Responses of Neurons in Primary Auditory Cortex (A1) to Pure Tones in the Halothane-Anesthetized Cat

Dina Moshitch,1 Liora Las,1 Nachum Ulavonsky,1 Omer Bar-Yosef,3 and Israel Nelken1,2
1Department of Neurobiology, The Alexander Silberman Institute of Life Sciences, 2Interdisciplinary Center for Neural Computation, and 3Department of Physiology, Hadassah Medical School, The Hebrew University, Jerusalem, Israel

Submitted 4 August 2005; accepted in final form 15 March 2006

Moshitch, Dina, Liora Las, Nachum Ulavonsky, Omer Bar-Yosef, and Israel Nelken. Responses of neurons in primary auditory cortex (A1) to pure tones in the halothane-anesthetized cat. J Neurophysiol 95: 3756–3769, 2006; doi:10.1152/jn.00822.2005. The responses of primary auditory cortex (A1) neurons to pure tones in anesthetized animals are usually described as having mostly narrow, unimodal frequency tuning and phasic responses. Thus A1 neurons are believed not to carry much information about pure tones beyond sound onset. In awake cats, however, tuning may be wider and responses may have substantially longer duration. Here we analyze frequency-response areas (FRAs) and temporal-response patterns of 1,828 units in A1 of halothane-anesthetized cats. Tuning was generally wide: the total bandwidth at 40 dB above threshold was 4 octaves on average. FRA shapes were highly variable and many were diffuse, not fitting into standard classification schemes. Analyzing the temporal patterns of the largest responses of each unit revealed that only 9% of the units had pure onset responses. About 40% of the units had sustained responses throughout stimulus duration (115 ms) and 13% of the units had significant and informative responses lasting 300 ms and more after stimulus offset. We conclude that under halothane anesthesia, neural responses show many of the characteristics of awake responses. Furthermore, A1 units maintain sensory information in their activity not only throughout sound presentation but also for hundreds of milliseconds after stimulus offset, thus possibly playing a role in sensory memory.

INTRODUCTION

The coding of frequency and level of pure tones in the auditory system has been intensively studied. Most pure-tone psychophysics can be accounted for by the response properties of subcortical neurons, from the auditory nerve and up to the inferior colliculus (IC) (Delgutte 1996; Ehret and Merzenich 1988; Moore 1982). In many respects, the representation of pure tones in primary auditory cortex (A1) does not seem to add much to the subcortical responses. For example, in A1 of rats under equithesin (pentobarbital/chloral hydrate) (Gaese and Ostwald 2001; Sally and Kelly 1988) and of cats under barbiturates (Brugge and Reale 1985), frequency-response areas (FRAs) have a rather uniform V-shape, although a minority of neurons have multipeaked FRAs (Sutter and Schreiner 1991). Most of the FRAs are sharply tuned (Schreiner and Sutter 1992), but not more than those of IC neurons (Ehret and Schreiner 1997). Furthermore, the neurons have phasic responses, consisting of a short burst of action potentials at a short latency after tone onset (Brugge et al. 1969; Eggermont 1991; Phillips and Sark 1991). These results suggest that information about pure tones is present in the firing rates of A1 neurons for only a few tens of milliseconds after tone onset, and that this information may be actually somewhat degraded relative to the information available in subcortical stations.

A number of studies addressed more complex characteristics of neuronal responses in A1. For example, deCharmes et al. (1996) suggested that correlation between neurons, rather than the absolute level of activity, carries information in auditory cortex about the presence of tones. However, other studies suggested that the main problem is actually technical—the use of anesthesia. Anesthesia strongly affects the responses of neurons in the central auditory pathway, from the dorsal cochlear nucleus (Young and Brownell 1976) to the auditory cortex (Gaese and Ostwald 2001; Sally and Kelly 1988; Schreiner and Sutter 1992; Sutter and Schreiner 1991, 1995). In awake animals frequency responses are more complex (Abeles and Goldstein 1972; deCharmes et al. 1998; Goldstein and Abeles 1975; Pelleg-Toiba and Wollberg 1989), a substantial number of them being multipeaked (Abeles and Goldstein 1972; Kadia and Wang 2003). The mean bandwidth of neurons in A1 of awake cats is about threefold larger than that reported under barbiturates anesthesia (Qin et al. 2003). Temporal response patterns in awake animals vary from phasic, as reported under barbiturates, to tonic, where the response is sustained with very little adaptation (Evans and Whitfield 1964; Frostig et al. 1983; Goldstein and Abeles 1975; Pfingst and O’Connor 1981; Qin et al. 2003; Recanzone et al. 2000; Shamma and Symms 1985; Wang et al. 2005).

In this study, we used a gas anesthetic, halothane. Although both gas anesthetics isoflurane and halothane have been previously used in cortex research, halothane shows weaker depressive effects in the primary visual cortex (Villeneuve and Casanova 2003) and a weaker suppressive effect on auditory-evoked responses (Antunes et al. 2003; Johnson and Taylor 1998; Villeneuve and Casanova 2003). The aim of this study is to characterize the responses of neurons in A1 of cats to pure tones under halothane anesthesia. We demonstrate a large variety of shapes of frequency-response areas and of temporal response patterns, which resembles in their richness the variety described in awake animals. Furthermore, we demonstrate that sensory information is present in the responses for hundreds of milliseconds after stimulus offset. This information is sufficient to fully identify pure-tone stimuli as long as they last.

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
and beyond their offset, and could be a correlate of sensory memory.

METHODOLOGY

Animal preparation

The data were collected from 27 healthy adult cats. The cats underwent a preliminary otoscopic examination to rule out external ear obstruction and middle ear infection. Surgical anesthesia was induced with xylazine [0.1 mg, administered intramuscularly (im)] followed by ketamine (100 mg, im). The cats received 0.1 mg of intramuscular atropine sulfate or atropine methyl nitrate. The radial vein was cannulated and the animals received a continuous infusion of lactated Ringer solution at a rate of 10 ml/h. Blood pressure and heart rate were continuously monitored with a cannula inserted into the femoral artery. Body temperature was kept at approximately 38°C using a heat pad. The trachea was cannulated and the cat received a mixture of oxygen and nitrous oxide (30/70%) with halothane (0.2–1.5%) for respiration. The halothane level was set so that mean blood pressure was about 100 mmHg. Breathing rate (set to about 30/min) and CO₂ levels (3–3.5%) were continuously monitored. Under these conditions, the cats did not have any paw-withdrawal and corneal reflexes and could usually be respirated without muscle relaxation. In case of respiratory resistance, muscle relaxation was induced with pancuronium bromide (0.05–0.2 mg given every 1–5 h, as needed) or vecuronium bromide (0.25 mg given every 0.5–2 h). The cats received 5 ml bicarbonate solution (8.4%) intravenously every 8–12 h, to control for the acidosis that always developed during the experiment. Experiments usually lasted between 48 and 84 h.

The temporal muscles were retracted to uncover the skull and the external auditory meati on both sides. The bullas were vented with a 30-cm-long polyethylene tube (PE90). The skull was opened above the middle ectosylvian gyrus. The dura was left intact. At the end of the experiments, the cats were killed with a lethal dose of pentobarbital (50–100 mg, intravenous) and perfused transcardially with saline followed by 500 ml of 4% formaldehyde. These methods were approved by the animal use and care committee of the Hebrew University–Hadassah Medical School.

The electrophysiological techniques and the acoustic stimulation are described in a previous paper (Bar-Yosef et al. 2002). Briefly, single neurons and multiunit clusters were recorded using two to four glass-coated tungsten electrodes simultaneously. The electrodes could be driven individually using either hydraulic drives (Kopf and Trent-Wells) or a four-motor drive (EPS, Alpha-Omega). The electrical activity was amplified (MCP8000, Alpha-Omega) and filtered between 200 Hz and 10 kHz.

