Nonlinear Spectrotemporal Sound Analysis by Neurons in the Auditory Midbrain

Monty A. Escabí^{1,2} and Christoph E. Schreiner¹

¹W. M. Keck Center for Integrative Neuroscience and University of California San Francisco/University of California Berkeley Joint Bioengineering Graduate Group, University of California, San Francisco, California 94143, and ²Department of Electrical and Computer Engineering, Biomedical Engineering Program, University of Connecticut, Storrs, Connecticut 06269

The auditory system of humans and animals must process information from sounds that dynamically vary along multiple stimulus dimensions, including time, frequency, and intensity. Therefore, to understand neuronal mechanisms underlying acoustic processing in the central auditory pathway, it is essential to characterize how spectral and temporal acoustic dimensions are jointly processed by the brain. We use acoustic signals with a structurally rich time-varying spectrum to study linear and nonlinear spectrotemporal interactions in the central nucleus of the inferior colliculus (ICC). Our stimuli, the dynamic moving ripple (DMR) and ripple noise (RN), allow us to systematically characterize response attributes with the spectrotemporal receptive field (STRF) methods to a rich and dynamic

The central nucleus of the inferior colliculus (ICC) is an obligatory station in the lemniscal auditory system that receives convergent inputs from numerous brainstem structures and sends its highly processed outputs to the auditory thalamus and, subsequently, to the primary auditory cortex. Neurons in the ICC are sensitive to systematic manipulations of temporal, spectral, binaural, and intensity stimulus attributes (Rees and Møller, 1983, 1987; Schreiner et al., 1983; Langner and Schreiner, 1988; Schreiner and Langner, 1988; Irvine and Gao, 1990; Kuwada et al., 1997; Ramachandran et al., 1999; Krishna and Semple, 2000). These properties have been studied extensively with pure tones, modulated tones, and noise stimuli; however, the overall capabilities of the ICC for processing dynamic, spectrally complex acoustic stimuli remain unknown. Clearly, because natural sounds have structurally rich acoustic spectra and can simultaneously vary along spectral, temporal, intensity, and aural acoustic dimensions, it is essential to understand how these are jointly processed and represented within the ICC.

The concept of a stimulus–response function or receptive field (RF) is a mathematical construct that describes the stimulus features that are encoded by a sensory neuron. A widely used RF description that measures the response of a neuron to pure tones

stimulus ensemble. Theoretically, we expect that STRFs derived with DMR and RN would be identical for a linear integrating neuron, and we find that \sim 60% of ICC neurons meet this basic requirement. We find that the remaining neurons are distinctly nonlinear; these could either respond selectively to DMR or produce no STRFs despite selective activation to spectrotemporal acoustic attributes. Our findings delineate rules for spectrotemporal integration in the ICC that cannot be accounted for by conventional linear–energy integration models.

Key words: inferior colliculus; spectrotemporal; receptive field; nonlinear; ripple; naturalistic; reverse correlation

of varying frequency and sound pressure level (SPL) is the frequency-tuning curve (FTC; Schreiner and Langner, 1988; Nelken et al., 1997; Ramachandran et al., 1999). Although this descriptor continues to be important, it cannot characterize the dynamic behavior of a neuron in response to an arbitrary, spectrally complex, time-varying stimulus. Consequently, secondary analyses are often used that measure the ability of a neuron to respond to other stimulus aspects, such as the ability to follow successively presented stimuli of different rates (Rees and Møller, 1983, 1987; Schreiner et al., 1983; Møller and Rees, 1986; Langner and Schreiner, 1988; Eggermont, 1999; Krishna and Semple, 2000).

Recently, the use of reverse correlation techniques to estimate the spectrotemporal receptive field (STRF) in the auditory system (Aersten et al., 1980; Yeshurun et al., 1985; Eggermont, 1993; Nelken et al., 1997; de Charms et al., 1998; Klein et al., 2000; Theunissen et al., 2000; Depireux et al., 2001; Miller et al., 2001, 2002) has allowed scientists to overcome some of the practical limitations posed by conventional auditory RFs and the stimuli used to derive them (e.g., pure tones and modulated tones). The STRF describes the stimulus–response function of an auditory neuron along both the spectral and temporal acoustic dimensions, to a rich stimulus ensemble, and makes no assumptions about independence of spectral and temporal response attributes.

Most RF methods, including the STRF procedure, operate under the assumption that the system under investigation integrates information, be it acoustic or visual, in an approximately linear manner. This requires that the spiking output of a sensory neuron be described as a linear or quasilinear function of its inputs. Although this is often a reasonable assumption, it may not always hold. For instance, direct STRF (referred to as spatiotemporal receptive field for visual neurons) approaches are readily

Received Sept. 18, 2001; revised Jan. 22, 2002; accepted Jan. 25, 2002.

This work was supported by National Institutes of Health Research Grants DC02260 and NS34835 and a grant from the Ford Foundation (M.A.E.). We thank two anonymous reviewers for many insightful comments and M. P. Stryker, J. A. Winer, and A. J. Doupe for comments on previous versions of this manuscript. We also thank L. M. Miller and H. Read for numerous discussions and help during experiments and M. Kvale for the use of his SpikeSort1.2 analysis tool.

Correspondence should be addressed to Monty A. Escabí, 260 Glenbrook Road, U-157, University of Connecticut, Storrs, CT 06269. E-mail: escabi@engr. uconn.edu.

Copyright © 2002 Society for Neuroscience 0270-6474/02/224114-18\$15.00/0

J. Neurosci., May 15, 2002, 22(10):4114-4131 4115

applicable for simple cells in the primary visual cortex (Jones and Palmer, 1987; DeAngelis et al., 1993, 1999; Victor and Purpura, 1998; Anzai et al., 1999; Reich et al., 2000) but fail for visual complex cells and neurons outside of VI (Emerson et al., 1987; Szulborski and Palmer, 1990; Livingstone et al., 2001). Other stimulus-dependent limitations are observed for sensory neurons in acoustically specialized animals, where central sensory neurons are often highly nonlinear and specifically tuned to behaviorally relevant vocalizations (Suga and Jen, 1976; Suga, et al., 1978; Margoliash, 1983; Doupe, 1997; Portfors and Wenstrup, 1999; Theunissen et al., 2000).

Theoretically, the STRF procedure requires the use of white noise as a probing stimulus. Practically, however, because sensory neurons in central stations respond to a limited range of spectrotemporal (spatiotemporal) modulations and are often inhibited by white noise, it is necessary to synthesize acoustic or visual sequences that are optimized for any particular station (de Charms et al., 1998; Klein et al., 2000). Often this is achieved with randomly arranged spectrotemporal tone pips in the auditory system (de Charms et al., 1998; Theunissen et al., 2000) and spatiotemporally interleaved bars or spots of light in the visual system (Emerson et al., 1987; DeAngelis et al., 1993, 1999; Anzai et al., 1999; Reich et al., 2000). Recently, some of the stimulusdependent limitations associated with such stimuli have been overcome with the use of natural sounds (Theunissen et al., 2000) in the avian auditory cortex homolog.

In this study, we recorded single-unit activity from neurons in the ICC of cats in response to dynamic spectrotemporally complex stimulus sequences. Our synthetic stimuli, the dynamic moving ripple (DMR) and ripple noise (RN), are designed to stringently satisfy a number of theoretical requirements for use with the reverse correlation STRF methods. Furthermore, these sounds share various properties with natural sounds that allow us to overcome some of the practical limitations of white noise, randomly interleaved tone pips, and other synthetic reverse correlation stimuli. Compared with natural signals, these stimuli offer the advantage that they can be parametrically manipulated, allowing for a systematic assessment of nonlinear response characteristics within the ICC. Our findings demonstrate the presence of distinct spectrotemporal nonlinearities in the ICC and identify possible mechanisms used for complex sound analysis, source segregation, and signal detection.

MATERIALS AND METHODS

Surgical preparation

Cats were initially anesthetized with a mixture of ketamine HCl (10 mg/kg) and acepromazine (0.28 mg/kg, i.m.). After an intravenous infusion line was inserted, a surgical state of anesthesia was induced with \sim 30 mg/kg Nembutal and maintained throughout the surgery with supplements. Body temperature was measured with a rectal probe and maintained with a heating pad at ~37.5°C. An incision was made in the intercartilaginous area of the trachea, and a tracheotomy tube was inserted. After performing a craniotomy, the ICC was exposed by removing the overlying cerebrum and part of the bony tentorium using a dorsal approach. On completion of the surgery, the animal was maintained in an areflexive state of anesthesia via continuous infusion of ketamine (2-4 mg \cdot kg⁻¹ \cdot h⁻¹) and diazepam (0.4–1 mg \cdot kg⁻¹ \cdot h⁻¹) in lactated Ringer's solution $(1-4 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1})$. The state of the animal was monitored (heart rate, breathing rate, temperature, and periodically checked reflexes) throughout the experiment, and the infusion rate was adjusted according to physiological criteria. Every 12 hr, the cat received an injection of dexamethasone (0.14 mg/kg, s.c.) to prevent brain edema and atropine to reduce salivation (0.04 mg \cdot kg⁻¹ \cdot d⁻¹, s.c.). All surgical methods and experiment procedures followed National Institutes of Health and US Department of Agriculture guidelines and were approved by the committee on animal research of the University of California, San Francisco.

Neuronal recording

Data were obtained from n = 81 single units in the central nucleus of the inferior colliculus of three anesthetized cats. One or two closely spaced parylene-coated tungsten microelectrodes (Microprobe Inc., Potomac, MD; 1–3 M Ω at 1 kHz) were advanced with a hydraulic microdrive (David Kopf Instruments, Tujunga, CA). Action potential traces were recorded onto a digital audiotape (CDAT16; Cygnus Technologies, Del-aware Water Gap, PA) at a sampling rate of 24.0 kHz (41.7 μ sec resolution) for off-line analysis. Off-line analysis consisted of digital bandpass filtering (0.3–10 kHz) and individually spike sorting the action potential traces using a Bayesian spike-sorting algorithm (Lewicki, 1994).

Electrode penetration trajectories were at ~20-30° relative to the sagittal plane. Electrodes were initially advanced through the external nucleus and onto the central nucleus while audiovisually determining single neuron and multiunit characteristic frequencies (CFs). The boundary between the external and central nucleus of the inferior colliculus (IC) was confirmed physiologically (Merzenich and Reid, 1974) by a reversal or discontinuity in the CF trend and by monotonically increasing CFs as a function of depth (over a range of ~1-20 kHz and ~1.5-5.0 mm relative to the surface of the IC), consistent with the central nucleus. All electrode recordings throughout the remainder of the experiment were taken from this physiologically defined region. Except for the depth and CF constraints, recording locations were randomly distributed within the ICC.

Acoustic stimuli

RN and DMR stimulus waveforms were designed on a digital computer using the MATLAB (Mathworks) programming environment. The spectrotemporal envelopes shown in Figure 1*C*,*D* define the energy modulations, in time and frequency, that are used to modulate a bank of sinusoidal carriers of frequencies f_k . As with natural signals, the envelope of these sounds is time-varying and probes spectral and temporal neuronal response preferences. Furthermore, analogous to various classes of natural signals (Fig. 1*A*,*B*), these sounds have unique short-term statistics (Fig. 2*D*,*E*) and yet their long-term statistics are identical (Fig. 2*D*,*E*, *far right*; see Stimulus correlation statistics). Therefore, both sounds satisfy the necessary requirement for use with the reverse correlation procedure that we use to estimate auditory spectrotemporal receptive fields (see Spectrotemporal receptive field).

