Figure 1*G* shows an electrolytic lesion made in the caudal end of the left BIN (AP, 2.5 mm) in the same animal.

In total, six of eight tracks in two marmosets were recovered histologically, showing that 79 of 96 neurons were located in the BIN and confirming the association of physiological response types with the BIN, as opposed to the SC or ICC. From these anatomical results, we feel justified in assuming that histologi-

cally unconfirmed recording locations with BIN-like response properties were in the BIN.

General response properties in marmoset BIN

Responses to diotic tones and VS stimuli were measured from 96 single neurons in the BIN. Of these, 39 were sorted from tetrode recordings. Most BIN neurons had spontaneous activity (83% had rates greater than 5 Hz). The mean spontaneous rate was 20 Hz (SD, 14 Hz). Thus it was usually possible to see inhibitory responses to the stimuli.

Diotic tone responses (6 octave range, ~70 dB SPL) measured in BIN neurons were highly variable. Some neurons (89 of 96) responded to tones, while others did not (7 of 96). BIN neurons that did respond to tones often responded over a range of frequencies covering several octaves. Both onset and sustained responses to tones were found. Some responses quickly adapted, while others did not. It was usually difficult to assign a best frequency to a neuron.

Tuning to spatial location was assessed using responses to VS stimuli, computed from the animal's individual HRTFs. In our sample of BIN neurons, four had very weak or no responses to VS stimuli and were not analyzed further. Figure 2 shows examples of the temporal responses to VS stimuli. The dot raster in Figure 2A and the peristimulus time histogram (PSTH) in Figure 2C show an example of sustained responses. This neuron has a short latency (9 ms) and was strongly tuned to the contralateral hemifield and weakly tuned to EL. This response pattern was typical of many neurons in the BIN.

An example of an onset response from a BIN neuron is shown in Figure 2, B and D. This neuron also has a short latency (15 ms). Similar onset responses (though often with shorter latencies) have been reported previously in many BIN neurons (Schnupp and King, 1997). The spatial tuning of this neuron is similar to that of the neuron in Figure 2, A and C.

The sustained versus onset nature of the temporal response was quantified by the onset index (Eq. 2). This index ranges from 1 for a pure onset response to 0 for a purely sustained response, and to negative values for a build-up response. The neuron in Figure 2A had an onset index of 0.07, and the neuron in Figure 2B had an onset index of 0.86. A histogram of onset indices from BIN neurons is shown in Figure 2E (black bars), along with a comparison histogram of ICC neurons from our previous study (Slee and Young, 2011). The values for ICC neurons are scattered around 0 (median 0.05), whereas there are many BIN neurons with nonzero index values consistent with a decrement in the response over time (median, 0.15; different from the ICC value $p \sim 10^{-4}$, rank sum). In fact, the distribution for BIN neurons appears to be bimodal. Figure 2F shows histograms of the average first-spike latencies of responses to the whole set of VS stimuli in both BIN and ICC. These distributions look very similar and are dominated by short latencies. The medians (BIN, 12 ms; ICC, 11 ms) are not significantly different. The similar latencies in these structures are in agreement with previous studies that suggest the BIN receives its predominant input from a very short connection from the ICC (Nodal et al., 2005). The short latencies found in the BIN provide additional evidence against the mistargeting of recordings to the dorsal cortex of the IC, where neurons often have much longer latencies (Lumani and Zhang, 2010).

Spatial tuning

The spike responses to VS stimuli (at an SPL ~20 dB above threshold) were used to construct VSRFs. Figure 3A shows the spike rates of an example BIN neuron (color scale) produced by

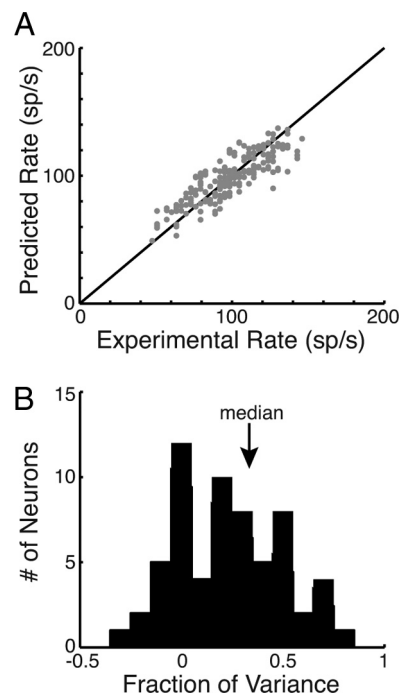


Figure 5. A, RSS model predicted rate (ordinate) versus experimental rate (abscissa) for the neuron in Figure 4B. B, Histogram of the f_v captured by the RSS models for novel RSS stimuli.

VS stimuli corresponding to various positions in space (Slee and Young, 2010). This neuron gives responses to stimuli on the contralateral side with maximal responses to stimuli at -60° AZ and 0° EL. These data are representative of the population of BIN neurons in that the spatial receptive field is located contralaterally and is broad in EL tuning. Figure 3B shows the driven rates (defined in the caption) of a population of 92 neurons versus AZ (averaged over ELs from -30° to 30° ; Fig. 3A, green horizontal lines). The red curve shows the population median. Most neurons had maximal rates in response to stimuli in the contralateral hemifield and had varying degrees of response in the ipsilateral hemifield; of 76 neurons with spontaneous activity, 70% were inhibited by stimuli in the ipsilateral hemifield. Figure 3C shows histograms of the best AZ, defined as the geometrical center by the following equation:

$$\text{Best AZ} = \langle \text{FR}_i / \text{FR}_{\max} \times \text{AZ}_i \rangle_i, \quad (4)$$

where $\langle \text{FR}_i / \text{FR}_{\max} \times \text{AZ}_i \rangle_i$ denotes the average over i ranging from the minimum to maximum AZ eliciting an FR >75% of the maximum relative to the minimum rate (Middlebrooks and Knudsen, 1984). The figure shows data from BIN neurons (red) as well as ICC neurons (blue) measured in our previous study (Slee and Young, 2011). The data in both regions are distributed around a peak at -60° AZ. However, in the BIN we found more neurons tuned at -90° AZ and in the far posterior region of the contralateral hemifield and fewer in the frontal field. The median best AZ of the population of BIN neurons was -71° , which was significantly more lateral than that of the ICC (-58° ; $p = 8.9 \times 10^{-5}$, rank sum test). Both of these values are near the AZ with the greatest ILD in the marmoset HRTF (-60°) (Slee and Young, 2010). The sharpness of azimuthal tuning was measured by the half-width (Fig. 3D). Across the population, we found that the tuning of BIN neurons was significantly broader (median half-width, 164°) than that of ICC neurons (median half-width, 141° ; $p = 1.9 \times 10^{-6}$).

