Encoding of Virtual Acoustic Space Stimuli by Neurons in Ferret Primary Auditory Cortex

Thomas D. Mrsic-Flogel, Andrew J. King, and Jan W. H. Schnupp

University Laboratory of Physiology, Oxford, United Kingdom

Submitted 21 July 2004; accepted in final form 11 January 2005

INTRODUCTION

The behavioral changes produced by lesions or pharmacologically inactivating the primary auditory cortex (A1) suggest that this cortical field plays an essential role in sound localization (Jenkins and Merzenich 1984; Kavanagh and Kelly 1987; Smith et al. 2004). Electrophysiological studies in different species have revealed that A1 neurons show varying degrees of sensitivity to sound-source direction (Brugge et al. 1994; Eisenman 1974; Imaging et al. 1990; Middlebrooks and Pettigrew 1981; Middlebrooks et al. 1994; Mrsic-Flogel et al. 2001; Poon and Brugge 1993; Recanzone et al. 2000; Samson et al. 1993). However, the neural computations underlying the spatial selectivity of these neurons have yet to be fully elucidated.

A more detailed description of the mechanisms that give rise to spatial sensitivity in A1 is necessary for improving our understanding of the cortical processing of sound-source location. By using individualized virtual acoustic space (VAS) stimuli, which encapsulate all the necessary cues for sound localization, we previously showed that the representation of sound direction by high-frequency [best frequencies (BFs), 4–30 kHz] A1 neurons in the ferret is highly dependent on the acoustic energy distribution associated with direction-dependent filtering of sounds by the head and external ears (Mrsic-Flogel et al. 2001, 2003; Schnupp et al. 2001). Furthermore, we have shown that the spatial sensitivity of most A1 neurons arises from a largely linear processing of these spatial cues (Schnupp et al. 2001). Thus the shape of the spatial response fields (SRFs) is well predicted by a simple filter model that integrates sound energy in VAS stimuli according to the neurons’ frequency–time response fields (FTRFs) (deCharms et al. 1998; Kowalski et al. 1996). This apparent linearity of spatial processing is perhaps surprising, considering that subcortical neurons can use nonlinear operations to achieve their selectivity to sound localization cues [e.g., processing of interaural time difference (ITD) in the medial superior olive (MSO)] (Goldberg and Brown 1969; Grothe and Park 1998; Spitzer and Semple 1995; Yin and Chan 1990) or of interaural level difference (ILD) by some superior colliculus neurons (Wise and Irvine 1983; Yin et al. 1985).

However, we also found that the SRFs of some A1 neurons were much more poorly predicted by a linear model than others (Schnupp et al. 2001). Because nonlinear interactions seem to be particularly associated with subcortical neurons that receive excitatory inputs from each ear, irrespective of their frequency tuning, we have now extended this approach to investigate whether the balance of excitatory and inhibitory influences from each ear determines whether cortical SRFs can be well predicted by a linear model. We also examined the effect of sound level on SRF parameters. A number of authors (Brugge et al. 1994; Middlebrooks and Pettigrew 1981; Rajan et al. 1990) noted that A1 SRFs typically expand with increasing sound level, and it is pertinent to ask whether these changes in SRF size are compatible with the view that most A1 cells perform a relatively simple, approximately linear filtering of binaural localization cues. A strictly linear model would predict that SRF shape and position remain constant, and responses should merely scale up or down as a function of sound level. We therefore assessed the effect of sound level changes on SRF position and size for each binaural class of neurons to see whether they were broadly compatible with a linear filter model.

Recent findings have shown that spike timing, in addition to spike count, could be used as a parameter for encoding sound-source direction in cat auditory cortex (Brugge et al. 1996; 1998; 1999). However, we showed that most A1 neurons encode sound-source direction with a relatively simple, approximately linear model, which integrates sound energy in VAS stimuli according to the SRF shape and position. We also examined the effect of sound level on SRF parameters. A number of authors (Brugge et al. 1994; Middlebrooks and Pettigrew 1981; Rajan et al. 1990) noted that A1 SRFs typically expand with increasing sound level, and it is pertinent to ask whether these changes in SRF size are compatible with the view that most A1 cells perform a relatively simple, approximately linear filtering of binaural localization cues. A strictly linear model would predict that SRF shape and position remain constant, and responses should merely scale up or down as a function of sound level. We therefore assessed the effect of sound level changes on SRF position and size for each binaural class of neurons to see whether they were broadly compatible with a linear filter model.

Address for reprint requests and other correspondence: J. Schnupp, University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK (E-mail: jan.schnupp@physiol.ox.ac.uk).
Furukawa and Middlebrooks 2002; Middlebrooks et al. 1998). In particular, response latencies were shown to vary with sound-source position (Brugge et al. 1996; Middlebrooks et al. 1998; Stecker et al. 2003). We therefore also looked for evidence of temporal coding in our recordings from ferret cortex and examined whether variations in response latency could be predicted by the linear model of A1 responses.

**METHODS**

Each experiment constituted 2 phases: acoustical [head-related transfer function (HRTF)] recordings from both ear canals, followed by electrophysiological recordings from A1. All surgical procedures were approved by the local ethical review committee and by the UK Home Office.

**HRTF recording**

The HRTFs of each animal were recorded at the beginning of each experiment. To perform these recordings the animals were deeply anesthetized with alphaxalone/alphadolone acetate (Saffan, 2 ml/kg administered intraperitoneal, Mallinckrodt Veterinary, Uxbridge, UK; supplementary doses given intravenously as required), and custom probe microphones were surgically implanted in each ear canal. The probe microphones consisted of miniature KE-4-211-2 microphone capsules (Sennheiser, High Wycombe, UK) with damped polythene probe tube attachments (20 mm long, 0.86 mm ID, 1.52 mm OD). The caudal aspect of the ear canal was exposed, and the microphone probe tube inserted about 2 mm deep into the canal through a small hole that was made as close to the cranium as practically possible. The animal was positioned on a small steel plate in an anechoic chamber. A stainless steel bar was fixed to the skull from behind with stainless steel screws and dental acrylic (Simplex Rapid, Austenal Dental, Harrow, UK) to support the animal’s head. All supporting structures were designed and positioned so as to keep the sound field around the head as free as possible. The incisions in the scalp were closed and care was taken to ensure that the external ears maintained their natural position, according to measurements made before surgery.