Dynamic moving ripple envelope. The DMR envelope is designed as a dynamic sinusoidal grating on a octave frequency and decibel amplitude axis. Two parameters defined the DMR envelope: the instantaneous ripple density, $\Omega(t)$, defines the number of spectral peaks per octave at a given time instant; and $F_m(t)$ defines the instantaneous modulation rate. The DMR spectrotemporal envelope is expressed as:

$$S_{\text{DMR}}(t, X_k) = M/2 \cdot \sin[2\pi\Omega(t)X_k + \Phi(t)], \qquad (1)$$

where M = 30 or 45 is the modulation depth of the envelope in decibels, $X_{\rm k} = \log_2(f_{\rm k}/f_1)$ is the octave frequency axis relative to the lowest stimulus frequency ($f_1 = 500$ Hz), and $\Phi(t) = \int_0^t F_{\rm m}(\tau) d\tau$ controls the time-varying temporal modulation rate, $F_{\rm m}(t)$. Spectral [$\Omega(t)$] and temporal [$F_{\rm m}(t)$] parameters are independent and slowly time-varying random processes (maximum rates of change, 1.5 Hz for $F_{\rm m}$ and 3.0 Hz for Ω). The time rate of change of both parameters was heuristically chosen so that they coincide with the observed range of values for similar acoustic features in speech and vocalizations (Greenberg, 1998). To guarantee that the stimulus space was covered in a statistically unbiased manner, both parameters were designed with uniformly (flat) distributed amplitudes in the intervals 0–4 cycles per octave for Ω and -350 to +350Hz for $F_{\rm m}$.

The time-varying stimulus parameters were generated in the MAT-LAB programming environment. First, the parameters were generated as a random sequence of normally distributed samples (randn function in MATLAB) using a sampling rate of 3 Hz for $F_m(t)$ and 6 Hz for $\Omega(t)$. These sequences had maximum frequency contents of 1.5 and 3 Hz, respectively (because the maximum signal frequency is half of the sampling frequency). To generate the acoustic sound waveforms at a sampling rate of 44.1 kHz (Eq. 4) it was necessary to resample both of the parameter signals to an equivalent sampling rate. Therefore, we upsampled both signals to 44.1 kHz using a cubic interpolation procedure (interp1 function in MATLAB; "cubic" option; upsampling factor, 14,700 for modulation rate and 7350 for ripple density). Next we needed to convert the parameter amplitudes from a normal to a uniform distribution so that the probability of occurrence of each parameter is statistically unbiased within the selected intervals. This normalization was performed with the error function:

$$erf(x) = 2/\sqrt{\pi} \int_0^x e^{-t^2} dt.$$

This function converts normally distributed amplitudes to uniformly distributed amplitudes over the interval -1 to +1 and a subsequent linear rescaling of the amplitudes to the selected interval. This operation had only a subtle effect on the spectrum of these signals and is necessary to guarantee that the signal parameters are statistically unbiased (flat distribution) within their predefined range.

Ripple noise envelope. The RN envelope is first generated as a linear superposition of L = 16 independently chosen DMR envelopes, $S_{\text{DMR}_i}(t, X_k)$:

$$\bar{S}_{\rm RN}(t, X_{\rm k}) = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} S_{\rm DMR_l}(t, X_{\rm k}), \tag{2}$$

where the sum is normalized so that the SDs of the RN and DMR are identical. Although this guaranteed that the average contrast of the DMR and RN envelopes be the same (i.e., identical SD), the RN amplitude distribution had long tails and resembled a Gaussian distribution, whereas the DMR envelope is approximately uniform and confined to the interval [-M/2, M/2]. Instances at the high- and low-intensity tails of the distribution of the RN envelope can therefore potentially activate undesirable intensity- and contrast-dependent nonlinearities. We overcame this possibility by compressing the RN envelope so that its amplitude statistics resemble those of the DMR. The compressed RN envelope is given by:

$$S_{\rm RN}(t, X_{\rm k}) = f[\bar{S}_{\rm RN}(t, X_{\rm k})], \qquad (3)$$

where $f(x) = M/2 \cdot erf(x/\sigma_{\text{DMR}})$ and $erf(\cdot)$ is the error function. This envelope covers a relative intensity range of [-M/2, -M/2] dB as for the DMR envelope. This procedure allows us to isolate spectrotemporal nonlinearities from intensity- or contrast-dependent ones. A second concern was that the $erf(\cdot)$ function significantly distorts the RN envelope by introducing high-frequency envelope modulation components, and this in turn could compromise experimental results. We found, both analytically (data not shown) and through simulation, that the ripple spectrum and spectrotemporal autocorrelation (see Fig. 2D,E, far right; shown for compressed RN) of the uncompressed and compressed RN were in close agreement (2.1% rms error for both the ripple spectrum and autocorrelation).

Acoustic waveform. From the DMR and RN spectrotemporal envelopes, the acoustic sound pressure waveforms, s(t), are constructed by modulating L = 230 sinusoidal carriers that are added together:

$$s(t) = \sum_{k=1}^{L} S_{\text{Lin}}(t, X_k) \cdot \sin(2\pi f_k t + \phi_k),$$
(4)

where ϕ_k is a randomly chosen phase $(0-2\pi)$, which gives s(t) a noise-like character, and $S_{\text{Lin}}(t, X_k)$ is a transformed version of the DMR or RN envelopes that describes the amplitude modulations in linear amplitude units. The linear envelope is bounded between $10^{-M/20}$ and 1. It is related to the decibel envelopes by (here we use S_{dB} in place of S_{DMR} and S_{RN}):

$$S_{\text{Lin}}(t, X_k) = 10^{\frac{S_{\text{dB}}(t, X_k) - M/2}{20}}$$
. (5)

Frequency carriers are geometrically spaced at a resolution of 43 carriers per octave: $f_k = \alpha \cdot f_{k-1}$ ($\alpha = 1.01617$) over a range of 5.32 octaves (500–20,000 Hz). Although the resultant power spectrum is not flat, this guarantees that the primary sensory epithelium is uniformly excited and equal energy is provided per unit octave.

Sound presentation. All recordings were made with the animal in a sound-shielded chamber (IAC, Bronx, NY), with stimuli delivered via a closed, binaural speaker system (electrostatic diaphragms from Stax). Single neurons or clusters of neurons were initially isolated audiovisually

by presenting pure tones, white noise, or both. FTCs were derived in two of the three experiments with a pseudorandom sequence of pure tones presented at 15 intensities and 45 geometrically spaced frequencies. In one experiment, rate-level functions were measured with the RN stimulus as a function of SPL and contrast. After these initial tests, DMR and RN stimuli were presented binaurally with an independent sound sequence for each ear. The DMR stimulus was presented for 10-20 min, followed by 10–18 min (full length presented for \sim 95% of the recording sites; identical stimuli for all experiments) of the RN at \sim 30–70 dB/ carrier greater than the neuron response threshold (as determined by the FTC or rate-level functions). Because the RN and DMR stimuli are each composed of 230 sinusoid carriers, the effective SPLs were 10 · $\log_{10}(230) = 23.6$ dB greater than these values (i.e., ~53–93 dB greater than threshold; SPL range, 75 ± 19 dB SPL, 64 ± 19 dB/one-third octave, or 51 \pm 19 dB/carrier). Both RN and DMR were presented at identical intensities and contrast so that they covered an identical range of amplitudes and fall well within the intensity response area of the neuron. Sixteen neurons were also tested with a short 5 sec segment of the DMR and RN that was presented 40 consecutive times. This was used to construct response rastergrams for each stimulus (see Fig. 10). Finally, for six neurons that did not respond to the RN, the DMR stimulus was again presented at the end of the recording session to verify that the given neurons were still responsive and to verify the stability of the electrode placement.

Stimulus correlation statistics

The long-term and instantaneous spectrotemporal correlation statistics of the RN and DMR stimulus constitute an essential aspect of the stimulus design and the experimental approach. These were evaluated in closed form and rigorously tested via simulation. Only a brief account is provided.

A spectrotemporal Gaussian window, $w_i(t, X)$, of SD $\sigma_x = 0.5$ octaves and $\sigma_t = 5$, 10, or 20 msec and centroid about $t = t_i$ was used to localize the RN or DMR spectrotemporal envelope, S(t, X). The instantaneous spectrotemporal autocorrelation function was obtained by evaluating the localized autocorrelation:

$$R_{\rm ss}(\tau,\xi|t_{\rm i}) = E[S(t,X)w_{\rm i}(t,X)S(t-\tau,X-\xi)w_{\rm i}(t-\tau,X-\xi)], \qquad (6)$$

where the expectation operator, $E[\cdot]$, is taken with respect to time, t, and the spectral distance variable, X [Eqs. 1, 3 are substituted for S(t, X)]. The variable t_i corresponds to the time instant when the autocorrelation is evaluated, and τ and ξ correspond to the temporal lag and spectral displacement, respectively.

In closed form the solutions for the RN and DMR are given by:

$$R_{\rm DMR}(\tau,\xi) = \sigma_{\rm DMR}^2 \cdot \cos(2\pi\Omega_i\xi + 2\pi F_{\rm m,i}\tau) \cdot R_{\rm ww}(\tau,\xi) \tag{7}$$

$$R_{\rm RN}(\tau,\,\xi) = \left[\sigma_{\rm RN}^2 \cdot \operatorname{sinc}(2\Omega_{\rm Max}\xi) \cdot \operatorname{sinc}(2F_{\rm Max}\tau) + e(\tau,\,\xi)\right] \cdot R_{\rm ww}(\tau,\,\xi), \quad (8)$$

where $\sigma_{\rm DMR}^2 = M^2/8$ and $\sigma_{\rm RN}^2 = M^2/12$ are the variance of the DMR and RN, respectively, and $R_{\rm ww}(\tau, \xi)$ is the autocorrelation function of the Gaussian window (which is itself a Gaussian window of SD $\sqrt{2\sigma_x}$ and $\sqrt{2\sigma_t}$). The parameters $\Omega_i = \Omega(t_i)$ and $F_{\rm m}$, $i = F_{\rm m}(t_i)$ are the instantaneous DMR parameters evaluated at t_i . Because the stimulus parameters dynamically vary with time at a nominal rate of 3 and 1.5 Hz (Fig. 2*A*,*B*), the DMR instantaneous spectrotemporal autocorrelation likewise varies with time (Fig. 2*D*). Accordingly, its spectrotemporal envelope is nonstationary at these time scales. The term $e(\tau, \xi)$ is a spectrotemporal noise term, and the parameters $\Omega_{\rm Max} = 4$ cycles per octave and $F_{\rm Max}$ =350 Hz are the maximum ripple parameters.

The long-term autocorrelation for both sounds was obtained by performing a time average of the instantaneous autocorrelation: $R_{SS}(\tau, \xi) = E[R_{SS}(\tau, \xi|t_i)]$ (*E*[·] is now evaluated with respect to t_i). The autocorrelation is identical in form for both sounds:

$$R_{\rm DMR}(\tau,\,\xi) = R_{\rm RN}(\tau,\,\xi) = \sigma_{\rm S}^2 \cdot \operatorname{sinc}(2\Omega_{\rm Max}\xi) \cdot \operatorname{sinc}(2F_{\rm Max}\tau) \cdot R_{\rm ww}(\tau,\,\xi). \tag{9}$$

The autocorrelations only differ in the SD by a multiplicative factor of 20% (RN, $\sigma_{\rm S} = \sigma_{\rm RN} = M\sqrt{12}$ dB; DMR, $\sigma_{\rm S} = \sigma_{\rm DMR} = M\sqrt{8}$ dB).

Spectrotemporal receptive field

STRFs are computed by averaging the pre-event spectrotemporal envelope. For a sequence of N neural events at times, t_n (sampled at 41.7 μ sec

resolution), contralateral and ipsilateral STRFs are obtained as [here we use $S_{dB}(t, X_k)$ in place of $S_{DMR}(t, X_k)$ or $S_{RN}(t, X_k)$]:

$$\mathrm{STRF}(\tau, X_{\mathrm{k}}) = 1/(\sigma_{\mathrm{s}}^2 \cdot T) \cdot \sum_{\mathrm{n}} S_{\mathrm{dB}}(t_{\mathrm{n}} - \tau, X_{\mathrm{k}}), \tag{10}$$

where T is the experimental recording time in seconds, τ is the temporal delay of the stimulus relative to the neural event time (0–100 msec), and σ_S^2 is the variance of the decibel spectrotemporal envelope for the DMR or RN. During the DMR and RN stimulus presentation, independent sound sequences were binaurally presented to each animal. This allowed us to independently estimate the contralateral and ipsilateral STRFs by replacing the contralateral and ipsilateral spectrotemporal envelopes into Equation 10 (Marmarelis and Naka, 1974).

Stimulus envelopes were sampled at 4.0 kilosamples/sec (temporal) and 43 samples per octave (spectral). The STRF is formally given in units of spikes per second per decibel. We use a rate-normalized version of the STRF, $STRF_r(\tau, X_k) = \sigma_s \cdot STRF(\tau, X_k)$, which corresponds to the average driven output produced at time 0, in units of spikes per second, for the average differential stimulus (decibels) presented within the receptive field of the neuron.

