Information conveyed by inferior colliculus neurons about stimuli with aligned and misaligned sound localization cues

Sean J. Slee and Eric D. Young
Biomedical Engineering Department, Johns Hopkins University, Baltimore, Maryland

Submitted 27 April 2011; accepted in final form 27 May 2011

Slee SJ, Young ED. Information conveyed by inferior colliculus neurons about stimuli with aligned and misaligned sound localization cues. J Neurophysiol 106: 974–985, 2011. First published June 8, 2011; doi:10.1152/jn.00384.2011.—Previous studies have demonstrated that single neurons in the central nucleus of the inferior colliculus (ICC) are sensitive to multiple sound localization cues. We investigated the hypothesis that ICC neurons are specialized to encode multiple sound localization cues that are aligned in space (as would naturally occur from a single broadband sound source). Sound localization cues including interaural time differences (ITDs), interaural level differences (ILDs), and spectral shapes (SSs) were measured in a marmoset monkey. Virtual space methods were used to generate stimuli with aligned and misaligned combinations of cues while recording in the ICC of the same monkey. Mutual information (MI) between spike rates and stimuli for aligned versus misaligned cues were compared. Neurons with best frequencies (BFs) less than ~11 kHz mostly encoded information about a single sound localization cue, ITD or ILD depending on frequency, consistent with the dominance of ear acoustics by either ITD or ILD at those frequencies. Most neurons with BFs >11 kHz encoded information about multiple sound localization cues, usually ILD and SS, and were sensitive to their alignment. In some neurons MI between stimuli and spike responses was greater for aligned cues, while in others it was greater for misaligned cues. If SS cues were shifted to lower frequencies in the virtual space stimuli, a similar result was found for neurons with BFs <11 kHz, showing that the cue interaction reflects the spectra of the stimuli and not a specialization for representing SS cues. In general the results show that ICC neurons are sensitive to multiple localization cues if they are simultaneously present in the frequency response area of the neuron. However, the representation is diffuse in that there is not a specialization in the ICC for encoding aligned sound localization cues.

head-related transfer function; cue integration; interaural level difference

SOUND LOCALIZATION is a computationally demanding task for the nervous system because the cochlea maps sound frequency along the sensory sheet (the basilar membrane) rather than location in space. Therefore, the direction of a sound must be computed in the brain from cues produced by the interaction of sound waves with the head and external ears. The most important localization cues are interaural time difference (ITD), interaural level difference (ILD), and pinna-based spectral shape (SS) cues (Blauert 1997; Middlebrooks and Green 1991), which are collectively measured by the head-related transfer function (HRTF; Musicant et al. 1990; Shaw 1982). The relevance of the HRTF is demonstrated by the fact that sounds filtered with HRTFs and presented over headphones, in a so-called virtual space (VS) experiment, produce an accurate perception of an externally localized source (Wightman and Kistler 1989b).

Previous research has shown that the sound localization cues are computed and analyzed largely separately in the brain stem: ITD in the medial superior olive (Goldberg and Brown 1969; Yin and Chan 1990); ILD in the lateral superior olive (Boudreau and Tsuchitani 1968; Guinan et al. 1972a, 1972b; Tollin and Yin 2002); and SS in the dorsal cochlear nucleus (Imig et al. 2000; Reiss and Young 2005; Young et al. 1992). The axons from these nuclei project to the central nucleus (ICC) of the inferior colliculus (IC) in partially overlapping patterns (Bruno-Bechtold et al. 1981; Lofts et al. 2010; Oliver et al. 1997), making the ICC an important site of cue convergence in the auditory pathway.

Recordings from the ICC in the mammal show that neurons generally respond to multiple sound localization cues (Chase and Young 2005; Delgutte et al. 1999; Devore and Delgutte 2010; Imig et al. 2000; Ramachandran and May 2002). This raises the questions of how the auditory system combines different location cues, and particularly whether some combinations of cues are more informative about sound location than others. A simple hypothesis is that ICC neurons provide more information about sound location with combinations of cues that would arise naturally at the eardrums from a real sound source. This could occur if the projections of inputs to the ICC containing information about different cues were aligned so that responses to corresponding cues converged in a facilitatory way. For example, some experience-based plasticity could strengthen responses to commonly heard combinations of cues. In support of this hypothesis, both human psychophysical studies (Middlebrooks 1999) and animal physiological studies (Sterbing et al. 2003) have shown that spatial coding is more accurate using individualized HRTFs.

The present study tested the hypothesis that neurons in the ICC are specialized to encode natural combinations of sound localization cues. Individual HRTFs were measured in marmoset monkeys and used to construct stimuli with natural and unnatural combinations of sound localization cues. The stimuli were presented to the same marmosets while recording from single neurons in the ICC. The results indicate that ILD cues are the most potent for modulating the firing rates of neurons with high best frequencies (BFs). When measuring the mutual information (MI) between the responses and the source location, some neurons have a preference (larger MI) for natural combinations of cues, but others prefer unnatural ones. Overall the results suggest that ICC neurons provide a mixed representation of sound localization cues and other features of...
stimuli, such as SSs, without a preference for stimulus combinations that occur naturally.

