

Responses of Neurons in Cat Primary Auditory Cortex to Bird Chirps: Effects of Temporal and Spectral Context

Omer Bar-Yosef,¹ Yaron Rotman,¹ and Israel Nelken²

¹Department of Physiology, Hebrew University-Hadassah Medical School, and ²Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem 91120, Israel

The responses of neurons to natural sounds and simplified natural sounds were recorded in the primary auditory cortex (AI) of halothane-anesthetized cats. Bird chirps were used as the base natural stimuli. They were first presented within the original acoustic context (at least 250 msec of sounds before and after each chirp). The first simplification step consisted of extracting a short segment containing just the chirp from the longer segment. For the second step, the chirp was cleaned of its accompanying background noise. Finally, each chirp was replaced by an artificial version that had approximately the same frequency trajectory but with constant amplitude. Neurons had a wide range of different response patterns to these stimuli, and many neurons had late response components in addition, or instead of, their onset responses. In general, every

simplification step had a substantial influence on the responses. Neither the extracted chirp nor the clean chirp evoked a similar response to the chirp presented within its acoustic context. The extracted chirp evoked different responses than its clean version. The artificial chirps evoked stronger responses with a shorter latency than the corresponding clean chirp because of envelope differences. These results illustrate the sensitivity of neurons in AI to small perturbations of their acoustic input. In particular, they pose a challenge to models based on linear summation of energy within a spectrotemporal receptive field.

Key words: auditory cortex; cats; natural sounds; electrophysiology; single neurons; frequency-modulated tones; bird chirps

Most experiments assessing response properties of neurons in the primary auditory cortex (AI) are performed using simple sound stimuli [tones, clicks, or broadband noise (BBN)] and somewhat more complex stimuli such as frequency- and amplitude-modulated tones (FM and AM tones). These stimuli are easy to manipulate, and the responses they elicit are simple to analyze. Such stimuli differ considerably from natural sounds, which are substantially more complex, both spectrally and temporally.

Recently, a growing number of studies have used more sophisticated sound stimuli as a basis for predictions of the responses to complex stimuli. Shamma and Versnel (1995), Shamma et al. (1995), Kowalski et al. (1996a,b), Versnel and Shamma (1998), Depireux et al. (2001), and Calhoun and Schreiner (1998) have shown that ripple spectra can be used to measure a neuronal spectrotemporal receptive field (STRF). STRFs have also been measured using random chords, as initially suggested by deCharms et al. (1998). Response predictions using linear summation of the stimulus energy weighted by the STRF are reported to be rather good (Schnupp et al., 2001).

Despite these successes, the question of how neurons in AI code complex natural sounds remains unresolved. For example, predicting the responses to a class of stimuli based on a characterization using unrelated stimuli (e.g., prediction of the responses to wideband stimuli based on narrowband characterization or vice versa) is often not very good (Wang et al., 1995; Rotman et al., 2001). Furthermore, some response properties of

AI neurons cannot be explained by energy summation models based on STRFs (Nelken et al., 1999). Last, STRFs in AI often have a temporal width of <100 msec. However, AI neurons have far longer memory for acoustic context, as shown for tones in broadband noise (Phillips, 1985) and in forward-masking paradigms (Calford and Semple, 1995; Brosch and Schreiner, 1997).

This study investigates the responses of neurons in AI to complex sounds using an alternative approach: instead of predicting the response to complex sounds from simple stimuli, the complex sound is simplified step by step. We hoped to generate simple stimuli that would evoke the same responses as the natural sounds but would be amenable to parametric manipulations.

We chose to work with bird chirps, which to a first approximation are FM tones. Relatively fast frequency transitions are common in animal vocalizations. Moreover, there is substantial information on responses to FM tones in AI neurons (Mendelson and Cynader, 1985; Heil et al., 1992a,b; Mendelson et al., 1993; Nelken and Versnel, 2000). The stimuli were simplified in three steps. The base version consisted of the chirps within a substantial temporal context. The first simplification consisted of using only a short segment containing the bird chirp. In the second step, the background noise was removed from the short segment. The third step consisted of using an artificial chirp that approximately follows the frequency trajectory of the clean chirp. Our results show that each simplification step exerts a substantial influence on the neuronal responses.

MATERIALS AND METHODS

Animal preparation. The data were collected from 10 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, i.m.) followed by ketamine (100 mg, i.m.). The cats received 0.1 mg of intramuscular atropine sulfate or atropine

Received Aug. 15, 2001; revised May 29, 2002; accepted June 17, 2002.

This work was supported by grants from the Israeli Scientific Foundation and the Human Frontiers Science Program.

Correspondence should be addressed to Israel Nelken, University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK. E-mail: israel@md.huji.ac.il.

Copyright © 2002 Society for Neuroscience 0270-6474/02/228619-14\$15.00/0

methyl nitrate. The radial vein was cannulated, and the animals received a continuous infusion of lactated Ringer's solution at a rate of 10 ml/hr. Blood pressure was monitored with a cannula inserted into the femoral artery. Heart rate was continuously monitored, and body temperature was kept at $\sim 38^\circ\text{C}$ using a heat pad. The trachea was cannulated, and the cat received a mixture of oxygen and nitrous oxide (30 and 70%) and halothane (0.2–1.5%) for respiration. Breathing rate, quality, and CO_2 levels were continuously monitored. In case of respiratory resistance, the cat was paralyzed with pancuronium bromide (0.05–0.2 mg given every 1–5 hr, as needed) or vecuronium bromide (0.25 mg given every 0.5–2 hr). The temporal muscles were retracted to uncover the skull and the external auditory meatuses on both sides. The bullas were vented with a 30 cm polyethylene 90 tube. The skull was opened above the middle ectosylvian gyrus. The craniotomy was located just lateral to the supra-sylvian sulcus and contained the superior tip of the posterior ectosylvian sulcus to ensure access to the low-frequency representation in AI. The dura was left intact. At the end of the experiments, the cats were killed with a lethal dose of pentobarbital (50–100 mg, i.v.) 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.

Electrophysiological recordings. Extracellular recordings were performed using one to four glass-insulated tungsten microelectrodes (laboratory-made). Their impedance was 0.2–0.5 M Ω at 1 kHz. Each electrode was independently and remotely manipulated using a hydraulic drive (Kopf) or a four-electrode electric drive (EPS; Alpha-Omega, Nazareth, Israel). The electrical signal was amplified (MCP8000; Alpha-Omega) and filtered between 200 Hz and 10 kHz. The spikes were sorted online using a spike sorter (MSD; Alpha-Omega). The system was controlled by a master computer, which determined the stimuli, collected and displayed the data on-line, and wrote the data to files for off-line analysis. At the end of some of the microelectrode tracts, electrolytic lesions were made. These lesions were used to recover some of the electrode tracts. Responsive neurons were encountered at all depths.

Acoustic stimulation. The cat was placed in a soundproof room (Industrial Acoustics Company 1202). Artificial stimuli were generated digitally at a rate of 120 kHz, converted to analog voltage (DA3-4; Tucker-Davis Technologies), attenuated (PA4; Tucker-Davis Technologies), and electronically switched with a linear ramp (SW2; Tucker-Davis Technologies). Natural stimuli and their modifications were prepared as digital sound files and presented in the same way, except that the sampling rate was 44.1 kHz. Stimuli were delivered through a sealed calibrated acoustic system (Sokolich) to the tympanic membrane. Calibration was performed *in situ* by probe microphones (Knowles) precalibrated relative to a Brüel & Kjær microphone. The system had a flat (± 10 dB) response between 100 Hz and 30 kHz. In the relevant frequency range for this experiment (2–7 kHz), the system was even flatter (the response varied by less than ± 5 dB in all but one experiment, in which the variation was ± 8 dB). These changes consisted of relatively slow fluctuations as function of frequency, without sharp peaks or notches.

Sound stimuli. The natural stimuli were taken from field recordings (Library of Natural Sounds, Cornell Laboratory of Ornithology, Ithaca, NY). To ensure that the small number of stimuli used here are representative of a large set of natural sounds, an initial statistical analysis of a large corpus of natural sound was performed. Of >2 hr of recordings of soundscapes and single animals, ~ 20 min of representative sections were chosen for detailed analysis (Nelken et al., 1999). All the chirps in these sections were extracted, and three parameters were measured: the direction of frequency change, the extent of the frequency change, and its rate. The rate of frequency change was estimated as the slope, in the time-linear frequency plane, of straight lines connecting turning points of the frequency trajectories. Therefore, chirps containing only one upward or downward frequency change were assigned a single rate of frequency change; chirps that contained both upward and downward segments were assigned multiple rates of frequency change. In Figure 1, the histograms of the frequency extent and the rate of change are presented. Clearly, the majority of the FM tones in this sample had a relatively short extent (1–2 kHz) and a relatively low rate (most rates, for both upward and downward FM tones, are <80 kHz/sec). Upward and downward sweeps were about equally probable.

For the experiment, six representative chirps were chosen from this sample (Table 1, Fig. 2). The frequency range of the selected stimuli was 2.6–5.6 kHz. Their rate of change was 10–95 kHz/sec, and their extent was 0.1–2.4 kHz. In some of the stimuli, FM direction changed during the stimulus. Other stimuli had only one FM direction.

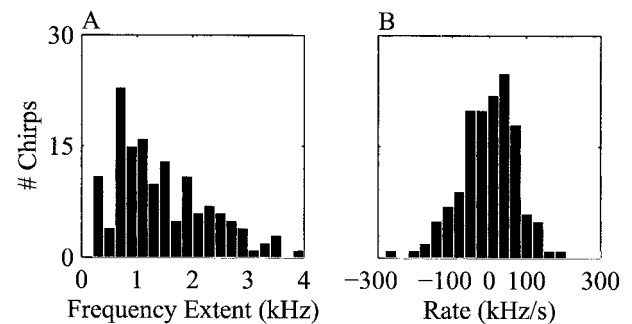


Figure 1. Histograms of parameters extracted from natural chirps. *A*, Chirp extent. *B*, Chirp rate. Positive values correspond to upward chirps, and negative values correspond to downward chirps.

Figure 2 displays the spectrograms and the waveforms of the six stimuli in four versions. To study the effect of the acoustic context, the six chirps were presented with 250 msec of the original recording before and after them (Fig. 2, bottom two rows), except for *Stimulus-6*, which was preceded by a segment of 180 msec. These stimuli are termed Long. The row on the right marked *Long (Full)* presents the full duration of the Long stimuli. The row above, marked *Long (Extract)*, displays magnified 100 msec segments of the Long stimuli around the selected chirp, which is marked by red lines. The chirps contained in the long stimuli are termed Natural, and they are shown, as used in the experiments, in the row marked *Natural* on the right. Two of the Long stimuli contained two of the six Natural stimuli each, and the other two contained one of the Natural stimuli each. *Stimulus-1* and *Stimulus-2* are contained in one Long stimulus, with a short interval consisting mostly of echoes between them. *Stimulus-3* and *Stimulus-4* are contained in another Long stimulus, in which *Stimulus-4* consists of the late part of *Stimulus-3*. *Stimulus-5* and *Stimulus-6* are contained each in a separate Long stimulus. *Stimulus-4* and *Stimulus-6* are sections of longer calls. They consist of one-directional sweeps and were chosen to include the simplest possible acoustic structures in the stimulus set.