Statistically significant STRF

We devised a procedure for measuring the statistically significant STRF by considering a null condition in which N randomly chosen spikes are put through Equation 10. This procedure consists of adding random sound waveforms to construct a control STRF from which statistical significance can be determined. Solutions for this procedure were derived analytically in closed form (data not shown). The distribution of amplitudes for the control STRF quickly approached a normal distribution (with as little as N = 50 spikes). Therefore, a simplification was made in which we determined the two-tailed probability of exceeding a threshold relative to the control STRF under the assumption of a normal distribution. The statistically significant portion of the STRF (p < 0.002) is obtained by keeping all values of the STRF that exceed 3.09 SD of the control noise STRF and setting all other values to 0. Analytically this is expressed as $|\sigma_s^2 \cdot T \cdot \text{STRF}(\tau, \xi)/\sqrt{N}| > 3.09 \cdot \sigma_s$. No smoothing was performed before or after thresholding. This procedure was tested against the analytically derived solutions, and we found that actual significance values were always slightly smaller (e.g., actual significance value of p < 0.0019 for N = 50).

To determine relative significance of STRFs, on an equal spike basis, we further evaluated significance by recomputing all STRFs using 100 action potentials and determining all pixel values that exceeded the p < 0.002 confidence intervals. For these pixel values, the average and maximum signal-to-noise ratio (SNR) was computed. Average and peak SNRs were computed as:

and

$$\mathrm{SNR}_{\mathrm{Max}} = \max(|\mathrm{STRF}_{100}|)/\sigma_{100} ,$$

 $\text{SNR}_{\text{Mean}} = \sqrt{E[\text{STRF}_{100}^2]/\sigma_{100}^2}$

where σ_{100} is the SD of the noise control STRF derived for 100 random spikes. Thus, for any given pixel, the SNR determines the number of SDs by which STRF pixels stand out above the noise.

Null hypothesis

Response nonlinearities are tested against the expected results for an ideal linear model neuron. Given that the long-term spectrotemporal autocorrelation functions for the DMR and RN are identical, it follows that for a purely linear neuron $\text{STRF}_{\text{DMR}} = \text{STRF}_{\text{RN}}$ (for proof, see Appendix A). Significant differences between the RN and DMR STRFs can be attributed to response nonlinearities. To quantify response differences, we use the statistically significant portion of the STRFs and use this to compute a number of response metrics for the DMR and RN: similarity index, rate and magnitude disparity index, and the phase-locking index (see below).

Quantifying DMR and RN response differences

Neural responses for DMR and RN were compared in three complementary ways. First the STRF similarity index (SI; DeAngelis et al., 1999; Reich et al., 2000) was used to quantify shape differences between STRF_{DMR} and STRF_{RN}. Using the STRF pixel values that exceeded the statistical significance threshold of p < 0.002 for either condition, we treated the STRFs as vectors (including significant contralateral and ipsilateral pixels). The vectorized RFs were then used to evaluate the similarity index:

$$SI = \frac{\langle RF_{DMR}, RF_{RN} \rangle}{\|RF_{DMR}\| \cdot \|RF_{RN}\|},$$
(11)

where RF_{DMR} and RF_{RN} are the significant STRFs, $\langle \cdot, \cdot \rangle$ is the vector inner product, and $\|\cdot\|$ designates the vector norm operator. The SI is numerically identical to the Pearson correlation coefficient.

We devised two metrics to evaluate differences in firing rate and driven activity independently of STRF shape. First we computed the rate disparity index (RDI):

$$RDI = s \cdot \left[\left(\frac{r_{DMR}}{r_{RN}} \right)^{s} - 1 \right] \cdot 100\%,$$
(12)

where $s = \text{sign}(r_{\text{DMR}} - r_{\text{RN}})$, and the mean spike rates for each condition are r_{DMR} and r_{RN} . The magnitude of the RDI is numerically equivalent to the percent change in firing rate between DMR and RN. Its sign tells us which condition, DMR or RN, had a higher firing rate (+, DMR; -, RN). To quantify differences in driven activity, we used a third metric, the magnitude disparity index (MDI). The MDI is identical in form to the RDI, where the mean firing rates, r_{DMR} and r_{RN} , are replaced by the rate-normalized STRF energies, E_{DMR} and E_{RN} , for the corresponding conditions. Here the STRF energy is computed as:

$$E = \sqrt{\sum_{k=1}^{L} \int \text{STRF}_{r}(\tau, X_{k})^{2} d\tau}.$$
 (13)

Because the response of the neuron could be fractionally distributed between the contralateral and ipsilateral ears, the energy of the contraand ipsi-STRFs was measured independently, and the cumulative sum was taken as:

$$E_{\text{Total}} = \sqrt{E_{\text{c}}^2 + E_{\text{i}}^2}$$

where $E_{\rm c}$ and $E_{\rm i}$ are the contra- and ipsi-STRF energies. The STRF energy measures phase-locked activity (units of spikes per second) and is equivalent to the average phase-locked output for a linear integrating neuron (for proof, see Appendix B).

Phase-locking index

The phase-locking index (PLI) quantifies the ability of a neuron to phase lock to the spectrotemporal envelope. This metric is obtained by dividing the peak-to-peak STRF amplitude (in spikes per second) by the mean spike rate, *r*:

$$PLI = \frac{\max(\sigma_{S} \cdot STRF) - \min(\sigma_{S} \cdot STRF)}{r} \cdot \frac{1}{\Delta}, \qquad (14)$$

and normalizing this quantity by a theoretically derived factor, Δ , that corresponds to the theoretical maximum peak-to-peak rate-normalized STRF amplitude (confining this index to the range of 0–1). For the DMR, $\Delta = \sqrt{8}$, and for the RN, $\Delta = \sqrt{12}$ (for proof, see Appendix C).

Frequency domain analysis: ripple transfer function and conditioned response histogram

As an alternative to the STRF, we further evaluated neuronal response preferences to DMR and RN in the frequency domain. These approaches are useful, because they can be used to quantify neural responses as a function of ripple frequency and temporal modulation rate parameters.

The ripple transfer function (RTF) is one such descriptor. It is obtained directly from the STRF by performing a two-dimensional Fourier transform on the statistically significant STRF (p < 0.002), discarding the phase, and keeping the magnitude (see Fig. 5*A*,*B*). From the RTF, the best ripple density and best modulation rate parameters were determined for all phase-locking neurons. These are chosen by the location in the magnitude response with the peak amplitude. In instances in which two responses are observed (for negative and positive modulation rates), the secondary response was selected only if its response magnitude exceeded 50% of the maximum response magnitude. Positive (negative) modulation rates designate downward (upward)-going stimulus features; however, because the STRF is a time-reversed version of

best stimulus of the neuron, this convention is flipped for the neuron and its RTF (positive, upward sweep; negative, downward sweep).

Although this approach was successfully applied for many neurons, other neurons did not show statistically significant STRFs; therefore, it was impossible to estimate their RTFs directly. We therefore approximate the probability distribution function of observing a given set of parameters given a spike at time t_n , $P(F_m, \Omega|t_n)$, by performing a spike-triggered average with respect to the time-varying DMR parameters, $\Omega(t)$ and $F_m(t)$:

$$P_{kl} = \sum_{n=1}^{N} I[k\Delta F_{m} \le F_{m}(t_{n}) \le (k+1)\Delta F_{m}] \cdot I[l\Delta\Omega \le \Omega(t_{n}) \le (l+1)\Delta\Omega],$$
(15)

where P_{kl} is the discrete version of $P(F_m, \Omega|t_n)$, and $I[\cdot]$ is the identity function. The identity function takes a value of unity whenever the condition inside its argument is satisfied. Otherwise, it assumes a value of 0. Thus for any given bin of P_{kl} , this conditioned response histogram (CRH) is incremented by +1 if and only if the instantaneous parameters, $F_m(t_n)$ and $\Omega(t_n)$, fall within the required intervals, $k\Delta F_m \leq F_m(t_n) \leq (k$ + 1) ΔF_m and $l\Delta\Omega \leq \Omega(t_n) \leq (l+1)\Delta\Omega$, at the time of the neuronal spike, t_n (see Fig. 5*C*,*D*). Bin width resolutions of $\Delta F_m = 15-35$ Hz and $\Delta\Omega =$ 0.2–0.4 cycles per octave were used. The exact position used to estimate the parameters relative to the neuronal spike time, t_n , did not alter the resulting histogram (tested for a time lag of 0–50 msec), because the parameters vary at a slow rate (1.5 and 3 Hz) compared with the integration time of ICC neurons (usually tens of milliseconds).

As for single units, it was also useful to characterize population responses in the frequency domain, and we therefore extended these methods to include population statistics. By averaging the RTFs of individual neurons, we estimated the population ripple transfer function (pRTF) for those neurons with significant STRFs. To avoid biasing the pRTF because of systematic differences in firing strength, the RTFs of individual neurons were equally weighted so that the cumulative area of each was exactly 1.

For neurons that did not produce statistically significant STRFs, a modified approach was applied. We normalized the CRH of each neuron so that its cumulative sum was exactly 1. An average was then taken over the entire population, thereby producing the "population" CRH (pCRH). To facilitate comparisons, the pCRH was interpolated using the interp2 function (spline option) in MATLAB to identical resolution as for pRTF.

RESULTS

We studied 81 single neurons with the intent of understanding how dynamic spectrotemporal signals are processed within the central nucleus of the inferior colliculus. Specifically, we address whether single neurons integrate spectrotemporal information according to a linear integration model and whether dynamic stimulus aspects significantly affect neuronal encoding. Our complex stimuli constitute an integral part of the experimental protocol, and we fully characterize several pertinent properties of the stimulus ensembles. By design, both test sounds have identical average statistics and, therefore, equally sample the relevant spectrotemporal stimulus dimensions for this study. As a firstorder test of evaluating spectrotemporal response nonlinearities, we compute and compare the spectrotemporal receptive field for each sound type. We also characterize higher-order response attributes that are not directly accessible with the STRF descriptor.

Stimulus statistics: average versus dynamic spectrotemporal characteristics of the dynamic moving ripple and ripple noise

To test the possibility that individual auditory neurons in the ICC are selective for structural features prevalent in natural sounds (Fig. 1A,B), complex broadband stimuli (Fig. 1C,D) were designed that allow us to systematically identify nonlinear processing capabilities of auditory neurons. These stimuli fulfill a num-



Figure 1. Synthetic sound sequence used for reverse correlation analysis (C, D) and some corresponding natural sound counterparts (A, kitten vocalizations; B, babbling brook). The DMR (C) is designed to mimic spectral profiles created by formants (spectral energy peaks) and temporal modulations in speech production and animal vocalizations. The ripple density parameter, $\Omega(t)$, corresponds to the number of energy peaks (cycles per octave) along the spectral axis at time t. The temporal modulation rate, $F_m(t)$, describes the repetition rate of the envelope in hertz. The second stimulus, the RN (D), has noise-like properties that uniformly cover the ripple dimensions. The DMR and RN are shown for a maximum temporal modulation rate of 70 Hz, although a value of 350 Hz was used for the experiments.

ber of theoretical and ecological constraints: first, both sounds were designed to stringently meet a number of necessary requirements for use with the STRF. Second, both sounds incorporate a number of pertinent acoustic stimulus attributes that are prevalent in various natural signals [e.g., spectral energy peaks, frequency modulation (FM) sweeps, and temporal modulations] and that determine important perceptual qualities (Plomp, 1970, 1983; Van Veen and Houtgast, 1983).

The DMR stimulus (Fig. 1C) is an extension of the rippled spectrum noise used to characterize spectral and temporal response properties in the ferret and cat auditory cortex (Schreiner and Calhoun, 1994; Kowalski et al., 1996; Klein et al., 2000). This sound is constructed so that its spectrotemporal envelope is dynamic and coherently modulated ("structured") in time and frequency. As for speech and animal vocalizations (Fig. 1A), the DMR has strong short-time spectrotemporal correlations. These are determined by two independent parameters that vary randomly in time: the temporal modulation rate, $F_{\rm m}(t)$, and ripple density, $\Omega(t)$ (see Materials and Methods; Figs. 1C, and 2). The temporal modulation parameter determines the number of onsets and offsets per unit time (units of hertz) (Fig. 1C, top right). At any given time, the DMR sound produces a sinusoidal energy excitation pattern along the sensory epithelium, where the number of peaks per octave frequency is determined by the ripple density at that instant (Fig. 1C, top right). To efficiently excite neurons in the range characteristic for vocalizations, these parameters continuously vary at a nominal rate of 3 Hz (ripple density) (Fig. 2A) and 1.5 Hz (temporal modulation rate) (Fig. 2B) (in speech, for instance, similar features change at a rate of \sim 2–8 Hz; Greenberg, 1998).