**MATERIALS AND METHODS**

**Animal preparation and recording.** 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 mo prior to surgery, the marmoset was adapted to sit quietly in the restraint device for the 3- to 4-h periods used for recording. Sterile surgery was performed under isofluorane anesthesia to mount a post for subsequent head fixation during recordings and to expose a small portion of the skull to provide access to the IC. The head post was surrounded by dental acrylic or Charisma composite, which bonded to the skull and also to a set of stainless steel screws that were embedded in the skull.

During a 2-wk recovery period, the animal was treated daily with antibiotics (Baytril) and the wound was cleaned and bandaged. Subsequently, a small (≈1-mm diameter) craniotomy was made in the skull. The location of the hole was based on stereotaxic coordinates as well as superficial landmarks on the skull that were marked during surgery (e.g., the bregma). The exposed recording chamber surrounding the craniotomy was covered with polysiloxane impression material (GC America, Alsip, IL) between recording sessions, and after many penetrations (usually >30) the hole was filled with a layer of bone wax and dental acrylic before another craniotomy was made to provide access to other regions of the same ICC. Multiple craniotomies were performed on the right hemisphere before the animal was killed and perfused for histological evaluation. All procedures were approved by the Johns Hopkins University Institutional Animal Care and Use Committee and conformed to National Institutes of Health standards.

Recordings were made in three marmosets, two females and one male. On each recording day, an epoxy-coated tungsten microelectrode (A-M Systems, impedance between 3 and 12 MΩ) was slowly advanced with a hydraulic microdrive (Kopf Instruments). 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. Depending on the angle of the dorsal approach to the ICC (which varied across craniotomies and sessions) the electrode traversed 1–1.2 cm of brain tissue before reaching the IC. Usually the electrode was in an approximately frontal plane at an angle of ∼0–6° medial or lateral and ∼0–10° anterior or posterior. Stimulus presentation, animal monitoring via video camera, and electrode advancement were controlled from outside a double-walled soundproof booth. Spike times of Schmidt-triggered single-neuron responses were recorded with respect to stimulus onset. Only well-isolated single neurons were studied. Sessions were terminated after 3–4 h, or earlier if the animal showed signs of discomfort. With these techniques, it was possible to collect data from several hundred electrode penetrations over the course of >12 mo in an animal.

Often, the external nucleus (equivalently, lateral cortex) or dorsal cortex of the ICC is encountered with our dorsal approach (Oliver 2004). The data reported here are from neurons considered to be in the central nucleus of the IC (ICC). We define “ICC-like” neurons on the basis of 1) short-latency, sharply tuned, nonhabituating responses and 2) a well-defined tonotopically organized frequency sequence as the electrode was advanced (Aitkin et al. 1975).

**Acoustic stimuli.** Sound stimuli were digitally generated in MATLAB (Mathworks), converted to analog signals (RP2.1, Tucker-Davis Technologies or National Instruments D/A board), and presented over calibrated headphone drivers (STAX) acoustically coupled to ear inserts (closed system). The acoustic system was calibrated in situ daily with a probe microphone, and stimuli were corrected for the calibration.

Binaural wideband noise bursts (50-ms duration) and pure tones of variable frequency and level were used to characterize neural activity as sound driven and to estimate the BF and threshold. BF was defined as the frequency at the center of the neuron’s response area at a sound level near threshold, usually a few decibels above threshold. Each neuron was characterized by its response map (Ramachandran et al. 1999). To obtain a response map, 200-ms tones (10-ms rise/fall) were presented (1/s) over a 2- to 4-octave range centered on the previously estimated BF, using 25 log-spaced frequencies per octave. This procedure was repeated for at least three sound levels, usually 20, 40, and 60 dB above the audiovisually determined estimate of BF threshold.

Each neuron was then studied with the sets of VS stimuli described in the next section. There were 32 different stimuli in a VS set. These had duration of 335 ms and were presented 1/s at a stimulus level ∼20 dB above the diotic binaural noise threshold measured with a noise filtered with the average HRTF. They were repeated as long as contact with the neuron was maintained (10–100 repetitions). The stimuli were presented in pseudorandom order within each presentation of a set of 32 stimuli. Each successive set of 32 stimuli was generated with a different sample of noise.

In some neurons, a spatial receptive field was determined with noise filtered by HRTFs from 135° (ipsilateral to the right ear) to −135° in azimuth (AZ) in steps of 15° and elevations (ELs) from −30° to 90° in steps of 15°. Each stimulus was presented once for a duration of 100 ms (with a stimulus repetition period of 600 ms) at ∼20 dB above the diotic binaural noise threshold. The AZ responses shown in Fig. 2A were averaged across ELs from −30° to 30°. Similarly, the EL responses shown in Fig. 2C were averaged across contralateral AZs from −120° to −60°.

VS stimuli were constructed by using HRTFs measured in both ears of each marmoset (individual HRTFs) to impose sound localization cues on samples of random noise. The HRTFs contained both aligned and misaligned cues as discussed in the next section. To present accurate VS stimuli, the transfer function of the acoustical delivery system and ear canal between it and the microphone must be removed (Wightman and Kistler 1989a). 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. 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 resampled from 8,192 to 32,768 points and multiplied by a random 32,768 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 335 ms, which included 10-ms linear onset/offset ramps.

In a subset of recordings from neurons with BFs <11 kHz the stimulus was again resampled so that acoustic features at 19 kHz were shifted to the neuron’s BF. This allows the responses of low-BF neurons to the notches in the marmoset HRTFs to be tested.