The Natural stimuli were further modified to create four other versions of each stimulus. The second version consisted of the clean bird chirp (Figs. 2, 3, *Main*). *Main* was extracted from the full Natural stimulus in the following way: a fast Fourier transform (FFT) was computed on 256-point frames. It was used to locate the approximate center frequency of the bird chirp at that frame. The exact frequency of the peak of the (continuous-frequency) Fourier transform was then located by maximizing the exactly interpolated FFT values:

$$F(\omega) = \sum F(\omega_k) e^{-i(N-1)(\omega-\omega_k)/2} \frac{\sin(N(\omega-\omega_k)/2)}{\sin((\omega-\omega_k)/2)},$$

where $F(\omega)$ is the continuous-frequency Fourier transform, N is the length of the FFT, and ω_k values are the FFT frequencies. This formula gives the values of the discrete Fourier transform, evaluated at frequency ω , in terms of the Fourier transform computed at the FFT frequencies. The amplitude and phase of the Fourier transform at the peak frequency were used to generate one sample of the *Main* stimulus, corresponding in time to the center of the FFT frame. The FFT frame was shifted by one sample, and the procedure was repeated for each sample of the natural sound. The *Main* stimuli are presented in Figure 2 (*Main*). The success of this procedure can be judged by the close similarity of the waveforms of *Main* and *Natural*, as presented in Figure 3. In particular, the onset segments of the two stimuli are very similar, and the remaining noise forms only a small perturbation of the waveform.

The third type of short stimulus (*Artificial*) was generated as a frequency-modulated, constant-amplitude tone whose frequency trajectory consisted of straight line segments interpolating between turning points of the frequency trajectory of *main*. The amplitude of the *Artificial* stimuli was set to the highest amplitude of the corresponding *main* stimulus. The *Artificial* stimuli are presented in Figure 2 (*Artificial*).

The last two stimuli were generated for testing the importance of the temporal modulation pattern of the *Main* stimulus, which can clearly be seen in the waveforms of Figures 2 and 3. They were generated by imposing the temporal envelope of each of the *Main* and *Artificial* versions on the other version. This manipulation resulted in the *MainEnv* stimuli, which have the frequency trajectory of the *Main* stimuli with the

Table 1. General description of the stimuli

Stimulus	Starting frequency (kHz)	Frequency range (kHz)	Slowest rate of modulation (kHz/sec)	Fastest rate of modulation (kHz/sec)	Direction ^a	Maximal amplitude (dB) ^b
1	2.6	2.6–4.5	24	50	+/-	-7.3
2	4.2	4.2–5.6	28	30	+/-	-3.9
3	4.2	3.0–5.4	48	95	+/-	-8.7
4	5.4	3.0–5.4	48 ^c		-	-8.7
5	4.1	4.1–4.9	10 ^c		+	-7
6	4.9	4.1–4.9	26 ^c		-	-13.8

^a+, Rising frequency; -, falling frequency.

^bWith respect to an arbitrary common level.

^cOnly one direction.

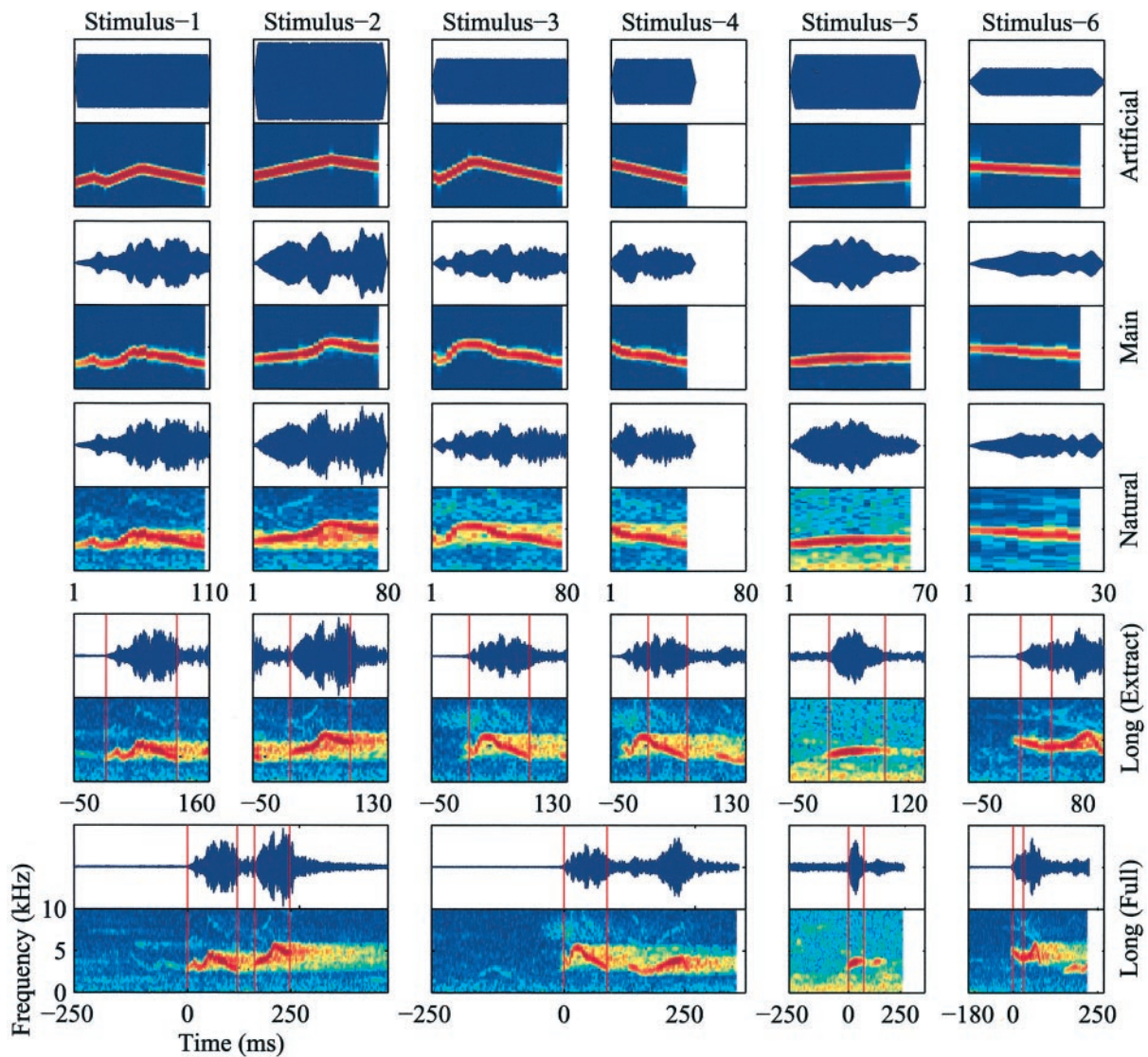


Figure 2. Stimuli used in the experiments. Each stimulus is displayed as a waveform (*top panel* of each pair) and as a spectrogram (*bottom panel* of each pair). The waveform is in analog-to-digital (A/D) units and has the same scale in all panels. Similarly, the values of the spectral energy are in decibels with respect to 1 A/D unit, and the color range is the same in all panels. The frequency axis is always 0–10 kHz. *Two bottom rows*, Long stimuli. *Long (Full)*, Entire time course of the Long stimuli, with the selected chirps marked by the red lines, except that *Stimulus-4* is part of *Stimulus-3* and is not marked separately. *Long (Extract)*, Magnified segment of the Long stimulus around the time of the chirp. The *top three rows* show three versions of the short stimuli: the *Natural* version, which is identical to the chirp that appears in Long; the *Main* version, which is the tonal component cleaned from the noise, including its frequency and amplitude modulation; and the *Artificial* version, which has a simplified frequency trajectory as Main and a constant temporal envelope.

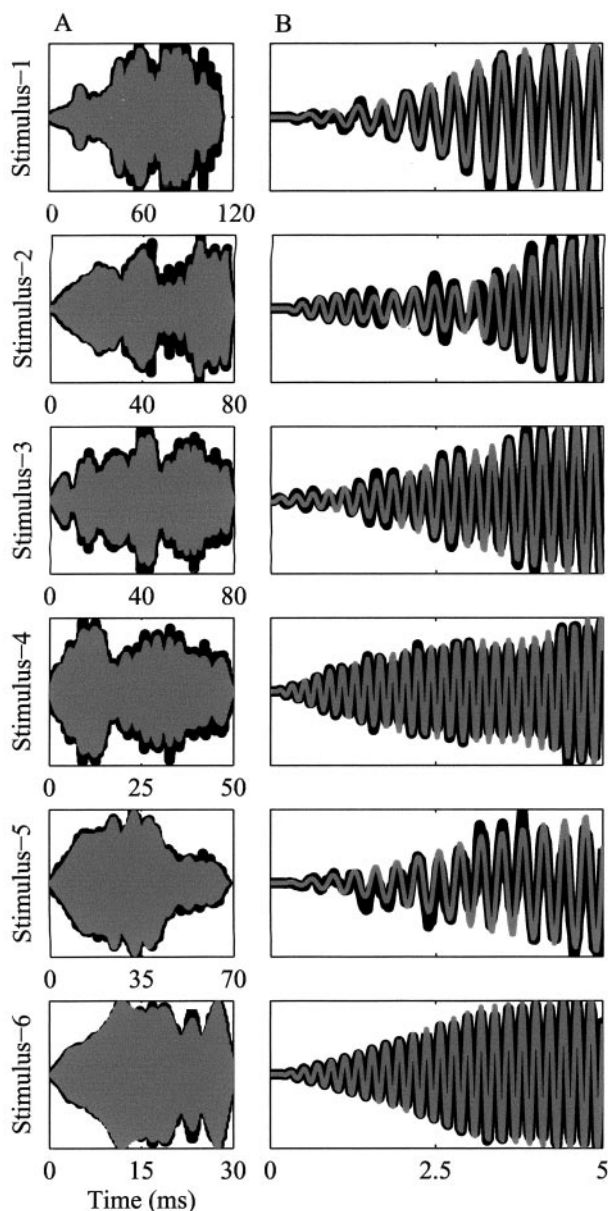


Figure 3. Comparison between the Natural and Main versions of all the stimuli. *A*, Waveforms of Natural (black, in the background) and Main (gray, in the foreground). Note the similarity in the temporal envelopes. *B*, Magnified view of the first 5 msec of each stimulus (Natural in black, Main in gray). The tonal component, Main, starts right at the beginning of Natural; there is no delay between the onset of the background noise and the onset of the tonal component.

constant temporal envelope of Artificial, and ArtEnv, which have the frequency trajectory of the Artificial stimuli with the temporal envelope of the corresponding Main stimuli (data not shown).