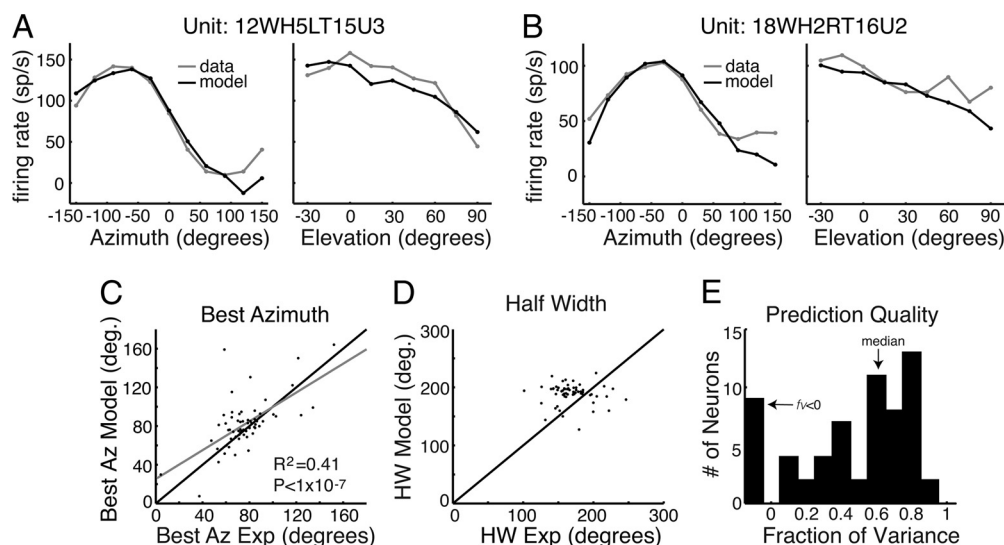


Figure 6. *A*, Firing rate versus AZ (left) and EL (right) for a BIN neuron (gray) and the first-order RSS model (black). *B*, Same as *A* for another BIN neuron recorded in a different marmoset. *C*, Best AZ of the model's predicted responses versus the experimental value. The gray line is a linear regression fit. *D*, Half-width of the model's responses versus the experimental value. *E*, Histogram of the f_v of the VS responses captured by the binaural RSS model.

Figure 3*E* shows the normalized driven rate of each neuron versus the EL of the VS stimulus (averaged over AZs from -120 to -60° ; Fig. 3*A*, vertical dotted lines). The red curve shows the population median. For EL, the range of maximal rates is spread more uniformly across space. In addition, only a few neurons showed suppression with EL. Most neurons are tuned to lower ELs, and the amount of rate modulation is less than for the AZ. Figure 3*F* shows histograms of the best EL in the BIN and ICC, defined as the geometrical center of ELs eliciting $>75\%$ of the maximum relative to the minimum firing rate. These distributions are similar in both regions. Both distributions peak at $\sim 0^\circ$ EL and then decline quickly at higher and lower ELs. For both distributions, the shape may be the result of a preference for a certain SS, but could also reflect the fact that ILDs are larger at lower ELs in the VS stimuli (Slee and Young, 2010, their Fig. 7*B, C*). The median best EL across the population in the BIN was 2° and was not significantly different from that in the ICC (6° ; $p = 0.84$). The median best ELs are likely influenced by the limited range of ELs measured (-30 to 90°). The tuning widths for EL were often so great that the half-width could not be defined.

Characterizing spectral tuning in the BIN with RSS stimuli

As mentioned above, diotic tone responses measured in BIN neurons were highly variable. We found that measuring spectral tuning using binaural RSS stimuli was much more reliable. RSS responses were measured in 62 BIN neurons. Figure 4*A* shows one example of the spectrum of a binaural RSS stimulus used in this study. Each dichotic stimulus consists of a sum of logarithmically spaced tones arranged in quarter-octave bins (16 tones per bin) with levels drawn independently for the two ears from a Gaussian distribution. The red and blue curves are the spectral shapes of one such stimulus in the two ears. As these spectral shapes vary, they provide a range of cues. The difference between the blue and red curves is the frequency-dependent ILD. Because the phase of each spectral component was the same in both ears, the RSS stimuli had a fixed ITD of 0. The firing rates in response to a collection of these stimuli were used to compute the parameters of a spectral weighting model (Eq. 1) relating firing rate to the spectrum of the stimuli. The first-order weights (Eq. 1,

w) measure the frequency-specific gain (spikes $\cdot s^{-1} \cdot dB^{-1}$) of the response produced by energy at various frequencies in each ear. Figure 4*B* shows an example of the first-order weight functions for a BIN neuron. The weights give a measure of the contralateral (red) spectral tuning, which is positive (excitatory), and the ipsilateral (blue) tuning, which is negative (inhibitory). The weights are significantly different from zero over almost the entire frequency range of the RSS stimuli (5 octaves, 1.15–36.7 kHz). They are approximately the inverse of each other, showing that this neuron is mainly sensitive to ILDs. However, the slight differences in the shape of the ipsilateral and contralateral weight functions provide sensitivity to binaural differences in SS cues.

Average weights with magnitudes at least one SD away from zero were considered significant. All of the 62 BIN neurons tested with RSS stimuli had at least one significant excitatory contralateral weight and one significant inhibitory ipsilateral weight. Figure 4*C* plots binaural RSS weighting functions for all BIN neurons in our sample and for three ICC neurons (bold, bottom right). For clarity, the insignificant weights have been set to zero. Most BIN neurons had weight functions similar to the example shown in Figure 4*B*. The weights were broadly tuned, were excitatory in the contralateral ear and inhibitory in the ipsilateral ear, and were asymmetric across ears. In contrast, two of the three ICC neurons (Fig. 4*B*, middle, bottom) are sharply tuned, with large contralateral weights near the best frequency. The lowest-frequency ICC neuron (top) has broader weights and has positive weights in both ears at lower frequencies. The median numbers of significant contralateral and ipsilateral weights in BIN neurons were 11.5 and 11, respectively, which correspond to spectral integration over ~ 3 octaves (Fig. 4*D*). There was no significant tendency for a greater number of weights in the ipsilateral or contralateral ear ($p = 0.35$, binomial test).