Figure 2. Stimulus dynamics and spectrotemporal correlation statistics of the DMR and RN. The DMR parameter trajectories $\Omega(t)$ (A; ripple density, 0-4 cycles per octave) and $F_{\rm m}(t)$ (B; modulation rate, -350-350 Hz) are shown for a short 15 sec segment. The spectrotemporal parameters efficiently cover the ripple space (C; shown for the 15 sec segments of A, B). The instantaneous correlation function of the DMR (D) and RN (E) are shown for three distinct time instants, t_1-t_3 [D; left to right, $\Omega(t_1) = 1$ cycle per octave; $F_m(t_1) = 0$ Hz; $\Omega(t_2) = 2$ cycles per octave; $F_{\rm m}(t_2) = 150$ Hz; $\Omega(t_3) = 0.15$ cycles per octave; $F_{\rm m}(t_3) = -60$ Hz]. The RN instantaneous correlation function consists of a narrow central peak and a noisy surround (E). The global autocorrelation is identical for both sounds, consisting of an impulse-like central peak of width 3 msec and one-fourth octave (D, E, far right).

By averaging 16 independently chosen DMR envelopes, we designed a second stimulus, the RN. This sound is locally weakly correlated ("unstructured"), resembling background and environmental noises such as wind and rain (Fig. 1*B*). Visually, its spectrotemporal envelope (Fig. 1*D*) has a noisy profile both along time and along the spectral axis and lacks coherent modulations as present in the DMR and many vocalization sounds (Voss and Clarke, 1975; Attias and Schreiner, 1998; Nelken et al., 1999; Theunissen et al., 2000).

To characterize and compare the instantaneous versus the average behavior of these stimuli and their suitability for the reverse correlation method, the spectrotemporal autocorrelation function was evaluated for each stimulus. Dynamic properties were evaluated over short intervals of 10, 20, and 40 msec, which are comparable with integration times for ICC neurons. Global correlation statistics were evaluated for the ensemble as a whole (consisting of a 20 min continuous sound segment; see Materials and Methods). Both the local (shown for 10 msec analysis interval) and global spectrotemporal autocorrelations are depicted in Figure $2D_{,E}$.

The local autocorrelation depicts the spectrotemporal modulations that are present at a given time instant over a 10 msec segment. For the DMR stimulus, these take the form of tapered oscillations at a characteristic ripple density, modulation rate, and frequency sweep direction (Fig. 2D). Comparing the DMR and RN, it is clear that the local stimulus statistics are markedly different. Although the DMR has strong local correlations over the defined 10 msec intervals, the RN lacks any definitive spectral and temporal oscillations (Fig. 2E). Accordingly, its local autocorrelation is qualitatively similar at all time instants, consisting of a narrow central peak with a noisy surround. Therefore, the RN appears to be stationary or locally time-invariant. By comparison, the DMR has local envelope statistics that are dynamic; that is, they continuously vary with time.

By averaging the instantaneous autocorrelation function over all 10 msec time instants, it is possible to characterize the average statistics for the DMR and RN stimulus ensembles, which are identical (Fig. 2*D*,*E*, *far right*). In both cases, the average spectrotemporal autocorrelation assumes a narrow impulse-like character, which is the essential requirement for deriving receptive fields with the reverse correlation method (Eggermont, 1993; Klein et al., 2000).

Linear spectrotemporal receptive fields for DMR and RN

Neuronal data were evaluated by computing the STRF for neurons in the ICC and comparing neuronal responses to the spectrotemporally structured (DMR) and unstructured (RN) sounds. The STRF is a mathematical construct that describes the integrating area of the neuron along time and along the sensory epithelium (i.e., the frequency axis) and that depicts the spectrotemporal arrangement of neuronal excitation (red domains) and inhibition (blue domains). Figure 3 illustrates the spike-triggered average procedure we use to derive STRFs in response to DMR and RN. The STRF procedure requires that the probing stimulus have an unbiased modulation spectrum (both in time and along the sensory epithelium) or, equivalently, an impulsive spectrotemporal autocorrelation function that fully covers the physiologically relevant limits. Both the RN and DMR were designed with this constraint in mind; by limiting the temporal modulation rate to 350 Hz and the ripple density to 4 cycles per octave, we should be able to characterize 90-95% of the neurons in the ICC (Langner and Schreiner, 1988; Krishna and Semple, 2000) without biasing their STRFs.



Figure 3. Spike-triggered average and the STRF. At each instant of an action potential, the pre-event sound segment (up to 100 msec before spiking) is extracted and averaged for the entire stimulus ensemble. *Red* regions indicate stimulus patterns that were likely to be present whenever a neural response occurred at delay of 0. *Blue* indicates stimulus patterns that tended to be off at a moment before spike initiation. Functionally, these are interpreted as excitation (*red*) and inhibition (*blue*).

This essential property, which makes the RN and DMR stimuli suitable for reverse correlation, also permits the identification of spectrotemporal response nonlinearities. Given that both stimuli have identical low-order statistics (matched in intensity, contrast, and average envelope modulations), it is expected that a linear integrating neuron would have an average neural response that is similar for the RN and DMR conditions. That is, because both the RN and DMR stringently satisfy the necessary requirements for reverse correlation, we expect that $STRF_{DMR} = STRF_{RN}$ if the neuron behaves as a linear integrator (see Materials and Methods and Appendix B for proof). By comparing DMR and RN responses, we find that 60% (n = 49) of the neurons in our ICC sample met this requirement (Fig. 4). For reference, pure tone FTCs are shown alongside the RN and DMR STRFs when available (Fig. 4A,D). A red bar designates the mean sound pressure level (per one-third octave) for DMR and RN.

Neurons in our sample showed a variety of preference to stimulus patterns in the DMR and RN, including suppressive side bands, obliquely oriented excitatory or inhibitory regions, and distinct temporal response profiles (e.g., on-off, off-on, and off-on-off). Typically, excitatory and inhibitory STRF features were consistent between DMR and RN, although in some cases, inhibitory features were less pronounced for the RN (Fig. 4*E*,*F*). DMR and RN firing rates were generally high (mean spike rate, 11.2 spikes/sec for DMR and 11.8 spikes/sec for RN) and significantly correlated [correlation coefficient, 0.85 \pm 0.08 (mean \pm SE)] for this subset of neurons. Likewise, all neurons had comparable STRF energies. The neuron of Figure 4*B*,*C*, for instance,

had a spike rate of 34.0 spikes/sec for the DMR and 36.2 spikes/ sec for the RN (difference, 6%) and comparable STRF energies (E_{DMR} , 2.6 spikes/sec; E_{RN} , 3.0 spikes/sec; difference, 13%). The presence of well defined, statistically significant STRFs (p < 0.002) for both DMR and RN indicates that neurons efficiently phase locked to the stimulus spectrotemporal envelope. To distinguish these functional properties from those of other neurons in our sample, we refer to these as type I responses.

Frequency domain RF analysis

Complementary to the STRF, we also evaluated neuronal data in the frequency domain to extract physiologically meaningful parameters from the STRF and to describe neuronal preferences in terms of low-pass and bandpass filtering (Depireux et al., 2001; Klein et al., 2000).

First, we converted the STRF to an RTF (Fig. 5A,B). The RTF maps a the preferences of a neuron as a function of the temporal (modulation rate) and spectral (ripple density) stimulus parameters (see Materials and Methods). Whether a neuron integrates spectral or temporal information in a low-pass or bandpass manner depends strongly on the spectrotemporal relationship between neural excitation and inhibition in its STRF. For instance, the neuron of Figure 4B,C, has an on-off temporal response pattern; therefore, its RTF resembles a bandpass filter along the temporal modulation axis (Fig. 6A) that is centered at a best temporal modulation rate (bTM) of 45 Hz. Likewise, along the spectral axis, this neuron has a weak but significant inhibitory region alongside an excitatory region. Therefore its response as a function of ripple density also has a bandpass response profile with the dominant response peak centered at a best ripple density (bRD) of 0.6 cycles per octave. Neurons that lack interleaved patterns of excitatory (on) and inhibitory (off) subfields in their STRFs generally have low-pass response characteristics (Fig. 4I,J) along the spectral and temporal dimensions. The STRF of this example is marked by an off-on-off temporal response pattern, but its spectral STRF patterns lack interleaved excitatory and inhibitory subfields. Accordingly, its RTF (Fig. 6B) shows a bandpass response pattern in time (bTM, 200 Hz) and a low-pass response pattern along the spectral axis (bRD, 0 cycles per octave).

In a second related approach, a CRH was used to evaluate neuronal selectivity by tabulating the number of action potentials as a function of ripple parameters (see Materials and Methods). Unlike the STRF and RTF, this method accumulates the stimulus parameters, as opposed to the averaging stimulus waveforms, and is therefore insensitive to spike timing jitter. Figure 5C,D illustrates this approach. Generally, we find that RTF and CRH are in close agreement (Fig. 6). However, the CRH also reflects nonspecific activity, that is, action potentials that fall outside the dominant RTF boundaries and presumably do not contribute to the construction of the STRF (Fig. 6A,B).

Nonlinear spectrotemporal receptive fields for DMR and RN

One question addressed in this study is whether ICC neurons require specific acoustic features to be efficiently activated and whether these features can be identified using the STRF method. One reason why it may be difficult to identify the preferred acoustic features of a neuron using a direct approach is because conventional reverse correlation stimuli (such as the RN or spectrotemporal tone pips) seldom contain isolated sound patterns during a typical recording period. As an example, the DMR stimulus has pronounced energy peaks and FM sweeps that



appear in isolation in its spectrotemporal envelope (Fig. 1*C*). These same features are much more subtle in the RN (Fig. 1*D*), because they are superimposed with other components. How do such stimulus characteristics affect the ability of a neuron to respond, and which of these stimuli is better suited for identifying neuronal preferences in central auditory stations? Presumably, if a neuron exhibits substantial nonlinearities, significant differences could be expected between DMR and RN.

Not all studied neurons responded equally well to the DMR and RN. A small but significant (14%; n = 11) subset of neurons responded selectively to the DMR stimulus (Fig. 7; *type II* neurons). In general, type II neurons had low firing rates to the DMR and little or no response to the RN. Average firing rates for either stimuli were significantly lower than for type I responders (mean DMR, 0.61 spikes/sec; t test, p < 0.003; mean RN, 0.13 spikes/sec; t test, p < 0.0025). Surprisingly, despite the low spike rates, STRFs derived with the DMR were highly significant ($p \le 0.002$) and exceptionally clean.

Figure 7 depicts typical responses for these neurons. Some neurons (Fig. 7B–E) responded to both the DMR (0.24 and 1.4 spikes/sec) and the RN (0.14 and 0.2 spikes/sec) sounds, although their DMR firing rate was significantly stronger. DMR STRFs were highly significant (p < 0.002), with well defined excitatory and inhibitory subfields. However, the RN STRFs of these type II neurons were weak, with no distinguishable boundaries and excitatory and inhibitory subregions. Furthermore, the DMR STRF energy was 725% (Fig. 7B,C; E_{DMR}, 0.100 spikes/sec; E_{RN}, 0.012 spikes/sec) and 1280% (Fig. 7D,E; E_{DMR}, 0.276 spikes/sec; E_{RN}, 0.020) stronger, respectively, than for RN. Although these neurons did respond weakly to RN, other neurons responded exclusively to the DMR (Fig. 7G, H, J, K). Again, these neurons had extremely low spike rates (0.45 and 0.11 spikes/sec, respectively) to the DMR and no response to the RN (0 spikes). These STRFs were constructed using 276 (Fig. 7G) and 139 (Fig. 7J) spikes for

Figure 5. Frequency domain response analysis. The auditory STRF (A; shown for RN) is used to compute the RTF (B; shown for RN) by applying a two-dimensional Fourier transform. The RTF depicts time-locked energy in the neural response as a function of temporal modulation rate, F_m , and ripple density, Ω . Red indicates parameter combinations that evoked a strong time-locked response, and blue indicates a weak response. The CRH (D) characterizes nonlinear neuronal responses that do not show up in the STRF. For each neural event, the spectral and temporal DMR parameters, $\Omega(t_k)$ and $F_m(t_k)$, are determined at the time instance of the neural spike, t_k . The values of Ω and F_m are then used to increment the corresponding bin in the histogram by +1 (D).