All stimuli were gated in the same way using a 3 msec, ramp-shaped rise and fall time. The use of such a short ramp was justified by the fact that each of the Natural stimuli had in fact a much longer rise time caused by the Natural temporal envelope. Thus, the levels of the Natural, Main, and ArtEnv versions at the end of the 3 msec ramp were much lower than the levels of the Artificial and MainEnv versions, by as much as 20–30 dB, depending on the stimulus. On the other hand, the total energy of the Natural, Main, and ArtEnv stimuli was lower only by ~6 dB than the total energy of the Artificial and MainEnv stimuli.

Experimental protocol. The microelectrodes were inserted into the low-frequency area of AI as described by Reale and Imig (1980). Each unit was characterized manually by determining approximately its best

frequency (BF) and its threshold to BBN bursts, all presented at a rate of 1/sec. Next the preferred aurality was determined using BBN rate level functions to the left (ipsilateral) ear alone, to the right (contralateral) ear alone, and to both ears together. The remainder of the experimental paradigm was performed at the preferred aurality. Frequency response area (FRA) was measured using a matrix of 45 frequencies logarithmically spaced from 0.1 to 40 kHz and 11 sound levels linearly spaced between 99 and 12 dB of attenuation [corresponding in most cases to a range of 10–100 dB sound pressure level (SPL), although because of the fluctuations in the acoustic calibration, some neurons have been tested down to ≤ 0 dB SPL]. All stimuli during this preliminary characterization phase were 115 msec long, with 10 msec linear rise and fall ramps. Finally, all versions of the natural stimuli were presented 20 times each in a pseudorandom order. The presentation level was always 20 dB above the neuron BBN threshold. Because of the large number of stimuli, not all neurons were tested with all chirps or all versions of each chirp. All stimuli, natural and artificial, were presented within trials whose duration was 1 sec. Stimulus onset was 200 msec after start of trial.

Data analysis. To compare results across neurons the neuronal responses were normalized for each neuron separately as:

$$\frac{\text{response rate} - \text{spontaneous rate}}{\text{maximal response rate} - \text{spontaneous rate}}$$

where the spontaneous rate of the neuron was estimated by its activity during the first 200 msec of each trial, just before stimulus onset, and the maximal response rate was taken over the responses to all natural stimuli in all their versions. Unless otherwise stated, the response rates were computed over an interval that consisted of the whole duration of the stimulus plus 10 msec after stimulus offset. In some cases different intervals were selected, as explicitly detailed in Results. Also, unless otherwise stated, the rates of the responses to Long were always estimated only from the interval containing the Natural stimulus.

In some of the scatter plots, it was necessary to check the reason for the large distribution of points around the equality line $x = y$. To test the effect of the estimation noise in each point on the width of the scatter, the following procedure was used. The orthogonal distance between each point and the diagonal was computed. Then, the SEs of each point along the abscissa and the ordinate were determined, and the variance along the orthogonal line connecting the point and the $x = y$ diagonal was calculated under the assumption that the true distribution is Gaussian with principal axes along the abscissa and ordinate. The square distance of the point from the diagonal was normalized by this variance. These normalized squared distances, one for each point in the scatter plot, are expected to have a χ^2 distribution with 1 df. The number of the points whose squared normalized distances were >2 (corresponding to a distance of >1.414 SEs from the diagonal) was compared with the expected number based on this χ^2 distribution.

To quantify the correlations between experimental variables, such as rates in response to two versions of the same stimuli, the simple correlation coefficient could not be directly used, because each neuron contributed multiple measurements (one for each stimulus), and each stimulus contributed multiple measurements (one for each neuron). To solve both problems, an approach similar in spirit to ANOVA has been used. To simplify the description, details will be given for the case of estimating the correlation between the responses to the Main and Natural stimuli. The same method was used, with the appropriate modifications, for all the statistical tests of correlation.

The response to Natural was modeled in two ways. First, it was modeled as a sum of an effect attributable to the specific neuron (and common to all the responses of this neuron) and an effect attributable to the specific stimulus id (and common to all the responses evoked by this stimulus). This model assumed no correlation at all with the responses to Main. Second, the response to Natural was modeled again as a sum of an effect attributable to neuron identity and an effect attributable to stimulus identity, but this time an additional factor was added, which was a linear dependency on the response to the Main version of the same stimulus. This procedure assumed that the ideal relation between Natural and Main had a common slope for all neurons; initial testing using ANCOVA showed that these slopes indeed were not significantly different from each other in the majority of cases. The same procedure was used even in the other cases, because it was conservative (if, under this assumption, a significant correlation can be shown, then a significant correlation will be present also under the more detailed model).

To judge the significance of the relationship between Natural and

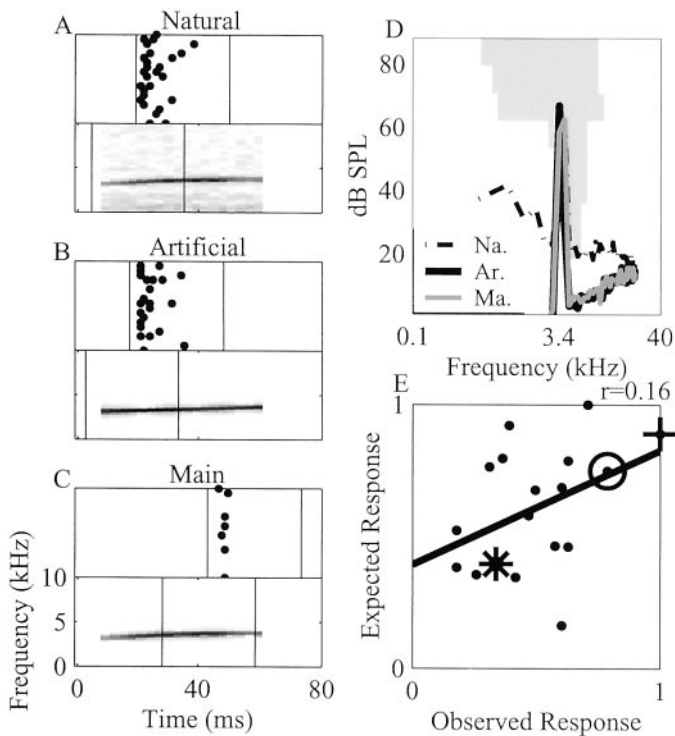


Figure 4. Generation of FRA predictions. *A–C*, Responses to the three short versions (*Natural*, *Main*, *Artificial*) of one stimulus. The estimated response onset is marked on the raster plot. Below the raster plots, the spectrograms of the three sounds are presented. The section of the sound that starts 15 msec before response onset lasts 30 msec and is assumed to have produced the response. *D*, FRA of this neuron, with each frequency level tile set to 1 if it is inside the tuning curve and to 0 otherwise. The power spectra of the three sections from *A–C* are superimposed. These power spectra were resampled at the frequencies at which the FRA was measured. The FRA prediction is the number of frequency level combinations that were traversed by the power spectra. *E*, Scatter plot of the FRA predictions versus the actual responses of the entire set of stimuli. Only a weak correlation is found for this neuron. The reported r is the square root of the increase in explained variance and is not significantly different from 0.

Main, the increase in the amount of explained variance attributable to the addition of the responses to *Main* was used as the basic test statistic. A standard F test for the significance of the increase in explained variance attributable to the addition of *Main* was performed (Sokal and Rohlf, 1995). The square root of the increase in explained variance corresponds to the absolute value of the usual correlation coefficient, except that it is adjusted for the effects of neurons and stimuli. In most cases, both the increase in explained variance and its square root (called the adjusted correlation coefficient below) are reported.

To quantify the relationships between the FRA and the responses to the natural sounds and their modifications, the procedure outlined in Figure 4 was used. First, the borders of the FRA were marked manually. Second, the response onset time (ROT) was determined for the response of each neuron and each stimulus separately. Then the number of spikes in the 30 msec after the ROT was determined, and the spectral energy of the sound segment starting 15 msec before the ROT and ending 15 msec after the ROT was computed. Because the onset time was different for each neuron–stimulus combination, the corresponding integrated spectral energy was also different. The value of 15 msec was used, because it was the typical latency of the neuronal response at BF. Finally, the spectral overlap between the stimulus spectrum and the FRA was estimated by counting the number of spectral bins in which the stimulus spectrum occurred within the FRA (Fig. 4*D*). The significance of the relationship between predictions and measured responses was judged by the increase in the explained variance of the responses, after adjustment for stimulus identity, that resulted from the addition of the prediction as regression variable (Fig. 4*E*). The prediction procedure could be per-

formed in various other ways, such as by summing the actual FRA rates at the bins where it intersected with the stimulus spectrum. However, the version used here produced higher correlation coefficients in most cases.

RESULTS

In total, 200 well separated neurons were recorded from 10 cats. Seventy-seven neurons were chosen for further analysis based on their stable response during the recording session (1–2 hr).

General characteristics of the neurons

The BFs of the neurons in this sample ranged from 1 to 15.5 kHz. The frequency range of the *Main* chirps was 2.6–5.6 kHz. Because of the width of the FRA, neurons with BFs between 2 and 7 kHz are considered as intersecting the frequency range of the stimuli. Most (43 of 77) of the neurons had BFs between 2 and 7 kHz. Neurons recorded in all layers are included here. There were no obvious differences in response characteristics as a function of depth.

FRAs were mostly narrowly tuned. Values of BF divided by the FRA bandwidth 20 dB above threshold were distributed between 0.3 and 4.5, with 40 of 77 of the values >1 . Thresholds at BF were between 0–50 dB SPL (mean \pm SD, 21 ± 11 dB SPL). Thus, the neuronal sample is typical of primary auditory cortex, as expected from the physiology and from the general anatomical recording location. Whenever histological reconstruction of the penetrations were performed, recording locations in AI were confirmed.

Neurons responded to the natural stimuli with response components dispersed throughout the stimulus duration and not restricted to stimulus onset, as is often the case when using barbiturate anesthesia (for responses to natural sounds under barbiturate anesthesia, see Rotman et al., 2001). To quantify this effect, the mean early activity (the first 45 msec of the stimulus) was compared with the late activity (at a time window from 45 msec until 10 msec after stimulus offset), for stimuli 1–3 and 5 (whose durations were >65 msec) and all their short versions. The early and late activities were not significantly different on the average [$F_{(1,1667)} = 0.7$; NS, four-way ANOVA on time section (early vs late) \times neuron \times stimulus id \times stimulus version].

Effect of acoustic context

Figure 5 presents examples of responses to *Long*, *Natural*, and *Main*. Each column presents the responses of one neuron. All the neurons had an FRA intersecting the frequency range of the stimuli (*top row*). On the row marked *Long (Full)*, the responses to the whole *Long* stimulus are presented. The row marked *Long (Extract)* presents again the response to the *Long* stimulus but at an extended time scale, corresponding to the time scale used to present the responses to *Main* and *Natural* in the *top two rows*. To simplify the descriptions, the response to *Natural* in the context of *Long* is termed below the response to *Long*.