We also measured the frequency tuning of the RSS weights. Figure 4*E* shows the normalized weight function averaged across all 62 units. It is clear that both the ipsilateral and contralateral weights are biased toward high frequencies. We quantified this by measuring the median of the weight distributions for each unit and ear. For both the ipsilateral and contralateral weights, there were significantly more neurons with median frequencies above

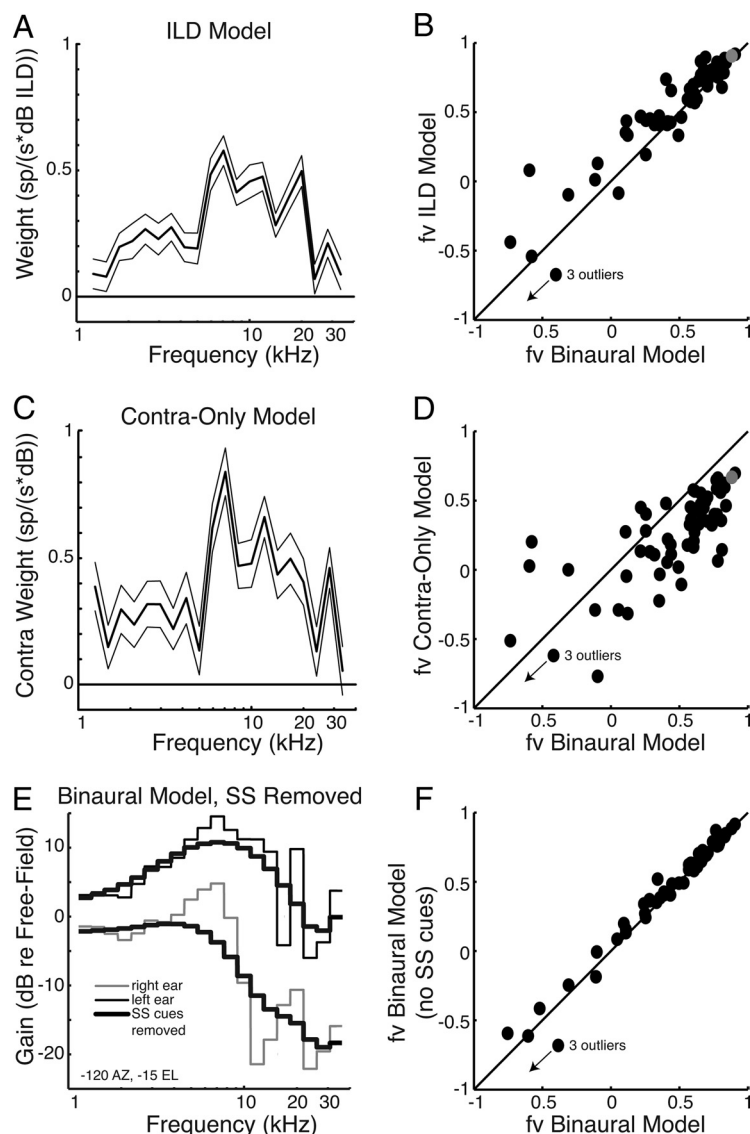


Figure 7. *A*, The ILD-only weighting function for the same neuron as in Figure 4*B*. Here the stimulus is described only by the ILDs, not the sound levels in the two ears independently. Note the different units on the ordinate. *B*, The f_V for VS stimuli captured by the ILD RSS model versus the f_V captured by the binaural RSS model. The solid line has a slope of 1. The model in *A* is indicated by a gray dot. Of 62 points, 48 are above the unity line ($p < 1e-5$, binomial). *C*, The contra-only RSS model for the same neuron. *D*, The f_V for VS stimuli captured by the contra-only RSS model versus the f_V captured by the binaural RSS model. The model in *C* is indicated by a gray dot. Of 62 points, 52 are below the unity line ($p < 1e-7$, binomial). *E*, Examples of HRTF magnitudes of VS stimuli in the left (black) and right (gray) ears presented to the binaural RSS model for one spatial position (-120° AZ, -15° EL). The thick black curves show the same stimuli but with SS cues reduced by smoothing (see Materials and Methods). *F*, The f_V for VS stimuli with SS cues removed captured by binaural RSS models versus the f_V for VS stimuli with SS cues intact captured by the binaural RSS model. Of 62 points, 44 are above the unity line ($p < 1e-3$, binomial).

the midpoint (47 of 62, $p = 2.9e-5$ and 46 of 62, $p = 8.8e-5$, respectively; binomial tests).

The quality of the linear binaural RSS models was measured by prediction tests. As described in Materials and Methods, 90% of the data were used to estimate the weights, and the resulting model was used to predict the responses to the remaining 10% of data. This was done for 1000 random samples of 90% of the model data. The fraction of variance (f_V) of the prediction (Eq. 3) was used to evaluate the quality of the fit. The f_V ranges from 1 for a perfect fit to negative values for poor fits. Figure 5*A* plots the predicted firing rate against the experimental rate responses for the model shown in Figure 4*B*. This model was among the most

accurate in the population as indicated by an f_V value of 0.88. However, the majority of the models were less accurate. The distribution of f_V values for the population of BIN neurons is shown in Figure 5*B*. The quality of the fits by the first-order model varied across neurons, with a median f_V value of 0.23. This degree of accuracy is comparable to first-order weight-function models of spectral processing in the ICC (our unpublished observations). It is also comparable to the prediction performance of spectrotemporal receptive-field models of midbrain neurons in previous studies (Eggermont et al., 1983; Lesica and Grothe, 2008; Versnel et al., 2009; Calabrese et al., 2011).

RSS models predict spatial tuning in the BIN

The responses to VS stimuli were computed from first-order RSS models as a cross-validation test of the models' accuracy. Figure 6, *A* and *B*, shows comparisons of the actual discharge rates to binaural VS stimuli (gray) and the predictions of the RSS models (black) for two neurons. The data are well fit by the model in these cases ($f_V = 0.58$ and 0.54). Note that this is a cross-validation test, in that the models were computed from responses to one kind of stimulus, and the test data are responses to a different stimulus.

In many cases, the RSS model predicted the best AZ, but in other cases it was clearly inaccurate (Fig. 6*C*). However, across the population the predicted and measured best AZs were significantly correlated ($R^2 = 0.41$; $p = 2.3e-8$). These differences are not surprising given that the AZ tuning is broad, which makes the measurement of the best AZ noisy. In most neurons, the half-width of the model tuning curve for the AZ was slightly greater than that measured in the neuron. The model half-width was significantly greater across the population (Fig. 6*D*; $p = 6.1e-7$, binomial test).

Figure 6*E* shows the distribution of f_V values for first-order model predictions of

spatial tuning in 62 BIN neurons for which responses to both RSS and VS stimuli were measured. The distribution (median, 0.58) shows that in most neurons the spatial tuning is accurately predicted from a linear binaural weighting of the spectrum in both ears. Thus, although the predicted best AZ and half-width were inaccurate in some cases, the models were still able to capture a large fraction of the variation in firing rate with spatial position. Note that RSS models predict responses to VS stimuli better than they do responses to RSS stimuli (median, 0.23; $p = 5.6e-5$, rank sum test). Presumably this reflects the fact that VS responses are quite smooth, compared to RSS responses, and therefore easier to predict.