Figure 6. Neuronal preferences determined with the RTF (*left column*) and CRH (*right column*) shown for the neurons of Figures 4*B*,*I* and 7*J*,*G* (*A–D*, respectively). The RTF and CRH depict the spectrotemporal frequency combinations (modulation frequency and ripple density) that preferentially activate a neuron. These can show either a low-pass or bandpass tuning profile along the temporal modulation or ripple density axis. Generally, neuronal tuning is similar for the RTF and CRH.

the DMR over a 10 and 20 min recording period, respectively. Nevertheless, their STRFs are as noise-free as those of type I responders that typically had thousands to tens of thousands of action potentials.

Interestingly, response characteristics for type II neurons are consistent with the idea that they are highly selective for some of the DMR stimulus features. The fact that we can compute highly significant DMR STRFs from very few spikes further suggests that the acoustic features leading to spike initiation must be precisely aligned in time and frequency; otherwise, STRFs would not accurately build up. To determine whether this is so, we recomputed all DMR STRFs using a subset of 100 randomly chosen action potentials for each neuron and determined the mean and maximum SNRs of those pixel values that exceeded a significance criterion of p < 0.002 (see Materials and Methods). The SNR of these conditioned DMR STRFs was approximately twice as strong for type II neurons (average maximum SNR, 8.7 for type II vs 4.0 for type I; paired t test, $p < 3.5 \times 10^{-5}$; average mean SNR, 4.7 for type II vs 2.8 for type I; paired t test p < t0.003). This suggests that the spectrotemporal waveforms added to compute the STRF are more consistent from spike to spike for type II neurons compared with type I neurons. Consequently, type II neurons are highly sensitive for particular stimulus features in the DMR stimulus, resulting in exceptionally clean STRFs that can be obtained with very few action potentials. Response specificity is also reflected in the CRH for these neurons. Compared with type I responses, CRHs for type II responses show highly localized peaks (Fig. 6*C*,*D*, *far right*) and lack nonspecific activity. Together, the low firing rates, high response specificity to the DMR, and unresponsiveness to RN demonstrate that these neurons are extremely nonlinear and highly selective for isolated spectrotemporal sound patterns.

It may be argued that the seemingly low spike rates and sparse responses of these neurons are simply attributed to stimulus levels near or below the response threshold of the neuron. We tested for this possibility in 6 of the 11 neurons by computing FTCs with pure tones (Fig. 7A, F, I). The FTCs are shown alongside DMR and RN STRFs, with a red line depicting the mean intensity per one-third octave during DMR and RN stimulation (mean \pm SE SPL per one-third octave, 69 ± 9 dB). In all cases, the DMR and RN intensity operating points were well above the response threshold of the neuron, thus arguing against potential thresholding effects. Many of these neurons had bandwidths exceeding one-third octave. Therefore, the actual energy exceeded the onethird octave estimate by up to 12 dB (Fig. 7G). For the five neurons for which frequency-tuning curves were not available, it is unlikely that these were near or less than the threshold, because DMR and RN were presented for these at moderately loud SPLs (58, 58, 78, 78, and 88 dB/one-third octave, respectively).

STRF shape, energy, and firing rate differences between DMR and RN

Differences in response activity between DMR and RN for type I and II responses were quantified with three metrics to independently assess STRF shape, mean firing rate, and STRF energy differences. STRF shape differences were quantified with the STRF SI (DeAngelis et al., 1999; Reich et al., 2000). The SI assumes values between -1 and 1. Values of 1 indicate that the STRFs have identical shapes. Values near -1 indicate that the STRFs have identical shapes but are of opposite polarity, and SI values near 0 occur only for STRFs that have nothing in common. The RDI and the STRF MDI quantify the percent change in mean firing rate and STRF energy between the RN and DMR. Values of 0 for the RDI indicate that the mean firing rates are identical ($r_{\rm DMR} = r_{\rm RN}$), whereas values >0 indicate that $r_{\rm DMR}$ > $r_{\rm RN}$. Values <0 indicate that $r_{\rm DMR} < r_{\rm RN}$. The magnitude of the RDI is numerically equivalent to the percent difference between $r_{\rm DMR}$ and $r_{\rm RN}$. The MDI is identical in form to the RDI, where the STRF energies, E_{DMR} and E_{RN} , are now substituted for the mean firing rate, r_{DMR} and r_{RN} , respectively. This metric therefore characterizes phase-locked or driven activity as depicted by the STRF.

In extreme scenarios, neurons either responded equally well to both sounds or responded only to the DMR. This was evident from both the similarity index statistics and the rate and magnitude disparity index. Figure 8*B* shows the similarity index distribution for all neurons that had significant STRFs for the DMR or RN stimulus conditions or both. The distribution of SI values is bimodally distributed. Most neurons (n = 49) had similar DMR and RN STRFs and therefore high SI values, >0.5 (mean SI, 0.75). These were classified as type I neurons. The remaining neurons (n = 11) had SI values of <0.5. Of these, two neurons had values of SI that were nearly 0.5 (0.48 and 0.49) (Fig. 7*D*,*E*), and six neurons had SI values that were identically 0 (Fig. 7*G*,*H*,*J*,*K*). Neurons in the latter subset responded to the DMR

Figure 7. Spectrotemporal receptive fields of neurons that responded specifically to the DMR sound (*B*, *D*, *G*, *J*, middle column) but responded weakly or had no response to the RN (*C*, *E*, *H*, *K*, right column). Frequency-tuning curves derived with pure tones are shown for reference (*A*, *F*, *I*, left column). Red lines designate the mean stimulus level (per one-third octave) used for DMR and RN. Significant STRF patterns are denoted by red contours. All neurons are shown at distinct spectral and temporal scales.

sound and produced statistically significant STRFs but did not respond to the RN sound. Therefore, although most neurons had similar integration areas for RN and DMR, significant differences in STRF shape were usually attributed to improper activation by RN. Consequently such neurons were classified as type II responses.

Although the SI index statistics are consistent with the observed response types of Figures 4 and 7, they do not tell us anything about the driven and average activity to these stimuli. The rate and STRF magnitude disparity index corroborate the results of Figures 4 and 7. Although most neurons (n = 49; type I) had RDI and MDI centered about 0 (|RDI| < 500% and |MDI| <500%; mean MDI, 49.1%; mean, 2.7%), a large subset of neurons (n = 11; type II) had values for either of these metrics that exceeded +500% (mean MDI, 896%; mean RDI, 712%; cluster verified post hoc, link tree cluster analysis) (Fig. 8A). Thus, the average activity or phase-locked activity of these neurons tended to be significantly higher for the DMR stimulus, consistent with type II response characteristics and the examples of Figure 7. Five of these neurons had MDI values between 500 and 1000%. For three neurons, the RDI values were <500%, and observable response differences manifested themselves only as a significant change in driven activity (MDI > 500%) (Fig. 7B,C). An additional six neurons had very large values of RDI and MDI, because they responded to the DMR stimulus but produced zero spikes for the RN sound (Fig. 7G, H, J, K). These are shown collectively as a single point centered about MDI of +1000% and RDI of +1000%.

STRF construction and the effects of phase locking

A basic requirement for computing the STRF is that the action potential linearly time lock or phase lock to the stimulus spectrotemporal envelope. Sinusoidal amplitude modulation studies have demonstrated that many ICC neurons phase lock to the stimulus modulation waveform (Rees and Møller, 1983, 1987; Møller and Rees, 1986; Langner and Schreiner, 1988; Krishna and Semple, 2000). Accordingly, a large percentage of neurons in this study phase locked to the spectrotemporal envelope and consistently produced statistically reliable STRFs (n = 61 of 81).

The remaining neurons (n = 20 of 81), failed to produce statistically reliable STRFs (p < 0.002) with a distinct spectrotemporal patterning (Fig. 9), despite a significant overall firing rate (mean firing rate, 7.5 spikes/sec). We labeled these neurons type III. One possible explanation is that these neurons were spontaneously firing and did not respond in a time-dependent manner to the DMR and RN. One of a number of possible alternatives is that these neurons responded selectively to energy fluctuations of the DMR and RN but did not linearly phase lock to their spectrotemporal envelope. Therefore, waveform averaging to estimate neuronal receptive fields would be of little use.

To test this possibility, we computed the CRH for these neurons. This procedure allows us to test whether type III neurons respond selectively to complex sound attributes even if they do not posses the necessary timing precision in their stimulusresponse alignment for producing STRFs. The CRH of all of these neurons revealed strong responses to particular stimulus parameter combinations (Fig. $9C_{2}F_{2}H$) despite the lack of linear time locking to the spectrotemporal envelope (resulting in no STRF in Fig. 9B,E or a very weak STRF in Fig. 9G). Thus, the responses of these neurons do not linearly follow the fast spectrotemporal modulations of the stimulus envelope (up to 350 Hz) but were able to track very slow changes of the stimulus parameters (1.5 Hz for the temporal modulation rate and 3 Hz for the ripple density) with changes in firing rate. On the basis of the STRF and mean firing rate alone, one would conclude that these neurons are only spontaneously firing without functional consequences for encoding stimulus information. However, the cumu-

Figure 8. Response statistics comparing the DMR versus the RN. The MDI and RDI quantify differences in mean firing rate and driven activity for the DMR and RN (A). Type II neurons have RDI and MDI values that exceed the 500% contour. STRF shape differences are quantified with the SI, which usually takes values from 0 (not similar) to 1 (very similar). The population similarity index distribution (B) is bimodally distributed, with the majority of neurons falling at ~0.7.

lative analysis of the stimulus–response relationship reveals that these neurons do respond selectively to pertinent stimulus parameters (Fig. 9C,F,H).

In the few instances in which significant STRFs (p < 0.002) were observed (n = 6 of 20) for type III neurons, these were diffuse and weak (Fig. 9G), despite the fact that the CRH was strong and tightly tuned (Fig. 9H). For comparison purposes, the color scale on all STRFs including those of Figures 4 and 7 are normalized so that the minimum and maximum values correspond to half of the mean firing rate (in the case in which the STRF amplitude exceeded these limits, the maximum absolute value of the STRF was used). Most neurons had STRF magnitudes that fell below this range of values, although these limits were often exceeded for neurons with type II responses (as is the case for all the neurons of Fig. 7). In the case in which the STRFs are absent, this observation indicates that the sound waveforms that were used to construct the STRF were not phase-aligned and, therefore, do not add constructively. In the case of type I responses, the sound waveforms are presumably moderately aligned, whereas for type II, they are tightly aligned (allowing the peak-to-peak rate of the STRF to exceed the mean firing rate of the neuron).

Examples depicting the different phase-locking scenarios for the three neural types are depicted in Figure 10B-D for a short 5 sec segment of the DMR stimulus. The type III neuron (Fig. 10D; same neuron as in Fig. 9A-C) had an elevated firing rate but showed no obvious correspondence between the occurrence of action potentials and the DMR stimulus spectrotemporal pattern (Fig. 10A, far right). The type I neuron of Figure 10B (same neuron as in Fig. 4A-C) had a high spike rate and a phasic response raster. Similarly, the type II neuron of Figure 10C (same neuron as in Fig. 7F, G) showed precisely aligned phasic response components; however, this neuron had a low spike rate to the DMR and no spontaneous background activity. By comparing its STRF (Fig. 7G), its response raster (Fig. 10C, far right), and the DMR stimulus sound pattern (Fig. 10A, far right), it is evident that the neuron responds specifically if the spectrotemporal sound patterns closely match the STRF of the neuron. This level of temporal specificity is less pronounced for the type I neuron (Fig. 10B) and absent for the type III neuron (Fig. 10D).