**Acoustic measurement of sound localization cues.** We measured individual HRTFs in each of the three 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 an arc stand in the center of an acoustically isolated chamber. 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.

VS stimuli with misaligned sound localization cues. The spatial positions of the HRTFs used to construct stimuli in the majority of experiments are shown on a sphere in Fig. 1Ai with the marmoset’s head centered at the origin. There were eight locations that varied in both AZ and EL. The coordinates of these positions were: (position no.: AZ, EL) = (1: 105, −15), (2: 75, 0), (3: 45, 15), (4: 15, 30), (5: −15, 30), (6: −45, 15), (7: −75, 0), (8: −105, −15).

There were four stimulus sets for each neuron. In the first, the stimuli contained naturally aligned cues, i.e., unmodified HRTFs were used to construct the VS stimuli for each of the eight positions shown in Fig. 1Ai. In the remaining three sets, the cues were interchanged so that cues were misaligned. In the most common case, the ITD and SS cues were unchanged and the ILD cues were interchanged with another stimulus location, as shown in Fig. 1, Aii, Aiii, and Aiv. The interchange occurred between cues one, two, or four stimulus positions distant. For example, in the “cues 2 off” stimulus set, stimulus 3 has the ITD and SS cues appropriate to the source position of stimulus 3 but the ILD cues appropriate to stimulus 1; the opposite is true: stimulus 1 has its own ITD and SS cues but the ILD cues of stimulus 3. Note that this procedure presents each possible stimulus feature (i.e., each ILD and each SS/ITD combination) an equal number of times.

To separate the ILD cues from the other two cues, the method in Fig. 1B was used. Figure 1Bi shows the HRTFs for the left (blue) and right (red) ears from two sound-source positions (positions 4 and 8). Figure 1Bii shows the ILD spectra for the two positions (blue), computed as the difference between the right and left curves in Fig. 1Bi. The ILDs show the effects of both overall level differences that vary slowly with frequency, presumed to be from general head-shadow effects, and the rapidly varying spectral notches of the SS cues, presumed to be from the local acoustics of the pinna.

To estimate the ILDs, we noted that at a fixed AZ the ILD spectrum was relatively constant over a range of ELs (~15° to 37.5°) except in

---

**Fig. 1.** A, i: positions of the 8 virtual space (VS) sound sources are shown in azimuth (AZ) and elevation (EL). ii–iv: Examples of cue misalignment for 1-position-off, 2-positions-off, and 4-positions-off stimulus sets. Interaural level difference (ILD) cues are interchanged between all possible pairs separated by the given number of stimulus positions, as illustrated for 1 case by the arrows. Spectral shape (SS)/interaural time difference (ITD) cues are kept unchanged. B, i: head-related transfer functions (HRTFs) in the left (blue) and right (red) ears for 2 positions, given as the magnitude of the Fourier transform of the HRTF. ii: ILD (blue) computed as the difference of the HRTFs. Dashed green lines are the smoothed ILD, computed as described in the text. iii: Cues-misaligned HRTFs for the 4-off case in which the ILDs of stimuli 4 and 8 were interchanged.
frequency regions with large spectral peaks or notches. Therefore, we averaged the ILD spectra across ELs at each AZ. These averages were then smoothed above 10 kHz with a broad triangular filter (6 kHz wide). The resulting ILD-only magnitudes for positions 4 and 8 are shown as green dashed lines (ILD4 and ILD8) in Fig. 1Bii. The modified HRTFs were computed by subtracting the smoothed ILD for each position from the ipsilateral HRTF and then adding the smoothed ILDs from the interchange position. The resulting HRTFs when the ILDs for positions 4 and 8 are interchanged are shown in Fig. 1Biii. Note that this procedure does not affect the phase spectrum, and so does not change the ITD cues (not shown), nor does it change the frequency location and general peak-notch characteristics of the SS cues.

In a subset of neurons we used the same method with the exception that we subtracted half the smoothed ILD from the ipsilateral ear and added half to the contralateral ear. This was done for the neurons presented in Fig. 3C in order to keep the average binaural intensity (ABI) the same for stimulus sets with only ILD or SS/ITD cues. The results were similar for both ILD processing methods.

Comparison of ILD and SS/ITD tuning. Tuning curves for various cues were constructed by plotting discharge rate versus stimulus AZ. Examples of these are plotted in Figs. 4, 5, 6, 8C, and 8E. To show ILD tuning, the data were ordered along the x-axis according to the AZ of the stimulus’ ILD. To show SS/ITD tuning, the data were ordered according to the AZ of the SS/ITD values. Of course, the AZs (and the tuning curves) are the same for stimuli with aligned cues, but they are different if the cues are misaligned. To evaluate the degree of tuning of the neuron to ILD versus SS/ITD, we reasoned that the tuning would be affected less by cue misalignment for the aspect of the stimulus that dominated the neuron’s tuning. The tuning curves for the aligned cues and for the cues misaligned by four positions were compared by cross-correlation, implemented by considering the tuning curves to be 8-dimensional vectors and computing their normalized dot-product, as

\[ v_{\text{aligned}} \cdot v_{\text{misaligned}}^{-4/8} = \left( \frac{1}{\|v_{\text{aligned}}\|^2} + \frac{1}{\|v_{\text{misaligned}}\|^2} + 4 \right) \]

where \( v \) is the tuning curve expressed as an 8-dimensional vector. This process was done twice, once for the tuning curve plotted with ILD AZ and once for the tuning curve with SS/ITD AZ.