Interaural level difference is the dominant cue for spatial tuning in the BIN

The example in the previous section (Fig. 4*B*) had similar ipsilateral and contralateral ear weight functions, which is expected if the neurons were only sensitive to ILD. To test the importance of monaural cues as opposed to ILD cues, first-order RSS models were computed using only the ILD as the stimulus level; the stimulus vector \mathbf{s} in Equation 1 contained only the differences between contralateral and ipsilateral levels. Thus, the information about monaural levels and the monaural SS cues was lost. Figure 7*A* shows the ILD RSS weighting function for the neuron shown in Figure 4*B*. The RSS weights are all positive and significant at all frequencies. This indicates that the neuron responds to positive ILDs (contralateral greater than ipsilateral) at all frequencies. Responses to VS stimuli were computed using these ILD-only weight-function models and the ILDs computed from HRTFs. Figure 7*B* shows the f_v values for prediction of VS data using the ILD model versus the full binaural model. Surprisingly, most points lie above the line of equality (48 of 62), indicating a better model fit for the ILD model than the binaural model ($p = 8.7 \times 10^{-6}$, binomial test), even though all the information in the ILD model is also available to the binaural model (see Discussion).

A contralateral-only RSS model was constructed by only considering the contralateral stimuli when calculating the RSS weights (Fig. 7*C*). The contralateral-only weight function is positive at all frequencies and has larger weights than the binaural weight function shown in Figure 4*B*. This model is excited by stimuli at all frequencies. Responses to the VS stimuli were computed using this model and the contralateral HRTFs. Figure 7*D* shows that the f_v values measured with the contralateral-only model are, on average, below those computed with the full binaural model (52 of 62; $p = 2.9 \times 10^{-8}$, binomial test). Together, these two analyses show that the information in the ipsilateral ear is used by BIN neurons.

A final manipulation was made by removing the SS cues in the VS stimuli used for prediction tests of the full binaural model developed with unmodified stimuli. Figure 7*E* shows an example of HRTFs in the left (thin, black) and right (gray) ears at one position (-120° AZ, -15° EL). Peaks and spectral notches are apparent in both functions. The functions were smoothed with a triangular filter to reduce the amplitude of SS cues (Fig. 7*E*, thick black curves; see Materials and Methods). The binaural RSS model was used to compute responses to the cues-removed VS stimuli, and the results were compared with the real neural responses to the VS stimuli with SS cues present (Fig. 7*F*). Most points lie on or just above the line of equality, indicating a slightly improved fit for model responses to VS stimuli without SS cues (44 of 62; $p = 6.5 \times 10^{-4}$, binomial test). The results in Figure 7, *B* and *F*, are unexpected because they suggest that SS cues act as noise rather than signal for models of BIN neurons (see Discussion).

ITD sensitivity is weak in the BIN

None of the RSS models above contained information about ITD, as phase is not incorporated in the RSS stimuli (see Materials and Methods). However, the VS stimuli presented to BIN neurons did contain natural ITD cues, which are encoded in the phases of the HRTFs. The fact that the RSS models accurately predicted VSRFs in most neurons (Fig. 6*E*) suggests that ITD is not an important cue for the spatial tuning of BIN neurons. This hypothesis was tested directly by measuring ITD tuning curves with broadband noise stimuli with static ITDs.

Figure 8*A* plots ITD tuning curves in three representative BIN neurons. Both Neurons 1 (black) and 2 (gray) had significant ITD

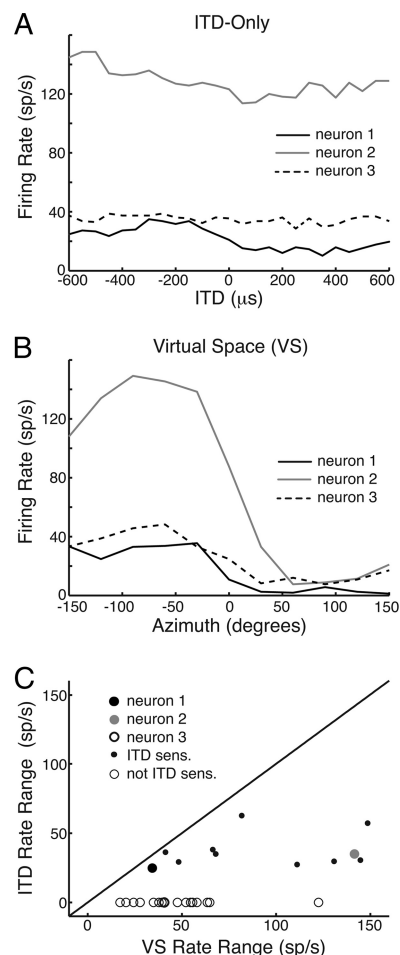


Figure 8. *A*, Firing rate versus the ITD of a noise carrier for three representative BIN neurons. The neuron represented by the dotted line was insensitive to ITD. *B*, Firing rate versus AZ of the virtual space stimuli for the same neurons in *A*. *C*, Firing rate range (maximum–minimum) for ITD stimuli versus VS stimuli in 28 BIN neurons.

tuning, whereas Neuron 3 (dotted) did not. In 28 BIN neurons tested, only 11 neurons showed significant modulation of firing rate by ITD ($p < 0.05$, ANOVA). Figure 8*B* shows AZ tuning curves for VS stimuli (with all three localization cues) in the same three neurons. Neuron 1 had similar tuning in both conditions, whereas Neurons 2 and 3 had much greater rate modulation with the other two cues (ILD and SS) present. Figure 8*C* plots the firing rate range for ITD versus that for VS cues in all 28 neurons tested (the ITD rate range is zero in 17 neurons). In general, the range of rate modulation due to ITD was usually much less than that caused by the VS stimuli containing all sound localization cues; the median [rate range_{ITD}/rate range_{VS stimuli}] for 11 ITD-sensitive neurons was 0.51 (range, 0.21–0.88). The fraction of variance captured by the binaural RSS model was not significantly different for ITD-sensitive neurons versus ITD-insensitive neurons ($p = 0.32$). These results suggest that ITD is a weak cue for the spatial tuning of BIN neurons. Together, the results of the RSS model manipulations are consistent with the hypothesis that linear processing of ILD underlies the spatial tuning of BIN neurons in the marmoset.