Sounds were generated digitally on-line, transformed into analog voltage (TDT DA3-4), attenuated (TDT PA4), and switched with a linear ramp of 10 ms (TDT SW2). They were presented to the animal through a sealed, calibrated system (constructed by G. Sokolich). Acoustic calibration using pure tones was performed in each ear of each animal. Because calibration curves were rather flat (±10 dB over the range 100 Hz to 30 kHz in most animals) without fast peaks or notches, changes in levels were not corrected on-line. For off-line data analysis, decibel (dB) attenuation settings were translated into dB SPL by using the appropriate calibration value at the characteristic frequency of each unit. Typically, 0 dB attenuation corresponded to 100-dB SPL.

Experimental protocol

The microelectrodes were inserted into the mid- and low-frequency areas of A1 (the posterior half of the middle ectosylvian gyrus) in the left hemisphere as described by Reale and Imig (1980). Neuronal activity was identified on the basis of spontaneous activity or responses to tones and broadband noise (BBN) stimuli. Spikes were separated on-line using a spike sorter (MSD, Alpha Omega). The quality of spike separation was assessed on-line. The spike sorter detects candidate spikes by computing the sum of squared differences (distance) from an eight-point template and the sampled signal from the electrode, looking for local minima in this distance measure. A spike is detected when a local minimum is lower from a manually set threshold, indicating a close fit to the template. In addition, these distances are collected and displayed on-line as a histogram. The shape of the histogram of distances was used to quantify spike separation quality, three levels of which are used here. When the histogram had a peak followed by a clear deep trough (at least half the height of the peak), indicating the presence of a well-defined class of spike shapes, the unit was considered well separated (q1, 916/1,828 units). Histograms with shallower troughs were considered as nonseparated activity composed of large spikes (q2, 420/1,828 units). When the histogram had no trough at all the unit was considered as nonseparated multiunit activity (q3, 492/1,828 units). Because about half the units described here do not represent well-separated single neurons, we prefer to use the term “unit” instead of “neuron” throughout the paper, with the understanding that units in the q1 class do probably represent the responses of single neurons.

Each unit was characterized manually by determining approximately its characteristic frequency (CF) and its threshold to BBN. Next, the preferred aurality (ipsilateral, contralateral, or diotic) was determined using BBN rate-level functions to the left (ipsilateral) ear alone, to the right (contralateral) ear alone, and to both ears diotically.

Frequency-responsiveness area (FRA) was measured at the preferred aurality (left ear: 75/1,828 units, right ear: 655/1,828 units, diotic: 1,098/1,828 units) using a matrix of 40–45 frequencies logarithmically spaced from 100 to 40,000 Hz and 8–11 sound levels equally spaced between 99- and 12-dB attenuation. Tones were presented once at each combination of frequency and level. Tone duration was 115 ms with 10-ms linear rise/fall time, presented at a rate of 1/s. In some cases (637/1,828 units), a second tone at a fixed frequency was added 75 ms after stimulus onset, to check two-tone interactions (a two-tone paradigm). In that case, only the first 70 ms of the responses were analyzed. Although aurality may certainly change response strength in complex ways (e.g., Reale and Kettner 1986; Semple and Kitzes 1993a,b), reports on the effects of aurality on FRA shape show relatively small effects (e.g., Mendelson and Grasse 1992). Indeed, none of the parameters discussed in the following text showed any clear difference between the units tested monaurally and those tested diotically (see Table 2S in the supplementary materials1). We therefore analyzed the responses regardless of their aurality.

After the measurement of the FRA, the units were further studied using other stimuli not reported here.

Data analysis

Statistical tests are considered significant at the 0.05 level, unless explicitly stated otherwise. Stricter significance levels were used when appropriate to correct for multiple comparisons. Variability is always reported as mean ± SD.

The statistical significance of the responses to pure tones was quantified by a paired t-test (P < 0.05). When single tones were used, the counting window consisted of the full 115-ms stimulus window and these counts were compared with counts during the 115 ms just preceding stimulus onset. For units tested with the two-tone paradigm, only the initial 70-ms poststimulus onset were used as a counting window, and the counts were compared with counts during the 70 ms just preceding stimulus onset. Responses to all frequency and level combinations were pooled together for the purpose of this test. Because many combinations of frequency and level did not elicit a response but were nevertheless included in the test, the test is conservative.

1 The Supplementary Material for this article (a Table) is available online at http://jn.physiology.org/cgi/content/full/00822.2005/DC1.
FIG. 1. Quantifying spectral and temporal aspects of the responses. A: frequency-response area (FRA, spike counts during the first 70 ms after stimulus onset as a function of frequency and level) of a well-separated unit. Color-scale saturation: 35 spikes/s. Tuning curve (standard algorithm; see METHODS) is represented by the white line. Magenta diamond is at the characteristic frequency (CF, 6 kHz) and minimum threshold (MT, 32-dB SPL). Magenta vertical lines are the boundaries of the central lobe (Fcl: 1 kHz, Fch: 10 kHz). Magenta horizontal line represents the width of the central lobe at 40 dB above MT (Fch\textsubscript{25}: 2.6 kHz, Fch\textsubscript{40}: 9.5 kHz, BW\textsubscript{c40}: 1.9 octaves). Dark green horizontal line represents the total width of the FRA at 40 dB above MT (Fch\textsubscript{25}: 2.5 kHz, Fch\textsubscript{40}: 23.8 kHz, BW\textsubscript{c40}: 3.2 octaves). Normalized level response vector is plotted on the right of the FRA (MR: 0.95, see text for details). Black line above the FRA is the frequency-response vector, whereas the arrow at the left marks the level from which it was derived (MMI\textsubscript{x}: 6, MMI\textsubscript{x}: 2; see text for details). Black frame surrounds the frequency/level combinations that were useful for describing more fully the current data set.

A tuning curve (TC) was extracted from the smoothed FRA by the use of an automatic algorithm (e.g., Fig. 1A, white line). A frequency/level bin was considered as having a response when its spike rate was larger than the spontaneous response +0.2(maximal response – spontaneous response) (as used by Sutter and Schreiner 1991). For each frequency the algorithm determined sequentially (from high levels to low levels) those levels to which the unit responded according to this criterion. When the algorithm found a nonresponding bin it continued to check the responses at lower levels, down to 25 dB below the lowest level with detected response, to account for possible misses. The lowest level that was judged to elicit a response was selected as the threshold for that frequency. All FRAs were inspected visually to verify the plausibility of the resulting tuning curve. The algorithm gave satisfactory results in about 90% of the cases.

For the remainder of the units, we could use the less-satisfactory results to keep the definition of the tuning curve as objective as possible, or we could correct the tuning curves at the price of losing some of the advantages of a fully automatic algorithm. To strike a balance between the two possibilities, we used for the remainder of the units two versions of the same algorithm, varying only in the range of levels tested below a nonresponding bin. Whereas in the standard version of the algorithm, termed later tc-25, this range was 25 dB, in the other two versions it could be larger (40 dB, tc-40) or smaller (only 10 dB, tc-10). In these cases, the most satisfactory tuning curve (as judged visually) was selected for further processing.

### Derived properties

Several response properties of each unit (defined in Table 1) were extracted from the FRA (see also Fig. 1A). Most of these properties are standard and we closely followed the definitions of Sutter and Schreiner (Schreiner and Sutter 1992; Sutter 2000; Sutter and Schreiner 1991, 1995) and of Suga (Suga et al. 1997).

In addition, we defined a number of features of the neural responses that were useful for describing more fully the current data set.