The neurons in Figure 5*A,B* had similar responses to *Long*, to *Natural*, and to *Main*, although the responses to *Long* were somewhat weaker than the responses to *Natural* and *Main*. The other neurons had different responses to the three stimuli. The neuron in Figure 5*C* responded only to *Natural* and to *Long*, with very weak response to *Main*, if at all. The neurons in Figure 5*D,E* responded only weakly to *Long*, probably because of a previous response to the background noise. The responses to *Natural* and *Main* had different temporal patterns in Figure 5*D* and a similar pattern in Figure 5*E*. There were significant continuous responses to the segment just before *Natural* in Figure 5*B,C,E*, although no clear acoustical component appeared inside the FRA of the neuron.

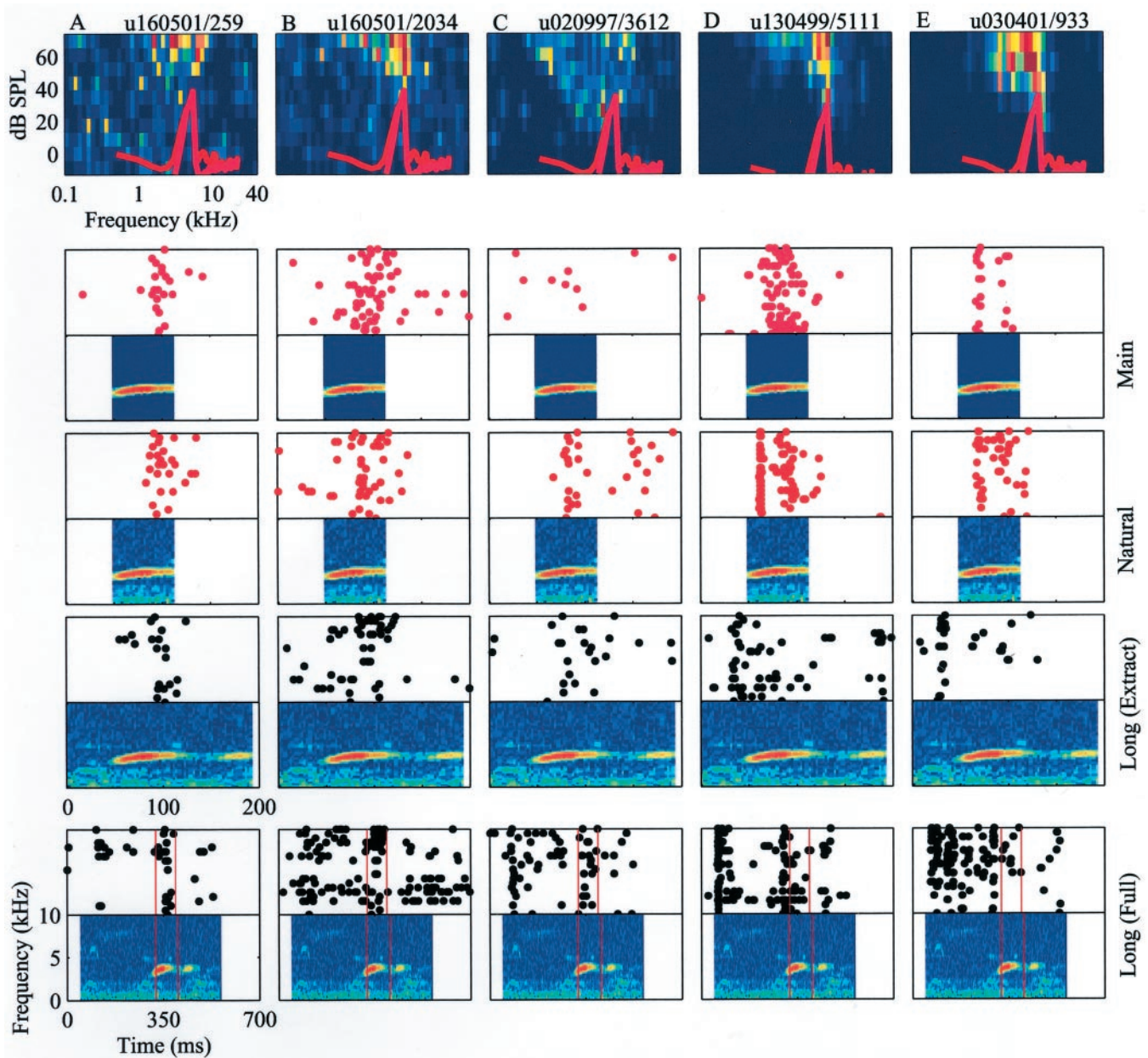


Figure 5. Responses of five neurons to the Long, Natural, and Main versions of stimulus 5. The FRAs of the neurons are displayed in the *top row* (BFs: A, 2.4; B, 5.4; C, 3.4; D, 5.4; E, 6.1 kHz). The *color scale* represents the firing rate, in which *blue* is 0 and *red* is the maximal rate of each neuron (A, 47; B, 86; C, 63; D, 109; E, 136 spikes/sec). The spectra of Main (*magenta*) and Natural (*red*) at the actual level in which they were presented is plotted at the *top* of the FRA. The responses are displayed as a *raster plot* above the *spectrogram* of the corresponding stimulus, plotted as in Figure 2. The *two bottom rows* display the responses to the Long stimuli. *Long (Full)*, Responses to the entire duration of Long. The *red lines* mark the borders of Natural. *Long (Extract)*, Magnified view of the segment in Long that contains the Natural stimulus. The responses to Natural and Main are displayed in the correspondingly marked rows. The time axes in the rows marked *Long (Extract)*, *Natural*, and *Main* are all aligned.

In Figure 6, the responses of one cell to all the stimuli are presented. The same diversity of responses to a single stimulus shown in Figure 5 across neurons can be seen in Figure 6 in the responses of a single neuron to the different stimuli. For example, the responses to the Long, Natural, and Main versions of *Stimulus-4* were somewhat similar. On the other hand, the robust onset response to the Natural version of *Stimulus-5* was essentially absent in the responses to the Long and Main versions of the same stimulus. In general, however, it appears that Long and Natural evoked similar responses, and these were different from the responses to Main. As also seen in Figure 5, the neuron in

Figure 6 responded during periods of background noise even when there were no clear acoustical components present inside its tuning curve.

These examples suggest that the acoustic context, both sequential (as in the transition between Long and Natural) and simultaneous (as in the transition from Natural to Main) plays an important role in determining the responses even to simple natural sounds. The effect of the sequential context, i.e., the relationships between the responses to Long and to the short stimuli, is considered first.

Figure 7 presents scatter plots of the normalized spike counts

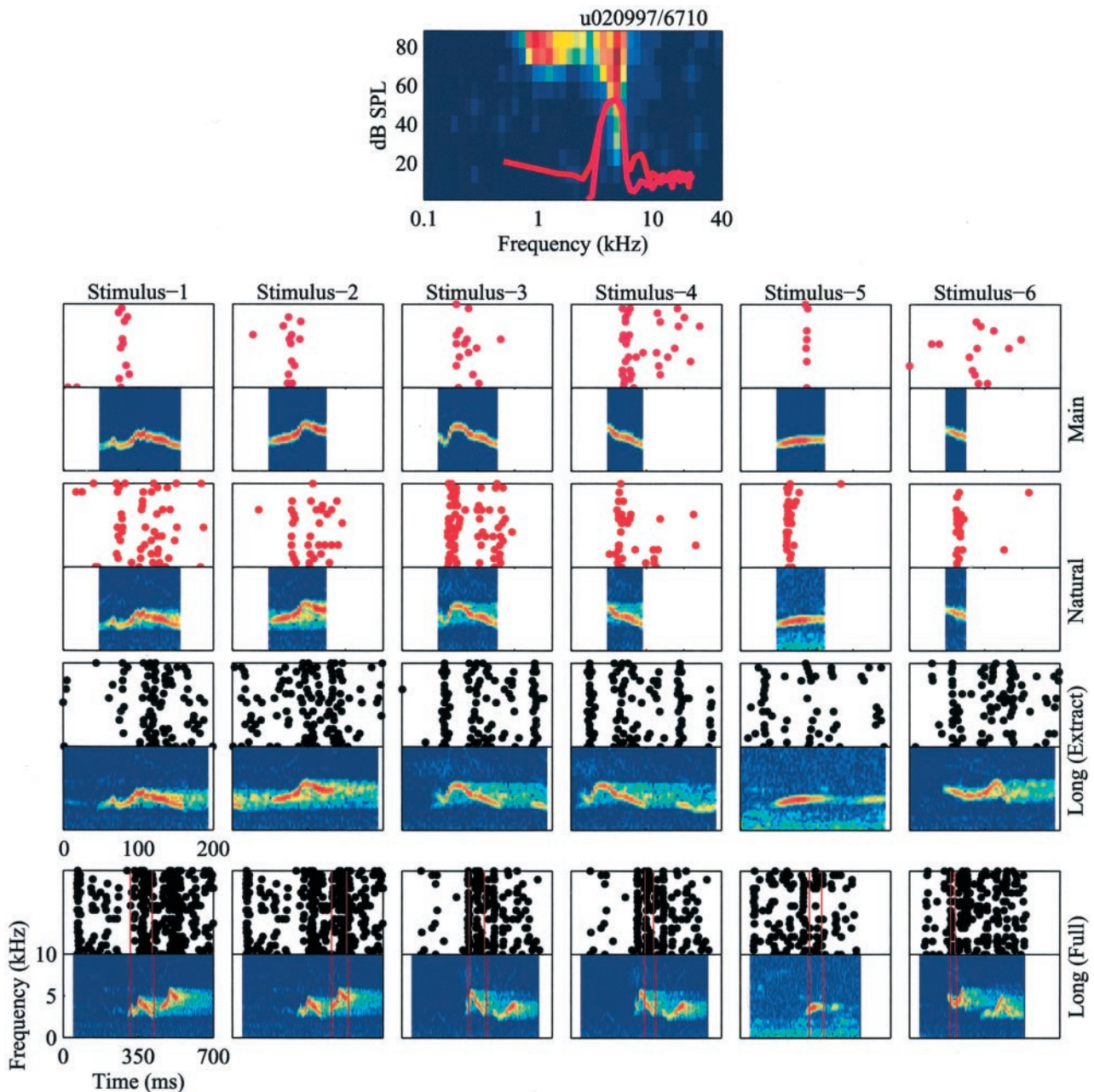


Figure 6. Responses of a single neuron to the Long, Natural, and Main versions of all six stimuli, in the same format as Figure 5. The BF is 4.7 kHz. The maximal firing rate is 260 spikes/sec.

over the whole neuronal population in response to Main, Natural, and Natural in the context of Long (referred to below as the responses to Long). Table 2 presents quantitative comparisons between these spike counts. Three-way ANOVA (version \times stimulus identity \times neuron) showed a highly significant main effect of stimulus version on response strength ($F_{(2,763)} = 30$; $p \ll 0.01$). Therefore, the differences between the average response strengths to Long and Main and to Long and Natural were tested *post hoc* separately using similar three-way ANOVA, and the main effect of version is reported in Table 2. Table 2 also reports the increase in explained variance attributable to the regression on Long (as described in Materials and Methods), its statistical significance, and its square root, the adjusted correlation coefficient, between the individual responses.