Nonlinear processing contributes little to the spatial tuning of BIN neurons

We tested for nonlinear processing in BIN neurons by comparing the predictive power of Equation 1 with and without the second-

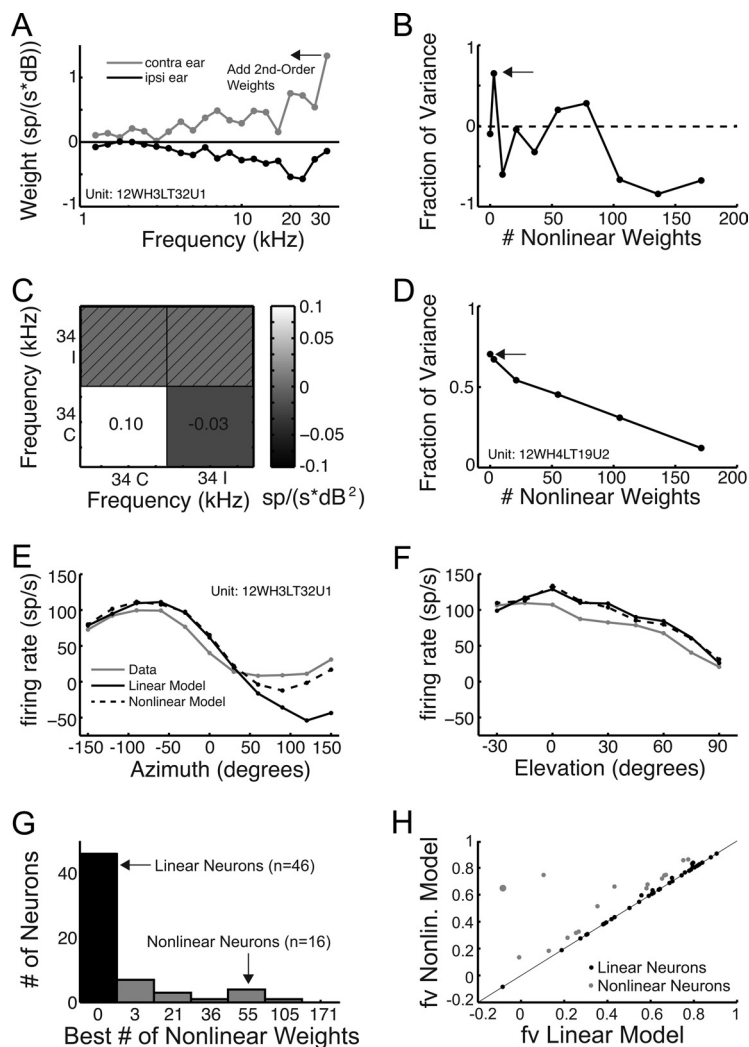


Figure 9. *A*, An example of a weighting function for an ILD RSS model illustrating how nonlinear weights are added around the peak linear weight. *B*, Fraction of variance versus the number of nonlinear weights included for the model neuron in *A*. The arrow indicates the best model. *C*, The nonlinear two-by-two weight matrix for the neuron in *A*. The grayscale indicates the value of each second-order weight. The hatched weights are insignificant or redundant. *D*, Same as *B* for a purely linear neuron. *E*, *F*, Firing rate versus AZ (left) and EL (right) for the neuron (gray), the linear RSS model from *A* (black), and the best nonlinear RSS model from *C* (dotted lines). *G*, Histogram of the best number of nonlinear weights for linear (black) and nonlinear (gray) neurons. *H*, The *fv* for VS stimuli captured by the nonlinear binaural RSS model versus the *fv* captured by the linear binaural RSS model. The model in *E* and *F* is indicated by the large gray dot.

order terms. The approach was to systematically add second-order interactions between frequencies and across ears, beginning with the frequencies near the peak linear weight (Fig. 9*A*; see Materials and Methods) and adding weights at successive frequencies. Second-order weights were kept in the model if they improved the ability of the model to predict responses to VS stimuli, which were not used in weight estimation. The reason for this cross-validation method is that the greater the number of weights, the better the model will be at fitting RSS responses, simply because it has a greater number of free parameters. However, as the number of weights increases, the model becomes overfit and begins fitting noise in the data. At this point, it begins to perform worse at predicting VS stimuli. With this method, we found the number of nonlinear weights capturing the greatest *fv* for predicting responses to VS stimuli.

Figure 9*A* shows linear weights for an example BIN neuron. The largest weight occurs at 33 kHz, so nonlinear weights were added relative to this weight (just below the maximum weight in

this case). Figure 9*B* shows that the *fv* value is largest for three nonlinear weights and thereafter oscillates and decreases. The three nonlinear weights for the best model are shown in the two-by-two matrix in Figure 9*C*. The second-order weight for 34 kHz tones in the ipsilateral ear (Fig. 9*C*, 34 I, 34 I, corresponding to the square of the energy at 34 kHz in the ipsilateral ear) is not significant. There are two significant weights for the frequency combinations (34 C, 34 C) and (34 C, 34 I). Figure 9, *E* and *F*, shows model predictions of the VSRF for the AZ and EL in this neuron for linear and nonlinear models. The best nonlinear model shows an improvement mostly by correcting the erroneously negative firing rates of the first-order model for positive AZ. This behavior occurred in several neurons.

In most BIN neurons, adding second-order weights only decreased the *fv* value, suggesting purely linear processing. An example of such behavior is shown in Figure 9*D*. The criterion we set for significant nonlinearity was an increase in *fv* value by at least 0.05 and a final *fv* value of at least 0.1. The histogram in Figure 9*G* shows the distribution of the best number of nonlinear weights for predicting VSRFs across the population. The majority of neurons were purely linear (46; Fig. 9*G*, black bar), although some neurons showed significant nonlinearities (16; gray bars).

The degree of nonlinear processing across the population is shown in Figure 9*H* by plotting the *fv* of the best nonlinear model versus the *fv* of the linear model. The linear neurons (black dots) lie near or on the line of unity, while the nonlinear neurons (gray dots) lie above it. The neuron from Figure 9*A* is represented by a large dot. It can be seen that although the nonlinearities led to improved prediction in some neurons, these improvements

were usually small (six neurons increased the *fv* value by at least 0.1, and three by at least 0.2). These results suggest that the spectral processing underlying the spatial tuning of BIN neurons is mostly first order.