#### COMPACTNESS. This parameter is defined as the area bounded by the TC (in terms of pixels) divided by the square of the length of the perimeter of the TC. The compactness was used to classify diffuse FRAs with many weak responses distributed over a large number of frequency/level combinations. Units with such FRAs could have highly significant responses, but their TCs had an irregular shape with possibly a large number of lobes, resulting in low compactness. In contrast, V-shaped FRAs (except for the few very narrow ones that had small area relative to their perimeter) had high compactness.

#### THE NORMALIZED LEVEL RESPONSE VECTOR. This parameter is an isofrequency cut through the FRA, calculated from the averaged response at the CF (determined from the central lobe; see Table 1 for definition) and the two adjacent frequencies. This rate-level function was then normalized by its maximum to give the normalized level response vector (black line to the right of Fig. 1A). The value of the normalized rate-level function at the second highest level tested (usually 80- to 90-dB SPL) was defined as the **monotonicity ratio**

<table>
<thead>
<tr>
<th>Table 1. Definitions of standard FRA properties</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minimum threshold (MT)</strong></td>
</tr>
<tr>
<td><strong>Characteristic frequency (CF)</strong></td>
</tr>
<tr>
<td><strong>Central lobe (CL)</strong></td>
</tr>
<tr>
<td><strong>F\textsubscript{x}</strong></td>
</tr>
<tr>
<td><strong>F\textsubscript{cl}</strong></td>
</tr>
<tr>
<td><strong>F\textsubscript{ch}</strong></td>
</tr>
<tr>
<td><strong>Bandwidth (BW\textsubscript{x})</strong></td>
</tr>
<tr>
<td><strong>Central lobe bandwidth (BW\textsubscript{cl})</strong></td>
</tr>
</tbody>
</table>

The subscript x indicates the level (10 or 40 dB above MT) at which these values were calculated.

J Neurophysiol • VOL 95 • JUNE 2006 • www.jn.org
(MR; Sutter and Schreiner 1995 used 80-dB SPL as their fixed reference level). This choice was explained by the fact that at the highest level tested, most units showed a significant reduction in activity. Using the highest level tested for defining nonmonotonicity would have made almost all units nonmonotonic, reducing the utility of this measure.

THE FREQUENCY-RESPONSE VECTOR. This parameter is an isolevel cut through the FRA, taken at the level at which the maximum number of spikes was elicited among all of the frequency/level combinations. The frequency-response vector was used to analyze potential multimodality of the responses. For that purpose, it was trimmed on both the high-frequency and low-frequency sides so that only the part of the function inside the TC was left (black line on top of the FRA in Fig. 1A). It was then transformed into the Fourier domain. For a unimodal frequency response, it was expected that the first Fourier component (at one cycle per length of the frequency-response vector) will have the largest intensity. For multimodal frequency responses, the Fourier components at two and three cycles would contribute a substantial amount of variance to the frequency-response vector. We have therefore defined the multimodality index (MMI) as the absolute values of the Fourier components at two and three cycles, normalized by dividing them by the absolute value of the Fourier component at one cycle of the frequency-response vector and expressed in dB (MMI₂ and MMI₃, respectively; a dB scale was used because these ratios spanned many orders of magnitude).

MUTUAL INFORMATION. We estimated the mutual information (MI) from the joint distribution of stimuli and responses. The stimuli were arranged first by levels and then by frequencies. Responses consisted of the number of spikes during the initial 70 ms of the stimulus. The joint distribution matrix of these two sets was remarkably sparse, since we had only one repetition per stimulus, resulting in high values of bias in the estimation of MI. We therefore used an adaptive method (Nelken et al. 2005) to estimate the MI of these two sets. First, the joint distribution matrix was used to estimate the raw MI and the bias as follows

\[
\text{raw MI} = \sum p(s, r) \log \frac{p(s, r)}{p(s)p(r)}
\]

\[
\text{bias} = \frac{(N_s - 1)(N_r - 1)}{2N \log 2}
\]

where \(N_s\) is the size of the stimulus set (\(N_s\) = number of frequencies × number of levels), \(N_r\) is the size of the response set (\(N_r\) = spike count), and \(N\) is the total number of stimuli (initially set equal to \(N_s\)). It is natural to assume that similar stimuli (with nearby frequencies and levels) elicited similar responses, and accordingly have similar conditional response distributions. Similarly, it is natural to assume that similar responses were distributed similarly across stimuli. Under these assumptions, the joint distribution matrix could be iteratively reduced by joining neighboring rows or columns, reducing \(N_s\) or \(N_r\), respectively. Specifically, at each iteration the stimulus class or the response class that had the lowest marginal probability was joined to that of its two immediate neighbors that had the smaller marginal probability. The raw MI and the bias were calculated as before for each of the reduced matrices. This reduction process generated a sequence of decreasing raw MI values (because joining two rows or columns of a joint-distribution matrix reduces the MI by the information-processing inequality; Cover and Thomas 1991) and a sequence of decreasing bias values (because the bias is roughly proportional to the number of bins in the joint distribution matrix). However, in the initial stages of the process the bias decreased much faster than the raw MI. Therefore the MI was estimated by the largest difference between the raw MI and bias. In this way, we performed explicitly the trade-off between high raw MI, which required a detailed view of the joint distribution matrix (which our sampling was unable to give), and low bias, which required a highly reduced joint distribution matrix.

Because of the severe undersampling of the original joint distribution matrix, simulations were performed to check how much MI was likely to be recovered from the responses. For the simulations, the measured FRAs (after smoothing) were used as the expectation values for Poisson generation of spike counts. For these models, it was possible to calculate the true MI (tMI). One random deviate with the appropriate expectation was generated from each frequency/level bin to generate a simulated FRA, and the MI was computed using the adaptive procedure described above. This procedure was repeated 10 times for each model FRA. The mean estimated MI was termed the simulated MI (sMI).

RESPONSE PATTERNS IN TIME. Only units that were presented with single tones and that had a globally significant responses (951/1,828 units) were analyzed for their response patterns in time. Because every frequency/level combination was repeated only once, it was not possible to build a peristimulus time histogram (PSTH) from the responses to a single stimulus. Nevertheless, to gain insight into the temporal response patterns, a PSTH was built from a set of ≤20 frequency/level combinations that was selected around the combinations that gave rise to the largest response. This block had to be contiguous and all responses in it were larger than half the strongest response. The frequency/level combinations were further limited so that all levels were not 10 dB below and not 20 dB above the level that evoked the largest response, and all frequencies were within half an octave of the frequency that elicited the largest response. This set of frequency/level combinations is named the core FRA later in the paper (encapsulated with black frame in Fig. 1A). Units that had fewer than five bins in the core were not included in the analysis (74/951 units).

The 410 ms after stimulus onset were divided into nine nonoverlapping time windows, the first seven with duration of 30 ms and the eighth and ninth windows with duration of 100 ms (Fig. 1B). Windows 1–4 occurred during stimulus presentation. The significance of the response at each time window was calculated by comparing it with the Poisson distribution whose expectation was equal to the mean number of spontaneous counts in the same window duration. Responses with \(P < 5 \times 10^{-5}\) were considered significant. This conservative cutoff point was used to avoid an excessive number of false alarms, given the extremely high number of comparisons involved in this analysis. Furthermore, the histogram of all individual \(P\) values for all time windows and units had a clear local minimum at this value. Twenty neurons had a globally significant response (tested over the whole duration of the tone stimuli) but their response was not significant in any of the individual time windows. These neurons were removed from the analysis of the temporal response pattern. The MI between the spike count and the stimuli was estimated for each time window as described above.

RESULTS

Overview

We started by studying parameters that describe the shape of the tuning curve. In addition to standard descriptors such as CF, minimum threshold, \(BW_{f0}\) and \(BW_{lp}\), we use here a new parameter, the compactness of the FRA. These parameters had wide distributions with low correlations between them, indicating the presence of a large variety of shapes of tuning curves.