Information analysis. Neural coding of VS stimuli was quantified with MI. MI between the spike rates (\( R \)) and the stimuli (\( S \)) was calculated for the four VS sets. The MI is a measure of the reduction in the uncertainty of the stimulus location provided by the neural response. The MI between a set of stimuli \( S \) and neural responses \( R \) is

\[ MI(S; R) = \sum_{s \in S} \sum_{r \in R} p(s, r) \log_2 \left( \frac{p(s, r)}{p(s)p(r)} \right) \]

where \( p(\cdot) \) is the empirical probability distribution over the stimuli and/or responses (Cover and Thomas 1991). Because each stimulus was presented with equal probability, the stimulus entropy was \( \log_2 8 = 3 \) bits. This value sets an upper bound on the MI, corresponding to perfect identification of each stimulus given the neural responses. The neural response \( R \) considered in this report is the number of spikes in the response (proportional to the average firing rate), calculated from the spike times occurring from the VS stimulus onset to 15 ms after stimulus offset.

Bias in the estimate of MI was reduced with a technique called the fill-in method (Chase 2006). This technique attempts to correct for the fact that empirical histogram bins containing no spikes are given zero probability, even though those responses may occur with some nonzero probability but were not observed because of limited sampling. The zero probabilities contribute greatly to MI because observing no spikes necessarily rules out some stimuli. The fill-in method operates with the assumption that the spike counts between the minimum and maximum do not actually have zero probability, so that a “count” of 0.5 is added to each histogram bin, as

\[ p(R = r | S = r) = \frac{N_{\text{obs}(r|s)} + 1/2}{Z} \]

where \( N_{\text{obs}(r|s)} \) is the number of times spike count \( r \) was observed while presenting stimulus \( s \). \( Z \) is a normalizing factor equal to the sum of the numerator over all values of \( r \) for the given \( s \):

\[ Z = N_{\text{rep}} + \frac{\max(r) - \min(r) + 1}{2} \]

\( N_{\text{rep}} \) is the number of repeats of the stimulus; in the present data set the median number of repeats was 32 (range = 9–100). The probability estimates of Eq. 2 were used to compute MI in Eq. 1, where the sum over \( r \) was taken from \( \min(r) \) to \( \max(r) \) only. Empirical calculations for simulated spike trains show that the fill-in method is effective in reducing bias (Chase 2006).

The bias was further reduced by bootstrapping (Efron and Tibshirani 1998). The median value of MI was 1.11 bits, and the median estimated bias was 0.15 bits. Without use of the fill-in method the median value of MI was 1.19 bits and estimated bias was 0.29 bits. All MI values reported here are corrected for bias.

RESULTS

General properties of frequency and spatial tuning in marmoset ICC. We studied the effects of cue misalignment in 69 “ICC-like” single neurons with BFIs ranging from 0.5 to 35 kHz. In 68 neurons we measured frequency response maps at multiple SPLs in order to classify them according to previous schemes (Ramachandran et al. 1999). Nearly all neurons had sharp tuning with a clear BF. The population reported here consisted of 34 type I, 21 type V, and 2 type O neurons, according to the definitions of Ramachandran et al. (1999). The remainder of the population (11 neurons) did not fit into this classification scheme. These included three neurons with only inhibition in their response maps and eight neurons with broad excitatory and/or inhibitory response maps.

In 46 of the neurons (17 type Vs, 23 type Is, and 6 with inhibitory or broad response maps), we measured VS receptive fields at an SPL ~20 dB above threshold. Figure 2A shows the normalized driven rate (as defined in legend) of each neuron versus the AZ of the VS stimulus (averaged over ELs from −30° to 30°). The red curve in Fig. 2A 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. Figure 2B shows a histogram of best AZ, defined as the AZ eliciting the maximal firing rate. The data are distributed roughly symmetrically around the peak at −60° AZ. The best AZ averaged across the population was −66 ± 32° (mean ± SD), slightly smaller than the AZ with the greatest ILD in the marmoset HRTF (Slee and Young 2010).

Figure 2C shows the normalized driven rate of each neuron versus the EL of the VS stimulus (averaged over AZs from −120° to −60°). The red curve in Fig. 2C 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. On average, neurons are tuned to lower ELs and the amount of rate modulation is less than for AZ. Figure 2D is a histogram of best EL, defined as the EL eliciting the maximal firing rate. The distribution peaks at
−30° EL and then declines gradually at higher ELs. The shape of this distribution may be the result of a preference for a certain SS but could also be affected by the general decrease in ILD with EL that is present in the VS stimuli (see Fig. 7, B and C, Slee and Young 2010). The best EL averaged across the population was 2° to 33° (mean ± SD). The average best EL is also likely influenced by the limited range of ELs measured (−30° to 90°).

In general, the spatial tuning properties in Fig. 2 are similar to those reported for the ICC in other species in both awake free-field (Zwiers et al. 2004) and anesthetized closed-field (Delgutte et al. 1999) conditions. Azimuthal spatial tuning is mostly broad and located in the contralateral hemifield. EL tuning is often weaker and more variable with both monotonic and tuned spatial responses.