The balance of ipsilateral and contralateral weights predicts the best azimuth

We created a simple RSS model to demonstrate how ILD-only processing can produce responses like those of BIN neurons [weight balance (WB) model]. As with the empirically derived RSS models, the WB model firing rate is computed from the average of linearly weighted sound levels in the ipsilateral and contralateral ears. The weights for the WB model are shown by lines in Figure 10*A*. The contralateral weight is fixed at 0.2, but the ipsilateral weight varies from -0.3 to 0.2 for different versions of the model. The ipsilateral and contralateral inputs are the average sound levels in quarter-octave frequency bins (taken from the HRTFs magnitudes). The output of the WB model is

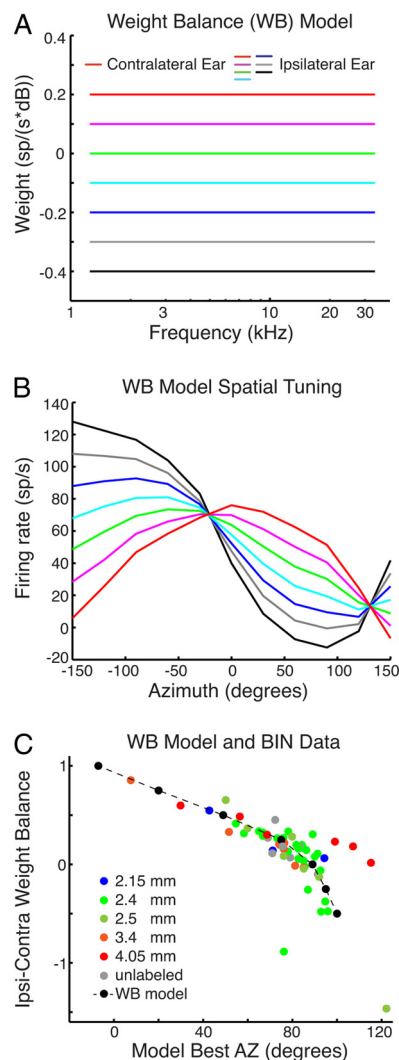


Figure 10. *A*, Contralateral weights are fixed, whereas the ipsilateral weights vary according to color for the different models. *B*, Firing rate versus AZ for the different WB models in *A* driven with VS stimuli. *C*, Ipsilateral–contralateral weight balance (W_{ic} , defined in Materials and Methods) versus best model AZ. Broken lines and black dots are from the WB models in *B*. Colored points are for BIN data measured in different craniotomies, identified in the legend. Positions range from posterior (blue) to anterior (red). Four outliers were removed from *C* and the correlation analysis: two neurons had small noisy weights, and two were tuned far posterior locations.

shown for various values of ipsilateral weight in Figure 10*B*. For approximately equal positive weights in both ears (Fig. 10*B*, red; an excitatory–excitatory model), the model gives a maximum response to azimuths near 0, where the sound levels in the two ears are equal and their sum is maximum. For large negative ipsilateral weights (Fig. 10*B*, blue; an excitatory–inhibitory model), the model responds at negative azimuths, where the stimulus intensity in the contralateral ear is greater than that in the ipsilateral ear, maximizing excitation relative to inhibition.

We compared the best azimuths of the WB models and of BIN neurons to their weight balance, W_{ic} , defined as follows:

$$W_{ic} = \sum(W_i + W_c)/2\sum W_c \quad (5)$$

where W_i and W_c are the contralateral and ipsilateral weights, respectively. These are constant across frequencies for the models in Figure 10*B*, but are the RSS weights for neurons. W_{ic} is negative when the ipsilateral inhibition dominates, zero when binaural

inputs are balanced, and positive when contralateral excitation dominates or the ipsilateral weights are positive (excitatory). Consistent with the weight functions in Figure 4*C*, the values of W_{ic} are mostly positive.

The dotted line in Figure 10*C* shows the relationship between W_{ic} for the WB model and the best AZ (from Fig. 10*B* for the WB model; from the data underlying Figs. 4*C*, 6*C* for the neurons). The WB model shows an approximately linear relationship between W_{ic} and the best AZ, except for large values of AZ. The colored data points are the values of W_{ic} and best AZ for the neurons encountered in recordings through six craniotomies, identified in the legend. W_{ic} has a significant correlation with the best AZ predicted for both the model and the data ($R^2 = 0.48$; $p < 1e-8$).

In five of the craniotomies, the electrodes were localized to the BIN histologically. The anatomical locations (indicated in the legend) do not provide clear evidence of a topographical map of the best AZ. However, the majority of neurons from the posterior craniotomies in both animals (green and blue symbols) had small W_{ic} values and are tuned lateral to 60° AZ. The red symbols from anterior craniotomies are more scattered but include two points from frontally tuned units with greater W_{ic} values, as predicted by the model.

Discussion

Summary of results

Our data show that BIN neurons are broadly tuned in frequency (median tuning width, ~3 octaves) with a bias toward high frequencies (Fig. 4). The temporal responses to spatial stimuli vary from onset to sustained, but onset responses are more prevalent in the BIN than in the ICC (Fig. 2*E*). These results are broadly consistent with a previous study in the BIN (Schnupp and King, 1997). However, the previous study found many fewer sustained responses, perhaps due to the use of anesthesia.

The azimuthal tuning in the marmoset BIN is broad and is contralaterally dominated. In contrast to the ICC, we found that the tuning is broader, with fewer neurons tuned to the frontal field and more tuned to the lateral field (Fig. 3). The EL tuning was also quite broad and was dominated by midline responses. These results are all consistent with a previous study in the BIN of anesthetized ferrets (Schnupp and King, 1997). Spatial tuning, spectral tuning, and spike latencies suggest that the BIN receives convergent input from the ICC, which is consistent with anatomical results (King et al., 1998).

Mechanisms of spatial tuning in the BIN

We used a model-based approach to investigate which cues are important for spatial tuning in the BIN. RSS models were able to accurately predict virtual space tuning curves in the majority of BIN neurons (Fig. 6*A*, *B*, *E*). In ~25% of neurons, model predictions were modest or poor ($f_v < 0.25$). This could reflect processes not captured by the RSS model such as temporal dynamics and higher-order nonlinearities. As a whole, RSS models predicted responses to VS stimuli better than responses to RSS stimuli (compare Figs. 5*B*, 6*E*). This result was obtained previously for auditory nerve fibers responding to similar stimuli (Young and Calhoun 2005). It probably reflects the highly structured spectral shapes of VS stimuli (with strongly correlated values at adjacent frequencies) compared to RSS stimuli. The model predicted one parameter of spatial tuning, the best azimuth, much better than chance, but did not do well with tuning half-width, which was systematically smaller than the predicted value.

Surprisingly, RSS models with only ILD cues outperformed the full models containing both ILD and monaural cues (Fig. 7).

This result may have come about because of the nature of the binaural data. Given finite data, the full binaural RSS model is optimized to fit parameters important for predicting spectral responses along with parameters that are important for VS responses. By removing spectral shape parameters that are essentially noise for spatial responses, the ILD-only analysis restricts the model to aspects of the signal that are actually used by the neurons. This argument may also explain why reducing SS cues does not degrade the performance of the model (Fig. 7F), i.e., the neurons are making little use of SS cues in their responses. The weak EL tuning found in these neurons is consistent with this hypothesis.

The contra-only model clearly underperforms the full model, which is consistent with numerous studies in the IC showing that most neurons are binaural (Davis et al., 1999; Delgutte et al., 1999).