Next, we studied the internal structure of the FRAs. We analyzed two sections of the FRA: an isofrequency section (the level function) and an isolevel section (the frequency-response vector). Whereas tuning curves had a very large variety of shapes, the internal structure of the FRAs was relatively...
TABLE 2. Significant responses as a function of separation quality

<table>
<thead>
<tr>
<th>Separation Quality</th>
<th>q1</th>
<th>q2</th>
<th>q3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsignificant</td>
<td>239 (26%)</td>
<td>96 (23%)</td>
<td>110 (22%)</td>
<td>445 (24%)</td>
</tr>
<tr>
<td>Significant</td>
<td>677 (74%)</td>
<td>324 (77%)</td>
<td>382 (78%)</td>
<td>1,383 (76%)</td>
</tr>
</tbody>
</table>

Values are means ± SD. Percentages are taken for each spike separation class separately.

uniform. Level functions had typically a single maximum separating an increasing from a decreasing limb, with a continuous range of nonmonotonicity. Frequency-response vectors were mostly unimodal or relatively flat.

Having quantified the shape and structure of FRAs, we asked how much information the units carry about the stimulus. This information was quantified by the mutual information (MI) between the set of frequency/level combinations used here and the neuronal responses. We found low correlations between the shape parameters of the FRA and the MI. In fact, the main determinant of the MI was the firing rate, independently of other features of the FRA.

Finally, we analyzed temporal response patterns to see how information about the stimulus develops with time during and after the stimulus. We found many units that had sustained, informative responses lasting throughout the stimulus and 300 ms or more after its offset. Tuning curves tended to be more compact early in the response and lost compactness later. However, it was still the firing rate that was the most important determinant of the amount of information carried by the responses.

In the remainder of this section we describe these key results in detail.

**Basic response properties**

The results are based on 1,828 units from 27 cats. Of these, 1,383 (76%) units responded significantly when tested over the whole duration of the tone stimuli. Units that did not respond significantly were not included in the population analysis. The distribution of significant responses among the three spike separation quality classes is summarized in Table 2. These proportions did not depend on the separation quality class ($\chi^2 = 3.1$, df = 2, n.s.).

The analysis of FRA shapes is based on the responses during the initial 70 ms of the stimulus. Of 1,383 units, 1,334 units responded significantly during this interval. The other 49 units had significant responses when tested over all 115 ms of the stimulus, but nonsignificant responses during the initial 70 ms of the responses. Because they represent a small class of units with weak responses, they are described only in the Temporal response patterns section.

**Characteristic frequency.** Of 1,334 units, 1,102 (83%) had a single CF, 192 (14%) had two CFs, and 40 (3%) had three CFs or more. The characteristic frequency of the units with a single CF had a geometric mean of 6 kHz and SD of 1.34 octaves. This reflects our tendency to record from the low- and midfrequency area of A1 (see METHODS).

There was no significant difference in the mean values between the spike separation quality classes for either the number of CFs or the mean of the CF of units with a single CF (see Table 3 for this and all other comparisons of parameter means across spike separation quality classes).

**Thresholds.** Figure 2, A–C presents FRAs of three units with different thresholds. Overall, thresholds had a mean of 40 ± 24-dB SPL, with 25% of the units having thresholds <24-dB SPL (Fig. 2D). There was a significant difference in the mean thresholds of the spike separation quality classes (Table 3). Post hoc comparisons showed somewhat lower thresholds of well-separated units compared with the thresholds of MUA (means of 38- and 43-dB SPL, respectively).

**The tuning bandwidth.** Figure 2, E and H presents four examples of units with different bandwidths ranging from narrow (Fig. 2E) to very wide (Fig. 2H). We quantified FRA width by the bandwidth of the tuning curve at both 10 and 40 dB above threshold, expressed in octaves (as in Schreiner and Sutter 1992). Neither of these bandwidths was correlated with the CF (see Supplementary materials). To some extent, this lack of correlation could be attributable to the limited range of CFs in the sample.

The distributions of bandwidths at 10 and 40 dB above threshold are displayed in Fig. 2, I and J, respectively. The

TABLE 3. Statistical test for equality of means across the spike separation quality classes

<table>
<thead>
<tr>
<th>Spike Separation Quality</th>
<th>q1</th>
<th>q2</th>
<th>q3</th>
<th>One-Way ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of CFs</td>
<td>1.2 ± 0.5</td>
<td>1.2 ± 0.48</td>
<td>1.22 ± 0.54</td>
<td>F(2,1331) = 0.19, P = 0.83</td>
</tr>
<tr>
<td>CF (kHz ± octaves)</td>
<td>5.96 ± 1.33</td>
<td>6.38 ± 1.3</td>
<td>5.87 ± 1.39</td>
<td>F(2,1099) = 0.6, P = 0.54</td>
</tr>
<tr>
<td>Threshold (dB SPL)</td>
<td>38 ± 21.34</td>
<td>39.88 ± 22.46</td>
<td>43 ± 23.10</td>
<td>F(2,1331) = 8.43, P &lt; 0.001</td>
</tr>
<tr>
<td>BW10 octaves</td>
<td>1.24 ± 1.3</td>
<td>1.13 ± 1.34</td>
<td>1.32 ± 1.53</td>
<td>F(2,1331) = 1.65, P = 0.19</td>
</tr>
<tr>
<td>BW40 octaves</td>
<td>3.86 ± 2.2</td>
<td>4.1 ± 2.8</td>
<td>4 ± 2</td>
<td>F(2,1192) = 1.65, P = 0.19</td>
</tr>
<tr>
<td>BW40 octaves</td>
<td>1.72 ± 1.15</td>
<td>1.7 ± 1.28</td>
<td>1.75 ± 1.18</td>
<td>F(2,1192) = 0.12, P = 0.88</td>
</tr>
<tr>
<td>Compactness</td>
<td>0.035 ± 0.019</td>
<td>0.033 ± 0.015</td>
<td>0.032 ± 0.016</td>
<td>F(2,1331) = 5.2, P = 0.005</td>
</tr>
<tr>
<td>MR</td>
<td>0.78 ± 0.24</td>
<td>0.82 ± 0.2</td>
<td>0.83 ± 0.2</td>
<td>F(2,359) = 3.59, P = 0.028</td>
</tr>
<tr>
<td>MM1</td>
<td>-8.83 ± 23.19</td>
<td>-5.3 ± 11.22</td>
<td>-8.5 ± 17.12</td>
<td>F(2,657) = 1.69, P = 0.18</td>
</tr>
<tr>
<td>MM1</td>
<td>-18.79 ± 35.36</td>
<td>-12.44 ± 19</td>
<td>-15.2 ± 23.93</td>
<td>F(2,657) = 2.5, P = 0.07</td>
</tr>
<tr>
<td>Rate within TC, spikes/s</td>
<td>22 ± 18</td>
<td>26 ± 17.8</td>
<td>30 ± 22</td>
<td>F(2,1331) = 18.6, P &lt; 0.001</td>
</tr>
<tr>
<td>Rate outside TC, spikes/s</td>
<td>3.4 ± 4.27</td>
<td>4.5 ± 5.3</td>
<td>6 ± 6</td>
<td>F(2,1331) = 32.7, P &lt; 0.001</td>
</tr>
<tr>
<td>Maximal rate, spikes/s</td>
<td>80 ± 48.7</td>
<td>96.6 ± 54.88</td>
<td>103 ± 63</td>
<td>F(2,1331) = 23.9, P &lt; 0.001</td>
</tr>
<tr>
<td>FRA Core rate, spikes/s</td>
<td>40 ± 30.81</td>
<td>44 ± 25.96</td>
<td>47 ± 35</td>
<td>F(2,874) = 4.23, P = 0.015</td>
</tr>
</tbody>
</table>

Values are means ± SD.
the BW40 distribution had a mean of 4.2 ± 1.4 octaves, whereas the BW40 distribution had a mean of 4 ± 2 octaves. There was a wide distribution of values of BW40, with 25% of the units having a value >5.4 octaves, a rather wide tuning (as in Fig. 2H), with BW40 >6 octaves), whereas only about 6% of the units had BW40 <1 octave. The presence of very wide FRAs led us to analyze separately the bandwidth of the central lobe of the FRAs (BWc; see Table 1) as well. The BW10 and BWc40 had means of 0.64 ± 0.57 and 1.73 ± 1.19 octaves, respectively. Figure 2K shows the distribution of BWc40.

