Single-Neuron Responses to Rapidly Presented Temporal Sequences in the Primary Auditory Cortex of the Awake Macaque Monkey

M. L. Phan1 and G. H. Recanzone1,2
1Center for Neuroscience and 2Section of Neurobiology, Physiology and Behavior, University of California at Davis, Davis, California

Submitted 6 July 2006; accepted in final form 22 November 2006

Phan ML, Recanzone GH. Single-neuron responses to rapidly presented temporal sequences in the primary auditory cortex of the awake macaque monkey. J Neurophysiol 97: 1726–1737, 2007. First published November 29, 2006; doi:10.1152/jn.00698.2006. One fundamental process of the auditory system is to process rapidly occurring acoustic stimuli, which are fundamental components of complex stimuli such as animal vocalizations and human speech. Although the auditory cortex is known to subserve the perception of acoustic temporal events, relatively little is currently understood about how single neurons respond to such stimuli. We recorded the responses of single neurons in the primary auditory cortex of alert monkeys performing an auditory task. The stimuli consisted of four tone pips with equal duration and interpip interval, with the first and last pip of the sequence being near the characteristic frequency of the neuron under study. We manipulated the rate of presentation, the frequency of the middle two tone pips, and the order by which they were presented. Our results indicate that single cortical neurons are ineffective at responding to the individual tone pips of the sequence for pip durations of <12 ms, but did begin to respond synchronously to each pip of the sequence at 18-ms durations. In addition, roughly 40% of the neurons tested were able to discriminate the order that the two middle tone pips were presented at durations of ≥24 ms. These data place the primary auditory cortex at an early processing stage of temporal rate discrimination.

INTRODUCTION

A major function of the auditory system is to process temporally complex stimuli, a common component of environmental sounds including human speech and animal vocalizations. The auditory cortex is necessary for the perception of temporally complex acoustic stimuli because lesions of this area result in deficits in the perception of auditory temporal sequences and durations (e.g., Diamond and Neff 1957; Scharlock et al. 1965). In macaque monkeys, lesions of either the left or both auditory cortices result in impairments in the monkey’s ability to discriminate conspecific vocalizations (Heffner and Heffner 1986). In humans, lesions in the temporal cortex were previously shown to result in compromised speech comprehension (Schnider et al. 1994) and the inability to order two successive tonal stimuli (Swisher and Hirsh 1972). Recent human psychophysics and imaging studies in learning impaired children, children classified as “poor readers,” and dyslexics indicated that they have a reduced ability to resolve the duration and rate of acoustic test stimuli, such as processed speech and tones, compared with control children (Merzenich et al. 1996; Nagarajan 1999; Tallal et al. 1996; Temple 2001; Wright et al. 1997). Those authors suggested that these deficits result from a general impairment in the ability to resolve auditory temporal information.

One manner in which single neurons can represent the temporal features of acoustic stimuli is by phase-locking to the envelope of the signal. Such a mechanism was demonstrated throughout the ascending auditory pathway (for reviews see Edelman et al. 1988; Lagner 1992; Moore 1989; Pickles 1992). Despite these high-fidelity representations of temporal information, it is still not clearly understood how these neural codes translate to a perception of order and distinct patterns (for review see Johnson 2000; Parker and Newsome 1998).

Initial inroads into the ongoing basis of the cortical computations of temporally complex stimuli were conducted in the somatosensory system (for review see, Romo and Salinas 2003). These studies indicated that the phase-locking of individual neurons are well correlated with the corresponding percepts of the individuals. In the auditory domain, similar studies pointed toward the same general conclusions (Brown and Maloney 1986; Moody 1994), although an overall rate code for high temporal frequencies was also proposed (Lu et al. 2001). To further investigate how auditory cortical neurons could potentially encode low-frequency temporal information, responses of single neurons to four tone-pip sequences were recorded in alert macaque monkeys. Preliminary results from this study were previously published in abstract form (Phan et al. 1998).

METHODS

Animals

This report is based on data collected from 167 single neurons recorded from the primary auditory cortex of two adult male rhesus monkeys (Macaca mulatta) weighing 7–12 kg over the course of the study. All protocols and procedures followed the guidelines outlined in the Ethical Treatment of Animals (National Institutes of Health) and were approved by the UC Davis animal care and use committee.

For all recording sessions, the monkey sat in a customized primate chair constructed to minimize acoustic reflections with its head fixed and pointed toward the center of a speaker array. All data were collected within a darkened double-walled sound booth (2.4 × 3.0 m, IAC, New York, NY) lined with sound-attenuating foam.

Before recording monkeys were implanted with a restraining head post and recording cylinder (see Pfingst and O’Connor 1980; Recanzone et al. 2000a). Two recording cylinders were implanted bilaterally in monkey A (the two surgeries were separated by 18 mo). One recording cylinder was implanted over the left auditory cortex of monkey L.