We quantified the phase-locking abilities of all neurons by computing the PLI (see Materials and Methods; Fig. 10E) for the DMR stimulus. This metric can assume values between 0 and 1 (observed range, 0-0.75), where 0 indicates no linear phase locking and 1 indicates maximal linear phase locking. Results for the population are consistent with the examples of Figures 4, 7, and 9. Type III neurons, which have no STRFs, had the lowest PLI values (mean PLI, 0.076 \pm 0.02; bootstrap p < 0.01 confidence interval; Fig. 9B, 0.028, E, 0.09, G, 0.093), and neurons with type II responses had the highest values (Fig. 7B, 0.46, D, 0.42, G, 0.64, J, 0.65). As postulated for type II responses, high PLI values (mean, 0.50 ± 0.13 ; bootstrap p < 0.01 confidence interval) suggest that the sound waveforms used to construct STRFs add constructively and are tightly aligned. In contrast, type I responders had intermediate PLI values (Fig. 4B, 0.22, E, 0.20, G, 0.14, I, 0.13; mean, 0.24 \pm 0.04; bootstrap p < 0.01 confidence interval), indicating that sound waveforms are moderately aligned.

Spectrotemporal filtering statistics

As noted previously, neuronal preferences to features of the DMR and RN depend strongly on the spectrotemporal arrangement and size of excitatory and inhibitory receptive field regions. These, in turn, determine the range of spectral and temporal preferences of each neuron and whether their filtering characteristics are bandpass or low-pass.

To evaluate the processing capabilities of all neurons and to characterize any systematic differences among type I-III responses, we measured the bRD and bTM parameters of each neuron (Fig. 11A). Because most neurons (77 of 81, $\sim 95\%$) responded symmetrically to upward-going (positive temporal modulation) and downward-going (negative temporal modulation) ripples, two values of the best parameters were extracted (one for each quadrant of the RTF; i.e., one for the positive and one for the negative modulation rate value). For type III responses, these were estimated directly from their CRH. bTM and bRD show a distinct covariation for time-locked responses (types I and II). There is a strong negative correlation between the absolute magnitude of bTM and bRD (type I, $r = -0.6 \pm 0.06$ bootstrap SE; $p < 1 \times 10^{-5}$; type II, $r = -0.5 \pm 0.1$ bootstrap SE; $p < 1 \times 10^{-5}$). Evidently, time-locking neurons that prefer fast temporal modulations also prefer stimuli with broad spectral features, and neurons that prefer slow temporal modulations can respond efficiently to stimuli with narrow or broad spectral features. This trend was significantly different for type III responses,

in which the absolute magnitude of bTM and bRD showed no correlation ($r = 0.1 \pm 0.16$ bootstrap SE; p > 0.78).

We characterized the overall spectrotemporal filtering capability of the ICC by averaging the DMR RTFs of individual neurons to estimate the pRTF (see Materials and Methods). The composite pRTF depicts a clear trend in the spectrotemporal filtering profile. Neurons with type I and II responses had similar response profiles (pRTF correlation coefficient, 0.765 \pm 0.015; p < 0.01bootstrap confidence interval) in which spectral resolution appears to be traded for temporal resolution (Fig. 11B,C). At low modulation rates, filtering profiles extended to intermediate ripple densities (up to two cycles per octave); however, at high modulation rates, neurons are sensitive only to low ripple frequencies. Overall, spectral filtering is low-pass, whereas temporal filtering characteristics are bandpass. This was true for both type I and II responses; however, the pRTF of type II responses is more compact, as evident from the 95th percentile contours (Fig. 11B,C, solid line). By direct comparison, the filtering profile of type III responses (Fig. 11D) is diffused and shows no systematic patterns as for type I and II responses (pRTF correlation coefficients: I vs III, 0.44 \pm 0.016; II vs III, 0.2695 \pm 0.015; p < 0.01bootstrap confidence interval). Accordingly, spectrotemporal filtering characteristics differ significantly between neurons with strong and weak phase locking.

DISCUSSION

Neurons in the central auditory system respond selectively to both spectral and temporal stimulus attributes (Rees and Møller, 1983, 1987; Schreiner et al., 1983; Langner and Schreiner, 1988; Schreiner and Langner, 1988; Nelken et al., 1997; Eggermont, 1999; Ramachandran et al., 1999; Krishna and Semple, 2000). Although this is well described for narrow-band stimuli, clicks, and modu-

lated tones, only a few studies have addressed how these acoustic dimensions are jointly processed by the brain. This is of interest, because natural sounds are composed of both spectral and temporal sound components, and because, in general, the response to complex stimulus ensembles cannot be extrapolated directly from the neuronal responses to simpler sounds. Our data demonstrate that $\sim 60\%$ of neural responses to complex auditory stimuli in the ICC are consistent with a linear integration model (type I neurons). However, we also identified conditions in which this model fails at fully describing neural responses. This result is true for type II and III neurons. Type II neurons phase lock well and respond selectively to spectrotemporal stimulus features. However, these neurons are not efficiently activated with the RN stimulus. In contrast, type III neurons do not phase lock tightly to the stimulus envelope, but they respond selectively to both spectral and temporal stimulus parameters.

Recent studies have demonstrated the usefulness of the STRF procedure for studying neural processing of complex sounds in the auditory cortex (de Charms et al., 1998; Klein et al., 2000; Miller et al., 2002) and its avian homolog (Theunissen et al., 2000). They have shown that it is indeed possible to approximate the stimulus–response function of some central auditory neurons as linear functions of their inputs. As for the related visual STRF, the auditory STRF can be used to identify the spectrotemporal features of the stimulus that a neuron prefers (de Charms et al., 1998). Because this method describes neural processing in terms of time and the sensory epithelium receptor surface, it has become a valuable and intuitive experimental tool. Despite its general utility, nonlinear aspects of neural integration are often difficult to identify and cannot be fully accounted for with the STRF (Young, 1998; Theunissen et al., 2000). Our results build

Δ

Stimulus

ON

OFF

Figure 10. Phase-locking statistics for the DMR stimulus. A, Left, A 5 sec segment of the DMR stimulus was presented to each neuron. Rastergrams (B-D, left) show individual response traces for 40 consecutive presentations for type I (B), II (C), and III (D) neurons. The occurrence of each action potential is shown as a single dot (1 msec resolution). A-C, Far right, Cutout (red) detailing the stimulus spectrotemporal envelope (A, far right) and the response rastergrams of each neuron (B-D, far right). E, The PLI measures the ability of a neuron to phase lock to the spectrotemporal envelope of a sound. A PLI of 0 indicates minimal linear phase locking, and a PLI of 1 indicates perfect phase locking. PLI distribution is skewed toward low values (mean PLI, 0.24) but extends over a broad range of 0-0.75.

on those findings, suggesting that the STRF is useful in various respects, but it does not fully account for all response nonlinearities. Therefore, to systematically identify nonlinear aspects of processing, complementary approaches and acoustic stimuli must be examined. For instance, approaches that extend the STRF method by performing a second-order reverse correlation (with respect to the spectrotemporal envelope) could be used. These methods, however, generally require a substantial amount of data and can easily fail, especially if the types of nonlinearities are not compatible with the approach (i.e., they must be of even order with respect to the envelope of the sound). Furthermore, results from any such methods may depend strongly on the stimulus ensemble used (Theunissen et al., 2000).

As outlined in the introductory remarks, limitations encountered with the STRF technique can be either stimulus-dependent or methodological in nature. Stimulus-dependent limitations are characterized by the inability of a stimulus to efficiently activate highly nonlinear sensory neurons. For instance, high-level auditory neurons in the avian forebrain in bats and other acoustically specialized animals exhibit complex nonlinearities and respond efficiently to acoustic features found in their vocalizations. Such high-level sensory neurons are likely optimized to analyze acoustic features and combinations found in natural signals (Suga and Jen, 1976; Suga et al., 1978; Margoliash, 1983; Doupe, 1997; Portfors and Wenstrup, 1999; Theunissen et al., 2000). Therefore, when these animals are studied using synthetic stimuli, such as conventional reverse correlation sequences, neural responses are generally weak, and, consequently, a quantitative evaluation of the stimulus response function is not possible (Theunissen et al., 2000). However, by presenting natural stimuli that contain pertinent stimulus correlations, some of these limitations can be overcome (Theunissen et al., 2000).

Our findings in the ICC further demonstrate the importance of the probing stimulus characteristics and how these may interact with the sensory system. The fact that significant response differences exist between DMR and RN is evidence that, for some neurons, a direct STRF procedure using conventional stimuli may be insufficient. Neurons with type II responses, for instance, cannot be characterized with RN stimuli, despite the fact that this stimulus contains the essential characteristics required to estimate auditory STRFs. The DMR, which is by all accounts a nontraditional reverse correlation sound, allowed us to characterize receptive field of these neurons. These differences suggest that high-order acoustic features in the DMR efficiently drove such neurons, whereas RN did not. By simply comparing responses between DMR and RN, we have therefore taken a

Figure 11. pRTFs and best ripple parameter statistics. A, Scatter plot of the bRD and bTM for the observed neural responses (triangles, type I; circles, type II; squares, type III). The pRTF of type I (B), II (C), and III (D) neurons depicts the filtering profiles of each neural response type. Black contours designate the 95th percentile interval (boundaries that account for 95% of the area under the pRT).

significant step toward identifying some of the acoustic features that are necessary to efficiently activate such neurons. At the same time, this comparison allows us to dissociate linear from nonlinear spectrotemporal interactions.

By controlling for multiple stimulus parameters (including SPL, contrast, and global second-order correlations), we show that instantaneous correlations of the probing stimulus are essential for activating some neurons. The DMR stimulus dynamically and efficiently probes the entire acoustic ripple space by providing maximal driving force over short periods (Fig. 2). Our finding that some neurons require strong time-limited correlations to activate them is not unexpected, because ICC neurons integrate stimulus information over a restricted temporal extent of less than \sim 50 msec. Similar processing principles are likely operative for natural signals (Theunissen et al., 2000); however, the large number of degrees of freedom necessary to describe natural stimuli makes their identification difficult.

Neurons with type II responses consistently produced highly significant STRFs despite low spike rates for the DMR. In fact, when the level of significance was compared for an equal number of action potentials, STRFs of neurons with type II responses were more significant and had a higher signal-to-noise ratio than for type I. Therefore, spectrotemporal acoustic features in the DMR were added effectively during the construction of these STRFs. This finding indicates that type II neurons precisely phase lock to particular stimulus features of the DMR, as reflected by the higher phase-locking index values (Fig. 10). These observations support the idea that type II neurons selectively respond to particular stimulus features within the DMR, whereas type I neurons integrate stimulus information in a quasilinear manner.

Although we cannot speak directly about the exact mechanisms underlying these nonlinear response characteristics, the general nature of the observed effect points to several possibilities. For instance, active engagement of inhibitory and excitatory neuronal inputs combined with intracellular thresholding in the ICC (Kuwada et al., 1997) could account for the low spike rates and observed differences between DMR and RN in type II neurons. If the inhibitory inputs are sufficiently strong, broadband stimulation would significantly reduce firing rates, because stimulus energy would almost always overlap inhibitory RF domains. This is especially true for RN, because it has a short but constant correlation width of \sim 3 msec and one-fourth octave. This possibility is supported by the fact that the fraction of inhibitory STRF energy was larger for type II responses than type I (mean \pm SE, 40 \pm 6 vs 36 \pm 8%; paired t test, p < 0.05). Even if subthreshold summation is strictly linear, a high intracellular reversal potential would drastically reduce overall spike rates. Under such conditions, the neuron would be most likely to fire only when the stimulus modulations precisely overlap the excitatory and inhibitory RF of the neuron. Such active engagement of excitation and disengagement of inhibition would allow the intracellular potential of the neuron to reach the spike initiation threshold. The strong instantaneous spectrotemporal correlations of the DMR stimulus could, under such circumstances, provide the necessary driving force to selectively activate and deactivate excitatory and inhibitory inputs. Preliminary modeling results (data not shown) are consistent with our findings and suggest that such mechanisms could serve as a general basis for selectivity enhancement, similar to feature selectivity mechanisms observed in other species and modalities (Casseday et al., 1994; Moore and Nelson, 1998; Bringuier et al., 1999).