COMPACTNESS. The overall shape of the FRA has been quantified in a number of ways in the literature (Sutter and Schreiner 1991). In our data, categorization into unimodal versus multimodal FRAs, or simple measures of bandwidth as described above turned out in many cases not to be very informative. This was ascribed to the presence of many units with rather diffuse FRAs having responses distributed over a large number of frequency/level combinations, without a clear border between response and no-response regions (Fig. 3A). Many of these units nevertheless had a highly significant mean response to sounds. Such units had highly irregular tuning curves having multiple lobes according to our criteria. However, describing these units as multilobed seemed to miss a crucial difference between them and the units that had a more compact FRA. To quantify this difference, compactness was defined as the area bounded by the TC divided by the square of its length (to have a dimensionless number; see METHODS).

With this measure, wideband units with diffuse FRAs had low compactness (Fig. 3A). Very narrowly tuned units could also have low compactness as a result of their very small area relative to their perimeter, although they were rather rare (only 6% of the units had BW40 <1 octave). The distribution of compactness was unimodal but rather wide (Fig. 3D) with a mean of 0.034 ± 0.017. About 20% of the units had a very diffuse shape with low compactness (0–0.02, Fig. 3A), 45% had two to four major lobes with medium compactness (0.02–0.036, Fig. 3B), and 35% had a single lobe, resulting in high compactness (0.036–0.12, Fig. 3C). There was a significant, although weak, difference between the mean compactness values of the separation quality classes (Table 3). Post hoc comparisons showed that well-separated units tended to be
The average response inside the tuning curve was 25 ± 19 spikes/s (about 1.77 spikes per 70-ms stimulus duration), compared with 4 ± 5 spikes/s (about 0.3 spike per 70-ms stimulus duration) outside the tuning curve (this number compares well with estimates of spontaneous activity in auditory cortex of awake cats; Vaadia et al. 1989). The maximal response of each unit within its tuning curve was obviously substantially higher, 90 ± 55 spikes/s on average (about 6.3 spikes per 70-ms stimulus duration) and the average rate in the core FRA, used for the analysis of the temporal response patterns later, was 43 ± 31 spikes/s (about 3 spikes per 70-ms stimulus duration). When considering only well-separated units, the rates were 22 ± 18 spikes/s within the tuning curve, 3.4 ± 4.3 spikes/s outside the tuning curve, a maximal response of 80 ± 49 spikes/s, and an average rate of 40 ± 31 spikes/s in the core FRA. There was a significant difference between the mean rates of the three separation quality classes (Table 3). Unsurprisingly, post hoc comparisons showed that well-separated units tended to have the lowest rates, small clusters had medium rates, and multi-units had the highest rates, although the differences were not very large—only 10–20% between well-separated units and multiunits (Table 3).

To study in more detail the responses inside the tuning curve, we studied the internal structure of the FRA by analyzing two one-dimensional cuts: the level function, summarizing the response as a function of level for frequencies close to the CF, and the frequency-response vector, summarizing the response as a function of frequency at the level that evoked the highest response.

Internal structure of the FRA

The parameters discussed above describe the contour of the FRA, but ignore a possible nontrivial internal structure. An important internal structure descriptor of the responses, which is completely ignored by the tuning curve, is the firing rate. Considering all units with significant responses, the average rate inside the tuning curve was 25 ± 19 spikes/s (about 1.77 spikes per 70-ms stimulus duration), compared with 4 ± 5 spikes/s (about 0.3 spike per 70-ms stimulus duration) outside the tuning curve (this number compares well with estimates of spontaneous activity in auditory cortex of awake cats; Vaadia et al. 1989). The maximal response of each unit within its tuning curve was obviously substantially higher, 90 ± 55 spikes/s on average (about 6.3 spikes per 70-ms stimulus duration) and the average rate in the core FRA, used for the analysis of the temporal response patterns later, was 43 ± 31 spikes/s (about 3 spikes per 70-ms stimulus duration). When considering only well-separated units, the rates were 22 ± 18 spikes/s within the tuning curve, 3.4 ± 4.3 spikes/s outside the tuning curve, a maximal response of 80 ± 49 spikes/s, and an average rate of 40 ± 31 spikes/s in the core FRA. There was a significant difference between the mean rates of the three separation quality classes (Table 3). Unsurprisingly, post hoc comparisons showed that well-separated units tended to have the lowest rates, small clusters had medium rates, and multi-units had the highest rates, although the differences were not very large—only 10–20% between well-separated units and multiunits (Table 3).

To study in more detail the responses inside the tuning curve, we studied the internal structure of the FRA by analyzing two one-dimensional cuts: the level function, summarizing the response as a function of level for frequencies close to the CF, and the frequency-response vector, summarizing the response as a function of frequency at the level that evoked the highest response.

Internal structure of the FRA

The parameters discussed above describe the contour of the FRA, but ignore a possible nontrivial internal structure. An important internal structure descriptor of the responses, which is completely ignored by the tuning curve, is the firing rate. Considering all units with significant responses, the average rate inside the tuning curve was 25 ± 19 spikes/s (about 1.77 spikes per 70-ms stimulus duration), compared with 4 ± 5 spikes/s (about 0.3 spike per 70-ms stimulus duration) outside the tuning curve (this number compares well with estimates of spontaneous activity in auditory cortex of awake cats; Vaadia et al. 1989). The maximal response of each unit within its tuning curve was obviously substantially higher, 90 ± 55 spikes/s on average (about 6.3 spikes per 70-ms stimulus duration) and the average rate in the core FRA, used for the analysis of the temporal response patterns later, was 43 ± 31 spikes/s (about 3 spikes per 70-ms stimulus duration). When considering only well-separated units, the rates were 22 ± 18 spikes/s within the tuning curve, 3.4 ± 4.3 spikes/s outside the tuning curve, a maximal response of 80 ± 49 spikes/s, and an average rate of 40 ± 31 spikes/s in the core FRA. There was a significant difference between the mean rates of the three separation quality classes (Table 3). Unsurprisingly, post hoc comparisons showed that well-separated units tended to have the lowest rates, small clusters had medium rates, and multi-units had the highest rates, although the differences were not very large—only 10–20% between well-separated units and multiunits (Table 3).

To study in more detail the responses inside the tuning curve, we studied the internal structure of the FRA by analyzing two one-dimensional cuts: the level function, summarizing the response as a function of level for frequencies close to the CF, and the frequency-response vector, summarizing the response as a function of frequency at the level that evoked the highest response.