ILD is a stronger sound localization cue than SS/ITD in high-BF ICC neurons. To estimate the relative strength of neurons’ responses to ILD versus SS/ITD, we compared the effects of cue misalignment on tuning curve shape for high-BF neurons (BF ≥ 11 kHz). The analysis was confined to high-BF neurons because all three sound localization cues are significant at these frequencies, so the largest effects of cue interaction occur in these neurons (shown below). We reasoned that if the responses were dominated by one cue then misaligning other cues would have a small effect on the shape of spatial tuning curves compared with misaligning the dominant cue. Figure 3A shows an example of a neuron for which responses are dominated by ILD. Figure 3A, left, shows the tuning curves for aligned (black) and four-position misaligned (blue) stimuli; tuning changes minimally between these two curves, for which the x-axis is appropriate to ILD. Figure 3A, right, shows the same data plotted on an x-axis appropriate for SS/ITD; now there is a large effect of misalignment. We quantified this effect by computing the vector dot-product of the aligned and misaligned tuning curves (considering them to be vectors in 8 dimensions), after normalizing the vectors by their norms. The correlations were 0.90 for ILD alignment versus 0.14 for SS/ITD alignment in this case, showing that the responses were largely determined by ILD.

Figure 3B plots the correlations described above for the population of neurons with BFs above 11 kHz. Both sets of correlations have a large range of values, showing that both ILD and SS/ITD can dominate the responses. However, in 33 of 36 neurons the ILD correlation is higher than the SS/ITD correlation. Figure 3B, inset, shows the distribution of the differences between correlations for ILD and SS/ITD tuning. Consistent with the main plot, the distribution has significantly more positive values than expected by chance. This result shows that tuning curve shape is dominated by ILD in most high-BF neurons in the population.

Information for stimuli containing only a single localization cue. In 10 neurons we tested responses to VS stimuli with only ILD or SS/ITD cues. These were synthesized as described in Fig. 1, except that the HRTFs were manipulated to eliminate ILD cues (by setting the green ILD function in Fig. 1B to zero) or to eliminate SS/ITD cues (by using only the green ILD

![Fig. 2. A: azimuthal tuning for all neurons, plotted as normalized rate (rate minus spontaneous rate divided by maximum rate minus spontaneous). Negative values show inhibition. Red line is the median of the data. B: best AZ computed as the AZ giving the maximum rate response. Negative values are in the contralateral field. C and D: same as A and B for EL.](http://jn.physiology.org/)

![CUE ALIGNMENT IN IC](http://jn.physiology.org/)
function with the average phase function in both ears). In these cases we computed MI as described above. Figure 3C plots $MI_{\text{ILD}}$ versus $MI_{\text{SS/ITD}}$. In 9 of 10 cases the data points lie above the unity line, which indicates that most neurons code ILD more strongly than SS/ITD.

Information can be larger, smaller, or unchanged by cue misalignment. More detailed examples of the effects of cue misalignment are shown in Figs. 4–6. The average firing rates to VS stimuli with variable cue alignments are shown in a low-BF neuron (BF = 1 kHz, type V) in Fig. 4A. Each curve plots the average firing rate ($\pm$ SD) versus the AZ of the ITD/SS cues. In this neuron, the response is primarily to ITD because both ILD and SS cues are small at this low frequency. The thick black curve in Fig. 4A corresponds to the aligned stimulus set where ITD/SS cues are aligned with ILD cues, whereas the other curves correspond to stimuli with increasing degrees of misalignment (described in legend). The tuning curves are nearly identical, as expected for a response dominated by ITD.

Fig. 3. A: example of spatial tuning curves for a neuron. These show discharge rate plotted vs. the AZ of the stimulus, using the AZ of the ILD at left and the AZ of the SS/ITD at right. In each case data are plotted for the cues-aligned stimuli (thick black lines) and 4-off misaligned stimuli (blue lines). Note that cue misalignment changes the tuning less for the ILD case than the SS/ITD case. B: cross-correlation (computed as normalized dot-product, see MATERIALS AND METHODS) between the cues-aligned and the cues-misaligned spatial tuning curves for 2 possible azimuthal choices, AZ appropriate to the SS/ITD cue (x-axis) or AZ appropriate to ILD (y-axis). Points are mostly above the diagonal line, meaning that the tuning curve showed less effect of cue misalignment when AZ was plotted as ILD (y-axis) as opposed to SS/ITD (x-axis). Red point is the data from the neuron in A. Inset: histogram of the difference between the x-axis and y-axis values of each point. The mean difference is 0.55, significantly different from zero ($P < 10^{-7}$ binomial test). C: the mutual information (MI) for ILD plotted against the MI for SS/ITD for 10 neurons studied with ILD-only and SS/ITD-only stimulus sets.

Fig. 4. A: spatial tuning for 1 low-best frequency (BF) neuron with cues aligned and 1, 2, and 4 positions away from alignment. The discharge rate during the stimulus is plotted against the AZ of the SS/ITD cues. The discharge rate is plotted as a function of the AZ of the SS/ITD cues. Error bars are 1 SD of the rates (BF = 1 kHz, type V, stimuli are 20 dB re noise threshold). B: effect of cue alignment on MI; error bars are 1 SD estimated from bootstrap. MI was slightly reduced by cue misalignment.
Figure 4 plots the bias-corrected MI between the distribution of firing rates and each VS set. The error bars show the SD of bootstrap MI estimates. The MI for each VS set is very similar, 1.4 bits, which indicates that this neuron is not sensitive to cue alignment.