A loudspeaker (Kef T27, KEF Audio, Maidstone, UK) mounted on a robotic hoop (radius 65 cm) was used to present broadband acoustic signals (512-point Golay codes; Zhou et al. 1992) from 433 different azimuth-elevation locations (60 to 85° azimuth, 60 to 160° elevation). Because “meridians” (lines of equal azimuth) come together at the poles, sampling at regular azimuth intervals would result in unnecessarily fine sampling near the poles. Consequently, elevations from −60 to +60° were sampled in 10° azimuth by 10° elevation steps, elevations +70 and +80° were sampled in 20° azimuth steps, and 2 samples spaced 180° apart in azimuth were taken at +85° elevation. Both the generation of the Golay code signals and the recording of the microphone signals were performed digitally using TDT system 2 A/D and D/A converters (sample rate of 80 kHz, Tucker–Davis Technologies, Gainesville, FL) and 30-kHz antialias filters. The microphone signals were analyzed for each stimulus direction, as described by Zhou et al. (1992), to calculate a spectral transfer function containing both the animal’s HRTF and the transfer characteristics of the loudspeaker and probe microphones. From this raw transfer function we calculated the animal’s directional transfer function (DTF). To generate the directional transfer function at each position by the mean transfer function from all positions. This procedure removes all those components from the raw transfer function that are independent of sound-source direction, including the microphone and speaker transfer functions, leaving only the directional component of the animal’s HRTF (Middlebrooks and Green 1990; Parsons et al. 1999). VAS stimuli were then generated from each DTF as described below.

**Preparation for electrophysiological recording**

On completion of the acoustical recording the probe tube microphones were removed, and custom insert earphones (M. Ravicz; MIT, Boston, MA) were sealed into the transected ear canal. The earphones were designed to deliver calibrated sound stimuli at a few millimeters from the eardrum. A craniotomy was performed to expose the left A1. In the ferret, A1 lies at the posterior tip of the ectosylvian gyrus (Kuly et al. 1986; Nelken et al. 2004; Versnel et al. 2002). The dura overlying A1 was removed, and the exposed cortex was protected by a well of dental acrylic built up around the craniotomy, which was filled with 2% agar in saline. A tracheal cannula was implanted for artificial respiration (7025 respirator, Ugo Basile, Milano, Italy), and the animal was transferred to an anesthetic regime of pentobarbital sodium (Sagatal; Rhône Mérieux, Harlow, UK; 2–3 mg·kg⁻¹·h⁻¹) and gallamine triethiodide muscle relaxant (Flaxedil; Sigma, Poole, UK, 20 mg·kg⁻¹·h⁻¹) administered by continuous intravenous infusion (Perfusor Secura FT infusion, B. Braun, Melsungen, Germany). Body temperature was held constant at 38 ± 1°C (using a custom-built thermostatic blanket and rectal probe). Inspired and expired CO₂ (47210A capnometer, Hewlett Packard GmbH, Boeblingen, Germany) and the EEG and ECG (custom-built amplifiers and monitors) were monitored throughout to ensure that the animal was maintained at a stable and adequate level of anesthesia.

**Generation of virtual acoustic space stimuli**

The ITDs were extracted from the microphone signals by cross-correlation of the impulse responses (Middlebrooks and Green 1990) after low-pass filtering (0–4 kHz). VAS stimuli were generated using a bank of digital minimum phase filters. Before each experiment, the transfer function of our closed-field speakers was calibrated by making probe-microphone measurements from an assembly of silicone tubing, plasticine, and ear-impression compound, which was designed to simulate the geometry of the ferret external ears. Amplitude corrections intended to compensate for the transfer characteristics of our in-ear headphones were added to the amplitude spectra of the DTFs for each stimulus direction, and minimum-phase filters were then calculated from the equalized amplitude spectra using the Hilbert transform. VAS stimuli consisted of short (20 ms) Gaussian noise bursts, which were convolved with the appropriate minimum-phase filters for each direction, and delayed to generate the appropriate ITD. Thus the VAS stimuli faithfully replicated the ILD, ILD, and monaural spectral pattern information contained in the DTF in the sounds delivered to the ear. VAS stimuli were generated afresh for every presentation (i.e., we did not use frozen noise).

**Electrophysiological recording and data analysis**

Single-unit activity was recorded extracellularly using conventional tungsten-in-glass electrodes. The electrode signals were band-pass filtered (500 Hz to 5 kHz), amplified (±10,000×), and digitized at 25 kHz. Units were isolated from the digitized signal by manually clustering data according to spike features such as amplitude, width, and area. To ensure adequate isolation of our single-unit data we also inspected interspike-interval (ISI) histograms. Clusters for which the ISIs did not reveal clear evidence of a refractory period were excluded from further analysis. Data acquisition and stimulus generation were controlled using BrainWare (Tucker–Davis Technologies).

To generate a spatial response field (SRF) map we typically presented VAS stimuli from 224 different positions. These positions sampled the available range (−160 to +160° azimuth, −60 to +85°...
element can be represented by a “point mass” at spherical coordinates
the apex to the center of the base. Therefore the weight of each SRF
distance between diagonal neighbors was
elevation) in 20° intervals, but were arranged so that the angular
positions above and behind the animal (along the “dateline” and at the
closer to the poles. The volume of the unit sphere is then
constituent pyramids be given by
accuracy of the dot product of
then the thresholded DTF
chord sequence stimuli (deCharms et al. 1998; Schnupp et al. 2001).
Frequency–time response fields and predicting SRFs
Binaural FTRFs were determined by reverse correlation to random
chorus sequence stimuli (deCharms et al. 1998; Schnupp et al. 2001).
Tone pips of 20 ms (5 ms risefall time) were started randomly at any of
60 frequency bands that spanned 0.5 to 29.857 kHz in 10th-octave
steps. The probability of a tone pip starting in any one frequency band
during any 5-ms time interval was set to values between 1 and 2%, so
that, on average, 2.4 to 4.8 tone pips would be on simultaneously at
any time during the random chorus sequence. Tone onsets in different
frequency bands and subsequent time intervals were statistically
independent. Two statistically independent random chord sequences
were used to trigger stimulation with levels that took between 15 and 20 min to collect.