Methodological limitations of the STRF are evident for type III neurons that, despite significant firing rates, showed no significant STRFs. We overcame these limitations by devising an alternate functional descriptor, the CRH. This consisted of performing a spike-triggered histogram with respect to the timevarying DMR parameters, as opposed to the stimulus spectrotemporal envelope. The fact that we do not obtain STRFs despite selective activation to both spectral and temporal attributes points to several mechanisms. On the one hand, dominant evenorder nonlinearity would render the STRF method useless, because this technique only characterizes suitable projections from odd-order nonlinearities. For example, a simple squaring operation would cause a linear neuron to phase lock to both stimulus onsets and offsets, the average of which is precisely zero. Therefore, the average pre-event stimulus (i.e., its STRF) would be zero. Such nonlinearities are well described for complex cells in the primary visual cortex, which have strong even-order nonlinearities and consequently do not produce linear spatiotemporal receptive fields (Emerson et al., 1987; Szulborski and Palmer, 1990). A number of alternate mechanisms, however, could also produce a similar result.

One such possibility is random spike-timing jitter, a mechanism likely responsible for loss of temporal synchrony at fast temporal modulation rates (Epping and Eggermont, 1986; Langner and Schreiner, 1988; Schulze and Langner, 1997; Krishna and Semple, 2000). If spike-timing jitter is comparable in its time scale with that of the preferred stimulus feature of the neuron (by as little as half of the STRF period), the spectrotemporal patterns that are added during the STRF computation would be randomly out of phase and would, therefore, not add constructively. This is especially true at high temporal modulation rates, at which the time scales for neural integration of fast stimulus features are at the limits of the internal precision of the spike generation mechanisms (on the order of a few milliseconds). Under such conditions, a small amount of jitter would abolish the STRF. If, on the other hand, a neuron prefers slow temporal modulations, a small amount of temporal jitter would distort or blur the STRF of the neuron but would not abolish it in its entirety. This possibility is supported by the fact that bTMs were significantly higher for neurons with type III responses than for those with type I and II responses (mean bTM, 190 vs 75 Hz; paired t test, $p < 2 \times 10^{-7}$). Our modeling results (data not shown) suggest that both of these mechanisms produce results identical in character to those observed for type III responses.

Although we have chosen to break up our data into functionally defined subgroups of neurons, our methods cannot distinguish between anatomically defined neural populations (Oliver and Morest, 1984) and functionally defined neural inputs into the ICC (Ramachandran et al., 1999). Our findings, however, show that the ability of a neuron to respond to DMR versus RN is ultimately reflected in other response properties, such as its phase-locking ability, its SNR, and even its preferred spectrotemporal parameters. Differences in the spectrotemporal filtering abilities of each neural type were determined from the best spectral and temporal parameters of each neuron or from the population transfer functions. Type I and II neurons had similar spectrotemporal preferences in which the preferred ripple density and modulation rate showed a strong negative correlation. Furthermore, the range of modulation rates and ripple densities were more restricted for type II neurons, indicating that the STRFs of these neurons were typically larger. Type III neurons, by comparison, showed no systematic filtering pattern; however, the observed modulation rates were significantly higher than for type I or II neurons. These differences in filtering ability argue for distinct coding strategies within the ICC according to differences in the spiking output (e.g., the degree of phase locking).

Advances in the STRF mapping techniques using natural sounds and other naturalistic stimuli (Klein et al., 2000; Theunissen et al., 2000) are providing the means to study complex nonlinearities that are necessary for the brain to efficiently processes sensory information from the outside world. Our findings delineate rules for spectrotemporal sound processing in the ICC that cannot be accounted for by linear integration models and that can, in general, not be characterized alone with narrowband stimuli, conventional reverse correlation stimuli, and direct STRF methods. Because of the dynamic spectrotemporal nature of natural sounds, such processing principles likely play an important role for natural sound analysis.

APPENDIX A

Nonlinear response characteristics are tested against the expected response of an idealized linear neuron. Because the RN and DMR both have identical autocorrelation functions, a hypothetical linear neuron would produce identical STRFs and RTFs for these sounds.

To prove this, we consider a multi-input, single-output linear filter bank (Marmarelis and Naka, 1974) as a model representation for auditory neuronal filtering. This representation is motivated by the fact that the primary sensory epithelium performs a spectrotemporal decomposition of incoming sounds, and consequently, all further processing along the auditory system is constrained by this output pattern.

The spectrotemporal filter bank model consists of a set of *L* octave spaced linear modulation filters, $[h_1(\tau), h_2(\tau), \ldots, h_L(\tau)]$, where $h_k(\tau) = \text{STRF}(\tau, X_k)$ is the impulse response of a linear filter centered about the frequency band X_k , and τ corresponds to the temporal lag of the filter. The expected firing rate of the neuron, r(t), is obtained by summing the firing rate contribution for each of the tonotopically arranged frequency channels:

$$r(t) = r_0 + \sum_{k=1}^{L} r_k(t),$$
(16)

where r_0 is the mean firing rate of the neuron (zero-order kernel), and the output of its *k*th frequency band is given by the convolution integral:

$$r_{\mathbf{k}}(t) = \int h_{\mathbf{k}}(\tau) s_{\mathbf{k}}(t-\tau) d\tau + e_{\mathbf{k}}(t), \qquad (17)$$

where $s_k(t) = S(t, X_k)$ is the modulation input to the *k*th filter channel, and $e_k(t)$ is a noise term that arises from measurement error and the internal noise of the neuron. For the nonlinear case, $e_k(t)$ contains the nonlinear response contributions that cannot be accounted for by the linear description (Klein et al., 2000). For practical reasons, we assume that $e_k(t)$ is statistically independent of the input, $s_k(t)$ and has 0 mean and SD of σ_{e_k} .

To compute the STRF from the experiment data, we perform a cross-correlation between the input and output. For the linear model neuron, this procedure is expressed as:

$$E[(r(t) - r_0) \cdot s_1(t - \sigma)] = \sum_{k=1}^{L} E[r_k(t) \cdot s_1(t - \sigma)]$$
$$= \sum_{k=1}^{L} \int E[s_k(t - \tau) \cdot s_1(t - \sigma)] \cdot h_k(\tau) d\tau$$
$$+ E[e_k(t) \cdot s_1(t - \sigma)]$$
$$= \sum_{k=1}^{L} \int R_{ss}(\tau - \sigma, X_k - X_l) \cdot h_k(\tau) d\tau,$$

(18)

where:

$$E[\cdot] = \lim_{T \to \infty} \int_{-T/2}^{T/2} \cdot dt$$

is the time average operator, and $R_{\rm SS}(\tau, \xi)$ is the stimulus average spectrotemporal autocorrelation function. For a sufficiently large recording period, T, the error cross-correlation $E[e_k(t) \cdot s_l(t - \sigma]$ approaches 0, because $e_k(t)$ and $s_l(t)$ are statistically independent and both have 0 means. Because the above equation is strictly a function of the stimulus long-term spectrotemporal autocorrelation function, $R_{\rm SS}(\tau, \xi)$, and is independent of the stimulus local statistics, an idealized linear neuron ought to produce identical STRFs for both sounds as hypothesized. This is expected, because the global spectrotemporal autocorrelation is identical for both stimuli (Eq. 9, Fig. 2).

To show that Equation 18 degenerates into a spike-triggered average, we consider the impulse-like correlation properties of the RN and DMR. If the spectrotemporal autocorrelation of the stimulus has the unique property that it has impulse-like characteristics, that is, $R_{\rm SS}(\tau, \xi) \approx \sigma_{\rm S}^2 \cdot \delta(\tau) \cdot \delta(\xi)$, then the spectrotemporal cross-correlation between the stimulus and the output simplifies to:

$$E[(r(t) - r_0) \cdot s_{\mathbf{l}}(t + \sigma)] = \sigma_{\mathbf{s}}^2 \cdot \sum_{k=1}^{L} \int \delta(\tau - \sigma) \cdot \delta(X_{\mathbf{k}} - X_{\mathbf{l}}) \cdot h_{\mathbf{k}}(\tau) d\tau$$
$$= \sigma_{\mathbf{s}}^2 \cdot h_{\mathbf{l}}(\sigma), \quad (19)$$

where σ_s^2 is to the variance of the spectrotemporal envelope. The spectrotemporal receptive field for the model neurons is instantly derived as:

$$\mathrm{STRF}(\sigma, X_{\mathrm{l}}) = h_{\mathrm{l}}(\sigma) = \frac{1}{\sigma_{\mathrm{s}}^{2}} \cdot E[(r(t) - r_{0}) \cdot s_{\mathrm{l}}(t - \sigma)]. \tag{20}$$

Therefore, for both RN and DMR, the STRF can be estimated by performing a cross-correlation between the response of the neuron, r(t), and each of its *L* inputs, $S_k(t)$, for k = 1, ..., L. Although the spectrotemporal correlation of the RN and DMR is not strictly an impulse (temporal correlation width, ~3 msec; spectral correlation width, one-fourth octave), it is in general significantly tighter than the spectrotemporal integration areas of ~95% of ICC neurons (Rees and Møller, 1987; Langner and Schreiner, 1988; Schreiner and Langner, 1988; Krishna and Semple, 2000) and can therefore be approximated as such.

For a spiking neuron with an action potential sequence $r(t) = \sum_i \delta(t - t_i)$ of *N* neuronal event times, t_i , Equation 20 can be easily expanded as a spike-triggered average:

$$h(\sigma, X_{\rm l}) = 1/(\sigma_{\rm s}^2 \cdot T) \cdot \sum_{i=1}^{L} s_{\rm l}(t_{\rm i} - \sigma).$$
 (21)

In practice, T corresponds to the experimental recording period (600–1200 sec for these experiments).

Therefore if the grand average spectrotemporal autocorrelation function of the stimulus has impulse-like properties, it is possible to estimate the STRF of the neuron via Equation 21. Other stimulus aspects, such as high-order statistics, and stimulus dynamics have no bearing on this result under the assumption of quasilinear integration. Both the RN and DMR satisfy the global correlation requirement and therefore should produce identical results for a linear integrating neuron.

APPENDIX B

Consider the linear model neuron of Equations 16 and 17, where the filter for the *k*th input channel is related to the STRF of the neuron by $h_k(\tau) = \text{STRF}(\tau, X_k)$. We would like to derive a metric that quantifies the energy in the response of the neuron that is captured by the STRF of the neuron. We do this by computing the expected output SD or, equivalently, the firing rate variance of the neuron that is predicted by its STRF. The predicted firing rate variance is expressed as:

$$\sigma_{\rm r}^2 = E[(r(t) - r_0)^2] = \sum_{i=1}^L \sum_{k=1}^L E[r_j(t) \cdot r_k(t)],$$
(22)

where r(t) is the predicted firing rate of the neuron (Eq. 16), r_0 is the mean firing rate of the neuron, and $r_k(t)$ is the predicted output for the *k*th filter channel. Substituting Equation 17 into the expectation of Equation 22 yields:

$$E[r_{j}(t) \cdot r_{k}(t)] = E[\int s_{j}(t - \tau_{1}) \cdot h_{j}(\tau_{1})d\tau_{1} \cdot \int s_{k}(t - \tau_{2}) \cdot h_{k}(\tau_{2})d\tau_{2}]$$

$$\iint E[s_{j}(t - \tau_{1}) \cdot s_{k}(t - \tau_{2})] \cdot h_{j}(\tau_{1}) \cdot h_{k}(\tau_{2})d\tau_{1}d\tau_{2}$$

$$\iint R_{ss}(\tau_{1} - \tau_{2}, X_{j} - X_{k}) \cdot h_{j}(\tau_{1}) \cdot h_{k}(\tau_{2})d\tau_{1}d\tau_{2}, \quad (23)$$

where $R_{\rm SS}(\tau_1 - \tau_2, X_j - X_k) = \sigma_{\rm S}^2 \cdot \operatorname{sinc}[2\Omega_{\rm Max}(X_j - X_k)] \cdot \operatorname{sinc}[2F_{\rm Max}(\tau_1 - \tau_2)]$ is the RN and DMR autocorrelation (Eq. 9). Given the arguments presented in Appendix A, the autocorrelation function is approximated by a spectrotemporal impulse, $R_{\rm SS}(\tau_1 - \tau_2, X_j - X_k) = \sigma_{\rm S}^2 \cdot \delta(X_j - X_k) \cdot \delta(\tau_1 - \tau_2)$. Substituting into Equation 23 yields:

$$E[r_{k}(t)^{2}] = \iint \sigma_{\mathrm{S}}^{2} \cdot \delta(\tau_{1} - \tau_{2}) \cdot h_{k}(\tau_{1}) \cdot h_{k}(\tau_{1}) d\tau_{1} d\tau_{2} = \sigma_{\mathrm{S}}^{2} \cdot \int h_{k}(\tau)^{2} d\tau,$$
(24)

for k = j and:

$$E[r_{\rm i}(t) \cdot r_{\rm k}(t)] = 0, \qquad (25)$$

for $k \neq j$. Combining with Equation 22, the firing rate variance that is captured by the STRF of the neuron is expressed as:

$$\sigma_{\rm r}^2 = \sigma_{\rm S}^2 \sum_{k=1}^{L} \int h_{\rm k}(\tau)^2 d\tau = \sum_{k=1}^{L} \int {\rm STRF}_{\rm r}(\tau, X_{\rm k})^2 d\tau.$$
(26)

The predicted firing rate energy, $E = \sigma_r$, is therefore computed directly from the STRF_r by computing its RMS value via Equation 26.