Figure 5A plots the responses of a high-BF neuron (BF = 16 kHz, type I) to the VS stimulus sets. Here, the ILD cues were misaligned relative to the SS/ITD cues. For each VS set the firing rate of the neuron changes nonmonotonically with AZ. However, the range of firing rates (from maximum to minimum) is greatest for the aligned stimulus set. As the degree of cue misalignment increased, the range of firing rates decreased monotonically, showing a preference for cue alignment in this neuron. This effect was greatest for VS positions in the ipsilateral sound field.

The MI in Fig. 5B decreases from 1.48 to 0.76 bits as the cues are misaligned by increasing amounts. This is expected from the flattening of the rate curves with misalignment in Fig. 5A. The difference in MI between the set with cues aligned and the set with cues misaligned by four positions is 0.72 bits (48% of the total MI). Unlike Fig. 4, both SSs and ILDs are large in this neuron’s frequency response area (as in the example HRTFs in Fig. 1B). This neuron shows a clear preference for natural, aligned cues in that the MI is greatest when sound localization cues are aligned.

Figure 6A shows data from another high-BF neuron (BF = 19.5 kHz, type O) tested with stimuli in which the ILD cues were misaligned relative to the SS/ITD cues. To make the effects of cue alignment more clear, this figure plots the variation of firing rate relative to the AZ of the ILD cues. For each VS set the firing rate of the neuron decreases as the AZ (of the ILD cues) changes from the contralateral to the ipsilateral hemifield. In this case, the VS set with the greatest degree of misalignment (4 positions) leads to the greatest range of firing rates. Consistent with this change of rate range, the MI increases from 0.96 to 1.43 bits (Fig. 6B), an increase of 50%, as the cues are misaligned. Again, both SSs and ILDs are present in the neuron’s frequency response area.

Some neurons had nonmonotonic changes in MI with cue alignment. The greatest MI could be found at any cue alignment: 6 of 51 neurons had the greatest MI with aligned cues, 12 of 51 neurons for cues one position off, 11 of 51 neurons for cues two positions off, and 7 of 51 neurons for four positions off. The largest group of neurons (15/51, 54.9%) did not show a single peak MI (as in the example in Fig. 4).
Greatest sensitivity to cue alignment is found in BF ranges where multiple cues overlap. The responses to cue alignment were examined across the tonotopic axis of the ICC. Figure 7 plots the difference between MI calculated in response to stimuli with aligned cues and stimuli with cues misaligned by four positions versus BF. The error bars are estimates of the SD of the MI difference, computed from bootstrap estimates of MI (see Fig. 7 legend). The data points with error bars that do not overlap the x-axis are considered to represent significant changes with cue misalignment (37/51 = 72.5% of neurons). Positive values indicate neurons with a preference for aligned cues, whereas negative values indicate neurons with a preference for fully misaligned cues. Figure 7, right, shows the data in a histogram collapsed across BFs (black bars). Two points are worth noting. First, there are nearly as many neurons with a significant preference for aligned cues (n = 17) as for misaligned cues (n = 20). In fact, cue misalignment is just as likely to increase as to decrease MI (P = 0.37 computed with a binomial test). Second, the range of the distribution of MI difference along the y-axis is significantly greater for BFs of 11 kHz and above (SD = 0.32 bits) compared with BFs below 11 kHz (SD = 0.16 bits, gray bars in Fig. 7, right; P = 0.01 computed with an F-test).

The approximate frequency ranges occupied by marmoset sound localization cues are plotted in Fig. 7, bottom (Slee and Young 2010). It can be seen that waveform ITD is the dominant cue below 2 kHz, ILD is dominant from ~2 to 11 kHz, and ILD and SS cues overlap above 11 kHz. The pattern of overlap of the cues at different frequencies seems to explain qualitatively the increase in the range of MI differences above 11 kHz, because it is only at high frequencies that there is strong neural sensitivity to more than one cue.

Does cue interaction represent a mechanism specific to sound localization or general spectral sensitivity? The greater variability of cue alignment responses for neurons with BFs above 11 kHz in Fig. 7 could result from a specialization for sound localization in high-BF neurons. However, the result can also be explained as normal frequency-selective responses to stimuli with different SSs like the high-frequency SS cues. To discriminate these possibilities, we resampled the VS stimuli to shift their spectra down by a factor of BF/19 kHz, so that SS cues were placed in the frequency response area of low-BF (<11 kHz) neurons. If the high-frequency responses are produced by spectral processing of SS cues, then the sensitivity to cue manipulation should be the same in neurons with BFs below 11 kHz when presented with cue-shifted stimuli.