To estimate the centroid position from a receptive field map, we
therefore simply iterate through all entries in the map matrix, calculate the “mass” of the corresponding pyramidal SRF element using Eq. 1
where \( d\theta \) and \( d\phi \) now equal the vertical and horizontal resolution of
each map entry, in our case 7.5°), and we accumulate the combined
centroid position over all SRF elements using Eqs. 2 and 3. Because
our SRF maps did not extend to elevations below –60° there is a risk
that this method will bias the centroid position estimates upward. We
therefore ran the iterations only over elevations from –60 to +60°.
The direction of the centroid vector thus summarizes the overall
directional preference of the SRF, whereas its length gives an indication
of how tuned the SRF is in that direction. The theoretical maximum length of the centroid (for an SRF tuned to a single sound
direction) is 3/4.

The dispersion of an SRF can be thought of as equivalent to the variance of the spike-weighted source directions around the centroid.
It was calculated as follows. Let \( \rho_i \) be the response of the neuron for
sound-source directions corresponding to the \( i \)th position of our SRF
map, normalized so that the total response over the entire SRF equals
one. Let the source direction for that map position be given by unit
vector \( \mathbf{p}_i = (\phi_i, \theta_i, 1) \) and let \( \mathbf{c} \) be a unit vector in the direction of the SRF
centroid. The angular distance between \( \mathbf{p}_i \), and \( \mathbf{c} \) is then equal to
the inverse cosine of the dot product of \( \mathbf{p}_i \) and \( \mathbf{c} \). The dispersion was therefore calculated by evaluating the sum

\[
D = \sum_{i} \arccos (\mathbf{p}_i \cdot \mathbf{c})^2 \cos (\phi_i)
\]

over all positions in our map. The term \( \cos (\phi_i) \) again serves to weight responses so as to compensate for the fact that parallels shrink nearer the poles.

Analysis of DTFs
Thresholding of broadband (8–22 kHz; the range encompasses
approximately the 10th to the 90th centile of unit BF distributions in our data set)
DTFs was carried out as follows: Let \( \epsilon(\theta, \phi) \) be the DTF gain for
stimulus direction with azimuth \( \theta \) and elevation \( \phi \), \( m \) be the maximum
\( \epsilon(\theta, \phi) \), and \( \tau \) be the threshold level relative to the maximum gain \( m \); then the thresholded DTF \( \tilde{\epsilon}(\theta, \phi) \) is given by

\[
\tilde{\epsilon}(\theta, \phi) = \max[\epsilon(\theta, \phi) - (m - \tau), 0]
\]

The calculation of the centroid vectors for the thresholded DTFs was the same as that for SRFs (see above).

Frequency–time response fields and predicting SRFs
Binaural FTRFs were determined by reverse correlation to random
chorus sequence stimuli (deCharms et al. 1998; Schnupp et al. 2001).
Tone pips of 20 ms (5 ms risefall time) were started randomly at any of
60 frequency bands that spanned 0.5 to 29.857 kHz in 10th-octave
steps. The probability of a tone pip starting in any one frequency band
during any 5-ms time interval was set to values between 1 and 2%, so
that, on average, 2.4 to 4.8 tone pips would be on simultaneously at
any time during the random chorus sequence. Tone onsets in different
frequency bands and subsequent time intervals were statistically
independent. Two statistically independent random chorus sequences
were presented simultaneously, one to each ear. FTRFs were deter-
mined by spike-triggered averaging of peri-pipe random chorus se-
quency segments. FTRFs were typically based on 500 to 1,500 spikes,
which took between 15 and 20 min to collect.

The centroid (or center of mass) was calculated by modeling the response field as a sphere of unit radius, whose “mass density” in each
direction was given by the observed response strength in the corre-
sponding direction. For example, an SRF might exhibit a number of peaks pointing in different directions that add up to a small total 50% area. We therefore used 2 further measures to quantify the spatial selectivity of the SRF:
the (dimensionless) SRF centroid vector and the dispersion of the SRF
around the centroid (in rad²).
We assigned each unit to a binaural response class based on whether the binaural FTRF indicated predominantly excitatory (E), inhibitory/suppressive (I), or no (O) input from the respective ear. When estimating the FTRFs by spike-triggered averaging we extended the peri-spike time interval to include ten 5-ms time steps past the time of occurrence of the spike. This “postspike” part of the FTRF cannot be causally related to the presence of the spike and, as a consequence of the random nature of the stimuli used, is statistically independent of the rest of the spike-triggered average. For the purpose of classifying the FTRFs we smoothed the spike-triggered averages slightly in both frequency and time by convolution with a 3-point moving average filter, and then used the noncausal part of the FTRF to estimate the amount of random variability that was to be expected in the causal part. If none of the peaks and troughs in the causal part of the FTRF extended beyond 4 SDs from the mean of the noncausal values, then the FTRF was considered unresponsive (O). Whether responsive FTRFs were classified as “E” or “I” depended simply on whether the largest deviation from the mean was positive or negative.

Spatial responses were then predicted by convolving the FTRFs for the left and right ear with the power spectrum vectors of the VAS stimuli to generate a (time-reversed) prediction for the response PSTH. By predicting the size of the response peak for each VAS position in turn we constructed predicted SRF maps for each unit.

Stimulus levels in these experiments were chosen arbitrarily, and varied from unit to unit. Our primary concern was to make good use of the available dynamic range of the digital signal processing equipment, and no attempt was made to match the sound levels of the VAS stimuli used to map the SRFs to that of the random chord sequences used to estimate the FTRFs.

RESULTS

A total of 127 acoustically responsive single units were characterized in A1 of 5 adult ferrets (3 male, 2 female).