ITD sensitivity is weak in the BIN, as was also found in the SC (Hirsch et al., 1985; Campbell et al., 2006). We found that only approximately one-third of the neurons tested were ITD sensitive, and in those neurons usually less than half of the spatial rate modulation was due to ITD. Weak ITD sensitivity could result from a simple linear pooling across (logarithmic) frequency. In the marmoset's ~6 octave hearing range, <2 octaves are dedicated to frequencies where fine-structure ITDs should be encoded. In addition, the BIN is biased toward high frequencies (Fig. 4), as is the SC (Hirsch et al., 1985).

Some BIN neurons showed nonlinear spectral processing that significantly contributed to spatial tuning. However, most BIN neurons linearly combined ILD across frequencies, consistent with previous studies in auditory cortex (Schnupp et al., 2001). Although in the barn owl ITD and ILD are nonlinearly combined within each frequency channel (Peña and Konishi, 2001), it has been suggested that this information is linearly combined across channels (Fischer et al., 2009). As suggested previously (McAlpine, 2005), these differences in within-channel processing likely arise because the barn owl has adapted a high-frequency ITD-sensitive pathway and sharp spatial tuning for nocturnal hunting.

Topographical map of auditory space

Previous work provides strong evidence for a topographic map of the best AZ in the BIN (Schnupp and King, 1997), although the map is less organized than in the SC. Our finding that the majority of neurons in the caudal region of the BIN were tuned to lateral and posterior fields is consistent with those results (Fig. 10C). In rostral craniotomies, we found some neurons tuned to the anterior field, but as a population they were more scattered than shown previously. There may be several reasons for these differences. In the present study, we did not control eye position (although recordings were made in the dark or at dim light levels), which is known to affect the auditory spatial map in the SC (Jay and Sparks, 1987). The previous study used anesthetized animals, and the eye position was well controlled. However, the difference might also arise from the difficulty of detecting a topographical map with the indirect histological localization of recording necessitated by the chronic preparation.

Although we did not obtain clear evidence of a topographical map of the AZ, our data-based modeling suggests that the best AZ is predicted well by the balance of ipsilateral and contralateral input. A similar result was reported in the SC of the cat, where the topographic map of the AZ is thought to arise from a parallel map of ILD sensitivity caused by changes from inhibitory to facilitatory binaural interactions (Wise and Irvine, 1985). In addition, a similar mechanism has been shown in the auditory cortex of the

bat (Razak, 2011), which suggests that a similar processing strategy may be repeated at multiple stages of the auditory system. The similar mechanism underlying AZ processing in both the BIN and SC provides additional evidence for its role as an intermediate processing stage between the ICC and the SC.

Implications of ILD processing across frequencies in the BIN

The dominance of ILDs and bias toward a broad range of high frequencies suggest a loss of sound information in the BIN. Specifically, decoding the spectrum of the stimulus and its frequency-dependent ITD is likely compromised relative to the ICC. While this initially seems like a poor strategy, it may contain benefits for some forms of spatial processing. A potential benefit is invariance of spatial tuning in reverberation. At a given location, reverberation distorts the spectral magnitude randomly within a frequency band, but these effects will tend to average out across bands (Rakerd and Hartmann, 1985). Reverberation is also known to distort sound localization cues in different ways (Devore and Delgutte, 2010; Kuwada et al., 2012), e.g., making ITD and ILD incoherent. In this case, dominance of the single most robust cue might be beneficial. This idea is consistent with a previous study showing greater sound localization accuracy in reverberation for high-frequency noise compared to low-frequency noise (Ihfeldt and Shinn-Cunningham, 2011). However, future studies are needed to test these possibilities directly.

References

- Aitkin L, Jones R (1992) Azimuthal processing in the posterior auditory thalamus of cats. *Neurosci Lett* 142:81–84. [CrossRef Medline](#)
- Bregman AS (1990) Auditory scene analysis. Cambridge, MA: MIT.
- Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SM (2011) A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS ONE* 6:e16104. [CrossRef Medline](#)
- Campbell RA, Doubell TP, Nodal FR, Schnupp JW, King AJ (2006) Inter-aural timing cues do not contribute to the map of space in the ferret superior colliculus: a virtual acoustic space study. *J Neurophysiol* 95:242–254. [CrossRef Medline](#)
- Chase SM, Young ED (2005) Limited segregation of different types of sound localization information among classes of units in the inferior colliculus. *J Neurosci* 25:7575–7585. [CrossRef Medline](#)
- Darwin CJ (2008) Listening to speech in the presence of other sounds. *Philos Trans R Soc Lond B Biol Sci* 363:1011–1021. [CrossRef Medline](#)
- Davis KA, Ramachandran R, May BJ (1999) Single-unit responses in the inferior colliculus of decerebrate cats. II. Sensitivity to interaural level differences. *J Neurophysiol* 82:164–175. [Medline](#)
- Delgutte B, Joris PX, Litovsky RY, Yin TC (1999) Receptive fields and binaural interactions for virtual-space stimuli in the cat inferior colliculus. *J Neurophysiol* 81:2833–2851. [Medline](#)
- Devore S, Delgutte B (2010) Effects of reverberation on the directional sensitivity of auditory neurons across the tonotopic axis: influences of interaural time and level differences. *J Neurosci* 30:7826–7837. [CrossRef Medline](#)
- DiCarlo JJ, Lane JW, Hsiao SS, Johnson KO (1996) Marking microelectrode penetrations with fluorescent dyes. *J Neurosci Methods* 64:75–81. [CrossRef Medline](#)
- Eggermont JJ, Aertsen AMHJ, Johannesma PIM (1983) Prediction of the responses of auditory neurons in the midbrain of the grass frog based on the spectro-temporal receptive field. *Hearing Res* 10:191–202. [CrossRef Medline](#)
- Fischer BJ, Anderson CH, Peña JL (2009) Multiplicative auditory spatial receptive fields created by a hierarchy of population codes. *PLoS ONE* 4:e8015. [CrossRef Medline](#)
- Hirsch JA, Chan JC, Yin TC (1985) Responses of neurons in the cat superior colliculus to acoustic stimuli. 1. Monaural and binaural response properties. *J Neurophysiol* 53:726–745. [Medline](#)
- Huang AY, May BJ (1996) Spectral cues for sound localization in cats: effects of frequency domain on minimum audible angles in the median and horizontal planes. *J Acoust Soc Am* 100:2341–2348. [CrossRef Medline](#)
- Ihfeldt A, Shinn-Cunningham BG (2011) Effect of source spectrum on