The differences in the response strength are described first. The responses to Long were weaker on average than the responses to Natural. The responses to Long were approximately equal on average to the responses to Main, although responses of individual neurons Main and Long could differ substantially, causing a large scatter around the diagonal in Figure 7. The weaker responses to Natural in the context of Long, relative to the responses to Natural, probably reflect some sort of neuronal adaptation. This adaptation can be simple fatigue or a more complex form of stimulus-specific adaptation [Movshon and Lennie, 1979 (visual system); Shu et al., 1993 (auditory psychophysics)]. The differences in the firing patterns in responses to Long and Natural (Figs. 5, 6) argue against simple fatigue, at least in some cases.

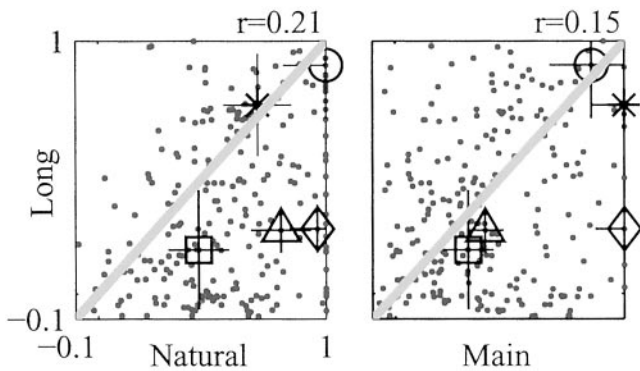


Figure 7. Scatter plots of the normalized spike count of Main and Natural versus Long. Each *point* represents the responses of one neuron to one stimulus, and there are a total of $n = 271$ neuron \times stimulus combinations. An $x = y$ equality line is plotted in gray. The responses displayed in Figure 5 are marked with the following symbols: *A*, circles; *B*, asterisks; *C*, squares; *D*, diamonds; *E*, triangles. Error bars indicate 1 SE.

If the adaptation is indeed simple fatigue, its cause would be the activity of the neuron just before the beginning of Natural, and its effect would be the reduction in the number of spikes evoked by Natural in the context of Long relative to the number of spikes evoked by Natural by itself. The cause of the fatigue was therefore quantified by the number of spikes in the 50 msec interval before the beginning of Natural. The effect of the fatigue was quantified by the difference in spike count between the responses to Long and to Natural. The correlation coefficient between the two was computed for each neuron separately. Stimuli 2 and 6 were not used for this analysis, because there is substantial acoustic energy preceding them. The correlation coefficients varied considerably across the population. Most of these correlation coefficients, however, were negative (44 of 68; $\chi^2 = 5.9$; $df = 1$; $p < 0.05$ against the null hypothesis of equal probability of positive and negative correlation coefficients), and the average correlation coefficient was negative ($t = -3.2$; $df = 67$; $p \ll 0.01$, one-tailed test against the null hypothesis of a mean ≥ 0). This finding argues against simple fatigue as the source of the adaptation.

Next, the shape of the scatter plots in Figure 7 is analyzed. The scatter plot of the responses to Natural and to Long is wedge-shaped, showing that although a weak response to Natural was associated with a weak response to Long, the reverse was not true in general. This is another manifestation of the adaptation described above, again hinting that the adaptation is not a simple firing rate fatigue. Both adjusted correlation coefficients between the responses to Long and to Natural and between Long and Main were significant, although the adjusted correlation between the responses to Long and Main was weaker. The adjusted correlation coefficients are, however, related to extremely small fractions of explained variance: only 4–5% of the total variance in the responses.

A possible reason for the low correlation coefficients is noise in the estimates of the normalized responses. To check the effect of estimation noise, the procedure outlined in Materials and Methods was used. In the scatter plot of the responses to Long against Natural, the expected number of points whose normalized squared distance from the diagonal was >2 was 42.9 of 273, whereas the actual number was 131 of 273 ($\chi^2 = 180$; $df = 1$; $p \ll 0.001$). In the case of the scatter plot between Long and Main, the expected number was 43.7 of 278, and the actual number was 121

of 278 ($\chi^2 = 136$; $df = 1$; $p \ll 0.001$). Thus, in both cases, estimation noise is not the only reason for the low correlation coefficients.

To illustrate the small effect of estimation noise, the points corresponding to the raw responses shown in Figure 5 are marked in Figure 7 together with their SEs. Clearly, the SEs of the points, which are far from the diagonal, do not cross it (*diamonds* in both panels, representing the data from Fig. 5*D*, *triangles* in the scatter plot of Long against Natural, representing the data from Fig. 5*E*).

Because most of the energy of Main and Natural lies between 2.5–6 kHz, neurons with FRAs intersecting this range might show a different pattern of correlation than those that do not. Therefore, the same tests were performed separately on the responses of neurons whose BFs were between 2 and 7 kHz and neurons whose BFs were outside of this range. The responses to Natural were correlated with the responses to Long both within and outside the stimulus frequency range, with the correlation outside this range being even slightly larger. In contrast, the correlation between the responses to Main and to Long was only significant within the stimulus frequency range. This correlation, although significant, was still small.

Responses to the short versions

Figure 8 presents the responses of five neurons to the Natural, Main, and Artificial versions of *Stimulus-1*. Figure 8*A–D* presents the responses of four neurons whose FRA intersected the frequency range of the stimulus. Under this condition, it might be expected that the responses to all three versions would be similar. This is not the case, as seen in Figure 8. In Figure 8*A*, the responses to Main and Natural were somewhat similar, consisting of an onset response followed by slowly decreasing activity throughout the rest of the stimulus, but the temporal pattern of the response to Artificial was different, consisting of a strong, well locked onset followed by an immediate return to a low firing rate. In Figure 8*B,C*, Natural and Artificial had a similar response. Main, the stimulus bridging the acoustic gap between them, evoked only weak responses. In Figure 8*D*, the responses to Natural and Main were more similar to each other than to Artificial, and Main elicited the strongest response. In contrast to these examples, the neuron in Figure 8*E* had a BF outside the frequency range of the chirps (1.7 kHz). At this low frequency, only Natural had any energy, and as expected, only Natural evoked a significant response.

The difference between the responses to Main and Natural is even more striking in Figure 9, in which the responses of one neuron to the Natural, Main, and Artificial versions of all six stimuli are presented. The FRA of this neuron and the spectral energy of the stimuli clearly overlapped. On the basis of the spectral energy, it could be predicted that all three versions of the same stimulus would evoke the similar responses. In fact, only one stimulus (*Stimulus-4*) had similar responses to Main and Natural. As a whole, the results produced an unexpected pattern.

Population analysis of response magnitudes

The responses displayed in Figures 8 and 9 suggest that, contrary to previous expectations, many neurons responded differently to Natural, Main, and Artificial. Figure 10 displays the scatter plots between the normalized responses to Natural, Main, and Artificial. The quantitative comparisons are reported in Table 3. As expected from the examples in Figures 8 and 9, the responses to Main were weaker on the average than the responses to either Natural or Artificial. The difference between the responses to

Table 2. Comparison between the normalized spike counts of Natural in the context of Long, Natural alone, and Main

Neurons	Natural-Long	Main-Long
All	$\Delta\text{resp} = 0.18$, $F_{(1,484)} = 46$; $p \ll 0.01$; $\Delta R^2 = 0.044$ ($r = 0.21$), $F_{(1,204)} = 18.9$; $p \ll 0.01$	$\Delta\text{resp} = 0.01$, $F_{(1,484)} = 0.12$; NS; $\Delta R^2 = 0.023$ ($r = 0.15$), $F_{(1,204)} = 9.4$; $p < 0.01$
2 kHz < BF < 7 kHz	$\Delta\text{resp} = 0.17$, $F_{(1,304)} = 30$; $p \ll 0.01$; $\Delta R^2 = 0.036$ ($r = 0.19$), $F_{(1,118)} = 9.8$; $p < 0.01$	$\Delta\text{resp} = 0.05$, $F_{(1,304)} = 2.6$; NS; $\Delta R^2 = 0.052$ ($r = 0.23$), $F_{(1,118)} = 15$; $p < 0.01$
BF < 2 kHz or 7 kHz < BF	$\Delta\text{resp} = 0.19$, $F_{(1,174)} = 17.5$; $p \ll 0.01$; $\Delta R^2 = 0.05$ ($r = 0.22$), $F_{(1,80)} = 7.8$; $p < 0.01$	$\Delta\text{resp} = -0.06$, $F_{(1,174)} = 1.6$; NS; $\Delta R^2 = 0.001$ ($r = 0.03$), $F_{(1,80)} = 0.19$; NS

Δresp , Average difference in normalized responses; ΔR^2 , increase in explained variance attributable to adding the responses to Long; r , adjusted correlation coefficient, square root of ΔR^2 .