General response properties

Previous electrophysiological (Phillips et al. 1988; Shamma et al. 1993) and optical imaging (Nelken et al. 2004; Versnel et al. 2002) studies have shown that A1 occupies the caudal aspect of the middle ectosylvian gyrus. The units recorded in our study were predominantly from the high-frequency, dorsal 80% of A1. Only 3 of the recorded units had BFs ≤5 kHz, whereas 50 units had BFs from 5 to 10 kHz, 38 had BFs from 10 to 20 kHz, and 36 had BFs from 20 to 30 kHz.

Most units recorded exhibited low or no spontaneous activity, but some did have pronounced spontaneous firing rates. Seventy units (about 55%) had a spontaneous firing rate of ≤1 Hz, 49 units (about 39%) had spontaneous rates from 1 to 5 Hz, whereas only 8 units (about 6%) exhibited spontaneous rates of ≥10 Hz. The majority of units (77/127, about 61%) responded to the VAS stimuli with a brief onset burst only, but units that presented multiple peaks in their PSTH were also commonly observed (46 units, about 36%). The number of peaks apparent in the PSTH of this class varied between 2 and 4. A small minority of units (4/127, about 3%) responded to VAS stimuli purely by a suppression of their spontaneous activity.

Of the units in our sample, 40 were classed as EE (predominantly excitatory inputs from both ears), 34 as EI (excited by contralateral but inhibited by ipsilateral stimulation), 51 as EO (monaural contralateral excitation), and 2 as OE (monaural ipsilateral excitation). Representative examples of FTRFs of these classes are shown in Fig. 1, A–C. The group of units classed as EI also included the small number mentioned above that exhibited purely inhibitory responses in their PSTHs when tested with broadband VAS stimuli. Perhaps surprisingly, the distributions of BFs in the 3 binaural classes were not significantly different, with EE-type and EI-type interactions being equally prominent over the full range of BFs represented in our sample.

Spatial tuning

Figure 2 gives illustrative examples of the SRFs recorded in this study. Despite the considerable diversity of SRFs, units exhibiting broad tuning to predominantly contralateral regions of space, like those shown in Fig. 2, A, D, and G, were very common. Units displaying SRFs of this type received excitatory input from the contralateral ear. Much less common were SRFs displaying 2 well-defined maxima (Fig. 2B; 3/127) or tuning to the ipsilateral side (Fig. 2C; 7/127). One might expect that ipsilateral SRFs correspond to units that are excited predominantly or exclusively through the ipsilateral ear. However, that was not necessarily the case. The binaural FTRF for the unit in Fig. 2C, for example, indicated excitatory inputs from both ears, whereas the multipeaked SRF in Fig. 2F was observed in a unit that appeared to be driven by the ipsilateral ear only. Figure 2, E, H, and I show examples of other SRF types that were only rarely encountered. The SRF in Fig. 2E was tuned to azimuths close to the animal’s midline (0° or ±180°), but largely insensitive to stimulus elevation. Figure 2H, in contrast, shows an SRF that was quite sharply tuned to frontal positions in both azimuth and in elevation. A clear elevation preference was also exhibited by the SRF in Fig. 2I, where the strongest responses were obtained for stimuli presented above the animal’s head. SRFs like those seen in Fig. 2, H and I are interesting because their structures are not easily explained in terms of the acoustical cues generated by the head and outer ears, but units like this are observed too rarely to facilitate a systematic study of their properties. Omnidirectional units (Fig. 2F) were also only very rarely observed.

Modeling spatial tuning

We previously showed that spectral cues are essential in shaping SRF structure of high-frequency A1 units (Mroe-Flogel et al. 2001, 2003; Schnupp et al. 2001). Using a linear model that assumes neurons integrate sound energy additively, we were able to predict SRF shape for most units quite accurately from the unit’s binaural FTRFs (Schnupp et al. 2001). Here we extended this analysis to investigate whether the accuracy of such linear predictions depends on the binaural response type of the unit. Figure 1, A–C shows observed and predicted SRFs, as well as the FTRFs from which the predictions were obtained, for a representative EO, EI, and EE cell, respectively. The match between the observed and predicted SRF was quantified by calculating the correlation coefficient (r, Fig. 1, A–C) from a comparison of the responses over all stimulus directions that evoked a measurable response in the observed SRF. Figure 1D summarizes the distribution of correlation coefficients for the 3 classes of cells in histogram form. For all 3 classes, these distributions are broad, indicating that the prediction accuracy of the linear model did vary consider-
ably from unit to unit. Whether the distributions varied systematically from one binaural response class to another is difficult to assess statistically because of the presence of “outliers” in the distributions for the EI and the EO units. Wilcoxon rank-sum tests on the data with all outliers included did not reveal any significant differences. When the outliers are excluded, however, the mean \( r \)-value for EE cells (0.54) is significantly smaller than that for EO cells (0.64, \( t \)-test, \( P = 0.034 \)). The mean for EI cells is slightly larger than that for EO units (0.66), but just fails to reach significance when compared with the EE cells (\( t \)-test, \( P = 0.064 \)) because of the relatively small sample size.

**SRF changes with sound level**

We found that the SRFs tended to expand, sometimes considerably, with increases in sound level. Figure 3, A–C shows a typical example of such an SRF expansion. (The data shown are from an EO-type unit.) In this case, the 50% response area of the SRF increased more than 10-fold as the sound level was raised by 20 dB. The SRF also appears to expand predominantly upward, resulting in a shift of the centroid from about \(-15\) to about \(+27^\circ\) in elevation. Figure 3D shows the linear prediction of the SRF for this unit. To facilitate a comparison between the predicted and the observed...
SRFs, the 50% response contours from Figure 3, A–C are shown superimposed on the predicted SRF. The 50% contours observed at the low, intermediate, and high sound levels lie very close to the 95, 74, and 53% contours of the predicted SRF, respectively. For this illustrative example at least, the observed changes in SRF size and position with sound level are therefore consistent with a “linear threshold” model: at near threshold sound levels, much of the predicted SRF structure remains below the unit’s spiking threshold, but as sound levels increase, more and more of the linearly predicted SRF structure is revealed in the observed data.