Internal structure of the FRA

The parameters discussed above describe the contour of the FRA, but ignore a possible nontrivial internal structure. An important internal structure descriptor of the responses, which is completely ignored by the tuning curve, is the firing rate. Considering all units with significant responses, the average rate inside the tuning curve was 25 ± 19 spikes/s (about 1.77 spikes per 70-ms stimulus duration), compared with 4 ± 5 spikes/s (about 0.3 spike per 70-ms stimulus duration) outside the tuning curve (this number compares well with estimates of spontaneous activity in auditory cortex of awake cats; Vaadia et al. 1989). The maximal response of each unit within its tuning curve was obviously substantially higher, 90 ± 55 spikes/s on average (about 6.3 spikes per 70-ms stimulus duration) and the average rate in the core FRA, used for the analysis of the temporal response patterns later, was 43 ± 31 spikes/s (about 3 spikes per 70-ms stimulus duration). When considering only well-separated units, the rates were 22 ± 18 spikes/s within the tuning curve, 3.4 ± 4.3 spikes/s outside the tuning curve, a maximal response of 80 ± 49 spikes/s, and an average rate of 40 ± 31 spikes/s in the core FRA. There was a significant difference between the mean rates of the three separation quality classes (Table 3). Unsurprisingly, post hoc comparisons showed that well-separated units tended to have the lowest rates, small clusters had medium rates, and multi-units had the highest rates, although the differences were not very large—only 10–20% between well-separated units and multiunits (Table 3).

To study in more detail the responses inside the tuning curve, we studied the internal structure of the FRA by analyzing two one-dimensional cuts: the level function, summarizing the response as a function of level for frequencies close to the CF, and the frequency-response vector, summarizing the response as a function of frequency at the level that evoked the highest response.
level, such that the response increased monotonically from threshold to the best level and decreased monotonically above the best level. Only 6% of the units (scattered throughout Fig. 4E) had a more complicated pattern of responses as a function of level. Thus the MR together with threshold and best level are in fact good descriptors of the behavior of most cortical neurons as a function of level around the CF.

**THE FREQUENCY-RESPONSE VECTOR.** FRAs could have multiple maxima in their responses as a function of frequency, even when the FRA did not have multiple lobes. We therefore tested the more compact FRAs (compactness > 0.036, n = 467, 35% of the total) for multiple maxima, using the MMI (see Methods). A large majority, 73% of these units, had MMI2 (Fig. 5D, top) and MMI1 (Fig. 5D, bottom) values below −2, corresponding to a frequency-response vector with one peak (as in Fig. 5A). Another 15% of the units had MMI1 or MMI3 values between −2 and 2, indicating the presence of multiple peaks riding over a major, wide peak (as in Fig. 5B). Only 12% of the units had MMI2 or MMI3 values ≥ 2 corresponding to FRAs with two or more clear peaks (as in Fig. 5C). Thus compact FRAs tended to be unimodal. On the other hand, units with lower compactness values tended to have two or more peaks, although many of them still showed one peak in their frequency-response vector (58% of the units with two to four major lobes and 43% of the units with diffuse FRA had MMI2 values less than −2). For comparison, a Gaussian-shaped frequency-response vector would have MMI2 = −6.65 and MMI1 = −18. Thus although most frequency-response vectors were unimodal, they had shallower slopes than expected from a Gaussian-like model.

**FIG. 5.** Frequency-response vectors of units with compactness values > 0.036 (n = 467). A-C: FRAs (gray-scale coded) and tuning curves (standard algorithm; gray line) of 3 units. Arrows at the left mark the level from which the frequency-response vector (black lines above the FRAs) were derived. A: a well-separated unit with one peak at its frequency-response vector, MMI2 = −10.6, MMI1 = −14.6. B: a small cluster with one wide peak of the frequency-response vector, MMI2 = −0.07, MMI1 = −6.6. C: a well-separated unit with 2 clear peaks in the frequency-response vector, MMI2 = −3.9, MMI1 = −0.5. Gray-scale saturation for A–C: 107, 60, and 60 spikes/s, respectively. D, top: distribution of the ratio between the 2nd harmonic and the 1st harmonic (MMI3). Bottom: distribution of the ratio between the 3rd harmonic and the 1st harmonic (MMI1). Histograms follow the conventions of Fig. 2.

**Correlations between FRA parameters.**

Correlations were computed between all pairs of FRA parameters. Because of the large number of data points, even small correlations could be significant. We therefore considered correlations < 0.3, explaining < 9% of the variance, as low (whereas even at a significance level of 0.01, a correlation of 0.1 would be considered as significant because of the large number of measurements).

One nonobvious correlation above the cutoff was the negative correlation between compactness and BW40 (r = −0.38; Fig. 6A). High compactness would be most easily achieved with wide, level-resistant tuning curves, suggesting that large BW40 values should be associated with high compactness, leading to positive correlation between the two variables. However, in practice the largest BW40 values were associated with rather diffuse FRAs, and thus the highest compactness was actually achieved by units with mid- and low-BW40 values.

Except for this case, however, and except for the obvious correlations between measurements that are strongly related to each other (see Supplementary materials), correlations between FRA parameters were rather low. Thus the correlation between the minimal threshold and the CF was essentially nonexistent (r = −0.01, P = 0.7; Fig. 6B), demonstrating the constant coverage of threshold levels as a function of frequency within the (admittedly, relatively narrow) frequency range tested here. Mean response rates were uncorrelated with the other parameters as well, except for a weak correlation with the monotonicity ratio (r = 0.27).

**Beyond the FRA**

Even after accounting for its internal structure, the FRA is still a reduced representation of the neuronal responses to tones. The FRA emphasizes the order and compactness of the responses in the frequency/level response plane, but a unit can carry information about the stimulus even when the FRA is diffuse. Furthermore, the FRA is generated by averaging responses over time and therefore it ignores temporal response patterns.

**Quantifying information about frequency and sound level.** The information about stimuli was quantified by the mutual information between stimuli and responses (see Methods and Nelken et al. 2005). First, the ability to extract any information from the extremely sparsely sampled experimental...
The spike count in a 70-ms time window during the spontaneous activity of the neuron was estimated. The MI values computed from the simulated responses of these models (sMI) were on average about half of the true MI of the generating model (tMI, Fig. 7A). It follows that the MI values are probably underestimated by a factor of about 2 the actual MI between pure-tone stimuli and responses, under the conditions studied here. The range of MI values estimated in the simulations matched those that were calculated from the real data.

Because MI estimation used a maximization procedure (see Methods), it was expected to lead to some amount of overestimation of the MI, resulting in positive information values even when there is no relationship whatsoever between stimuli and responses. To check the significance of the MI values derived from the actual data, the MI was also computed from spike counts in the 70-ms segments preceding stimulus onset (MIspan). The MI computed from the responses was about eightfold larger (0.14 ± 0.1 bits/trial) than that computed from the prestimulus segments (0.018 ± 0.019 bit/trial, t = 45, df = 1333, P ≪ 0.05; see Fig. 7B, top). Because the average number of spikes per stimulus in the 70-ms segment considered here was <1, the MI per spike was higher, with a mean of 0.27 ± 0.2 bit/spike (Fig. 7C).

The correlation between MI and rate was unusually high for the data described here (Fig. 7D, r = 0.53). On the other hand, the correlations between the MI and parameters quantifying the shape of the FRA were generally low. The only noteworthy correlation was a weak one between the monotonicity ratio and the MI (Fig. 7E, r = 0.3). Thus standard parameters characterizing the shape of the FRA are not very indicative of the amount of information carried about the stimulus—a wide range of FRA shapes carry the same amount of information (although they presumably carry different types of information about the stimuli). In particular, the MI was only weakly correlated with the compactness (Fig. 7F, r = 0.21), supporting the claim that similar levels of information about frequency/level combinations could be carried both by units with well-defined FRAs and by units with rather diffuse FRAs.