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.
Data collection

At the start of each recording session, a plastic grid was placed into the recording chamber (Crist et al. 1988). Electrodes (FHC) were introduced into the cerebral cortex through a guide tube placed into a selected grid location. Search stimuli consisted of noise, tones, and clicks presented through either a speaker located directly contralateral to the recording cylinder or a frontal location (±30° in azimuth, 0° in elevation). All speakers were located 1 m from the center of the interaural axis. Once a neuronal response was encountered, neural waveforms were amplitude-filtered and stored (50-kHz sampling) for off-line sorting using customized software. The latency, threshold, best frequency at threshold [characteristic frequency (CF)] and bandwidth at 10 or 40 dB above threshold was audiovisually determined for the recording site. This allowed us to rapidly assess the basic physiological response properties used in classifying the cortical area and to set the stimulus frequencies. All neurons had response characteristics consistent with being within the primary auditory cortex (A1) including these response parameters, the progression of CF, and the location within the recording cylinder (Recanzone et al. 2000a). This was verified histologically by Nissl, myelin, and cytochrome oxidase staining of the superior temporal gyrus in both monkeys.

Stimuli

Stimulator generation and data collection were controlled by Tucker Davis Technologies (Gainesville, FL) hardware and software interfaced with a personal computer. Stimuli consisted of 20 different patterns of four tone pips. The pip durations (2-ms linear rise/fall) were the same as the interpip intervals and varied from 6 to 30 ms in 6-ms steps. Each pip intensity was 60 dB SPL (A-weighted) with a ±2-dB variation between pip sequences. Frequencies of the first and fourth pips of the sequence were near the CF of the neuron under study. Four different sequences were presented for each stimulus duration by varying the frequency of the second and third tone pips. The frequency of these pips differed from the CF by either 5 or 20% and were presented in either order, for example, the second pip was [CF + (CF/20)] and the third was [CF – (CF/20)] or the reverse. Because sequential presentation of identical tones was previously shown to result in decreases in firing rate with each presentation (Brosch and Schreiner 1997; Calford and Semple 1995; Evans and Whitfield 1964; Ulanowsky et al. 2004; Werner-Reiss et al. 2006), these stimuli were chosen to prevent rapid habituation to each stimulus element, as well as to investigate the possibility that single neurons could encode the temporal order of these sequences.

Behavioral paradigm

The monkey performed a go/no-go sound localization task (see Recanzone et al. 2000b for details) during the recording of neural activity. Each trial consisted of a series of four to seven presentations of the four-pip sequences (S1) delivered from the speaker 90° contralateral to the recording site, followed by one presentation (S2) from a speaker located directly in front of the monkey (0°). Tone sequences were presented with a minimum of 800 ms between them. The monkey was trained to depress a lever to initiate a trial and then to release the lever when the stimulus changed location to receive a fluid reward. Within the S1 series, each sequence could vary in temporal rate (pip duration) and frequency but not in order. The order of the S2 sequence was different from that of the S1 and was always presented with a pip duration of 30 ms. Each recording session consisted of 15–20 randomly interleaved presentations of each stimulus duration, order, and frequency difference (about 20 min). Only responses to stimuli presented from 90° are reported here.

Data analysis

Poststimulus time histograms (PSTHs) were constructed using 1-ms time bins. The latency measured for each cell was defined as the median latency to the first spike that was >10 ms from stimulus onset pooled across all S1 stimulus presentations. This provided a reliable estimate even for neurons with relatively high spontaneous activity. To quantify the degree of phase-locking or synchronization to each tone pip of the sequence the vector strength (VS) was calculated using 1-ms time bins (Goldberg and Brown 1969; Mountcastle et al. 1969; Talbot et al. 1968). Each spike was assigned a vector with a length of 1.0 and phase assigned as the time of occurrence relative to the (pip + pip-interval) duration (e.g., 12 ms for the 6-ms pip duration and interval). These vectors were then summed and divided by the total number of spikes. Thus these values range from 1 (perfect synchrony) to 0 (no synchrony). To determine the statistical significance of these responses a Rayleigh test of uniformity, 2n(VS)2, was calculated for each neuron, where VS is the vector strength and n is the number of spikes used to calculate the vector strength (Buenen and Rhode 1978; Lu and Wang 2000; Mardia and Jupp 2000). The Rayleigh test is an approximation of a chi-square test constrained with 2 degrees of freedom, appropriate for circular distributions (Mardia and Jupp 2000; Zar 1999). Rayleigh test values >13.8 indicated that the synchronization of neuronal activity reflected a sample of an oriented distribution that is not attributed to chance (P < 0.01; Buenen and Rhode 1978).