Only a very small proportion of our sample exhibited large expansions in their SRF size or shifts in SRF centroid that were poorly predicted by the model. The large majority of cells behaved rather like the example shown in Fig. 3. Of the cells in our sample 87% showed a clear expansion of their SRFs with increasing sound level. This expansion was typically accompanied by comparatively modest shifts in the SRF centroid, usually in the upward direction, that were broadly compatible with the linearly predicted receptive field structure. (Also compare Fig. 5 below.)

Figure 4 summarizes, for our entire data set, the dependency of SRF 50% area, dispersion, and centroid length (see Methods) on sound level relative to unit threshold: 50% area and dispersion generally increased, and centroid length decreased, with increasing sound level. Note that 50% areas rarely exceeded the size of one hemifield (6.28 rad²), even at the highest sound level. Values as high as 10 rad² were highly unusual.

Figure 5 shows the changes in SRF centroid direction and length as a function of sound level for units of different binaural type. Centroid azimuth, elevation, and length are plotted against sound level above unit threshold in the 1st, 2nd, and 3rd row of plots, respectively. The 3 columns of plots show data for units with EO, EI, or EE binaural response properties, respectively. In each plot, data points obtained from the same unit are connected by lines. These data suggest that centroids of EO and EI units tend to have long centroid vectors (3rd row), indicative of sharp spatial tuning. (Indeed at sound levels of 5–10 dB above unit threshold, their centroid vector lengths frequently approached the 0.75 “ceiling” value attained when the unit responds to only one of the 224 tested sound locations.) At these low sound levels, the centroids of EI and EO units invariably pointed toward azimuths near −50° (1st row) and elevations just below 0° (2nd row). With increasing sound level, these centroid vectors became shorter and their locations became more variable, but tended to shift to higher elevations and slightly more contralateral azimuths. In all but 2 cases, the centroids remained within the contralateral side of space (negative azimuth values).

The effect of sound level on the centroids of EE units did not appear to follow the trends observed for EO and EI units. EE units had short centroids (i.e., broader spatial tuning), even at low sound levels. Their centroid azimuths were located close to or in front of either the contralateral or the ipsilateral interaural axis and did not shift systematically as a function of sound level. The centroid elevations of EE cells also lacked the upward trend with increasing sound level observed among EO cells.

To assess the statistical significance of these differences across binaural class, we performed t-test comparisons on the change in centroid length per dB increase in sound level for the data plotted in the bottom row of Fig. 5. These slopes did not differ significantly when EO and EI units were compared (P > 0.09), but the differences between the EO and EE cells were
highly significant ($P < 0.0002$). The changes in centroid elevation as a function of sound level also differed significantly between EO and both EE ($P < 0.00004$) and EI units ($P < 0.03$). No differences were found between the 3 binaural classes for changes in centroid azimuth with sound level.

If the spatial tuning of high-frequency A1 neurons does conform to the predictions of a linear threshold model, then some of the trends illustrated for EO and EI units in Fig. 5, like the clustering of near-threshold SRF centroids around azimuths close to $-50^\circ$ and elevations just below the horizon and the tendency for centroids to shift up and slightly backward with increasing sound level, should be directly attributable to the directional acoustic properties of the contralateral ear, which provides the predominant excitatory input for these cells.

Figure 6 shows the overall gain in sound level produced by the outer ear DTF for broadband (8–22 kHz) signals. Data are...
shown for the right (contralateral) ear of 2 of the 5 ferrets used in this study (Fig. 6, A and B). The DTFs for the other 3 animals were similar. The DTFs have their maximum gain at elevations just below 0° and at azimuths between −50 and −60°. The sharply tuned SRFs observed for both EO and EI units at near-threshold sound levels (Fig. 5) were centered on this maximum gain direction, the “acoustic axis” of the pinna. As sound levels increase, more and more of the DTF filtered stimuli are thought to exceed unit threshold and become audible to the unit. If we assume that the units respond at these higher sound levels in a threshold-linear fashion, then we can attempt to predict the likely position of unit SRF centroids by “thresholding” the DTFs at different levels below the maximum gain and calculating centroid vectors for the thresholded DTFs (see METHODS).

The white triangle, circle, and square superimposed on the DTF plots in Fig. 6, A and B show acoustical centroid directions that were obtained in this way for DTFs thresholded at 0.5, 15.5, and 30.5 dB below the maximum gain, respectively. These thresholded DTF centroids show a similar trend to that observed for the effects of sound level on the SRF centroids of EO and EI cells. Raising the level (or equivalently, for the DTFs, lowering the threshold) resulted in higher centroid elevations and small shifts in centroid azimuths toward the contralateral side. Because of their frequency tuning, A1 units will presumably sample acoustic energy over a relatively narrow frequency range, whereas the DTF shown in Fig. 6 is based on broadband signals. Nevertheless, the correlation between the neuronal and acoustic centroid directions suggests that the SRFs of EO and EI units are shaped by the directionality of the contralateral external ear. This conclusion is further supported by the fact that normalizing and averaging unit SRFs for these binaural response classes (Fig. 6, D and E) produces a function that closely resembles the broadband DTF energy distribution. The average SRF of the EE units, perhaps unsurprisingly, matches the acoustic energy distribution from the contralateral ear less well (Fig. 6F).

**Predicting SRF changes with sound level**

Figure 7 shows predicted changes in SRF centroid direction and length with level. Centroids were calculated at different levels below the predicted maximal response. These plots are equivalent, and should be compared with, those obtained from observed SRFs in Fig. 5. For all 3 binaural classes, the length of the centroid vector tended to diminish and its direction to shift upward as predicted SRFs were thresholded at progressively lower values below the response maximum (i.e., with increasing area of the predicted SRFs). For EO and EI units the predicted changes in SRF centroid length and position closely resembled the actual changes observed in Fig. 5. However, for EE units, observed SRF centroid trends differed from those predicted by the linear model. In particular, the centroid vector lengths of observed EE-type SRFs were always short (implying broad spatial tuning even at low sound levels), even though the linear threshold models predicted that these centroid vectors, like those of EO and EI units, should be long at near threshold levels. The predictions for centroid directions in both azimuth and elevation and how they change with increasing level were also generally poor for EE cells. Thus threshold linear modeling appears to generate more accurate predictions for EO and

---

**Fig. 6.** Correlation of directional transfer function (DTF) energy and average A1 population activity. A and B: broadband (8–22 kHz) DTF energy for the right (contralateral) ears of 2 ferrets used in this study. Contour lines are drawn for every 5 dB, starting at 0.5 dB below the maximum gain. White triangle, circle, and square superimposed on the DTF plots show the centroid directions obtained for the DTF energy function thresholded at 0.5, 15.5, and 30.5 dB below the maximum, respectively. C: average normalized SRF calculated over the entire data set. D–F: average normalized SRFs for EO, EI, and EE cells, respectively. Contour lines in normalized response steps of 0.2 from 0.15 to 0.95.
EI units over a range of sound levels than for EE units, which suggests that the latter class may exhibit greater nonlinearities in their responses.