- sound localization in an everyday reverberant room. *J Acoust Soc Am* 130:324–333. [CrossRef Medline](#)
- Jay MF, Sparks DL (1987) Sensorimotor integration in the primate superior colliculus. II. Coordinates of auditory signals. *J Neurophysiol* 57:35–55. [Medline](#)
- Jiang ZD, King AJ, Moore DR (1993) Topographic projection from the brachium of the inferior colliculus to the space mapped region of the superior colliculus in the ferret. *Br J Audiol* 27:344–345.
- King AJ, Jiang ZD, Moore DR (1998) Auditory brainstem projections to the ferret superior colliculus: anatomical contribution to the neural coding of sound azimuth. *J Comp Neurol* 390:342–365. [CrossRef Medline](#)
- Knudsen EI (1982) Auditory and visual maps of space in the optic tectum of the owl. *J Neurosci* 2:1177–1194. [Medline](#)
- Knudsen EI (1987) Neural derivation of sound source location in the barn owl. An example of a computational map. *Ann N Y Acad Sci* 510:33–38. [CrossRef](#)
- Knudsen EI, Konishi M (1978) A neural map of auditory space in the owl. *Science* 200:795–797. [CrossRef Medline](#)
- Kuwada S, Bishop B, Kim DO (2012) Approaches to the study of neural coding of sound source location and sound envelope in real environments. *Front Neural Circuits* 6:1–12.
- Lesica NA, Grothe B (2008) Dynamic spectrotemporal feature selectivity in the auditory midbrain. *J Neurosci* 28:5412–5421. [CrossRef Medline](#)
- Lu T, Liang L, Wang X (2001) Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. *J Neurophysiol* 85:2364–2380. [Medline](#)
- Lumani A, Zhang H (2010) Responses of neurons in the rat's dorsal cortex of the inferior colliculus to monaural tone bursts. *Brain Res* 1351:115–129. [CrossRef Medline](#)
- McAlpine D (2005) Creating a sense of auditory space. *J Physiol* 566:21–28. [CrossRef Medline](#)
- Middlebrooks JC, Knudsen EI (1984) A neural code for auditory space in the cat's superior colliculus. *J Neurosci* 4:2621–2634. [Medline](#)
- Mills AW (1958) On the minimum audible angle. *J Acoust Soc Am* 30:237–246. [CrossRef](#)
- Moiseff A (1989) Binaural disparity cues available to the barn owl for sound localization. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 164:629–636. [CrossRef](#)
- Nelson PC, Smith ZM, Young ED (2009) Wide-dynamic-range forward suppression in marmoset inferior colliculus neurons is generated centrally and accounts for perceptual masking. *J Neurosci* 29:2553–2562. [CrossRef Medline](#)
- Nodal FR, Doubell TP, Jiang ZD, Thompson ID, King AJ (2005) Development of the projection from the nucleus of the brachium of the inferior colliculus to the superior colliculus in the ferret. *J Comp Neurol* 485:202–217. [CrossRef Medline](#)
- Oliver DL, Beckius GE, Bishop DC, Kuwada S (1997) Simultaneous anterograde labeling of axonal layers from lateral superior olive and dorsal cochlear nucleus in the inferior colliculus of cat. *J Comp Neurol* 382:215–229. [CrossRef Medline](#)
- Palmer AR, King AJ (1982) The representation of auditory space in the mammalian superior colliculus. *Nature* 299:248–249. [CrossRef Medline](#)
- Peña JL, Konishi M (2001) Auditory spatial receptive fields created by multiplication. *Science* 292:249–252. [CrossRef Medline](#)
- Rakerd B, Hartmann WM (1985) Localization of sound in rooms, II: The effects of a single reflecting surface. *J Acoust Soc Am* 78:524–533. [CrossRef Medline](#)
- Razak KA (2011) Systematic representation of sound locations in the primary auditory cortex. *J Neurosci* 31:13848–13859. [CrossRef Medline](#)
- Roffler SK, Butler RA (1968a) Localization of tonal stimuli in the vertical plane. *J Acoust Soc Am* 43:1260–1266. [CrossRef Medline](#)
- Roffler SK, Butler RA (1968b) Factors that influence the localization of sound in the vertical plane. *J Acoust Soc Am* 43:1255–1259. [CrossRef Medline](#)
- Schnupp JW, King AJ (1997) Coding for auditory space in the nucleus of the brachium of the inferior colliculus in the ferret. *J Neurophysiol* 78:2717–2731. [Medline](#)
- Schnupp JW, Msrac-Flogel TD, King AJ (2001) Linear processing of spatial cues in primary auditory cortex. *Nature* 414:200–204. [CrossRef Medline](#)
- Schumacher JW, Schneider DM, Woolley SM (2011) Anesthetic state modulates excitability but not spectral tuning or neural discrimination in single auditory midbrain neurons. *J Neurophysiol* 106:500–514. [CrossRef Medline](#)
- Schwarz DM, Zilany MS, Skevington M, Huang NJ, Flynn BC, Carney LH (2012) Semi-supervised spike sorting using pattern matching and a scaled Mahalanobis distance metric. *J Neurosci Methods* 206:120–131. [CrossRef Medline](#)
- Shinn-Cunningham BG (2008) Object-based auditory and visual attention. *Trends Cogn Sci* 12:182–186. [CrossRef Medline](#)
- Slee SJ, Young ED (2010) Sound localization cues in the marmoset monkey. *Hear Res* 260:96–108. [CrossRef Medline](#)
- Slee SJ, Young ED (2011) Information conveyed by inferior colliculus neurons about stimuli with aligned and misaligned sound localization cues. *J of Neurophysiology* 106:974–985. [CrossRef](#)
- Strutt JW (1904) On the acoustic shadow of a sphere. *Philos Trans R Soc Lond Ser A* 203:87–89. [CrossRef](#)
- Strutt JW (1907) On our perception of sound direction. *Philos Mag* 13:214–232.
- Versnel H, Zwiers MP, van Opstal AJ (2009) Spectrotemporal response properties of inferior colliculus neurons in alert monkey. *J Neurosci* 29:9725–9739. [CrossRef Medline](#)
- Wightman FL, Kistler DJ (1989) Headphone simulation of free-field listening. I: Stimulus synthesis. *J Acoust Soc Am* 85:858–867. [CrossRef Medline](#)
- Wise LZ, Irvine DR (1985) Topographic organization of interaural intensity difference sensitivity in deep layers of cat superior colliculus - implications for auditory spatial representation. *J Neurophysiol* 54:185–211. [Medline](#)
- Yin TC, Kuwada S, Sujaku Y (1984) Interaural time sensitivity of high-frequency neurons in the inferior colliculus. *J Acoust Soc Am* 76:1401–1410. [CrossRef Medline](#)
- Young ED, Calhoun BM (2005) Nonlinear modeling of auditory-nerve rate responses to wideband stimuli. *J Neurophysiol* 94:4441–4454. [CrossRef Medline](#)
- Zhou B, Green DM, Middlebrooks JC (1992) Characterization of external ear impulse responses using Golay codes. *J Acoust Soc Am* 92:1169–1171. [CrossRef Medline](#)