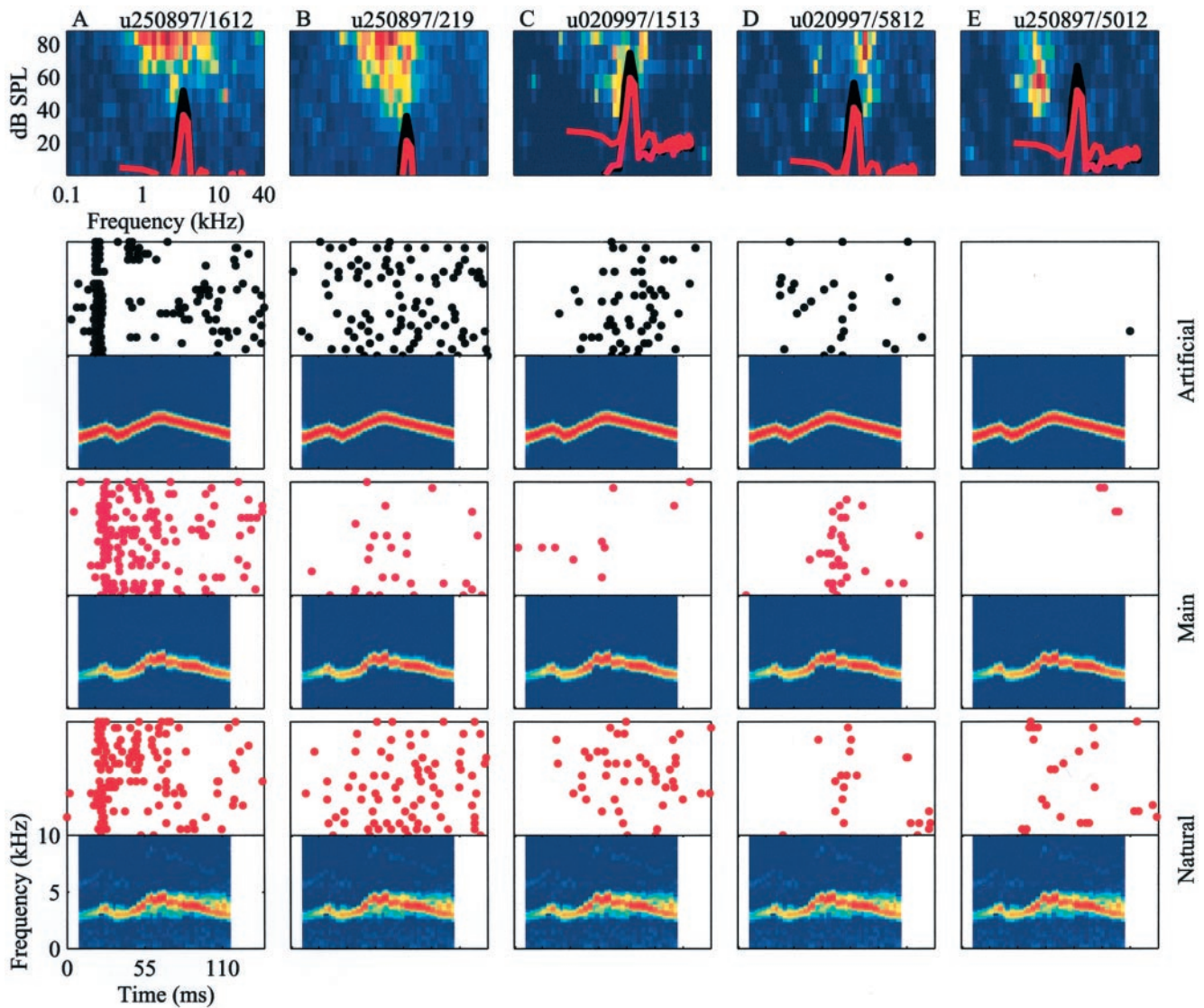


Figure 8. Responses of five neurons to the Natural, Main, and Artificial versions of stimulus 1. The *top row* presents the FRA of five neurons (BFs: *A*, 2.6; *B*, 2.7; *C*, 2.8; *D*, 5.3; *E*, 1.7 kHz). The *color scale* represents the firing rate, in which *blue* is 0 and *red* is the maximal rate of each neuron (*A*, 116; *B*, 128; *C*, 65; *D*, 180; *E*, 166 spikes/sec). The spectra of Natural (*red*), Main (*magenta*), and Artificial (*black*) at the actual level in which they were presented are plotted at the *top* of the FRA. The responses are displayed as *raster plots* above the *spectrogram* of the corresponding stimulus. The time axes in all rows are aligned.

Natural and Artificial was not significant on the average, although the responses of individual neurons to the two versions could be very different.

The correlations between the responses to all pairs of versions were significant but low. The correlation between the responses

to Natural and Main was lower than the correlations between the responses to Artificial and either of the other two stimuli, although acoustically, Main is more similar to Natural than Artificial.

One possible explanation for these results is that the estimates

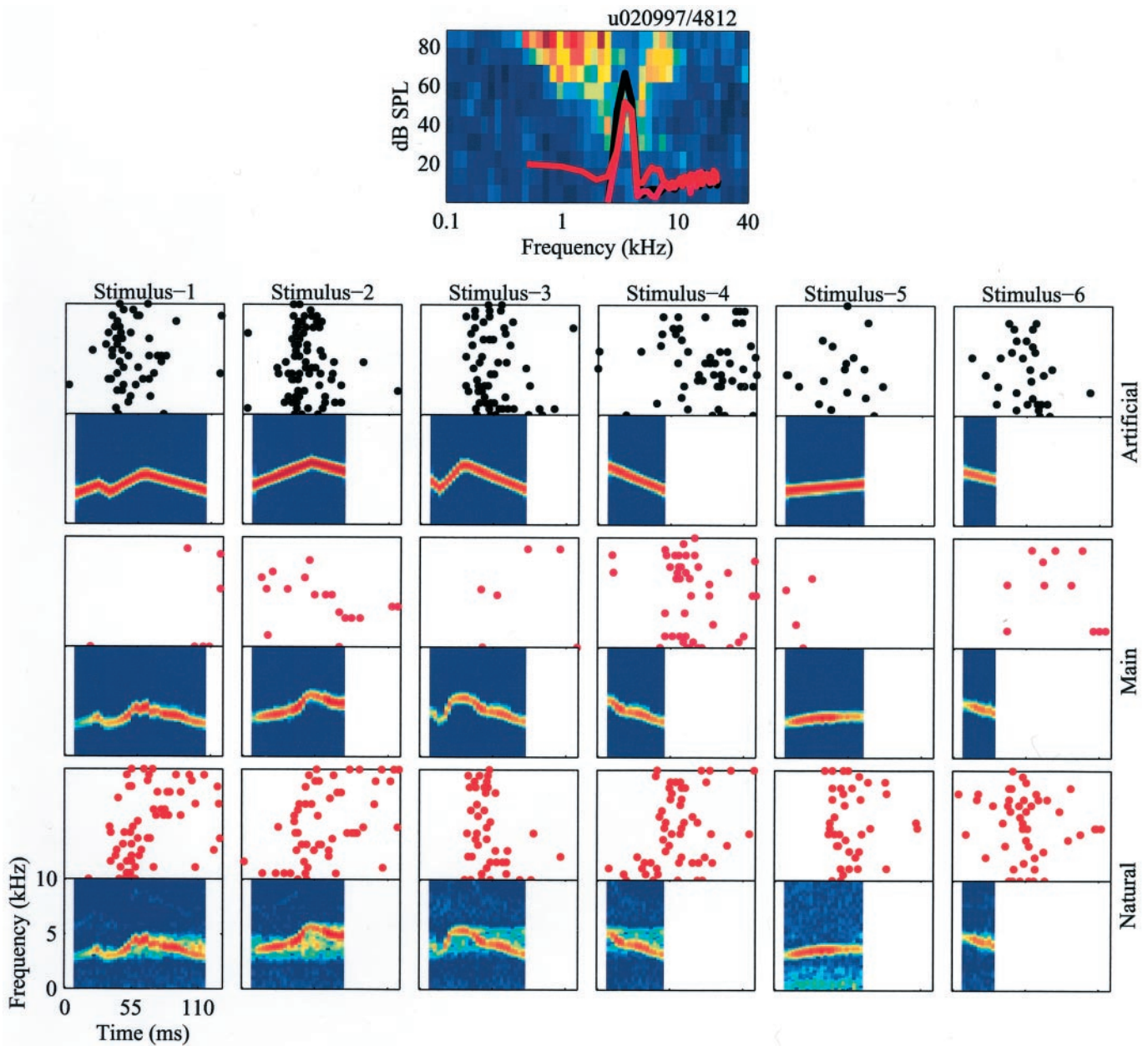


Figure 9. Responses of a single neuron to the Natural, Main, and Artificial versions of all six stimuli, in the same format as Figure 4. The BF is 4.8 kHz. The maximal firing rate is 245 spikes/sec.

of the response magnitude have large statistical variability, contributing to the scatter of the points in Figure 9. To check the effect of estimation noise, the procedure outlined in Materials and Methods was used. The expected number of points, whose normalized squared distance from the diagonal was more than two, was always significantly smaller than the actual number (Fig. 10, *A*, 36.4 vs 91 of 232, $\chi^2 = 81$, $df = 1$; *B*, 36.4 vs 102 of 232, $\chi^2 = 117$, $df = 1$; *C*, 50.1 vs 162 of 319, $\chi^2 = 249$, $df = 1$; all results $p \ll 0.001$). The relatively small effect of estimation noise is illustrated in Figure 10 by the symbols marking the data from Figure 8.

Another possible explanation for the weak average responses to Main and to the low correlations between the responses to Main, Natural, and Artificial is that neurons whose FRA did not intersect the frequency range of the stimuli responded differently

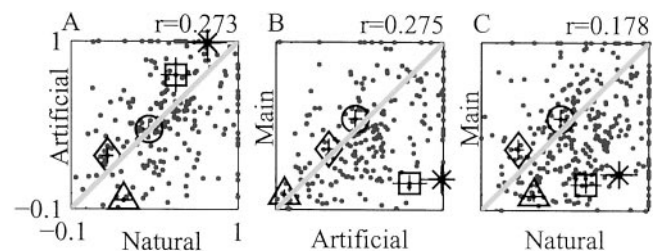


Figure 10. Scatter plots of the normalized spike count of the responses to the Natural, Main, and Artificial versions of all stimuli. Individual points are plotted for each neuron \times stimulus condition. An $x = y$ equality line is plotted in gray. The responses displayed in Figure 7 are marked with the following symbols: *A*, circles; *B*, asterisks; *C*, squares; *D*, diamonds; *E*, triangles. Error bars indicate 1 SE. The numbers of points are 232 in *A* and *B* and 319 in *C*.

Table 3. Comparison between the normalized spike counts of Natural, Main, and Artificial

Neurons	Natural-Main	Artificial-Main	Natural-Artificial
All	$\Delta\text{resp} = 0.09$, $F_{(1,561)} = 43$; $p \ll 0.01$; $\Delta R^2 = 0.03$ ($r = 0.18$), $F_{(1,239)} = 32$; $p \ll 0.01$	$\Delta\text{resp} = 0.12$, $F_{(1,471)} = 25$; $p \ll 0.01$; $\Delta R^2 = 0.08$ ($r = 0.27$), $F_{(1,179)} = 32$; $p \ll 0.01$	$\Delta\text{resp} = 0.03$, $F_{(1,471)} = 4.2$; NS; $\Delta R^2 = 0.08$ ($r = 0.27$), $F_{(1,179)} = 31$; $p \ll 0.01$
2 kHz < BF < 7 kHz	$\Delta\text{resp} = 0.12$, $F_{(1,282)} = 26$; $p \ll 0.01$; $\Delta R^2 = 0.04$ ($r = 0.19$), $F_{(1,133)} = 17$; $p \ll 0.01$	$\Delta\text{resp} = 0.14$, $F_{(1,315)} = 25$; $p \ll 0.01$; $\Delta R^2 = 0.12$ ($r = 0.35$), $F_{(1,111)} = 25$; $p \ll 0.01$	$\Delta\text{resp} = -0.05$, $F_{(1,282)} = 0.1$; NS; $\Delta R^2 = 0.19$ ($r = 0.43$), $F_{(1,111)} = 43$; $p \ll 0.01$
BF < 2 kHz or BF > 7 kHz	$\Delta\text{resp} = 0.12$, $F_{(1,240)} = 20$; $p \ll 0.01$; $\Delta R^2 = 0.02$ ($r = 0.15$), $F_{(1,100)} = 13$; $p < 0.01$	$\Delta\text{resp} = 0.01$, $F_{(1,183)} = 1.7$; NS; $\Delta R^2 = 0.03$ ($r = 0.18$), $F_{(1,62)} = 6.4$; $p < 0.01$	$\Delta\text{resp} = 0.11$, $F_{(1,183)} = 16$; $p \ll 0.01$; $\Delta R^2 = 0.006$ ($r = 0.07$), $F_{(1,62)} = 1.0$; NS

Δresp , Average difference in normalized responses; ΔR^2 , increase in explained variance attributable to adding the responses to Long; r , adjusted correlation coefficient, square root of ΔR^2 .

to Main and Natural or to Artificial and Natural, because the two pairs of stimuli differed in their spectral composition near the BF of these neurons. The correlations were therefore calculated separately for the subpopulation of neurons with BFs between 2 and 7 kHz and for the subpopulation of neurons whose BFs were outside this range (Table 3). Indeed, the correlations between responses of neurons inside the stimulus frequency range were higher, for all pair-wise comparisons, than the correlations between the responses of neurons outside the stimulus frequency range. However, the smallest correlation inside the stimulus frequency range was still between Natural and Main, despite the acoustic similarity between the two stimuli. Furthermore, the responses to Natural were significantly larger than the responses to Main both inside and outside the stimulus frequency range.