This process is clarified by the example in Fig. 8, which shows the response map of an IC neuron (Fig. 8A) and the magnitudes of the HRTFs used to construct VS stimuli (Fig. 8B). The BF of the neuron (7.2 kHz) is indicated by the red line. It is evident that most spectral cues are well above the BF of this neuron, with the exception of a single notch centered at 9 kHz. Figure 8C shows the rate responses of this neuron to the standard VS stimulus sets (as in Figs. 4–6). It responds in the contralateral hemifield and is relatively insensitive to cue alignment (with an MI difference of 0.03 bits, a 2.6% increase from the aligned-stimulus value of 1.18 bits). This result is consistent with the neuron’s narrowly tuned response map centered at 7.2 kHz and the lack of cues in the stimuli other than ILDs at frequencies near the BF. Figure 8D shows the result of stimulus resampling; there are now substantial spectral cues at frequencies near BF (red line). With these stimuli, the neuron has a greater sensitivity to the alignment of sound localization cues (Fig. 8E), with an MI difference of 0.27 bits, a 25.3% increase over the aligned-cues value of 1.18 bits. In this neuron, stimulus resampling reveals a preference for cue misalignment similar to that seen in the neuron shown in Fig. 6.

Figure 9 shows the MI differences from neurons with BFs <11 kHz studied with resampled stimuli (red data points) compared with the data of Fig. 7 (blue data points). The histograms in Fig. 9, right, compare the results for all high-BF neurons (BF ≥ 11 kHz, blue bars) and for low-BF neurons with resampled stimuli (red bars). As noted with natural stimuli, most neurons (19/21) respond to resampled VS stimuli with significant decreases (5) or increases (14) in MI as cues are misaligned. MI is more likely to increase when cues are misaligned in the resampled stimuli (P = 0.03 with a binomial test).

Stimulus resampling also increased the range of the MI changes in low-BF neurons. The variability of MI differences in response to resampled stimuli in low-BF neurons is significantly greater (SD = 0.39 bits) than for natural stimuli in neurons with these BFs (SD = 0.16 bits, P = 0.01 F-test). More important, the variability of low-BF neuron responses to resampled stimuli is not significantly different.
from high-BF neuron responses to natural stimuli (SD = 0.32, \( P = 0.15 \) F-test). These results suggest that the range of MI differences in high-BF neurons is a result of ordinary spectral processing of the stimuli containing SS cues, and not a specialization of IC neurons to respond to particular SS cues for location.

DISCUSSION

Summary of results. The present study was designed to evaluate the means by which different sound localization cues are combined by neurons in the ICC. Consistent with previous studies (Delgutte et al. 1999; Devore and Delgutte 2010; Sterbing et al. 2003), most ICC neurons convey information about virtual sound source location and ILD is the strongest cue for AZ in high-frequency neurons.

MI was used as the primary measure for spatial coding capacity in this study. MI has the advantage that it is agnostic about tuning curve shape or the assumed decoding strategy. It quantifies the potential information that an optimal decoder could capture from the spike response rates. Therefore, no assumptions about the form of the coding need to be made with MI, which allows a direct measure of changes in spatial coding with cue misalignment. Because we used spike rate as the response measure, the MI measures mainly the range over which discharge rate varies relative to the standard deviation of rate.

The specific hypothesis offered here was that responses are stronger, in the sense that they provide more information, if the different cues are aligned as opposed to misaligned. The results do not support that hypothesis in a simple way. Although neurons that respond more strongly with cue alignment are seen, there are just as many neurons that do the opposite, respond more strongly to cue misalignment (Figs. 7 and 9).

Ordinarily the marmoset ear does not produce strong SS cues over the frequency range below 11 kHz. Because frequency tuning is sharp in most ICC neurons, each neuron only samples the HRTFs in a small frequency band. Thus effects of cue alignment or misalignment were not large for neurons with BFs below 11 kHz in Fig. 7; there is little cue overlap at those frequencies, and ILDs or ITDs dominate the responses, depending on BF. The problem of integrating multiple cues then
is mainly an issue for ICC neurons at high BFs where SS and ILD cues overlap.

However, several studies in the ICC show that many neurons also respond to the ITDs in the stimulus envelopes, which are available at all frequencies (Batra et al. 1993; Devore and Delgutte 2010; Griffin et al. 2005; Joris et al. 2006; Yin et al. 1984). In most of the neurons studied here, the stimulus ITD cues were aligned with the SS cues. If neurons contained a large fraction of information about envelope ITDs, we would likely have seen greater effects of cue alignment or misalignment in the frequency region dominated by ILDs (~3–11 kHz in the marmoset). The results presented in Fig. 7 suggest this is not the case. In high-BF neurons strong envelope sensitivity would have produced a preference for SS/ITD over ILD in some neurons. For the great majority of neurons this also was not the case (Fig. 3). Overall our results are consistent with previous studies showing a less prominent role for envelope ITDs relative to ILDs with unmodulated stimuli (Devore and Delgutte 2010).

**Role of spectral sensitivity.** When the stimuli were treated to shift SS cues into the frequency region of neurons’ BFs (Fig. 8a), effects of cue alignment or misalignment similar to those in high-BF neurons appeared (Fig. 9). That is, the variability of MI differences with cue misalignment increased and was indistinguishable for neurons with BFs above and below 11 kHz. This suggests that neurons in the high-BF region of ICC, which receive stimuli with multiple cues, are not specialized to process multiple cues relative to neurons found in the low-BF region.

The neurons stimulated with shifted HRTFs showed a statistically significant preference for misalignment of cues (histogram, Fig. 9, right). For these neurons the features in the HRTF around 19 kHz were shifted into the BF region under study. To investigate a stimulus origin for this result, we computed average ILDs in a 1/4th-octave band centered at 19 kHz for the HRTFs measured in the three monkeys used in this study. We found a greater range of ILDs for the fully misaligned set of HRTFs compared with the aligned set in two of the latter two monkeys, with the greater range of ILDs for misaligned cues. This suggests that the details of the shifting procedure for the HRTFs may have produced the significant preference for misaligned cues.