APPENDIX C

The theoretical maximum peak-to-peak amplitude of the ratenormalized STRF is obtained by considering perfectly aligned sound waveforms that add constructively. This theoretical value is used as a reference normalization factor for the phase-locking index. Consider the amplitude normalized STRF:

$$STRF_{n} = \frac{STRF_{r}(\tau, X_{k})}{r} = \frac{\sigma_{s} \cdot STRF(\tau, X_{k})}{r}$$
$$= \frac{1/(\sigma_{s} \cdot T) \cdot \sum_{n} S(t_{n} - \tau, X_{k})}{r}, \quad (27)$$

where Equation 10 is substituted for $STRF(\tau, X_k)$. Because STRF is described in units of spikes per second, STRF_n is unit-less. Substituting the measured firing rate, r = N/T, we have:

$$\mathrm{STRF}_{\mathrm{n}}(\tau, X_{\mathrm{k}}) = \frac{1}{N} \sum_{\mathrm{n}} \bar{S}(t_{\mathrm{n}} - \tau, X_{\mathrm{k}}), \qquad (28)$$

where $\bar{S} = S/\sigma_{\rm S}$ is a ripple envelope with unit variance. The maximum peak-to-peak amplitude of $\overline{S}(t, X_k)$ is $\sqrt{8}$ for the DMR, because the peak-to-peak amplitude of $S(t, X_k)$ is M, and because $\sigma_{\rm S} = M/\sqrt{8}$. For the RN, $\sigma_{\rm S} = M/\sqrt{12}$; therefore, the maximum peak-to-peak amplitude is $\sqrt{12}$. The theoretical maximum peakto-peak amplitude of STRF_n is obtained as:

$$\Delta = \max(\text{STRF}_n) - \min(\text{STRF}_n), \quad (29)$$

under the assumption that the N spectrotemporal waveforms used to construct the STRF are perfectly aligned. This yields:

$$\Delta = \max\left(\frac{1}{N}\sum_{n}\bar{S}\right) - \min\left(\frac{1}{N}\sum_{n}\bar{S}\right) = \max(\bar{S}) - \min(\bar{S}) = \sqrt{8},$$
(30)

for the DMR and $\Delta = \sqrt{12}$ for the RN.

REFERENCES

- Aersten AMHJ, Olders JHJ, Johannesma PIM (1980) Spectro-temporal receptive fields in auditory neurons in the grass frog: analysis of the stimulus-event relation for tonal stimulus. Biol Cybern 38:235–248. Anzai A, Ohzawa I, Freeman RD (1999) Neural mechanisms for pro-
- cessing binocular information: I. Simple cells. J Neurophysiol 82:891–908.
- Attias H, Schreiner CE (1998) Temporal low order statistics of natural
- Adviss H, Schreiner CE (1998) Tempora fow order statistics of natural sounds. Adv Neural Inf Process Syst 10:27–33.
 Bringuier V, Chavane F, Glaesr L, Frégnac Y (1999) Horizontal propagation of visual activity in the synaptic integration field of are 17 neurons. Science 283:695–699.
 Casseday JH, Ehrlich D, Covey D (1994) Neural tuning for sound dusted by the field birth.
- ration. Role of inhibitory mechanisms in the inferior colliculus. Science 264:847–850.
- DeAngelis GC, Ohzawa I, Freeman RD (1993) Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex: II. Linearity of temporal and spatial summation. J Neurophysiol 69:1118-1135
- DeAngelis GC, Ghose GM, Ohzawa I, Freeman RD (1999) Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. J Neurosci 19:4046-4064.
- de Charms 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 the ferret primary auditory cortex. J Neurophysiol 85:1220-1234.
- Doupe AJ (1997) Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. J Neurosci 17:1147–1167. Eggermont JJ (1993) Wiener and Voltera analyses applied to the audi-
- tory system. Hear Res 66:177-201.
- Eggermont JJ (1999) The magnitude and phase of temporal modulation transfer functions in cat auditory cortex. J Neurosci 19:2780–2788. Emerson RC, Citron MC, Vaughn WJ, Klein SA (1987) Nonlinear di-
- rectionally selective subunits in complex cells of cat striate cortex. J Neurophysiol 58:33–65.
- Epping WJM, Eggermont JJ (1986) Sensitivity of neurons in the auditory midbrain of the grassfrog to temporal characteristics of sounds. II. Stimulation with amplitude modulated sound. Hear Res 24:55-72.
- Greenberg S (1998) Speaking in shorthand: a syllabic-centric perspective for understanding pronunciation variation. In: Proceedings of the

ESCA Workshop on Modeling Pronunciation and Variation for Automatic Speech Recognition, pp 47–56. Kekrade, The Netherlands. Irvine DRF, Gao G (1990) Binaural interaction in high-frequency neu-

- rons in the inferior colliculus of the cat. Effects of variations in sound pressure level on sensitivity to interaural intensity differences. J Neurophysiol 63:570-591.
- Jones JP, Palmer LA (1987) The two dimensional spatial structure of simple receptive fields in cat striate cortex. J Neurophysiol 58:1187–1211.
- Klein DJ, Depireux DA, Simon JZ, Shamma SA (2000) Robust spectro-temporal reverse correlation for the auditory system: optimizing stimulus design. J Comp Neurosci 9:85-111.
- Kowalski Ň, Depireux DA, Shamma SA (1996) Analysis of dynamic spectra in ferret primary auditory cortex: I. characteristics of single unit responses to moving ripple spectra. J Neurophysiol 76:3524-3534. Krishna BS, Semple MN (2000) Auditory temporal processing: re-
- sponses to sinusoidally amplitude-modulated tones in the inferior col-liculus. J Neurophysiol 84:255–273.
- Kuwada S, Batra R, Yin TCT, Oliver DL, Haberly LB, Stanford TR (1997) Intracellular recordings in response to monaural and binaural stimulation of neurons in the inferior colliculus of the cat. J Neurosci 17:1565–1581.
- Langner G, Schreiner CE (1988) Periodicity coding in the inferior col-liculus of the cat: I. Neuronal mechanisms. J Neurophysiol 60:1799–1822.
- Lewicki MS (1994) Bayesian modeling and classification of neural signals. Neural Comput 6:1005-1029.
- Livingstone MS, Pack CC, Born RT (2001) Two-dimensional substructure of MT receptive fields. Neuron 30:781-793.
- Margoliash D (1983) Acoustic parameters underlying the response of song-specific neurons in the white-crowned sparrow. J Neurosci 3:1039-1057.
- Marmarelis PZ, Naka KI (1974) Identification of multi-input biological systems. IEEE Trans Biomed Eng 21:88–101. Merzenich MM, Reid MD (1974) Representation of the cochlea within
- the inferior colliculus of the cat. Brain Res 77:397-441
- Miller LM, Escabí MA, Read HL, Schreiner CE (2001) Functional convergence of response properties in the auditory thalamocortical system. Neuron 32:151–160
- Miller LM, Escabí MA, Read HL, Schreiner CE (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J Neurophysiol 87:516–527.
- Møller AR, Rees A (1986) Dynamic properties of the responses of single neurons in the inferior colliculus of the rat. Hear Res 24:203–215. Moore CI, Nelson SB (1998) Spatio-temporal subthreshold receptive
- fields in the vibrissa representation of rat primary somatosensory cortex. J Neurophysiol 80:2882–2892. Nelken I, Kim PJ, Young ED (1997) Linear and nonlinear spectral
- integration in type IV neurons in the dorsal cochlear nucleus. II. Predicting responses with the use of nonlinear models. J Neurophysiol 78:800-811.
- Nelken I, Rotman Y, Yosef OB (1999) Responses of auditory-cortex neurons to structural features of natural sounds. Nature 37:154-157.
- Oliver DL, Morest DK (1984) The central nucleus of the inferior col-liculus in the cat. J Comp Neurol 222:237–274.
- Plomp R (1970) Timbre as a multidimensional attribute of complex tones. In: Frequency analysis and periodicity detection in hearing (Plomp R, Smoorenburg GF, eds), pp 397-414. Linden, The Netherlands: Sijthoff.
- Plomp R (1983) The role of modulations in hearing. In: Hearing: physiological bases and psychophysics (Klinke R, Hartmann R, eds), pp 270–276. New York: Springer. Portfors CV, Wenstrup JJ (1999) Delay-tuned neurons in the inferior
- colliculus of the mustached bat: implications for analysis of target distance. J Neurophysiol 82:1326-1338.
- Ramachandran R, Davis KA, May BJ (1999) Single-unit responses in the inferior colliculus of decerebrate cats: I. Classification based on frequency response maps. J Neurophysiol 82:152-163.
- Rees A, Møller AR (1983) Response of neurons in the inferior colliculus of the rat to AM and FM tones. Hear Res 10:301–330.
- Rees A, Møller AR (1987) Stimulus properties influencing the response of inferior collicular neurons to amplitude-modulated sounds. Hear Res 27:129-143
- Reich DS, Mechler F, Purpura KP, Victor JD (2000) Interspike intervals, receptive fields, and information encoding in primary visual cortex. J Neurosci 20:1964-1974.
- Schreiner CE, Calhoun BM (1994) Spectral envelope coding in the cat primary auditory cortex: properties of ripple transfer function. Audi-tory Neurosci 1:39–61.
- Schreiner CE, Langner G (1988) Periodicity coding in the inferior col-liculus of the cat: II. Topographic organization. J Neurophysiol 60:1799-1822
- Schreiner CE, Urbas JV, Mehrgardt S (1983) Temporal resolution of amplitude modulation and complex signals in the auditory cortex of the

cat. In: Hearing: physiological bases and psychophysics (Klinke R, Hartmann R, eds), pp 169–175. New York: Springer. Schulze H, Langner G (1997) Periodicity coding in the primary auditory

- Schulze H, Langner G (1997) Periodicity coding in the primary auditory cortex of the Mongolian gerbil (*Mariones unguiculatus*): two different coding strategies for pitch and rhythm? J Comp Neurophysiol 181:651–663.
- Suga N, Jen PH (1976) Disproportionate tonotopic representation for processing CF-FM sonar signals in the mustache bat auditory cortex. Science 194:542–544.
- Suga N, O'Neil WE, Manabe T (1978) Cortical neurons sensitive to particular combinations of information bearing elements of bio-sonar signals in the mustache bat. Science 200:778–781.Szulborski RG, Palmer LA (1990) The two-dimensional spatial struc-
- Szulborski RG, Palmer LA (1990) The two-dimensional spatial structure of nonlinear subunits in the receptive fields of complex cells. Vision Res 30:249–254.
- Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20:2315–2331.
- Van Veen TM, Houtgast T (1983) On the perception of the spectral envelope. In: Hearing: physiological bases and psychophysics (Klinke R, Hartmann R, eds), pp 277–281. New York: Springer.
- Victor JD, Purpura KP (1998) Spatial phase and the temporal structure of the response to gratings in V1. J Neurophysiol 80:554–571.
- Voss RV, Clarke J (1975) 1/f noise in music and speech. Nature 258:317–318.
- Yeshurun Y, Wollberg Z, Dyn N, Allon N (1985) Identification of MGB cells by Voltera kernels: I. Prediction of responses to species specific vocalizations. Biol Cybern 51:383–390.
- Young ED (1998) What's the best sound. Science 280:1402-1403.