**TEMPORAL RESPONSE PATTERNS.** Because each stimulus was repeated only once, it was not possible to analyze in detail the temporal response patterns to any specific stimulus. A rough quantification of the response patterns as a function of time was achieved by pooling the responses to a restricted set of combinations, the core FRA surrounding the combination with the strongest response, to build a PSTH (see Methods). Only units that were tested with a single tone, having significant responses and more than five stimuli in their core FRA, were included in this analysis (n = 857).

Although the stimulus duration was 115 ms, a longer segment of 410 ms starting at the onset of the stimulus was analyzed. It was divided into nine time windows and the significance of the responses was tested in each time window separately, at a high significance level (see Fig. 1B and Methods). This process produced a binary vector of length 9, showing whether the responses were significant in each time window (Fig. 8, main plot). Only about 9% of the units had pure onset responses, defined as significant responses in the first or second time window (but not both) and nonsignificant responses in all the other time windows (Fig. 8A). Some sustained response was seen in a large number of units, with a 341/857 (40%) having significant responses in all the time windows during the stimulus after response onset (time windows 1–4, covering 120 ms, or time windows 2–4 covering 90 ms; Fig. 8, E, G, and H). Many units had, in addition, significant responses after stimulus offset, with 27.7% of the units having significant responses during time windows 5 and 6 (early offset, ≤60 ms after stimulus offset; Fig. 8, G and H) and 12.9% of the units having significant responses as late as time windows 7–9 (60–300 ms after stimulus offset; Fig. 8H). These responses, hundreds of milliseconds after stimulus offset, are called here very late responses (VLRs). In fact, 9% of the units had significant responses that started at the first or second time windows and lasted uninterruptedly up to and including the last time window (through responses, bottom-most units in the main plot of Fig. 8H). On average, units had significant response in 4.1 ± 2.6 time windows.

The groups of temporal patterns defined above are summarized in Table 4, separately by spike separation class. The
was no significant concentration of a specific temporal pattern in any of the spike separation classes (for the five temporal pattern classes: \( \chi^2 = 0.68, 6.58, 1.91, 0.89, 1.69; \) df = 2; \( P > 0.01 \) in all cases, to adjust for multiple comparisons).

**Changes in FRA Shape with Time.** The FRA shape could change over time. An example is shown in Fig. 9. This well-separated unit had significant responses during time windows 1–8 (Fig. 9A). The FRAs built from the responses at each time window are displayed in Fig. 9B (the top left FRA is computed from the overall response). The main change in the FRA with time was a reduction in its compactness after stimulus offset. Figure 9C presents the population distribution of the compactness as a function of time window number, for all the units with significant responses at that time window. In this plot, the abscissa corresponds to compactness values, the ordinate corresponds to the time window number (with time advancing from top to bottom), and the color represents the probability of observing a compactness value in the FRAs computed at each time window. There was a significant difference between the means of the compactness at the different time windows [one-way ANOVA, \( F(8,4740) = 13.32, P < 0.05 \)]. Post hoc comparisons showed that FRAs computed for the first time window had on average higher compactness (mean = 0.036) than the rest of the time windows. Compactness during the stimulus and the immediate offset responses (time windows 2–5) had an intermediate value (mean = 0.029). After stimulus offset compactness decreased still further (mean = 0.027 for time windows 6–9). Thus the FRAs generally became more diffuse with time.

To check whether the responses at each time window carried information about the stimulus, we calculated the MI for each time window separately. Figure 9D represents the results in the same format used for the compactness, except that an additional row (at the bottom of the plot) represents the distribution of MIs computed from prestimulus time segments, to emphasize the significance of the estimated MI.

### Table 4. Temporal patterns at the three separation quality classes

<table>
<thead>
<tr>
<th>Spike Separation Class</th>
<th>Onset</th>
<th>Sustained</th>
<th>Early offset</th>
<th>Late offset</th>
<th>Through</th>
</tr>
</thead>
<tbody>
<tr>
<td>q1</td>
<td>37 (9.34%)</td>
<td>175 (44.2%)</td>
<td>113 (28.5%)</td>
<td>55 (13.88%)</td>
<td>41 (10.35%)</td>
</tr>
<tr>
<td>q2</td>
<td>18 (7.8%)</td>
<td>87 (37.82%)</td>
<td>68 (29.56%)</td>
<td>30 (13%)</td>
<td>20 (8.69%)</td>
</tr>
<tr>
<td>q3</td>
<td>23 (9.95%)</td>
<td>79 (34.19%)</td>
<td>56 (24.24%)</td>
<td>26 (11.25%)</td>
<td>17 (7.35%)</td>
</tr>
<tr>
<td>Total</td>
<td>78 (9.1%)</td>
<td>341 (39.8%)</td>
<td>237 (27.7%)</td>
<td>111 (13%)</td>
<td>78 (9.1%)</td>
</tr>
</tbody>
</table>

Values are means ± SD. Temporal patterns: Onset: significant responses in first or second time window (but not both) with no significant response at any other window; Sustained: response at least in time windows 1–4 or 2–4 (with possible responses in other windows); Early Offset: significant response at least in time windows 5 and 6 (with possible responses in other windows); Late Offset: significant responses throughout time windows 7–9 (with possible responses in other windows); Through: significant responses in time windows 1–9 or 2–9. Percentages are taken for each spike separation class separately. The temporal patterns are not mutually exclusive.
ANOVA, means of the MI at the different time windows [one-way values. There was a significant difference between the means of the MI at the different time windows [one-way ANOVA, $F(8,4836) = 86$, $P \ll 0.05$]. MI values were larger during the first two time windows after stimulus onset (mean = 0.116 bit/trial) and dropped somewhat for later time windows (time windows 3–8, mean = 0.066 bit/trial). However, even for the last time window (200–300 ms after stimulus offset) the values of the MI (0.042 ± 0.036 bit/trial) were significantly higher than the comparison values calculated from spontaneous activity (0.023 ± 0.022 bit/trial; $t = 9.2$, df = 296, $P \ll 0.05$).

**DISCUSSION**

Under halothane anesthesia, cat auditory cortex has a substantially richer behavior than previously described in anesthetized animals. Both spectral and temporal properties are extremely varied: the FRA bandwidth spans a much larger range of values, with some units having diffuse FRAs but highly significant and informative responses to tones. The temporal patterns of the responses are also much richer than previously described, with many units responding throughout the stimulus duration and for hundreds of milliseconds beyond. In other respects, cat auditory cortex under halothane is the same as that described under barbiturates and ketamine anesthesia.

**Anesthesia effects**

The two main differences between the results described here and previous characterizations of cat auditory cortex are 1) the bandwidth of the resulting FRAs and 2) the much richer temporal response patterns described here. Because the methods used to acquire and analyze the FRAs were almost identical to previous studies (Schreiner and Sutter 1992; Sutter and Schreiner 1991, 1995), it is probable that the main cause for these differences is the difference in anesthesia. Whereas most previous studies of auditory cortex have been performed under barbiturates (Sally and Kelly 1988; Schreiner and Sutter 1992; Sutter and Schreiner 1991, 1995) and ketamine anesthesia (DeWeese et al. 2003; Read et al. 2001), here we used halothane.

Many of the features that we describe here are usually associated with recording in awake animals rather than with recordings under anesthesia. For example, Gaese and Ostwald (2001) described wider frequency responses in recordings from awake rats compared with recordings from anesthetized rats. The bandwidths reported in the alert cat (Qin et al. 2003) are about threefold larger on average than those reported under barbiturate anesthesia (Schreiner and Sutter 1992; Sutter and Schreiner 1991). Neurons with phasic response in the alert cat reached a bandwidth of 6 octaves at their best level (Qin et al. 2003). In particular, the presence of sustained responses has often been cited as a major difference between anesthetized and awake recordings (Evans and Whitfield 1964; Pfingst and O’Connor 1981; Qin et al. 2003; Recanzone et al. 2000; Shamma and Symmes 1985; Wang et al. 2005; Zurita et al. 1994). In the data reported here, the
majority of the units had responses beyond stimulus onset, sometimes lasting far beyond stimulus offset for at least a subset of the tones.