RESULTS

The data of this report are based on 167 neurons recorded from the primary auditory cortex in three hemispheres of two monkeys. Figure 1 shows the frequency distribution of the median first-spike latencies and the characteristic frequencies of the sampled neurons. The median first-spike latency (Fig. 1A) ranged from 16 to 64 ms, with a mean of 32 (±10.3), which is in excellent agreement with previous studies of awake macaque A1 latencies (Recanzone et al. 2000a). The solid bars in Fig. 1A show the shortest median first-spike latency for all neurons recorded from that grid location on a given day. Because we often recorded from more than one neuron within a grid location on a given day, the open bars show the latencies of all other neurons recorded from on that experimental day. Different neurons recorded from a single grid location on a given experimental day were usually recorded from different depths, but could also be different neurons isolated from the same electrode location. The distribution of the characteristic frequency of these neurons is shown in Fig. 1B in 0.5-octave bins. The sample was biased toward neurons with higher CFs, which is the result of more extensive investigation of the caudal regions of A1. Some neurons were recorded with CFs near 500 Hz, which is in the border region between A1 and R (e.g., Merzenich and Brugge 1973; Morel et al. 1993; Recanzone et al. 2000a). These data, coupled with the CF gradient in each monkey and the histological verification, indicate that the vast majority, and likely all, of the neurons in this report were well within the primary auditory cortex. The recordings done in these monkeys went well beyond A1 as part of different studies, which verified the medial, lateral, and caudal borders of A1 with the belt fields (Recanzone et al. 2000a; see Kaas and Hackett 2002). Thus the region of A1 was evenly sampled across its extent.

The responses to each of the 20 stimulus types are shown for a representative neuron in Fig. 2. Each row shows the response to a single tone-pip duration sequence and each column shows the responses to the different frequency differences between the second and third pips in the sequence shown across the top of each column. All PSTHs are shown on the same timescale.
seek to explore the relationship between firing rate and both the frequency differences in the sequences and the tone pip durations, the firing rate for all neurons was calculated. Because the firing rate is expected to increase with increased stimulus durations (e.g., from 6- to 30-ms pip durations), the data were normalized by the stimulus duration (4 pips + 3 intervals). Further, the median first-spike latency was added to the stimulus duration of the four tone-pip sequence (e.g., for the 6-ms sequence the time would be [6 ms × (4 pips + 3 intervals)] + median latency) and the average firing rate for each of these stimuli was calculated over this time period. With this analysis, there is a higher firing rate for the shortest tone-pip-duration stimuli, as indicated by a leftward shift in the frequency distributions with increasing pip duration shown in Fig. 3A. There was a relatively uniform distribution of firing rates across neurons, although some neurons showed much higher firing rates than the rest, giving rise to a rightward tail in the distributions. The median firing rate across neurons is shown in Fig. 3B. These functions are not linear because the slope between the firing rates for the 6- to 12-ms tone-pip-duration stimuli is much greater than the slope between firing rates for the 24- and 30-ms tone-pip-duration stimuli, reflecting the relatively constant onset response to the first tone pip of the sequence (e.g., see Fig. 2). Statistical analysis showed similar trends, with a strong effect of tone duration (ANOVA; main effect of tone duration; $F = 12.48; P < 0.01$) but no effect of the frequency of the second and third pips ($F = 1.67; P > 0.01$). Tukey analysis showed a significant difference between all possible combinations of pip durations except between the firing rates for 24- and 30-ms tone durations.

The preceding analysis shows that most neurons had a clear and usually robust response to these stimuli. Although the distribution of firing rates across neurons was fairly similar, there were neurons that had either very high or very low firing rates. Previous studies in awake marmosets indicated that high temporal rate stimuli (corresponding to the shortest tone-duration stimuli of this study) could be encoded by a rate code in some A1 neurons (Lu et al. 2001). In contrast, other A1 neurons encoded the temporal structure of the stimuli by the periodicity of the response. Although our sample appeared fairly uniform across neurons, it could be that there were similar subpopulations of neurons that showed a stronger relationship between firing rate and the temporal envelope than others. To address this possibility, we subdivided the response based on the firing rates of the neurons to the 6-ms-duration stimuli. The firing rates for the 25% of neurons that had the lowest (L) and highest (H) firing rates to the 6-ms-duration stimuli were compared with the remaining neurons nearest the median of the population (M) using the same analyses shown in Fig. 3. These results are shown in Fig. 4. Although there is an offset between the sets of lines, the slopes are generally similar. We quantified this impression by calculating the regression line of these data and comparing the slopes with 95% confidence intervals. The slopes for the median and lowest firing rate cells were nearly identical ($-0.0105$ and $-0.0107$ spikes$^{-1}$s$^{-1}$ for lowest and median, respectively). The greatest slope was for the high-activity population ($-0.0206$ spikes$^{-1}$s$^{-1}$), although the 95% confidence intervals encompassed the slopes for the other two populations ($-0.010$ to $-0.040$). This difference was almost entirely eliminated if the 6-ms data were excluded (slopes of $-0.002$, $-0.003$, and
plots, the timescales are adjusted such that the stimuli span the
characteristic frequency by 5% and the third was greater than the
CF by 5%. In these stimulus trains in which the second tone pip was
less than the CF, the periodicity of their responses. Because the
carrier frequency was generally very high (see Fig. 1), periodicity
failed to reach statistical significance, as indicated by
the Rayleigh value for all neurons at each of the five tone-pip
duration sequences. The vertical dashed line shows the cutoff value for statistical
significance at the 0.01 level. For all tone-pip durations,
periodicity is statistically significant (see METHODS). Distributions of
the Rayleigh value for all neurons at each of the five tone-pip-
duration sequences are shown in Fig. 7A. These data are
restricted to the frequency sequences where the second pip is
the CF plus 5% and the third pip is the CF minus 5%, although
the other three frequency sequences showed a similar pattern.
The vertical dashed line shows the cutoff value for statistical
significance at the $P < 0.01$ level. For all tone-pip durations,
there was a substantial population of neurons where the peri-
ocidity failed to reach statistical significance, as indicated by
the large distribution at values $\approx 5$. However, there was a clear
trend for the number of neurons showing statistically signifi-
cant periodicity to increase with increasing pip duration, as
evidenced by the initial peak in the distributions becoming
progressively smaller and the far-right tail of the distributions