Temporal coding

Previous studies have suggested that auditory cortical neurons may encode spatial information not only by variations in firing rate with stimulus position but also by the temporal pattern of spike discharges (Brugge et al. 1996; Middlebrooks et al. 1998).

As described above, the majority of A1 units recorded in the present study responded to VAS stimuli with a single brief burst of spikes. As previously attempted in the cat by Brugge et al. (1996), we examined these responses for possible systematic changes in response latency. The first spike latency metric used by Brugge and colleagues did not seem the most appropriate metric to use because we commonly observed considerable trial-to-trial variability in the response. Figure 8 illustrates this with responses of 3 units as a function of sound azimuth in the form of raster plots. The example shown in Fig. 8A was not spontaneously active and responded to the noise stimuli with a brief burst of spikes. There is some indication that the onset latency for this unit increased as the stimulus was moved away from the center of the SRF. However, the response latency in each of the 5 stimulus repetitions was quite variable. In addition, response failures were common, particularly near the edges of the SRF. For many of our units, response variability and spontaneous activity made it impossible to decide by visual inspection of the raster plots whether response latency varied systematically as a function of sound-source direction. This is true for the units whose responses are shown in Fig. 8, B and C. In neither case was there an obvious latency gradient across the SRF, although the further analysis described below revealed that one of the units showed an appreciable dependency of response latency on sound azimuth (Fig. 8B), whereas the other did not (Fig. 8C).

To avoid latency estimates being confounded by relatively high spontaneous discharge rates we excluded from the following analysis all units in which the initial peak discharge rate (as seen in the pooled PSTHs) did not exceed the spontaneous discharge rate by a factor of at least 10, which left 97 (out of 127) units. Instead of first spike latency, we used a mean spike time metric to estimate the latency of this response, where the times of all spikes occurring 10–40 ms after stimulus presentation were averaged.
To obtain more reliable response latency estimates, responses from 5 neighboring stimulus directions were pooled to yield 25 response pools, as shown in Fig. 9A. For each pool containing more than 3 spikes, we calculated and analyzed the mean spike time and the total spike count. For 39/97 units, there was a significant inverse relationship between these parameters (Fig. 10). Examples of such cases are shown in Fig. 9, B and C, which represent the units whose raw responses are shown in Fig. 8, A and B, respectively. These units exhibit a trend commonly observed elsewhere in the auditory system, that is, the highest spike counts were associated with the shortest response latencies (Heil and Irvine 1996). The majority of units (58/97), however, showed no significant relationship between response strength and response latency (Fig. 10). Examples of such cases are shown in Fig. 9, D (same unit as in Fig. 8C) and E. For these units response latencies were found to be relatively constant across space. Only 7 units behaved like the one seen in Fig. 9F, showing substantial changes in response latency with sound-source direction that were independent of spike count.

Irrespective of whether systematic changes in response latency constitute a neural code per se, it is important to appreciate that they are likely to play a major role in shaping the dynamic properties of an SRF. A simple way of revealing these dynamic properties is to construct SRF maps from spike counts that are constrained such that only spikes within certain latency ranges are included. For the data shown in Fig. 9B, for example, we might expect the shortest latency spikes to form a fairly sharply tuned SRF centered on $-50^\circ$ azimuth, whereas later spikes would form a ring around the early SRF. This is confirmed in Fig. 11A, which shows how the SRF for this unit changes in size and shape when different response windows are selected. For the unit in Fig. 9F, in contrast, one might expect to see a contralateral SRF for the earliest part of the response.

**FIG. 9.** Comparison of latency and spike counts for sound source location. A: pooling of data for statistical analysis. Dots indicate the virtual sound directions sampled during SRF recording. Circles indicate how source directions were pooled for response onset latency analysis. Note that adjacent response pools overlap in azimuth. B–F: response onset latency analysis for 5 different units. Left panels in each pair show poststimulus-time-by-sound-azimuth histograms for the spike data pools centered at 0° elevation. Superimposed on the histograms is a plot of mean spike time vs. azimuth for each pool (connected circles). Right panels plot mean spike time vs. spike count for all data pools. Also shown are the number of points in the regression ($N$), the correlation coefficients ($r$), and their significance level ($P$). Binaural response types and BFs for the units shown are B: EO, 11 kHz; C: EI, 8 kHz; D: EE, 8 kHz; E: EO, 7 kHz; F: EE, 29 kHz.

**FIG. 10.** Histogram of sample correlation coefficients for response latency vs. response spike count. Portion of histogram bars colored in black corresponds to units with a value of $r$ that is statistically significant at the 1% level.
and an ipsilateral SRF for the later spikes. Again, this is confirmed in Fig. 11B.

Differences in the spatial selectivity of onset and late components of a neuron’s response have also been observed by measuring ILD functions in the cat midbrain (Heil 1998) and could be related to neuronal selectivity for stimulus motion (Jenison et al. 2001). For example, the unit whose SRF is shown in Fig. 11B would be expected to show maximal discharge probability when a stimulus is present first in the contralateral and then in the ipsilateral hemifield. However, the predicted optimal stimulus velocity for this unit would be substantial, in the order of 2,000° s⁻¹ [for comparison, ferret head-orienting movements have angular velocities in the range of 500–1,000° s⁻¹ (C. H. Parsons, O. Kacelnik, and A. J. King, unpublished observation)]. Alternatively, selectivity for lower, ecologically more plausible velocities could be achieved by
units that exhibit secondary peaks in their PSTH, which typically occurred >100 ms after the initial onset response, provided the later peaks display spatial tuning that differs from that of the initial response peak. However, we frequently observed that, whereas the SRFs for these later peaks might be more or less broadly tuned than those for the initial response, they tended to overlap substantially or entirely with the initial response SRF (Fig. 11, C and D). Another possibility is that multiple peaks in PSTH could be useful for processing reverberant stimuli.