The somewhat higher correlation between the responses of Natural and Artificial could be the result of spectral splatter because of the fast rise time of Artificial. The onset of Artificial would be, according to this argument, similar to the onset of Natural, because both are composed of a central peak with a relatively wide band of spectral energy around it. Although no direct tests were performed to refute this possibility, spectral splatter is probably insufficient to explain the results. First, the bandwidth of the splatter, even for a 3 msec rise time, is only ~ 600 Hz, narrower than the bandwidth of the noise component in Natural (even when considering only the echoes). Second, although the strengths of the responses were equal on average, their temporal patterns were often different (Figs. 8, 9). Many of the response components were late and are probably not directly affected by the spectral splatter at the onset of the stimuli. Third, spectral splatter would selectively enhance the correlation between early responses to Natural and Artificial, but in fact the correlation between the early responses (0–45 msec) was lower than the correlation between the late responses [early, $\Delta R^2 = 0.01$ ($r = 0.12$), $F_{(1,180)} = 13$, $p < 0.01$; late, $\Delta R^2 = 0.11$ ($r = 0.33$), $F_{(1,180)} = 67$, $p \ll 0.01$].

Response predictions based on the FRA

Figure 11 shows a histogram of the fraction of variance in the responses to Main, Natural and Artificial that is explained by the regression on the FRA predictions (as illustrated in Fig. 4) after adjustment for the effect of the stimulus identity. The explained variance fractions were generally small, most being < 0.2 . Furthermore, only a few were significantly > 0 (white bars).

Neurons inside the stimulus frequency range did not have stronger correlations with the FRA predictions. For example, of the six neurons with significant correlations, only two had their BF inside the stimulus frequency range. Furthermore, the prevalence of small correlations was larger in the population of neurons inside the stimulus frequency range (inside, 28 of 35, 80%; out-

side, 14 of 20, 70%), although this difference is not significant statistically.

It could also be that for neurons within the stimulus frequency range, the correlations between FRA predictions and responses were low because the range of values of either the predictions or the responses was small. In such cases, the true relationship between the two measures cannot be ascertained. However, the correlation between the range of FRA predictions for each individual neuron and the fraction of explained variance for each neuron was actually negative although not statistically different from 0 ($r = -0.14$; $df = 53$; NS). Similarly, the correlation between the range of responses and the fraction of explained variance was positive but also not significant ($r = 0.19$; $df = 53$; NS). Furthermore, the scatter plot of explained variance versus the range of responses had a wedge shape, so that although neurons with small ranges of responses showed the expected small correlation with FRA predictions, neurons with large ranges of responses could have either large or small correlations with FRA predictions. Thus, a small range of responses could not fully explain these results.

Effect of the temporal envelope on the responses

Main and Artificial have a similar spectrum. The major difference between them is their amplitude modulation: whereas Artificial has fast rise and fall times and a constant envelope, the temporal envelope of the Main version of all stimuli has a much slower intrinsic rise time. To test the effect of the temporal envelope on the strength of the responses, we compared the responses to the

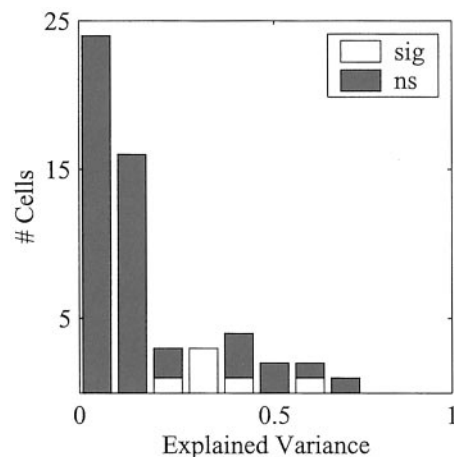


Figure 11. Histogram of the adjusted fraction of variance in the responses that is explained by the FRA predictions. Gray bars correspond to values that are not statistically significant at the $p = 0.05$ level; white bars correspond to values that are significant.

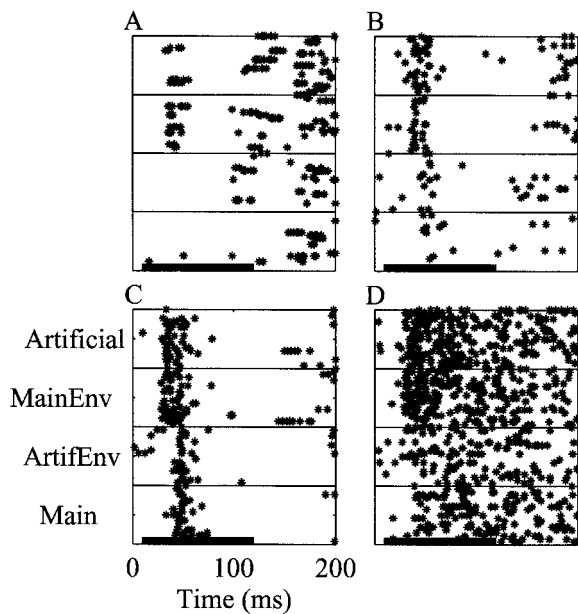


Figure 12. Responses of four neurons to Main, Artificial, MainEnv, and ArtifEnv. The BFs of the neurons are as follows: *A*, 2.8; *B*, 5.8; *C*, 6.3; *D*, 7 kHz. The responses are displayed as *raster plots*. The period of stimulus presentation is denoted by a *thick line* at the bottom of the response to Main.

stimulus pairs Main–ArtEnv and Artificial–MainEnv, which have the same envelope (Fig. 12). In these examples, the stimuli with same temporal envelope evoke similar response patterns. Artificial and MainEnv, with their fast rise time and constant envelope, tended to evoke an earlier and stronger onset response than Main and ArtEnv, with their naturally slower rise time.

These results are quantified in Figure 13 for the responses of the 14 neurons tested with these stimuli. Relatively large adjusted correlation coefficients were present between the responses to stimuli with the same envelope (Fig. 13*A,D*), whereas the adjusted correlation coefficients between the responses to stimuli with the same frequency trajectory, but different temporal envelopes, were weaker (as expected from Fig. 12, see Fig. 13*B,C*). The same pattern was also seen in the distribution of the onset latencies (Fig. 13*E*), in which Artificial and MainEnv, with the fast rise time, evoked earlier responses than Main and ArtEnv.

DISCUSSION

The aim of this study was to test the similarity between the responses to a class of natural stimuli, bird chirps in their acoustic context, and the responses evoked by simplified versions of the same stimuli. Our results show that each of the three simplification steps performed here significantly affected the neuronal responses.

Effect of the acoustic context

Natural sound stimuli, such as the bird chirps used here, always occur within an acoustic context. This context had a major effect on responses to the clean chirp (Main) in AI. The effect of sequential context was apparent in the differences in responses between Natural when presented alone and Natural in the context of the Long stimuli. The effect of simultaneous context is apparent in the differences in responses between Natural and Main. Although the acoustic structure of the two versions in the frequency band containing the chirps is very similar, the two versions evoked often very different responses.

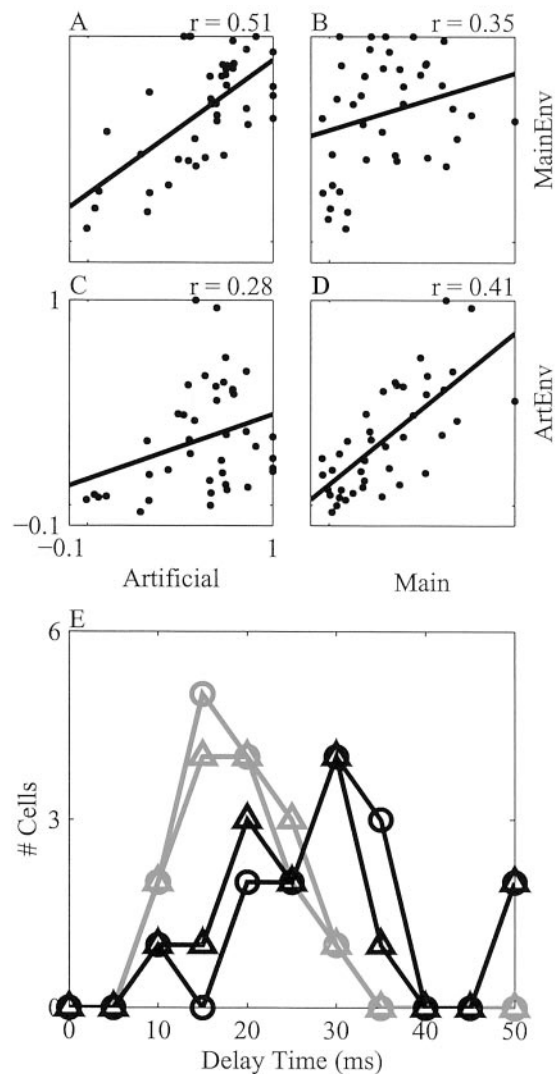


Figure 13. Onset latency and response to Main, Artificial, MainEnv, and ArtEnv. *A–D*, Scatter plots of the responses to Main, Artificial, MainEnv, and ArtEnv. The results represent responses of 14 neurons to 3 stimuli (1, 3, 5), with an individual count for each neuron \times stimulus condition. *E*, Distributions of the onset latency of Main (black circles), Artificial (gray triangles), ArtEnv (black triangles, similar to Main), and MainEnv (gray circles, similar to Artificial).

Previous work (Phillips, 1985; Phillips et al., 1985; Phillips and Hall, 1986) compared the responses to pure tones when they were gated together with BBN and when the BBN started 250 msec before the tone. When both sounds were gated together, the neurons responded in the same way as they responded to the single sound in the mixture that evoked the stronger response by itself. In contrast, responses to tones in continuous BBN were very similar to the responses to pure tones, except that thresholds were raised (by a remarkable 1 dB for each increment of 1 dB in masker level). Brugge et al. (1998) and Furukawa and Middlebrooks (2001) described a similar phenomenon for spatial receptive fields in AI.

On the basis of these experiments, it might be hypothesized that the responses to Main and Natural were different because Natural is formed by the addition of wideband noise to a tone-like stimulus, Main. If so, the response to the embedded chirp in the context of Long should be similar to the response to Main.

However, the results show that the response to Natural embedded in Long was more correlated with the responses to Natural than with the responses to Main.

This discrepancy could stem from the different structure of the noise stimuli used in this study and the noise used by Phillips and colleagues (Phillips, 1985; Phillips et al., 1985; and Phillips and Hall, 1986). Natural background noise has a nontrivial statistical structure (Nelken et al., 1999), whereas BBN does not. This could account for the continuous response to the preceding segment in Long, seen in Figures 5 and 6, a phenomenon that is not seen with BBN. The data indicate that these differences are not attributable to simple fatigue caused by previous firing of the neurons but may be stimulus-specific (Movshon and Lennie, 1979; Shu et al., 1993).

Effects of the temporal envelope

One seemingly puzzling finding in this study was the substantial difference between the responses to Main and Artificial, although their spectral content is similar. The origin of this discrepancy was the difference in the temporal envelopes of the two stimuli, as shown by switching their envelopes. This finding is consistent with those of studies of the effect of rise time on neuronal responses (Heil, 1997a,b; Phillips, 1998; Fishbach et al., 2001). These studies found that the time of the first spike and the strength of the response are correlated with the rate of change of the linear onset ramp. In the stimuli used here, the rise time was 3 msec for both, but the final level of the ramp is higher for Artificial, which therefore has a steeper slope. As a result, the responses to Artificial were earlier and stronger than the responses to Main on average. However, the neurons studied here also responded later during the stimulus, and not all the differences between Main and Artificial can be accounted for by the differences in the onset responses (Fig. 12*D*). These differences may be attributable to further effects of the postonset temporal envelope fluctuations. These results may be related to data reported by Lu et al. (2001), who used ramp and damp stimuli to show strong effects of changes in the temporal envelope on neural responses.

Relationships to models of AI neurons

Two approaches to the characterization of neurons in AI have been suggested in the literature. The standard approach is based on the measurement of a large set of neuronal characteristics based on simple sounds such as parameters related to the frequency response area (e.g., BF and width) or on somewhat more complex stimuli such as FM sweeps (directional preference and velocity preference) or AM tones (modulation transfer function; for review, see Schreiner, 1998). The second, more recent approach for the characterization of AI neurons is based on the measurement of STRF as the basic tool for estimating neuronal responses. Neither of these approaches provides a satisfactory account for our results.

Some of the results presented here fit well with predictions based on the responses to simple sounds. The prime example is the relationship between the responses to Main and Artificial, which is almost fully accounted for by the structure of their temporal envelopes. However, this approach would also predict that the responses to Main and Natural would be similar. Both stimuli have most of their energy within the same narrow frequency band, and the additional background present in Natural is extremely weak (the background spectral level is 17 dB below the main peak on average). Both stimuli have the same FM trajectory and temporal envelope. Thus, if a characterization based on the

FRA and on the responses to FM and AM tones is sufficient, the responses to the two sets of stimuli should have been similar. Our results, which show high diversity between Main and Natural even for neurons intersecting the frequency range of the stimuli, refute this prediction. The failure of FRA predictions (Fig. 11) is a quantitative statement of the same finding. The other approach for the characterization of AI neuron is based on the STRF (Kowalski et al., 1996a,b; deCharms et al., 1998; Depireux et al., 2001; Schnupp et al., 2001) as the basic descriptor. The STRF summarizes the sensitivity of a neuron to all stimuli provided that the neuron operates, at least approximately, as a linear filter in the time–frequency domain. The stimuli used in this study are not rich enough spectrally and temporally for usefully estimating the STRF. However, published examples of STRFs predict that stimuli such as Main and Natural or, in many cases, Main and Artificial would evoke similar activity as long as the main features of the STRF intersect the stimulus frequency range.

In fact, all the energy of Main and most of the energy of Natural are within the stimulus frequency range. This range would be the source of the dominant contribution to any prediction based on the STRF when the STRF intersects the stimulus frequency range. The energy outside the stimulus frequency range cannot influence the responses much, because the stimulus energy outside this range is low, and the STRF weights multiplying this energy are small. Thus, the responses to Main and Natural should be similar when a neuron is well described by the STRF. As shown here, this is definitely not the case.

The argument for similarity of the responses to Main and Artificial is based on the published temporal structure of STRFs in mammals. These are often quite sluggish [in the ferret (Depireux et al., 2001); in the monkey (deCharms et al., 1998)], with a best temporal modulation frequency of ~10 Hz. However, the Main and Artificial versions of all stimuli are shorter than ~100 msec; therefore, their envelopes fluctuate only on faster time scales. Because the frequency trajectories of Main and Artificial are essentially identical, STRFs should predict similar responses to the two versions, contrary to the actual results.

It may well be that the STRFs, with their typically sluggish time course, describe slow dynamics of neural responses, whereas the data presented here are related to substantially faster dynamics with different properties. A similar separation of time scales is also required to explain the effects of changes in stimulus rise time on neuronal responses. To model these results, it is necessary to use time constants (1–10 msec; Fishbach et al., 2001) that are much faster than the dynamics of most published STRFs in the auditory cortex.

The data presented here suggest that an essential level of complexity emerges in the responses of AI neurons when tested with natural mixtures of tonal stimuli and noise, as seen by comparing the responses to Main and Natural or to Long and Natural. These phenomena are difficult to explain by the responses to tonal components alone (Main), although the tonal components dominate the stimuli acoustically. This finding is similar in nature to the physiological comodulation-masking release described by Nelken et al. (1999), which is essentially the complementary finding. Whereas here the response to a strong tonal stimulus is modified in a substantial way by the addition of a weak background sound, in the findings of Nelken et al. (1999), the responses to a strong noise stimulus are substantially modified by the addition of a weak tone. We believe that neither of these findings can be easily accounted for by current models of spectrotemporal integration in AI.

Extreme sensitivity of AI neurons to their acoustic input

The main result of this study is the demonstration of extreme sensitivity of AI neurons to small perturbations in their acoustic input. Standard models do not predict such extreme sensitivity. This sensitivity is seen in the considerable differences between the responses to Natural in the context of Long, to Natural alone, to Main, and to Artificial. We can only partially identify the exact acoustic features that are responsible for this sensitivity: only in one case, the responses to Main and Artificial, could the differences in the responses be explained by a simple acoustic difference. The specification of the acoustic determinants of the other differences is left for future work.

REFERENCES

- Brosch M, Schreiner CE (1997) Time course of forward masking tuning curves in cat primary auditory cortex. *J Neurophysiol* 77:923–943.
- Brugge JF, Reale RA, Hind JE (1998) Spatial receptive fields of primary auditory cortical neurons in quiet and in the presence of continuous background noise. *J Neurophysiol* 80:2417–2432.
- Calford MB, Semple MN (1995) Monaural inhibition in cat auditory cortex. *J Neurophysiol* 73:1876–1891.
- Calhoun BM, Schreiner CE (1998) Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics. *Eur J Neurosci* 10:926–940.
- deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for cortical neurons. *Science* 280:1439–1443.
- Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85:1220–1234.
- Fishbach A, Nelken I, Yeshurun Y (2001) Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. *J Neurophysiol* 85:2303–2323.
- Furukawa S, Middlebrooks JC (2001) Sensitivity of auditory cortical neurons to locations of signals and competing noise sources. *J Neurophysiol* 86:226–240.
- Heil P (1997a) Auditory cortical onset responses revisited. I. First-spike timing. *J Neurophysiol* 77:2616–2641.
- Heil P (1997b) Auditory cortical onset responses revisited. II. Response strength. *J Neurophysiol* 77:2642–2660.
- Heil P, Rajan R, Irvine DR (1992a) Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. I: Effects of variation of stimulus parameters. *Hear Res* 63:108–134.
- Heil P, Rajan R, Irvine DR (1992b) Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. II: Organization of response properties along the “isofrequency” dimension. *Hear Res* 63:135–156.
- Kowalski N, Depireux DA, Shamma SA (1996a) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J Neurophysiol* 76:3503–3523.
- Kowalski N, Depireux DA, Shamma SA (1996b) Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *J Neurophysiol* 76:3524–3534.
- Lu T, Liang L, Wang X (2001) Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. *J Neurophysiol* 85:2364–2380.
- Mendelson JR, Cynader MS (1985) Sensitivity of cat primary auditory cortex (AI) neurons to the direction and rate of frequency modulation. *Brain Res* 327:331–335.
- Mendelson JR, Schreiner CE, Sutter ML, Grasse KL (1993) Functional topography of cat primary auditory cortex: responses to frequency-modulated sweeps. *Exp Brain Res* 94:65–87.
- Movshon JA, Lennie P (1979) Pattern-selective adaptation in visual cortical neurones. *Nature* 278:850–852.
- Nelken I, Versnel H (2000) Responses to linear and logarithmic frequency-modulated sweeps in ferret primary auditory cortex. *Eur J Neurosci* 12:549–562.
- Nelken I, Rotman Y, Bar Yosef O (1999) Responses of auditory-cortex neurons to structural features of natural sounds. *Nature* 397:154–157.
- Phillips DP (1985) Temporal response features of cat auditory cortex neurons contributing to sensitivity to tones delivered in the presence of continuous noise. *Hear Res* 19:253–268.
- Phillips DP (1998) Factors shaping the response latencies of neurons in the cat’s auditory cortex. *Behav Brain Res* 93:33–41.
- Phillips DP, Hall SE (1986) Spike-rate intensity functions of cat cortical neurons studied with combined tone-noise stimuli. *J Acoust Soc Am* 80:177–187.
- Phillips DP, Orman SS, Musicant AD, Wilson GF (1985) Neurons in the cat’s primary auditory cortex distinguished by their responses to tones and wide-spectrum noise. *Hear Res* 18:73–86.
- Reale RA, Imig TJ (1980) Tonotopic organization in auditory cortex of the cat. *J Comp Neurol* 192:265–291.
- Rotman Y, Bar-Yosef O, Nelken I (2001) Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex. *Hear Res* 152:110–127.
- Schnupp JW, Mrcic-Flogel TD, King AJ (2001) Linear processing of spatial cues in primary auditory cortex. *Nature* 414:200–204.
- Schreiner CE (1998) Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. *Audiol Neurootol* 3:104–122.
- Shamma SA, Versnel H (1995) Ripple analysis in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary spectral profiles. *Auditory Neurosci* 1:255–270.
- Shamma SA, Versnel H, Kowalski N (1995) Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra. *Auditory Neurosci* 1:233–254.
- Shu ZJ, Swindale NV, Cynader MS (1993) Spectral motion produces an auditory after-effect. *Nature* 364:721–723.
- Sokal R, Rohlf F (1995) *Biometry*, Ed 3. New York: Freeman.
- Versnel H, Shamma SA (1998) Spectral-ripple representation of steady-state vowels in primary auditory cortex. *J Acoust Soc Am* 103:2502–2514.
- Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74:2685–2706.