As a final analysis we investigated whether there was a
significant correlation between the characteristic frequency and
the firing rate of the neurons. Regression analysis showed a
statistically significant correlation in only one comparison: the
6-ms-duration stimulus with the smallest difference in tone
frequency. Although this comparison did reach statistical
significance ($F = 9.33; P < 0.01$), characteristic frequency could
account for only a fraction of the variance in firing rate ($r^2 = 0.03$).

Temporal precision in the responses

Whereas the firing rates were fairly uniformly distributed
across the sample, a second way in which the tone-pip dura-
tions within a sequence could be encoded by the neurons is in
the periodicity of their responses. Because the carrier fre-
quency was generally very high (see Fig. 1B), the periodicity
would be expected to be related to the envelope of the stimulus
sequences. Tuning to the stimulus envelope is evident for
tone-pip durations $>18$ ms in Fig. 2. Examples of such tuning
are shown for a different example neuron in Fig. 5 for the
stimulus train in which the second tone pip was less than the
CF by 5% and the third was greater than the CF by 5%. In these
plots, the timescales are adjusted such that the stimuli span the
same distance in each raster. Again, there is little if any
periodicity in the response corresponding to each tone pip in
the sequence for pip durations of 6 and 12 ms (Fig. 5, A and B).
Periodicity is first noted for 18-ms pip durations (Fig. 5C) and
increases for the 24- and 30-pip-duration sequences (Fig. 5, D
and E). A similar trend is shown for a third neuron that had a
relatively low firing rate in Fig. 6. Although the firing rate was
lower for this neuron, the appearance of the periodicity of
the response is similar to that shown for the neurons in Figs. 2
and 5.

To compare the periodicity of the responses across neurons,
the Rayleigh test was applied to all neurons using the duration
of the tone pip plus the interpip interval as the period. This
statistic takes into account both the vector strength and the
overall firing rate of the neuron and tests whether the periodic-
ity is statistically significant (see METHODS). Distributions of
the Rayleigh value for all neurons at each of the five tone-pip-
duration sequences are shown in Fig. 7A. These data are
restricted to the frequency sequences where the second pip is
the CF plus 5% and the third pip is the CF minus 5%, although
the other three frequency sequences showed a similar pattern.
The vertical dashed line shows the cutoff value for statistical
significance at the $P < 0.01$ level. For all tone-pip durations,
there was a substantial population of neurons where the peri-
ocidity failed to reach statistical significance, as indicated by
the large distribution at values $\approx 5$. However, there was a clear
trend for the number of neurons showing statistically signifi-
cant periodicity to increase with increasing pip duration, as
evidenced by the initial peak in the distributions becoming
progressively smaller and the far-right tail of the distributions

\begin{figure}
\centering
\includegraphics[width=\textwidth]{response_figures}
\caption{Responses to all stimuli in a representative neuron. Poststimulus time histograms (PSTHs) derived from 20 presentations of each stimulus. Each column shows the response to one of 4 frequency sequences presented. Each row shows the response for different tone pip durations (6, 12, 18, 24, and 30 ms from top to bottom, respectively). Horizontal bars under each PSTH show the period that the tone was presented. This neuron had a CF of 20 kHz and stimulus frequency sequences had the middle 2 pips at CF $\pm$ 5% (left 2 columns) or CF $\pm$ 20% (right 2 columns). Regardless of the stimulus frequency sequences presented, overall activity and the periodicity of the response increased with increasing tone pip duration.}
\end{figure}
becoming progressively larger from 6-ms (gray) to 30-ms (black) pip durations. The percentage of neurons that showed statistically significant periodicity by this metric was 3.66, 23.45, 49.77, 64.61, and 73.57% for 6-, 12-, 18-, 24-, and 30-ms pip durations, respectively, where 1.0% of neurons are expected to show statistically significant tuning by chance.

Figure 7B shows the medians of the population for each frequency sequence as a function of the pip durations. For the 6- and 12-ms-duration stimuli, most neurons had very low values of the Rayleigh statistic, with the median response well below our criteria for statistical significance. However, around 18 ms the median across the population approached statistical significance (with nearly 50% showing significant periodicity; see above) and most neurons did show statistically significant periodicity in their responses for tone durations of 24 and 30 ms.

One problem inherent in this analysis is that the vector strength, and therefore the Rayleigh statistic, can be dominated by the response to the first pip in the sequence. For example, a neuron that has only a phasic response to the first tone pip, and does not respond at all to the subsequent three pips, could have a high Rayleigh statistic even though it clearly did not have a periodic response. Within our sample, there were several instances where this occurred to varying degrees. To quantify the effect of such phasic responses to the first pip of the sequence, we compared the overall activity to the second through fourth tone pips to the activity to the first pip in the sequence. In this analysis, the activity during the latency period plus the first tone pip and interpip interval was calculated as a percentage of the total firing rate to each stimulus. The distribution of these percentages is shown in Fig. 8A, where the neuronal responses from each of the four stimulus-frequency orders are pooled. As expected from the example neurons shown in Figs. 2, 5, and 6, the majority of the responses occurred during the first bin for tone-pip durations of 6 ms (gray line), as indicated by the high tail of the distribution near 100. This tail disappears and the peak of the distribution moves leftward with increasing pip durations, indicating that the response to the first pip of the sequence has a progressively smaller dominance of the total response. To determine how this may have influenced the proportion of neurons showing statistically significant periodicity in their responses, the Rayleigh statistic was calculated using only the period corresponding to the responses to the second through fourth tone pips (Fig. 8B). Although there was a decrease in the proportion of neurons with significant peri-

---

**Figure 3.** Firing rate as a function of stimulus time. A: frequency distributions of the firing rate normalized by the tone stimulus duration during the 4-pip sequence (see RESULTS). Each colored line shows the frequency distribution for each of the sequences that varied by the frequency of the 2nd and 3rd pips in the sequence (red: CF +20% then CF -20%; green: CF -20% then CF +20%; gray: CF -5% then CF +5%; black: CF +5% then CF -5%; for the 2nd and 3rd pips, respectively; see columns in Fig. 2). All frequency distributions are plotted on the same scale. Shortest pip duration condition tended to have higher firing rates using this metric. B: median firing rates of the population of neurons shown in A. There is a consistent decrease in firing rate as pip duration increases from 6 to 24 ms, but no difference between 24 and 30 ms.

**Figure 4.** Median firing rates for neurons divided into high (H), medium (M), and low (L) firing neurons (see text for details). Different pip sequence types are color coded as in Fig. 3. Note that the slopes of these sets of lines are very similar, indicating that neurons behave in a similar manner regardless of their overall firing rate.
To determine the relationship between the tuning to all four pips versus the last three pips, we performed a regression analysis of these two parameters. We restricted the analysis to neurons that showed statistically significant tuning under at least one condition (Table 1). Because there were very few neurons with significant tuning in the 6- and 12-ms tone-duration cases, it was unsurprising that there was no significant correlation between these two measures. Once the tone-pip durations reached or exceeded 18 ms, there was significant correlation, with the $r^2$ values progressively increasing from

---

**Fig. 5.** Single-neuron rasters to these stimuli. A–E: rasters from a single session from a neuron with a CF of 12 kHz when the 4 pips were presented at 12.0, 11.4, 12.6, and 12.0 kHz for pips 1–4, respectively. In each plot, the row of rasters represents a single trial and each tic mark represents a single action potential. Note that each raster has a different x-axis so that the stimulus on-time spans the same distance on each plot. Actual pip duration is given at the top of each raster. Note the lack of periodicity for the short-duration stimuli, the first appearance at 18-ms tone-pip durations, and clear periodicity for the longer pip duration stimuli.

---

**Fig. 6.** Example neuron with a low firing rate. Same conventions as in Fig. 5, except this cell had a very low firing rate. Similar results as the neuron shown in Fig. 5 are seen here, with the first evidence of periodicity seen at 18-ms pip durations. CF of this neuron was 8 kHz.

---
To address this, we measured the latency of the peak response to each of the four tone pips presented in each sequence and the latency to the peak response to pips two to four were normalized to the latency of the response to the first pip. Only neurons that had a reliable peak response to the first pip were included in this analysis, the results of which are shown in Fig. 9. Each panel in Fig. 9 shows the frequency distribution of the peak

0.53 to 0.77 from 18- to 30-ms tone-pip durations. These data indicate that, although there was an overall reduction in the Rayleigh statistic when considering only the last three tone pips, the temporal discrimination of many cells was maintained at a high rate when the onset response was not considered.

A second concern with the Rayleigh analysis is that differences in the latency to each individual tone pip could artificially decrease the value. For example, if the latency to the first pip in the sequence was shorter (or longer) than the latency to the subsequent tone pips, this temporal mismatch would decrease the vector strength and therefore the Rayleigh statistic.

To address this, we measured the latency of the peak response to each of the four tone pips presented in each sequence and the latency to the peak response to pips two to four were normalized to the latency of the response to the first pip. Only neurons that had a reliable peak response to the first pip were included in this analysis, the results of which are shown in Fig. 9. Each panel in Fig. 9 shows the frequency distribution of the peak
latency to the second (red), third (green), and fourth (blue) tone pip relative to the peak latency to the first pip (0) across two-pip and interpip-interval period. For all stimulus durations except 6 ms, the peak latencies are distributed very close to 0-ms latency, with the vast majority of comparisons within 15% of the first-spike latency relative to the pip duration. The means of these distributions were not statistically significantly different from a population with a mean of zero (all \( P \) values >0.05). These results indicate that there is little difference in the latency of the response to each of the four pips in the sequence. The 6-ms-duration stimuli were more difficult to compare because there were fewer cases where a clear peak was evident during the short periods corresponding to the second to fourth pips given the phasic response of the neurons to these stimuli (e.g., Figs. 2, 5, and 6).

In summary, there was very little synchronization to the individual tone pips in the sequence for short-duration tones (6 and 12 ms), the population of neurons began to show significant synchronization for tone durations at about 18 ms, and clear synchronization at the longer tone durations tested (24 and 30 ms).

As with the firing rate, we also compared the temporal response properties of these neurons with the characteristic frequency. Because the shorter-duration stimuli had few neurons that reached statistical significance, we restricted our analysis to those conditions in which the tone-pip durations were 30 ms and the neurons showed statistically significant tuning. This analysis revealed that there was no correlation with the characteristic frequency and the value of the Rayleigh statistic (\( r^2 \) values ranged from 10^{-5} to 0.02; all \( P > 0.01 \)). Thus the characteristic frequency does not account for any of the variance in the degree of synchronization as measured by the Rayleigh statistic.

Temporal order discrimination

The stimuli used in these experiments differed not only in the tone duration, but also in the frequency and order of the second and third pips in the sequence. This was partly to prevent the neurons from habituating to these tone pips, but also to allow the opportunity to determine whether the responses of the neurons could differentiate between the different tone-pip sequences. To address this issue we compared the firing rates between stimulus sequences of different frequencies and orders of presentation at each pip duration for each neuron. For clarity, we will consider sequences in which the second pip was lower than the CF by 5% (and the third pip greater than the CF by 5%) as “downward-small” sequences with small frequency differences [downward-small (DS)]. The reversed order, in which the second pip was 5% greater than CF (and the third 5% less) will be considered upward-small (US).

Similarly, sequences will be identified as downward-large (DL) and upward-large (UL) for sequences where the second pip was less than or greater than the CF by 20%, respectively.

### TABLE 1. Results from regression analysis of Rayleigh statistics

<table>
<thead>
<tr>
<th>Pip Duration, ms</th>
<th>( r^2 )</th>
<th>Slope</th>
<th>( F )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0</td>
<td>0.059</td>
<td>-0.230</td>
<td>1.77</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>12.0</td>
<td>0.026</td>
<td>0.146</td>
<td>3.71</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>18.0</td>
<td>0.526</td>
<td>0.345</td>
<td>350.4</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>24.0</td>
<td>0.740</td>
<td>0.598</td>
<td>1,224</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>30.0</td>
<td>0.770</td>
<td>0.573</td>
<td>1,582</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

**FIG. 9.** Latency distributions as a function of the pip number within a sequence. A–E: distributions for each of the 5 pip durations. Each color represents the percentage of comparisons for the latency to each of the 2nd to 4th pips in the sequence relative to the latency to the 1st pip of the sequence. In each case, the distribution is clustered near a relative latency difference of 0 ms, indicating that the latency of the response to each pip in the sequence is largely unaltered.
Analysis of the overall firing rate across neurons to each of the six possible comparisons between the sequences revealed that very few neurons showed any significant differences in the overall firing rate (Fig. 3; repeated-measures ANOVA; P < 0.05). At best, roughly 10% of the neurons showed a significant difference in firing rate and only at the longest tone durations tested. It could be, however, that the overall rate is not different but the responses to the different pips in the sequence does vary with the type of sequence presented. Figure 10 shows such a neuron to the US and DS stimuli presented with 30-ms-tone durations. This neuron is representative of many in which the overall firing rate is similar between the two conditions but there is a very clear difference in the temporal features of the response, with a greater response to the lower frequency [CF – (CF/20) or 10.45 kHz] than to the higher frequency, regardless of the position of this stimulus in the sequence. Such neurons could therefore encode the temporal order of the stimulus sequence. To investigate how many cells showed such a response, we conducted the same analysis of firing rate but restricted the analysis times to those corresponding to each of the individual pips in the sequence. There were very few neurons that showed significant differences in the response to the first pip, as expected, because the frequency of the first pip was invariant across sequences (ANOVA; P < 0.05 for 0–4.8% of neurons across comparisons and tone duration; percentage expected by chance = 5%). A similar finding was also noted for the response to the fourth pip in the sequence (range 0–3.6% across comparisons). This implies that the frequency of the preceding tone pips had very little influence on the response to the CF tone presented as the fourth pip in the sequence. In contrast, there were relatively large numbers of neurons that did show a difference in the response to the second or third tone pips in the sequence depending on the frequency and order of presentation of the individual pips (Fig. 11). There were very few neurons that showed significantly different responses at the shortest tone-pip durations (6 and 12 ms). For comparisons that differed in the order of presentation only (left panels) far more neurons showed significant differences in the responses to either the second or third pips for the larger frequency difference (DL vs. UL) than for the smaller frequency difference (DS vs. US). For comparisons where only the frequency differed but the order remained the same (middle panels), there were generally fewer neurons showing significantly different responses, again only for the longer tone durations. Finally, and somewhat surprisingly, changing both order and frequency had little additive effect compared with the large frequency difference where only the order was changed (compare DL vs. UL in the left column to the results in the right column). Regression analysis of the individual neurons that showed this effect did not reveal any suggestion that these differences were correlated with the CF of the neuron under study ($r^2 = 0.03; P > 0.05$).

**DISCUSSION**

The results from these experiments indicated that near the pip/interpip duration of 18 ms (about 28 Hz), the neurons were able to reliably extract distinct tones within a four-tone-pip sequence. Further, a large minority of neurons were also able to discriminate between the different tone-pip sequences based on the order of presentation, frequency differences, or both, with the longest pip intervals tested (24 and 30 ms) revealing close to 40% of tested neurons having significantly different firing patterns. This would indicate that, perceptually, the threshold for discriminating the individual tone-pip elements should be near 18 ms, with the threshold for determining the order of presentation slightly greater. To our knowledge there have been no equivalent psychophysical studies in macaques and the support for our observations from human psychophysical experiments are mixed. This is most likely explained by the different paradigms used. For example, Hirsh (1959) found that subjects could discriminate the order of long tone durations (500 ms) with temporal separations as short as 2 ms. A second study found 200-ms durations were necessary to discriminate order using different stimuli (Barsz 1988; Warren 1974a), although with training, discriminating the order of noise, tones, and buzz sequences can occur at ≤10 ms (Warren 1974b).
when either the 2nd or 3rd tone pip was considered.

stimuli, but fairly common for the longest pip duration stimuli, particularly
where both the order and frequency differed (US vs. DL or UL vs. DS). Overall
intervals between pips and the fact that we presented four pips
monkey, however, are in the relatively short duration and
ple. Key differences between this and previous studies in the
did see both response facilitation and attenuation in our sam-
2005; Brosch et al. 1999; McKenna et al. 1989). We certainly
focused on attenuation and/or enhancement of a response when
the sequence was different [downward-large vs. upward-large (DL vs. UL) or
order did not (US vs. UL or DS vs. DL). Rightmost column: comparisons where both the order and frequency differed (US vs. DL or UL vs. DS). Overall
it was rare that such differences were observed for the shortest pip duration stimui, but fairly common for the longest pip duration stimuli, particularly when either the 2nd or 3rd tone pip was considered.

Previous studies in the auditory nervous system using dis-
crete stimuli (as opposed to amplitude-modulated stimuli) focused on attenuation and/or enhancement of a response when a previous stimulus was presented (e.g., Bartlett and Wang 2005; Brosch et al. 1999; McKenna et al. 1989). We certainly
did see both response facilitation and attenuation in our sam-
ple. Key differences between this and previous studies in the
monkey, however, are in the relatively short duration and
intervals between pips and the fact that we presented four pips

instead of two. This allowed us to investigate the limits of
temporal resolution that these neurons could encode rapid
sequences. Our thresholds are in reasonably good agreement
with those seen for amplitude-modulated stimuli seen in a
variety of species and auditory areas (e.g., for review see
Langer 1992), although we rarely saw synchronized responses
much below 18-ms-pip durations (e.g., ≥28 Hz), indicating
that A1 neurons in alert macaque monkeys have difficulty in
reliably following such fast temporal sequences.

Recent studies is the alert marmoset also indicated that A1
neurons could potentially use two different coding schemes to
encode temporal rate information. For low-frequency stimuli,
neurons can use a temporal code by firing in synchrony with
the stimulus. At higher frequencies, the neurons can then shift
to a rate code where the overall numbers of action potentials is
related to the stimulus rate (Lu et al. 2001). Our results are in
agreement with that study in a few respects. As in the marmo-
set, we did not observe reliably synchronous responses for
intervals <18 ms and the degree of synchrony increased with
increasing interval duration. However, we saw little if any
evidence that there were two distinct populations of neurons.
This may be explained by the limited range of temporal rates
that we investigated (16–83 Hz). If this is the case, then a
neural population that encodes high temporal rates in macaques
must have cutoff rates near ≥100 Hz.

We used four tone pips with varying frequencies in the
middle two pips to see whether we could gain some insights
into how auditory cortical neurons could potentially differen-
tiate between these different sequences. One potential diffi-
culty is that we defined the stimuli based on the characteristic
frequency that was assessed qualitatively by the experimenter.
It is possible that using a stimulus that was not accurately
centered at the characteristic frequency, or was centered on a
subpeak of the tuning curve, could have changed the firing rate.
We doubt that this was the case because most neurons re-
sponded reliably to each of the tone pips in the sequence (e.g.,
Figs. 2, 5, 6, and 10), although it remains a possibility. It is also
possible that those neurons that did show differences between
the stimuli were the most sharply spectrally tuned neurons; one
frequency was inside the tuning curve and the other was not.
Although we did not quantitatively assess the spectral tuning of
these neurons, we feel that this is unlikely. Because there was
no real difference between the DL versus UL comparison (20%
greater vs. 20% smaller) and the US versus DL comparison
(20% greater vs. 5% smaller), this would indicate that 35–40%
of neurons have bandwidths of roughly 20% from the CF at
this intensity level, which corresponds to a Q-value of 0.40.
Previous studies in which spectral bandwidths in A1 were
measured under nearly identical conditions (and from these
same two monkeys) indicated that only 11/413 neurons had a
Q-value in this range at 40 dB above threshold (see Fig. 12G
of Recanzone et al. 2000a).

Anecdotally, it was extremely difficult for human observers
to differentiate between the different elements of the sequences
at the fastest rates, but they could for the slowest rate se-
quences. It is unlikely that the different sequences could be
differentiated by a simple rate code because we saw very little
difference between the overall firing rates between the different
sequence orders. However, given recent results from Werner-
Reiss et al. (2006), which reported that neuronal responses to
the second auditory stimulus were often weaker when com-

**FIG. 11.** Percentages of neurons with significantly different responses to
the 2nd (A), 3rd (B), or either 2nd or 3rd (C) pip in the sequence depending on
the sequence presented. Leftmost column: comparisons when only the order of
the sequence was different [downward-large vs. upward-large (DL vs. UL) or
DS vs. US]. Middle column: comparisons where the frequency changed but the
order did not (US vs. UL or DS vs. DL). Rightmost column: comparisons where both the order and frequency differed (US vs. DL or UL vs. DS). Overall
it was rare that such differences were observed for the shortest pip duration stimui, but fairly common for the longest pip duration stimuli, particularly when either the 2nd or 3rd tone pip was considered.
pared with the preceding stimuli, it is possible that a rate code might be used if the durations of the interstimulus interval (ISI) are in the range of hundreds of milliseconds instead of tens of milliseconds. Likewise, using stimulus and ISI rates slower than ours (230 vs. 736 ms), Ulanovsky et al. (2004), using an oddball paradigm, reported that neurons responded more strongly to the same frequency when it was the deviant stimulus than when it was the standard. Thus within the context of stimulus history and its potential role in differentiating sequences, it is possible for neurons to distinguish among the different sequences if the rates are sufficiently slow. However, at the level of our analysis a temporal code could indeed account for perceptual differences (Figs. 10 and 11). These results imply that the synchronization observed in the neuronal responses is a good indicator of the ability of these neurons to detect each element of these tone-pip sequences. These results also predict that there would be little perceptual gain by changing both the order and frequency once the frequency differences become large. These data also indicate that any type of “feature extraction” of such stimuli must occur at stages subsequent to the primary auditory cortex in the ascending auditory pathway.

ACKNOWLEDGMENTS

The authors thank the California National Primate Research Center for expert veterinary care and D. Guard, P. Geiger, and K. Su for participation in these experiments.

Present address of M. L. Phan: Psychology Department, Rutgers University, 152 Frelinghuysen Rd., Piscataway, NJ 08854.

GRANTS

This research was funded in part by National Institute on Deafness and Other Communication Disorders Grant DC-02371, the Mind Institute, Sloan Foundation, and Klingenstein Foundation.

REFERENCES


learning impaired children improved with acoustically modified speech.