Putative latency codes are not predicted by the linear model

Because the strength of cortical responses can often be successfully predicted from the units’ FTRF it is pertinent to ask whether the putative response timing codes just described are also predictable. Because the linear prediction is effectively based on a weighted summing of contributions from the neuron’s various aural and frequency channels, this predicted response will exhibit substantial changes in response latency only if corresponding relative delays are manifest across the different frequency or aural channels of the FTRF. However, the binaural FTRFs in our sample only very rarely exhibited excitatory features with substantially different latencies (see Fig. 1). Consequently, it is not surprising that an analysis of the SRFs predicted by our binaural linear filter model generally failed to predict the systematic changes in response latency across the field described above (data not shown).

Discussion

In this study, we have used VAS stimuli to map the SRFs of high-BF single units in ferret A1. Although these SRFs exhibited considerable diversity, the majority of units were, in agreement with previous reports, tuned to a broad region of space within the contralateral hemifield. For example, at near-threshold sound levels, high-frequency neurons in cat A1 tend to be tuned to directions in space that correspond to the acoustic axis of the pinna (Brugge et al. 1994; Middlebrooks and Pettigrew 1981; Rajan et al. 1990). Using VAS stimuli, Brugge et al. (1996) found that, under monaural stimulation, the shape of the SRF is determined by the spatial distribution of sound pressure produced by the spectral filtering of the external ear. This is consistent with our previous finding that, for most ferret A1 neurons, SRF shape can be predicted from their binaural spectrotemporal profiles and the filtering properties of the head and external ears (Schnupp et al. 2001). However, because the match between the linearly predicted and observed SRFs varied from unit to unit, it was important to investigate whether any departure from linearity can be explained by the unit’s binaural response properties or sound level.

Binaural response properties

Because of time constraints during the experiments we based our classification into binaural response classes purely on the binaural FTRF data for each unit. The observed distribution of contralateral/ipsilateral response types (EO, 40%; EE, 32%; EI, 27%) is nevertheless in good agreement with previous reports of binaural classes in ferret A1 (Kelly and Judge 1994). The FTRFs, being linear descriptors of the behavior of the neuron, do not readily reveal any nonlinear facilitatory or inhibitory binaural interactions. However, evidence for such interactions has led other authors to further subdivide the binaural response classes of A1 neurons (Irvine et al. 1996; Reale and Kettner 1986; Rutkowski et al. 2000; Zhang et al. 2004). The large proportion of units classified as EO in this study should therefore not be interpreted to mean that the responses of these neurons are based entirely on monaural inputs. In fact, recent research from our laboratory (Campbell et al. 2003) and others (Zhang et al. 2004) suggests that truly monaural responses may not occur at all in A1. Thus the units classified here as EO would probably exhibit nonlinear influences from the ipsilateral ear if their binaural interactions had been studied in more detail. Nevertheless, linear predictions (which for EO units are effectively based on the contralateral ear alone) tended to produce good approximations for the SRFs of these cells. In fact, no significant differences were found among the 3 binaural classes when we examined the correlation between the observed and predicted SRFs.

Effects of sound level

We also investigated how accurately our linear model predicted sound-level–dependent changes in the SRFs. As in other species (Brugge et al. 1994, 1996; Imig et al. 1990; Middlebrooks and Pettigrew 1981; Rajan et al. 1990), we found that the SRFs of ferret A1 units were prone to expand, sometimes quite considerably, with increasing sound level. The centroids of the majority of SRFs in our study nevertheless remained close to the acoustic axis in the contralateral hemifield over a range of sound levels (±30–40 dB above threshold in many cases). Although the SRF centroids of EO and EI cells tended to shift somewhat with increasing sound level, the direction of this shift correlated with the centroid shift of the contralateral broadband DTF at different values below the maximum gain. This suggests that it is often possible to account for shifts in the SRF centroid position by modeling A1 neurons as linear threshold devices. In other words, as the sound level increases, progressively more of the linearly predicted SRF exceeds the unit’s firing threshold and the centroid shifts in the direction of the shallowest slope of the DTF.

The observations made in Figs. 3, 5, and 7 suggest that this is a reasonable assumption for cells that exhibited EI-type or EO-type binaural FTRFs, but less so for EE-type cells. Thus the model accurately predicted level-dependent changes in the SRFs of units that are dominated by excitation from the contralateral ear only, but fared less well for units receiving excitation from both ears. This suggests that A1 neurons receiving EE inputs can exhibit more substantial nonlinearities in their binaural response properties than their EO and EI counterparts. Given the nonlinear binaural interactions reported for cells that receive excitatory inputs from both ears throughout the auditory pathway (Goldberg and Brown 1969; Grotte and Park 1998; Irvine et al. 1996; Kувада and Yin 1983; Rutkowski et al. 2000; Spitzer and Semple 1995; Yin and Chan 1990), significant departures from a largely linear model might not be unexpected for these cells. In fact, one might wonder why the behavior of EE cells did not depart more dramatically from the predictions of the linear model; Fig. 2, C and D show that linear predictions for EE cells were often quite good. However, EE cells are often associated with sensitivity
to ITDs based on a highly nonlinear coincidence detection mechanism. Although there is psychophysical evidence that ITD detection can occur at the relatively high frequencies that the neurons in this study were tuned to (Henning 1974), ITD sensitivity declines at high frequencies (Macpherson and Middlebrooks 2002). Consequently, it may well be that the SRFs of EE units with BFs of only a few hundred Hz or less would be much less well predicted from the unit’s FTRF.

**Representation of space in A1**

Although A1 neurons tend to exhibit sharp spatial tuning near threshold, their SRF centroids are usually located near the acoustic axis, so that only a subset of possible directions are represented. At higher sound levels, spatial selectivity becomes less specific because neurons start responding to a progressively greater region of space, and the directions of peak activity often shift. At these suprathreshold sound levels, binaural (EI) interactions can restrict the SRFs to a single hemifield (Brugge et al. 1994). Nevertheless, this expansion in SRF size contrasts with the relative constancy of psychophysical performance over a wide range of sound levels (Macpherson and Middlebrooks 2000).

In carnivores and primates, one of the clearest consequences of damage to or inactivation of auditory cortex is an impairment in the ability to localize sound (Heffner and Heffner 1990; Heffner and Masterton 1975; Jenkins and Masterton 1982; Jenkins and Merzenich 1984; Kavanagh and Kelly 1987; Smith et al. 2004). Although sometimes less pronounced, similar deficits have been described when A1 was specifically targeted (Jenkins and Merzenich 1984; Kavanagh and Kelly 1987; Smith et al. 2004). However, our experimental and modeling data imply that, for the majority of high BF A1 neurons, the SRFs arise from a simple linear interaction between the acoustical properties of the ears and the frequency sensitivity of the neurons. This is further supported by the changes in SRF structure that result when ferrets are presented with VAS stimuli based on measurements made from other animals of the same (Mrsic-Flogel et al. 2001; Schnupp et al. 2001) or different ages (Mrsic-Flogel et al. 2003). Under these conditions, the SRFs shift in ways that are directly related to intersubject differences in the acoustical cues and predictable from the linear filter model.

Because our data are based on recordings from barbiturate-anesthetized ferrets, we cannot rule out the possibility that the high incidence of linear processing of spatial cues may reflect the depressive effects of this anesthetic on both evoked and spontaneous neuronal activity. However, it has been reported that the linearity of FTRFs in ferret A1 is the same under barbiturate and ketamine anesthesia (Kohn et al. 1996). Moreover, although SRFs recorded in A1 of awake cats tend to be more sharply tuned and less dependent on sound level than those recorded in anesthetized animals, they are still broad, often encompassing an entire hemifield (Mickey and Middlebrooks 2003). Although it is possible that nonlinear processing may be more evident in the awake animal, our findings indicate that differences exist among A1 neurons and that those differences may be related to their binaural response properties.

Together, these results suggest that further processing of spatial information, beyond that taking place at subcortical levels, is either a function performed by the minority of A1 neurons whose SRFs are poorly approximated by linear modeling or that the role of A1 could be that of a general purpose gateway to more specialized cortical areas (Griffiths et al. 2004). Although neurons devoted exclusively to spatial or other forms of auditory processing have not been found, there is growing evidence that certain higher cortical areas may play a more specialized role in spatial hearing (Recanzone et al. 2000; Stecker et al. 2003; Tian et al. 2001).

**Temporal coding**

Because of the size and level dependency of SRFs based on spike count, alternative ways of encoding sound location have been proposed. Neurons in auditory cortex tend to respond to brief noise bursts from varying positions in space with relatively brief bursts of spikes, and it is thought that the latency of these bursts may form an important part of the cortical neural code for sound location. Direction-dependent variations in first spike latency have been reported in several areas of the cat’s auditory cortex (Brugge et al. 1996; Furukawa and Middlebrooks 2002; Stecker et al. 2003). Because these latency gradients can be maintained over a range of sound levels, it has been suggested that spike timing may provide a level-independent way of encoding stimulus direction (Reale et al. 2003). Moreover, studies based on maximum likelihood estimation methods suggest that latency gradients within ensembles of A1 units could provide sufficient information to code for sound-source direction and account for the acuity of spatial localization in both cats and humans (Jenison 1998). It has been proposed that, at least in some cortical areas, other aspects of the temporal discharge pattern may also contribute to the cortical coding of auditory space (Middlebrooks et al. 1998). Although it remains uncertain how temporal information is decoded at higher levels, these reports clearly indicate that spike timing could make a substantial contribution to spatial coding in the cortex.

Our data from ferret A1 are broadly consistent with this notion, in that we often observed a variation in the latency of the onset response across the SRF. About 40% of our sample showed an inverse relationship between spike count and onset latency, with the strongest responses associated with the shortest latencies, a trend that is observed at many levels of the auditory system (Irvine 1992). We also found units with the potential to encode sound-source direction through spike latency independently of firing rate. However, these units represented a minority (<10%) of our sample, and a large proportion of cells showed no reliable relationship between stimulus location and response latency.

It is important to note that the binaural linear filter model, although often successful at predicting changes in spike counts, failed to predict the latency gradients exhibited by many of the A1 units recorded in this study. Although some temporal variation in the SRF is likely to stem from differences in sound level that are associated with different virtual directions (effective levels typically increase as one approaches the acoustic axis of the pinna), these “intensity cues” cannot explain the range of latencies observed. Indeed, by comparing the onset latencies of responses to the same virtual direction over a 30-dB range (this range is nearly the same as that encountered across the DTF), we found latency differences (≤5 ms) that were smaller than those across the SRF (>5 ms; Fig. 9).
Moreover, response latency changes could, in many cases, not be explained in terms of linear interactions between the spectrotemporal structure of the stimulus and the neuron’s FTRF. Consequently, these putative temporal codes for sound-source location must arise from the network properties of the auditory pathway. Whereas origins of response latency coding are already apparent in the auditory nerve (Heil and Irvine 1997), the mechanisms of its elaboration in the ascending auditory pathway remain unknown. Future studies will be needed to investigate the neural mechanisms that give rise to these putative temporal codes and to confirm their role in shaping auditory spatial perception.

Acknowledgments

We thank I. Nelken for insightful comments on our data.

Present address of T. D. Mrsic-Flogel: Max-Planck-Institute of Neurobiology, Am Klopferspitz 18a, 82152 Martinsried, Germany.

Grants

This work was supported by Defeating Deafness and the Biotechnology and Biological Sciences Research Council (Dunhill Medical Research Trust Fellowship and project grant No 43/S19595 to J.W.H. Schnupp) and the Wellcome Trust (Wellcome Prize Studentship to T. D. Mrsic-Flogel and Wellcome Senior Research Fellowship to A. J. King).

References