A previous study compared responses under isoflurane, another inhalation anesthetic, and barbiturates (Cheung et al. 2001). In that study, the gas anesthetic was given without N2O, which potentiates the effects of the anesthetic and leads to a reduction in the required anesthetic gas concentration. As a result, the concentration of the anesthetic gas was very high (1.7–2.7%). Under these conditions, a strong depression in cortical activity was found, resulting in higher thresholds, lower spontaneous activity, and impaired ability to follow trains of clicks.

The reason for the striking differences between the results of Cheung et al. with isoflurane and the data presented here may be related to the dose-dependent cardiodepressive effects of isoflurane, which are substantially more potent than those of halothane (Hardman et al. 1996). Possibly as a result, EEG and auditory-evoked responses are suppressed to a greater degree by isoflurane than by halothane (Antunes et al. 2003; Johnson and Taylor 1998). Similarly, greater suppression of cortical activity under isoflurane was found in single units in the visual cortex (Villeneuve and Casanova 2003). These findings strongly suggest that the difference between the results of Cheung et al. (2001) and our results are attributed both to the use of isoflurane rather than halothane and to the rather high level of isoflurane that they used. We suggest that the systemic blood pressure—and, as a result, perfusion of the brain tissue—is better maintained when using halothane in the O2/N2O mixture than when using isoflurane in pure oxygen, as in Cheung (2001).

We conclude from these considerations that halothane anesthesia is in fact maintaining to a large extent the properties of awake responses to sounds. The results presented here suggest that excitatory inputs are much more dominant under halothane than under barbiturates or ketamine, leading to wider tuning widths and to longer and more sustained responses. There is a general resemblance of this much more active cortical state and the stimulus-driven active state described by Miller and Schreiner (2000) under ketamine anesthesia. However, whereas Miller and Schreiner (2000) had to use a very rich set of sounds (moving ripples) to evoke this state, here it seems to be present even under tonal stimulation. Thus the use of halothane anesthesia is a viable alternative to recordings in awake, nonbehaving animals, at least when considering the responses of A1 neurons.

**FRA shapes and temporal response patterns**

The data presented here differ from other studies in two major ways. First, the range of FRA shapes described here is very large. All the parameters that we estimated showed a wide distribution, and furthermore the correlations between different parameters were generally low. As a result, essentially any combination of shape parameters could be achieved. In particular, we have documented the presence of diffuse FRAs, which do not have the standard V-shape but may nevertheless be highly informative about the stimuli. It is tempting to hypothesize that studies under barbiturates and ketamine (e.g., Read et al. 2001) emphasize the functional architecture of the input to auditory cortex, whereas our data represent more faithfully the responses of A1 neurons after processing by the cortical network.

The second way in which these data are different from other studies is the presence of responses past sound onset. Within the limitations of our data, it seems that many neurons had sustained responses for at least some of the stimuli used here. The presence of sustained responses has often been cited as the main difference between the responses of awake versus anesthetized animals (see most recently Wang et al. 2005). We have documented in a number of papers the presence of sustained responses under halothane anesthesia in response to a number of different stimuli: bird songs and their modifications (Bar-Yosef et al. 2002); fluctuating noise bands (Las et al. 2005); and pure tones (in oddball paradigms: Ulanovsky et al. 2003, 2004). The current paper, however, is the first to quantify the amount of sustained responses to pure tones under these conditions.

**Single cells versus multiunits**

After adjusting for multiple comparisons there are only three response properties that vary significantly with the degree of spike separation: threshold, compactness, and firing rates. Well-separated units had lower thresholds, tended to be slightly more compact, and had lower firing rates than those of multiunits. Behaviors of the compactness and of the firing rates are consistent with the idea that multiunits reflect the responses of a number of single neurons with different response properties and therefore potentially “smear” any response parameter. Overall, however, all these effects are not very large and we believe that they in fact emphasize the high local homogeneity of auditory cortex in response to these simple sounds, rather than a possible dispersion in local response properties to pure tones. This finding is compatible with the conclusions of Eggermont (1983) and particularly of Sutter and Schreiner (1992) because we tended to record from central and dorsal A1 rather than from the less homogeneous ventral part.

**Encoding of tones in the responses of A1 neurons**

The data presented here demonstrate some unexpected features of the mechanisms by which neurons in auditory cortex encode the identity of a pure tone stimulus. The first is the presence of an appreciable amount of information about tones throughout and after sound stimulation: in the initial 70 ms of the response, the mean MI between stimuli and responses was estimated at about 0.27 bit/spike. Because this value represents about half of the true information according to the simulations reported here, the actual information is probably about 0.5 bit/spike. This estimate is close to the information per spike in the neural responses to a set of 15 bird songs and their modifications, collected under the same conditions by Nelken (2005) (0.43 bit/spike; in fact, some of the data in this paper were collected during the same experiments). This finding suggests that single spikes may carry the same amount of information when encoding complex stimuli and when encoding simple stimuli.

A second unexpected feature of the data is the finding that the FRA shape by itself was not a major determinant of the pure-tone–encoding capability of an A1 neuron: the correlation between MI and parameters that quantify global aspects of
the FRA, such as bandwidth or compactness, was rather low. Thus both units with classical FRA shapes and units with diffuse, although significant, responses carried similar amounts of information about the stimuli. These considerations support the conclusion of Schreiner (1998) that tones are encoded in a combinatorial way by populations of neurons in A1 (see also Phillips et al. 1994).

Finally, the most surprising result of this paper is the demonstration of an appreciable amount of late responses, lasting hundreds of milliseconds after stimulus offset. These late responses carry information about the stimulus. The origin of the late activity is unclear at the moment. It could mirror late activity in lower stations but also sustained activity states in the cortex itself, or any combination of these sources. It is tempting to speculate that this late activity represents a correlate of sensory memory, residing in some kind of semistable cortical activity states (Amit and Brunel 1997; Goldberg et al. 2004; Kenet et al. 2003), and that the idiosyncratic, diffuse late responses at the single-neuron level may be a reflection of the global structure of these states.

As suggested many times before, a tone would evoke responses in a very large population of neurons (Schreiner 1998). Our results suggest that because of the large variety of FRA shapes, different frequency/level combinations will evoke activity in overlapping but nonidentical populations (Phillips 1995). Many neurons will respond throughout the stimulus and beyond. Furthermore, the decrease in compactness of the FRA after stimulus onset suggests that the responding populations to each tone/level combination will change as a function of time.

An ideal observer of the responding population may therefore be able to identify the stimulus rather easily (with about 0.27 bit/trial on average, and stimulus entropy of <9 bits, it may be enough to observe <100 neurons to identify the frequency/level combination on a trial-by-trial basis). Even more important, the ideal observer may be able to do that throughout stimulus duration and beyond. Finally, because the responding population changes with time, the ideal observer may even be able to identify the time elapsed since stimulus onset. Thus instead of the picture of an auditory cortex encoding pure tones rather poorly, this paper demonstrates that auditory cortex has the ability to encode pure tones in a rich and highly informative manner.

Acknowledgments

The authors thank E. Tomer for technical assistance and N. Taaseh for critical comments.

Present addresses: N. Ulanovsky, Department of Psychology, University of Maryland, College Park, MD; O. Bar-Yosef, Department of Pediatrics, Safra Children’s Hospital, Sheba Medical Center, Tel-Hashomer, Israel.

Grants

This work was supported by grants administered by the Israel Science Foundation, the US–Israel Binational Science Foundation, and the German–Israeli Foundation.

References