**Effects of cue misalignment in perceptual studies.** Many psychophysical studies have examined lateralization judgments by using HRTFs containing misaligned cues. For broadband sounds, when the HRTF magnitude (ILD/SS cues) points in one direction and HRTF phase (ITD) at low frequencies points in another, perception follows the ITD (Macpherson and Middlebrooks 2002; Wightman and Kistler 1992). A similar experiment with high-pass sounds usually results in the perception following the ILD, although some listeners show an effect of phase (presumably envelope ITD cues; Macpherson and Middlebrooks 2002; Wightman and Kistler 1992). In another study (Macpherson and Middlebrooks 2002) the monaural HRTF ipsilateral to a target source was interchanged with an HRTF from a nontarget source. At the same time the HRTF in the other ear was altered to keep the target ILD spectrum the same. The results strongly suggested that perceived source location followed the target ILD rather than the SS cues. The physiological results in the present study provide a general substrate for these psychophysical studies in that ITDs dominate in low-BF neurons (as in the example of Fig. 4) and ILDs dominate in high-BF neurons (Fig. 3). The weaker contribution of SS cues was likely influenced by the fact that the VS stimulus set used in this study contained a greater range of AZs (210°) than ELs (45° or 52.5°).

Other psychophysical literature has focused on cue weighting in different listening conditions. One study demonstrates rapid reweighting of cues by a monaural earplug and suggests that cues are reweighted depending on their reliability (Van Wanrooij and Van Opstal 2007). Another related study shows that reverberation changes the magnitudes of cues and suggests they are reweighted accordingly (Rakerd and Hartmann 1985). Yet another study measured EL judgments by using HRTFs corresponding to different ELs in the two ears (Macpherson and Sabin 2007). In this case the perceived EL was a weighted sum of each position and the relative weights could be changed by adding biasing ITDs and ILDs, which favored one ear or the other. Finally, there is evidence that with multiple competing sources listeners attend to the cues corresponding to the target location, i.e., the one to which attention is directed (Darwin 2008; Kidd et al. 2005; Shinn-Cunningham 2008).

**Representation of sound location in the ICC.** The perceptual results discussed above show that the auditory system has the ability to change the set of cues on which it depends for computing the location of an object, depending on listening.
conditions. Generally the weights given different cues vary with the likely reliability of a cue as judged from other information, as, for example, when the dominance of low-frequency ITD cues in localization is lost for high-pass stimuli. The representation of sound localization cues in the ICC seems to correspond to this flexibility in that neurons respond to multiple cues, but not in a way that is constrained by a simple representation of auditory space. Cues could be combined in a way that corresponds to natural ear acoustics, the hypothesis that motivated this work. Such a system would be a necessary step in the formation of constrained spatial receptive fields, in which neurons respond only to sources in a small region of space. Such responses are seen in the superior colliculus (King and Hutchings 1987; Middlebrooks and Knudsen 1984) and the nuclei that form a path from the ICC to the superior colliculus (Binns et al. 1992; Schnupp and King 1997). However, they have not been described in the ICC or in auditory cortex where spatial receptive fields are generally large. As studied here, the spatial responses in ICC seem to correspond more to those in the higher auditory centers than to those in the superior colliculus. This work is thus consistent with the general finding of a broad representation of localization information in which cues are mixed with little sorting of cues according to their consistency with particular sources.

This mixed pattern of cue combinations certainly provides flexibility to the auditory system when it is desirable to change the weighting of different cues. It may also be useful in other computational tasks such as using spatial separation for binaural unmasking (Hirsh 1948). A stimulus source could be identified in the presence of one or more spatially separated maskers by identifying the frequency regions in which binaural cues are aligned. Presumably those regions contain signals from sources at high enough signal-to-noise ratios to allow them to dominate the sound at those frequencies. Thus activity in “cue-alignment” neurons, those that provide more information about aligned-cue stimuli, could serve as a substrate for binaural unmasking. The opposite is true; activity in “cues-misalignment” neurons would represent responses to an unresolved mixture of sources. Similar comments apply to other difficult listening situations, like reverberant environments where the activity in cue-alignment neurons would be the likely best estimate of first-wave front information, which is often used for localization in the presence of reverberation (Wallach et al. 1949).

The approach used here may be useful in exploring cue interaction in other brain areas. A limited number of studies report spatial auditory maps in the external nucleus of the IC (ICX) (Binns et al. 1992) and in the nucleus of the brachium of the IC (BIN) (Schnupp and King 1997). In particular, the BIN provides the main source of input to the superior colliculus (Nodal et al. 2005), where the existence of a auditory space map has been long established (Middlebrooks and Knudsen 1984; Palmer and King 1982). It seems likely that the BIN contains more cue-alignment neurons than the ICC, and future investigation will be important to answer questions about cue representation in these nuclei.

ACKNOWLEDGMENTS

We thank Ron Atkinson, Jay Burns, Phyllis Taylor, and Qian Gao for technical assistance and Jenny Estes and Judy Cook for assistance with animal care.

GRANTS

This work is supported by National Institute on Deafness and Other Communication Disorders Grants DC-00115, DC-00023, and DC-05211.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES


