Gabor Analysis of Auditory Midbrain Receptive Fields: Spectro-Temporal and Binaural Composition

Anqi Qiu, Christoph E. Schreiner and Monty A. Escabí

You might find this additional information useful...

This article cites 52 articles, 25 of which you can access free at:
http://jn.physiology.org/cgi/content/full/90/1/456#BIBL

This article has been cited by 5 other HighWire hosted articles:

Nonlinear Modeling of Auditory-Nerve Rate Responses to Wideband Stimuli
E. D. Young and B. M. Calhoun
J Neurophysiol, December 1, 2005; 94 (6): 4441-4454.
[Abstract] [Full Text] [PDF]

Adaptive Stimulus Optimization for Auditory Cortical Neurons
K. N. O'Connor, C. I. Petkov and M. L. Sutter
J Neurophysiol, December 1, 2005; 94 (6): 4051-4067.
[Abstract] [Full Text] [PDF]

The Contribution of Spike Threshold to Acoustic Feature Selectivity, Spike Information Content, and Information Throughput
M. A. Escabi, R. Nassiri, L. M. Miller, C. E. Schreiner and H. L. Read
J. Neurosci., October 12, 2005; 25 (41): 9524-9534.
[Abstract] [Full Text] [PDF]

Delayed Inhibition in Cortical Receptive Fields and the Discrimination of Complex Stimuli
R. Narayan, A. Ergun and K. Sen
J Neurophysiol, October 1, 2005; 94 (4): 2970-2975.
[Abstract] [Full Text] [PDF]

Naturalistic Auditory Contrast Improves Spectrotemporal Coding in the Cat Inferior Colliculus
M. A. Escabi, L. M. Miller, H. L. Read and C. E. Schreiner
J. Neurosci., December 17, 2003; 23 (37): 11489-11504.
[Abstract] [Full Text] [PDF]

Updated information and services including high-resolution figures, can be found at:
http://jn.physiology.org/cgi/content/full/90/1/456

Additional material and information about *Journal of Neurophysiology* can be found at:
http://www.the-aps.org/publications/jn

This information is current as of September 26, 2006 .

Gabor Analysis of Auditory Midbrain Receptive Fields: Spectro-Temporal and Binaural Composition

Anqi Qiu,[1] Christoph E. Schreiner,[3] and Monty A. Escabí[1,2]

[1]*Biomedical Engineering Program and* [2]*Department of Electrical and Computer Engineering, University of Connecticut, Storrs, Connecticut 06269-2157; and* [3]*W. M. Keck Center for Integrative Neuroscience, University of California, San Francisco, California 94143*

Qiu, Anqi, Christoph E. Schreiner, and Monty A. Escabí. Gabor analysis of auditory midbrain receptive fields: spectro-temporal and binaural composition. *J Neurophysiol* 90: 456–476, 2003; 10.1152/jn.00851.2002. The spectro-temporal receptive field (STRF) is a model representation of the excitatory and inhibitory integration area of auditory neurons. Recently it has been used to study spectral and temporal aspects of monaural integration in auditory centers. Here we report the properties of monaural STRFs and the relationship between ipsi- and contralateral inputs to neurons of the central nucleus of cat inferior colliculus (ICC) of cats. First, we use an optimal singular-value decomposition method to approximate auditory STRFs as a sum of time-frequency separable Gabor functions. This procedure extracts nine physiologically meaningful parameters. The STRFs of ~60% of collicular neurons are well described by a time-frequency separable Gabor STRF model, whereas the remaining neurons exhibited obliquely oriented or multiple excitatory/inhibitory subfields that require a nonseparable Gabor fitting procedure. Parametric analysis reveals distinct spectro-temporal tradeoffs in receptive field size and modulation filtering resolution. Comparisons between an identical model used to study spatio-temporal integration areas of visual neurons further shows that auditory and visual STRFs share numerous structural properties. We then use the Gabor STRF model to compare quantitatively receptive field properties of contra- and ipsilateral inputs to the ICC. We show that most interaural STRF parameters are highly correlated bilaterally. However, the spectral and temporal phases of ipsi- and contralateral STRFs often differ significantly. This suggests that activity originating from each ear share various spectro-temporal response properties such as their temporal delay, bandwidth, and center frequency but have shifted or interleaved patterns of excitation and inhibition. These differences in converging monaural receptive fields expand binaural processing capacity beyond interaural time and intensity aspects and may enable colliculus neurons to detect disparities in the spectro-temporal composition of the binaural input.

INTRODUCTION

Auditory neurons are unique for their ability to process rapidly varying stimuli and track changes in the stimulus spectrum. Neurons in central auditory stations are highly sensitive to dynamic variations in the temporal, spectral, intensity, and aural composition of the sensory stimulus (Goldberg and Brown 1969; Irvine and Gago 1990; Krishna and Semple 2000; Kuwada et al. 1997; Langner and Schreiner 1988; Ramachandran et al. 1999; Rees and Møller 1983). Although numerous studies have evaluated the response characteristics to structur-ally simple stimuli, only a handful of studies have analyzed the joint spectral, temporal, and/or binaural receptive field arrangements responsible for this response diversity (Depireux et al. 2001; Miller et al. 2002; Sen et al. 2001).

Auditory receptive fields are typically derived with isolated pure tones that are presented at varying frequencies and intensities or by measuring neural sensitivity to narrowband time-varying stimuli (e.g., Krishna and Semple 2000; Langner and Schreiner 1988; Ramachandran et al. 1999; Rees and Møller 1983). Recently, the auditory spectro-temporal receptive field (STRF), a linear model representation of the integration area of a neuron, has expanded these classical methods. The auditory STRF has the advantage that it simultaneously describes spectral and temporal stimulus attributes that preferentially activate a neuron and can be used to identify the spectral arrangement and temporal dynamics of neural excitation and inhibition of a neuron during dynamic broadband stimulation (Aersten et al. 1980; deCharms et al. 1998; Depireux 2001; Escabí and Schreiner 2002; Klein et al. 2000; Miller et al. 2002; Nelken et al. 1997; Sen et al. 2001; Theunissen et al. 2000). In particular, the STRF technique is useful for predicting neuronal response patterns to complex auditory stimuli, including natural sounds (Aersten et al. 1980; Klein et al. 2000; Sen et al. 2001; Theunissen et al. 2000), and can accurately account for spatial selectivity profiles that contribute to sound localization (Schnupp et al. 2001).

In the visual system, the direct counterpart of the auditory STRF is the spatio-temporal receptive field. Here the spectral dimension (which extends along the primary sensory epithelium receptor surface of the cochlea) is replaced by spatial dimensions along the retinal sensory epithelium (Cai et al. 1997; DeAngelis et al. 1995; De Valois and Cottaris 1998; Shamma 2001). Visual neurophysiologists have used Gabor and Gamma functions as quantitative descriptors of visual STRFs (Cai et al. 1997; DeAngelis et al. 1993a, 1999; Jones and Palmer 1987a,b). Advantages for fitting visual STRFs by quantitative functions include: improved estimates of the spatio-temporal structure of visual response areas and the removal of estimation noise. Furthermore, these model STRFs can be used to study the arrangements of excitatory and inhibitory neural inputs and to extract physiologically meaningful parameters from neural data (DeAngelis et al. 1993a, 1999). Al-

though it has been suggested that auditory and visual STRFs have remarkably similar time-varying structure (deCharms et al. 1998; Shamma 2001), only a few studies have quantitatively evaluated the spectro-temporal structure of auditory STRFs (Depireux et al. 2001; Escabí and Schreiner 2002; Miller et al. 2002; Sen et al. 2001). However, these studies did not quantitatively compare the structure of the auditory STRF directly with their visual counterpart.

In this study, we present a time-frequency Gabor STRF model to fit auditory STRFs in the central nucleus of cat's inferior colliculus (ICC). Spectral and temporal Gabor functions are used to model spectral receptive field (SRF) and temporal receptive field (TRF) profiles of ICC neurons, respectively. Each STRF is then fitted by a weighted sum of products of time-frequency separable Gabor functions. From the definition of a Gabor function, nine physiologically meaningful parameters are extracted: the center frequency, the best ripple density, the best temporal modulation frequency, the peak latency, the bandwidth of the SRF profile, the response duration, the response strength, and the spectral and temporal phases. These parameters are used to quantify spectral, temporal, and time-frequency response characteristics to dynamic moving ripple stimuli (Escabí and Schreiner 2002; Miller et al. 2002). This Gabor STRF model is a direct extension of receptive field models used to study the structure of visual receptive fields in the primary visual cortex (DeAngelis et al. 1993a,b, 1999) and provides a basis for comparing the structure of auditory and visual STRFs. In particular, we apply this methodology to compare STRF properties of contra- and ipsilateral inputs to ICC neurons. We demonstrate specific aural STRF differences that suggest binaural filtering mechanisms beyond intra-aural time and level sensitivity.

MATERIALS AND METHODS

Electrophysiology

Physiological recording methods have been presented in detail elsewhere (Escabí and Schreiner 2002). Briefly, cats ($n = 4$) were initially anesthetized with a mixture of ketamine HCl (10 mg/kg) and acepromazine (0.28 mg/kg im). A surgical state of anesthesia was induced with ~30 mg/kg pentobarbital sodium (Nembutal) and maintained throughout the surgery with supplements via an intravenous infusion line. Body temperature was measured and maintained at ~37.5°C. The overlying cerebrum and part of the bony tentorium was removed to expose the ICC via a dorsal approach. During the unit recordings, animals were maintained in an areflexive state 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 solution ($1–4 \ mg \cdot kg^{-1} \cdot h^{-1}$). The infusion rate was adjusted according to physiologic criteria (heart rate, breathing rate, temperature, and peripheral reflexes). All surgical methods and experiment procedures follow National Institutes of Health and U.S. Department of Agriculture guidelines.

Neural data was acquired from $n = 99$ single units in the ICC with parylen-coated tungsten microelectrodes (Microprobe, Potomac, MD; 1–3 MΩ at 1 kHz) that were advanced into the central nucleus with a hydraulic microdrive (David Kopft Instruments, Tujunga, CA). Action potential traces were recorded onto a digital audio tape (Cygnus Technologies CDAT16; Delaware Water Gap, PA) at a sampling rate of 24.0 kHz (41.7-μs resolution) and spike sorted off-line with a Bayesian spike sorting algorithm (Lewicki 1994).

Acoustic stimuli

Dynamic moving ripple (DMR) stimuli (Escabí and Schreiner 2002) were presented with the animal in a sound-shielded chamber (IAC, Bronx, NY) with stimuli delivered via a closed, binaural speaker system (electrostatic diaphragms from Stax). The Dynamic Moving Ripple sound is specifically designed to dynamically activate the primary sensory epithelium and to probe the physiologically relevant range of spectral and temporal stimulus modulations of neurons in an unbiased fashion. Sounds were presented binaurally with an independent sound sequence to each ear—from which independent contra- and ipsi-lateral STRFs were computed via spike-triggered averaging (Escabí and Schreiner 2002).

In three experiments, the DMR stimulus was presented for a period of 10–20 min (Escabí and Schreiner, 2002). In one experiment, a two-repeat 4-min sequence of the DMR (8 min total) was presented. In all experiments, stimuli covered the same range of spectral and temporal parameters and were presented at ~30–70 dB above the neurons response threshold.

Gabor STRF model

STRFs were decomposed into a superposition of time-frequency separable functions from which we could model and fit each component by a spectro-temporal Gabor function (product of Gaussian and cosine; Fig. 3). Measured STRFs were first decomposed using a singular value decomposition (SVD) (Depireux et al. 2001; Press et al. 1995; Theunissen et al. 2000) into a sum of separable STRF components ($STRF_i$)

$$STRF(t, x) = U \cdot S \cdot V^* = \sum_i STRF_i(t, x)$$

$$S = diag(\sigma_1, \sigma_2, \cdots \sigma_i, \cdots), \quad \sigma_1 \geq \sigma_2 \geq \cdots \sigma_i \geq \cdots > 0 \quad (1)$$

where U and V are unitary orthogonal matrixes containing the temporal and spectral receptive field profiles of each STRF component (Fig. 3, B and C; *top* and *right*); S is a diagonal matrix with real, non-negative elements, σ_i, in descending rank order according to energy; and * denotes the Hermitian transpose. Each STRF component, $STRF_i$, is obtained by the vector product

$$STRF_i(t, x) = \sigma_i \cdot u_i \cdot v_i^* \quad (2)$$

where σ_i is the ith singular value of $STRF(t, x)$ and determines the energy of the ith STRF component. u_i and v_i are the ith unitary orthogonal vectors of U and V, respectively. Conceptually, these correspond to the spectral and temporal receptive field profiles of each component STRF (e.g., shown on the *top* and *right* of Fig. 3, B and C). The dominant spectral and temporal receptive field profiles, u_1 and v_1, account for ~80% of the total STRF energy, and we therefore use these to quantify spectral and temporal response characteristics throughout.

According to the SVD procedure, every $STRF_i$ component is time-frequency separable (although the entire STRF may be nonseparable). Therefore each component can be modeled by the product of a spectral and a temporal waveform, which we approximate by a Gabor function. Thus the fitted STRF model is expressed as a weighted sum of a finite set of N of statistically significant separable Gabor components (typically, $N = 1$ or 2)

$$STRF_m(t, x) = \sum_{i=1}^{N} STRF_{im}(t, x) = \sum_{i=1}^{N} sign \cdot K_i \cdot G_i(x) \cdot H_i(t) \quad (3)$$

where $STRF_m(t, x)$ (e.g., in Fig. 3F) is the fitted STRF model. $STRF_{im}(t, x)$ (e.g., in Fig. 3, D and E) is the fitted $STRF_i$ component. K_i, $G_i(x)$, and $H_i(t)$ correspond to the response strength, the fitted and normalized SRF profile, and the fitted and normalized TRF profile of the ith STRF component, $STRF_i$. The modeled spectral and temporal profiles, $G_i(x)$ and $H_i(t)$, assume the form of a Gabor function (see

Eqs. 11 and *13,* respectively) each with an independent set of spectral and temporal parameters. Finally, the variable *sign* assumes a value of 1 or -1 and is included in the model to designate the type of STRF, which can be dominantly excitatory $(+)$ or inhibitory $(-)$, respectively. The optimal parameters of the Gabor-STRF model are determined iteratively by minimizing the mean square error between the model and the real data (Press et al. 1995).

Level of noise

Auditory STRFs are estimated from real neural data by a spike-triggered average method (Escabí and Schreiner 2002) that is inherently noisy. Measurement noise corresponds to random deviations from the expected STRF that would result from an infinite amount of averaging. These variations result from unexpected variations in the neural response and from finite data averaging due to the finite experiment recording periods (Klein et al. 2000; Theunissen 2000). Therefore to minimize the effects of noise, it is necessary to consider only those independent time-frequency components of the Gabor STRF model that significantly contribute to the STRF's energy and structure.

To determine the maximum number of independent dimensions of the STRF that contribute to its structure (*N* in *Eq. 3*), it is essential to quantify the STRF noise level. Singular values that exceed the measured noise level typically contribute significantly to the neural response and should therefore be incorporated into the Gabor STRF model; alternately, singular values that fall below the noise level contribute largely to the noise and can therefore be ignored. A significant noise level ($P < 0.01$) was determined empirically via a bootstrap STRF re-estimation procedure for a random Poisson firing neuron of identical spike rate as the neuron under investigation. Twenty-five randomly constructed STRFs, STRF$_r$ (e.g., Fig. 4*A*), were simulated by correlating a random Poisson spike train of firing rate, λ, with the dynamic moving ripple noise stimulus. The first singular value (σ_{r1}) of each random-STRF, STRF$_r$, was obtained directly by performing a SVD. For each of the 25 trials (shown by vertical red circles in Fig. 4*B*), the measured level of noise was randomly distributed. Therefore the desired threshold noise level for a specific spike rate (solid line in Fig. 4*B*) was determined as the sum of the mean of σ_{r1} and 2.57 times its SD ($P < 0.01$). The mean $\pm$ SD of σ_{r1} were calculated from the 25 simulated samples by a bootstrap resampling technique (Efron and Tibshirani 1993). All first-order STRFs considered here were above the estimated noise level.

Similarity index

The Gabor STRF model can potentially account for much of the structure of collicular receptive fields, however, the utility of the model needs to be quantitatively evaluated. We devised three metrics to validate the goodness of fit of the model. We evaluated the goodness of fit of SRF and TRF profiles independently and for the entire STRF.

To compare the receptive field structure of the model and data, we devised the spectral similarity index (SI$_s$), temporal similarity index (SI$_t$) and spectro-temporal similarity index (SI). The spectral SI, SI$_s$, accounts for differences in shape between original and model SRF profiles; SI$_t$ is used to compare the original and model TRF profiles; the spectro-temporal SI, SI, measures shape differences between original and model STRFs. Individually these metrics correspond to a correlation analysis performed between the model and original data (DeAngelis et al. 1999; Escabí and Schreiner 2002; Miller et al. 2002) and can be expressed as

$$SI_s = \frac{\langle SRF, SRF_m \rangle}{\|SRF\| \cdot \|SRF_m\|} \tag{4}$$

$$SI_t = \frac{\langle TRF, TRF_m \rangle}{\|TRF\| \cdot \|TRF_m\|} \tag{5}$$

$$SI = \frac{\langle STRF, STRF_m \rangle}{\|STRF\| \cdot \|STRF_m\|} \tag{6}$$

where $\langle , \rangle$ corresponds to the vector correlation, and $\| \cdot \|$ designates the vector norm operator. Because the STRF is formally defined by a two-dimensional matrix of spectral and temporal samples, *Eq. 6* could not be evaluated directly since it requires vector inputs. Therefore the statistically significant samples of the STRF that exceeded a significance criterion of $P < 0.002$, were converted into a unidimensional vector, from which the SI was determined using *Eq. 6* (Escabí and Schreiner 2002).

Because all three similarity indices are effectively correlation coefficients between the real data and model waveforms, they assume a value of one whenever the waveforms inside their arguments are identical in shape, zero if the waveforms have nothing in common and negative one if the waveforms have identical shapes but differ by a negative sign.

Normalized mean square error

A fourth metric was defined that quantifies the relative difference in energy between the fitted (STRF$_m$) and the measured STRF (STRF). The normalized mean square error (MSE) is defined as the energy of the difference STRF normalized by the energy of a measured STRF (DeAngelis et al. 1999)

$$MSE = \frac{\sum_s \sum_t (STRF_m - STRF)^2}{\sum_s \sum_t STRF^2} \tag{7}$$

The MSE assumes values between zero and one, where lower MSE values are indicative of a properly fitted STRF.

Temporal asymmetry index

Initial evaluation of the temporal receptive field envelope revealed that timing profiles of ICC neurons are characterized by sharp transient onset. We therefore quantitatively evaluated the structure of the temporal response envelope. To evaluate the degree of temporal asymmetry in the TRF profile, we define an asymmetry index (α_t) as the skewness of the temporal envelope (Bliss 1967)

$$\alpha_t = \frac{\int (t - \mu_t)^3 \cdot E_t(t) \cdot dt}{[\int (t - \mu_t)^2 \cdot E_t(t) \cdot dt]^{3/2}} \tag{8}$$

where μ_t is the mean or centroid of the temporal envelope, $E_t(t)$, measured at the center frequency (x_0) of the neuron and normalized for unit area. A temporal asymmetry index of zero is observed only for TRF envelopes with perfectly symmetric envelopes about the mean point, μ_t. A α_t significantly less than 0 indicates that the TRF profile is skewed to the right; and a α_t significantly greater than 0 indicates the TRF profile is skewed to the left.

Separability index

An inherent aspect of the Gabor model is that it is composed of multiple receptive field components, each of which is a time-frequency separable function. If the receptive field contains only one singular value, the receptive field is time-frequency separable; that is, it can be described by a multiplicative product of a temporal and spectral receptive field profile as in *Eq. 2*. Hypothetically, such a neuron would encode spectral and temporal information independently. If, alternately, the receptive field has multiple significant singular values, the receptive field will exhibit time-frequency inseparable structure. This can manifest as obliquely oriented STRF features or multiple asymmetrically aligned excitatory and inhibitory receptive field subregions. Neurons with such receptive field arrangements most likely prefer sound stimuli with dynamically changing

frequency components, and, consequently, the spectral and temporal dimensions for such neurons cannot be treated independently of each other. This effect becomes more pronounced if the higher-order singular values account for a large proportion of the receptive field energy. Thus we can define a separability index by considering the proportion of energy provided by first singular value in relationship to the cumulative energy of the higher-order singular values. We define the separability index (α_d) as

$$\alpha_d = \frac{\sigma_1^2 - \sum_{i=2}^{N} \sigma_i^2}{\sigma_1^2 + \sum_{i=2}^{N} \sigma_i^2} \qquad (9)$$

where σ_1 and σ_i are the first- and higher-order singular values of the STRF (*Eq. 1*), and N is the number of statistically significant singular values used in the Gabor STRF model. Conceptually, α_d is defined as the normalized energy of the first singular value (relative to the total energy of the model STRF) minus the normalized energy of the higher-order singular values. Separability index values range from 0 to 1; where 1 corresponds to a perfectly separable STRF and values close to zero designate a highly inseparable receptive field arrangement.

RESULTS

We studied in 99 single neurons how dynamic stimuli are encoded in the ICC by identifying structural characteristics of the auditory STRF. Our dynamic moving ripple stimulus (DMR) is a broadband sound that efficiently probes spectro-temporal attributes of the acoustic space (Escabí and Schreiner 2002). It is characterized by a dynamically changing spectrum with widespread spectral fluctuations over a broad range of resolutions (0–4 cycles/octave). Superimposed on this spectral variability, the DMR exhibits temporal energy fluctuations over a wide range of modulation frequencies: 0–350 Hz. Its statistically unbiased properties makes the stimulus directly applicable for the study of auditory receptive fields during dynamic stimulation. We combined STRF measurement techniques with a spectro-temporal Gabor model to study the structural properties and binaural arrangements of inferior colliculus STRFs. This model allows us to extract nine physiologically meaningful STRF parameters. To determine whether the Gabor model is well suited for describing auditory STRFs, we first fitted each contralateral STRF to the Gabor model and found the optimal parameters of each receptive field. Next, we independently characterized spectral and temporal receptive field profiles as well as the arrangement of excitation and inhibition of each neuron in order to determine how these dimensions contribute to the STRF. Finally, we use the Gabor STRF model to characterize and compare ipsi- and contralateral receptive field arrangements. By studying the spectral and temporal parameters of the contralateral and ipsilateral STRFs, we identify how the spectro-temporal arrangement of excitation and inhibition contribute to the formation of binaural response properties seen in the inferior colliculus.

Structure of the spectral receptive field

The spectral receptive field (SRF) profile is a model representation of the frequency integration area of auditory neurons (Calhoun and Schreiner 1998; Kowalski et al. 1996; Miller et al. 2002; Schreiner and Calhoun 1994; Versnell and Shamma 1998). This descriptor can be used to quantify neuronal responses to sounds with complex spectra (such as for formant transitions in speech and spectral resonances in animal vocalizations) and to study the receptive field arrangement of excitation and inhibition along the cochleotopic dimension of the stimulus. Most studies using this descriptor largely focused on qualitatively identifying general integration properties (such as the arrangement of spectral excitation and inhibition) and only for stimuli with static temporal characteristics. By slicing the STRF at a fixed latency (solid lines in Fig. 1, *B* and *C*) we can study the dynamic behavior of the SRF profile for complex stimuli with time-varying structure. Specifically, we would like to identify a model representation of the STRF that quantitatively captures the general characteristics of the SRF profile and its associated dynamics. When the latency is >40 ms, there is no discernible SRF structure for the STRF shown in Fig. 1A. At shorter latencies, however, SRF profiles can exhibit pure excitation, inhibition, or an alternating arrangement of excitation and inhibition. The phase of SRF profiles changes continuously so that the excitatory bandwidths and center frequencies change with increasing latency. Consequently, there is no direct analytic equation to model the SRF profile at all latencies.

One step toward solving this problem is to break up the SRF profile into an envelope and a carrier component via the Hilbert transform (Cai et al. 1997; Daugman 1985; DeAngelis et al. 1993a, 1999; Jones and Palmer 1987a,b; Marcelja 1980). The envelope, $E_s(x)$, is computed by the vector sum of the SRF profile, SRF(x), and its Hilbert transform, H[SRF(x)]

$$E_s(x) = \sqrt{\text{SRF}(x)^2 + H[\text{SRF}(x)]^2} \qquad (10)$$

Example spectral envelopes of a single neuron are shown as dashed lines at two latencies in Fig. 1, *B* and *C*. The Hilbert transforms of each envelope, H[SRF(x)] (Fig. 1, *B* and *C*), are represented by the dotted lines and are obtained by shifting the phase of all frequency components of SRF(x) by 90° (solid lines in Fig. 1, *B* and *C*). Conceptually, the Hilbert transform isolates the fine carrier structure from the coarse envelope structure of the STRF.

Although the SRF profile depends strongly on the latency of the STRF, the spectral envelope assumes a nearly invariant structure at all latencies. The envelopes of the SRF profiles (dashed lines in Fig. 1, *B* and *C*) are approximately Gaussian functions and can be conveniently defined by their bandwidth and center frequency. The bandwidth of the SRF profile is defined as the width of the envelope at a response level that is $1/e$ relative to the absolute maximum of the envelope, capturing ~85% of the energy in a Gaussian the SRF envelope. The center frequency is defined as the peak value of the spectral envelope. As expected for the SRF profiles of Fig. 1, *B* and *C,* the measured bandwidths and center frequencies along the excitatory and inhibitory cross-sections are in close agreement: bandwidth = 1.00 and 0.89 octaves (octave is defined as $\log_2$ (f/f_r), f_r = 500 Hz is a reference frequency), respectively; center frequency = 4.37 and 4.42 octaves.

The spectral receptive field structure was modeled at each time point as the product of a Gaussian envelope and a sinusoidal carrier. Qualitatively, the Gaussian function defines the center and extent over which the neuron integrates spectral information, whereas the sinusoid carrier component is necessary to account for the interleaved patterns of excitation and inhibition. This functional form of the SRF profile, a Gabor
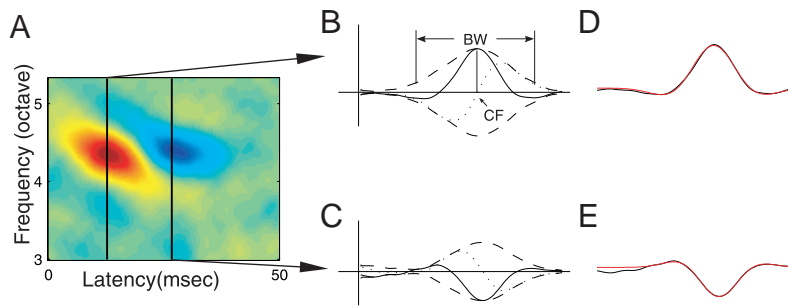
FIG. 1. Spectral receptive field (SRF) profile analysis. A typical inferior colliculus spectro-temporal receptive field (STRF) showing obliquely oriented excitatory and inhibitory subregions (*A*). Two SRF profiles taken along the excitatory (*T* = 12.7 ms) and inhibitory (*T* = 26.7 ms) spectral cross-sections (solid lines in *B* and *C,* respectively). Their Hilbert transform (H[SRF(x)]) are represented by dotted lines and their spectral envelope, $E_s(x)$, by dashed-lines (*B* and *C*). The neuron's center frequency (CF) is determined from the peak of the SRF envelope. Typically, the CF is close to the peak of the SRF profile (as in *B*), although these may differ depending on the arrangement of spectral excitation and inhibition (as in *C*). The bandwidth of the SRF profile, BW, is measured directly from the spectral envelope. The range of frequencies covered by the BW account for ~85% of the energy of the SRF envelope. Measured SRF profile (red line) and Gabor fitted SRF profile (black line) are typically in close agreement (*D* and *E*).

function, is a direct extension of the receptive field models used to study spatio-temporal integration in the visual system (Cai et al. 1997; Daugman 1985; DeAngelis et al. 1993a; Jones and Palmer 1987a,b; Marcelja 1980). The Gabor function can capture numerous receptive field aspects and can be used to extract physiologically meaningful parameters directly from the neuron's receptive field.

At each time point, the SRF profile was fitted by a Gabor function taking the general form

$$G(x) = K \cdot e^{-[2(x-x_0)/BW]^2} \cos [2\pi \cdot \Omega_0 \cdot (x - x_0) + P] \qquad (11)$$

where K, x_0, BW, Ω_0, and P are free parameters. The parameter K models the strength of the spectral response in unit of spikes · $s^{-1} \cdot dB^{-1}$. x_0 is the center frequency or the central position of the SRF envelope in units of octaves; BW is the bandwidth of the SRF which accounts for the spectral extent of the receptive field; Ω_0 is the best ripple density (units of cycles/octaves) that models the distance between the excitatory and inhibitory lobes; P is the spectral phase of the SRF profile with respect to the center frequency of the Gaussian envelope. This parameter accounts for the alignment of excitation and inhibition relative to the peak of the SRF envelope. The optimal parameters in *Eq. 11* can be obtained by minimizing the mean square error between the Gabor function and the measured SRF profile (Press et al. 1995). Example SRF profiles (Fig. 1, *D* and *E*) and optimal-fitted results are shown in Fig. 1, *D* and *E* at two latencies of the STRF. Fitted profiles (continuous red lines) and the measured SRF profiles (continuous black lines) are in close agreement.

Structure of the temporal receptive field

The structure of the temporal receptive field (TRF) profile was analyzed using a similar functional descriptor as for the SRF profile. The TRF profile obtained by slicing through the STRF at a particular frequency has an alternating arrangement of excitation and inhibition. The TRF profiles of collicular neurons typically have short excitation (or inhibition) followed by long inhibition (or excitation) (e.g., solid line in Fig. 2*B*), and their envelopes are, therefore, not symmetric about the peak point. For example, the envelope of the TRF profile shown by the dashed line in Fig. 2*B* is not symmetric about the peak of the temporal envelope (vertical line) because it has a

sharp onset and slower off-response. Because of this temporal asymmetry, the TRF profile is not well described by a symmetric Gabor function.

The degree of temporal asymmetry was measured for all contralateral responsive neurons in our ICC sample (*n* = 93 of 99) with an asymmetry index, α_t (see METHODS). The TRF profile in Fig. 2*B* is skewed to the left and it therefore has a positive asymmetry index (0.935). Figure 2*C* (blue histogram) illustrates the distribution of asymmetry indices, obtained for the dynamic moving ripple sound. The population distribution shows a bias toward positive values (mean ± SD: 1.93 ± 1.64;
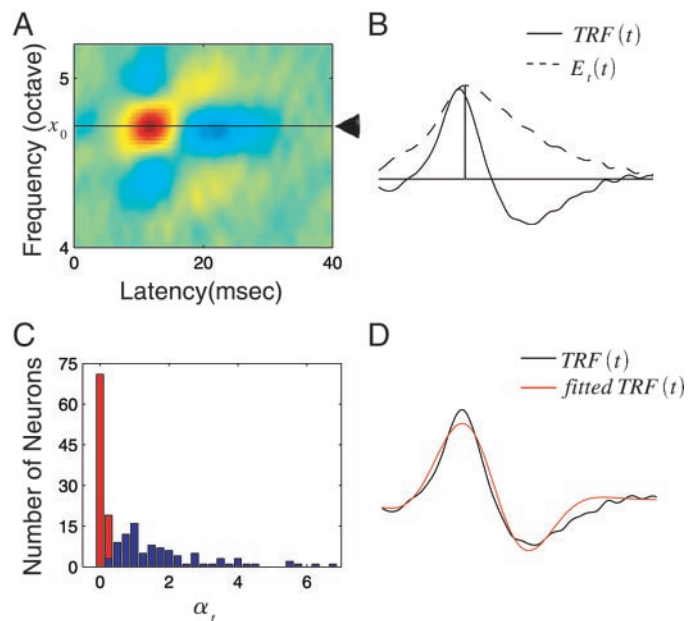

FIG. 2. Asymmetry analysis of the temporal receptive field (TRF) profiles. *A*: typical STRF showing a short excitatory onset response and a long inhibitory offset response. The TRF profile is obtained by taking a temporal cross-section about the center frequency (x_0) (solid line in *B*) and its envelope is extracted with the Hilbert transform (dashed line in *B*). The envelope shows a strong asymmetry about its peak point, which is designated by the vertical line. *C*: the distribution of asymmetry index (α_s) for our sample of neurons is displaced toward positive values (blue histogram). After performing a time-warping transformation, temporal envelopes are nearly symmetric and the asymmetry indices are tightly distributed about 0 (red histogram). *D*: the TRF profile of *A* (black line) was fitted with a skewed Gabor function (red line) which takes into account the temporal asymmetry of the TRF profile.

observed range: 0.30–9.7; t-test, $P < 0.001$), indicating that the temporal envelopes and TRF profiles are skewed toward zero delay. Accordingly, the temporal responses profiles of most ICC neurons exhibit a short primary response (excitatory or inhibitory) followed by a long secondary response of opposite sign (inhibitory or excitatory, respectively). Such timing differences between the onset and offset of the receptive field are consistent with asymmetric preferences to ramped auditory stimuli observed both physiologically (Lu et al. 2001) and psychoacoustically (Neuhoff 1998; Patterson 1994).

Considering the observed temporal asymmetry, we modified the Gabor model so that it accounts for the observed timing profiles by incorporating a time-warping factor that skews the time axis and allows us to model the TRF with a symmetric Gabor function (DeAngelis et al. 1999). The time-skewing function was defined as

$$T = 2 \cdot \arctan(\beta \cdot t) \qquad (12)$$

where β is the skewing factor (observed range: 0.45–0.68), t is the uncompressed time-axis, and T is the corrected temporal axis. The TRF profile is then fitted by a Gabor function of the form

$$H(t) = K \cdot e^{-[2(T-T_0)/D]^2} \cdot \cos[2\pi \cdot F_{m_0} \cdot (T - T_0) + Q] \qquad (13)$$

where K, T_0, D, F_{m_0}, and Q are free parameters. K corresponds to the strength of the temporal response; T_0 is the peak latency of the TRF profile; D reflects the time-skewed duration of the response; the best temporal modulation frequency is described by F_{m_0}; and Q is the phase of a sinusoid component about T_0. During the fitting procedure, each parameter was adjusted iteratively until the optimal parameters in *Eqs. 12* and *13* are found by minimizing the mean square error between the model and the measured TRF profile (Press et al. 1995). An example fitted TRF profiles is illustrated in Fig. 2D. The fitted TRF profile (solid red line) captures the structure of the measured TRF profile (solid black line). Further analysis of the entire population confirms the validity of the temporal receptive field asymmetry and the appropriateness of the time-skewing parameter. We recomputed the asymmetry index of all neurons using the time-warped TRF profiles (Fig. 2C; red histogram), which resemble symmetric Gaussian functions (not shown). The time-warped asymmetry indices were near zero (time-warped mean $\pm$ SE = 0.083 ± 0.014) and were significantly smaller than for the unwarped TRF (time-unwarped, 1.93 ± 0.17; paired *t*-test, $P = 1$). Thus the time-warping factor accurately accounts for the observed temporal receptive field asymmetry observed for all ICC neurons.

Gabor-STRF model

The analysis of the TRF and SRF profiles shows that the temporal and spectral receptive field dimensions of auditory neurons can in principle be independently approximated by temporal and spectral Gabor functions. Does this approach generalize for the STRF? Can we model the auditory STRF by a product of Gabor TRF and SRF profiles? If so, what conditions must be satisfied?

In terms of time and frequency response interactions, auditory STRFs can be divided into two fundamental types: separable and inseparable (Adelson and Bergen 1985; DeAngelis et al. 1995; Depireux et al. 2001; Miller et al. 2002; Reid et al.

1991; Sen et al. 2001). Time-frequency separability of the STRF occurs whenever the STRF can be described as the product of a SRF profile and a TRF profile, in which case the SRF and TRF profiles are independent of each other. If a separable STRF is taken into the Fourier domain, the ripple transfer function (RTF) is symmetric about the zero temporal modulation frequency axis (Depireux et al. 2001; Escabí and Schreiner 2002; Miller et al. 2002; Sen et al. 2001). However, inseparable STRFs cannot be broken down into two independent time and frequency functions. The representations of these STRFs in the Fourier domain can therefore show conspicuous asymmetries (Depireux et al. 2001; Escabí and Schreiner 2002; Miller et al. 2002; Sen et al. 2001).

Many auditory STRFs have some inseparable features, including, time-frequency oriented subregions or multiple asymmetrically aligned excitatory and inhibitory receptive field components. Such structural features may be necessary to encode specific structural components in natural signals, such as consonant-vowel transitions in speech, and to dynamically track changes in the frequency spectrum of complex signals, such as frequency-modulated sweeps.

In the previous discussions, we showed that it is relatively easy to model auditory receptive fields by independent Gabor profiles (spectral and temporal) if they are time-frequency separable; however, this procedure is not directly applicable for inseparable STRFs. One way to overcome this difficulty is to first decompose an inseparable STRF (Fig. 3A) into several separable STRF components (Fig. 3, *B* and *C*). Each of the separable STRF components can then be fitted by a time-frequency separable Gabor (Fig. 3, *D* and *E*). Finally, the fitted resultant STRF is approximated by the sum of each separable fitted STRF component (see METHODS, *Eq. 3;* Fig. 3). This procedure is realized using a singular value decomposition (SVD) to determine numerically the smallest number of independent time-frequency dimensions of the STRF (Depireux 2001; Press et al. 1995; Theunissen 2000).

We determined the number of independent STRF components required for the Gabor STRF model numerically by finding those components that exceed a significance criterion of $P < 0.01$ (Fig. 4C). Figure 4C describes the relationship between the measured spike rate and the level of the noise for dynamic moving ripples. The level of the noise increases as function of the spike rate. The magnitude of the first (red *), second (blue ◇), and third (green ○) STRF singular values are plotted against the noise-threshold level; of which 100% of the first STRF components exceeded the noise level. By comparison, only 39.7% of the second, 7.5% of the third STRF components exceeded the significance criterion (solid black line in Fig. 4, *B* and *C*). The total energy contribution of the first and second singular value components accounts for 78.9 ± 15.7 and $6.2 \pm 5.0\%$ of the STRF energy, respectively. The third component, however, only contributes $2.3 \pm 1.8\%$ of the total STRF energy. Therefore the first and second singular values are typically sufficient for describing the spectro-temporal structure of ICC receptive fields.

Validating the Gabor STRF model

As with any model, its overall utility ultimately depends on its ability to account for observed empirical results. Specifically, we are interested in determining how well the separable
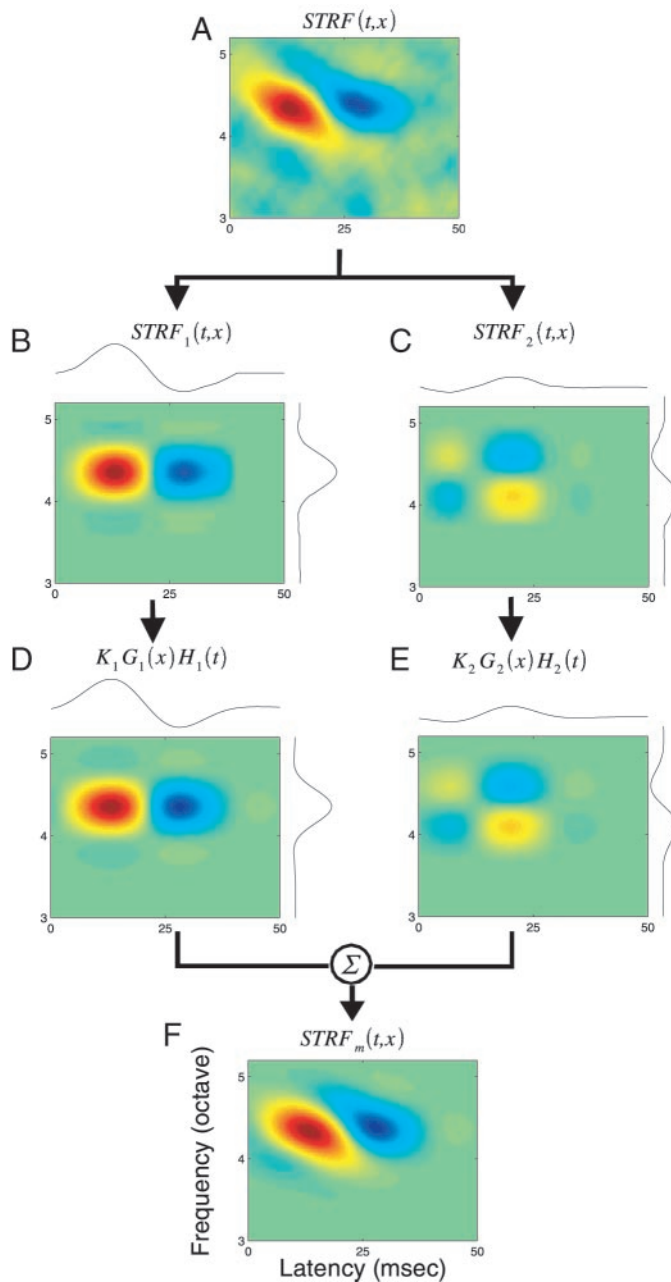
FIG. 3. Schematic diagram of the Gabor STRF model. A singular value decomposition procedure (SVD; see METHODS) is used to decompose the measured STRF into a weighted sum of separable STRF components (STRF$_1$, STRF$_2$, . . . ; B and C; shown for the 1st 2 components only). The SRF profile at the peak latency and the TRF profile at the center frequency of each separable STRF component are illustrated on the right and top of B and C, respectively. SRF and TRF profiles are then individually fitted by a Gabor function (D and E; top and right waveforms). Each separable STRF component is described by the product of 2 Gabor functions [$G_i(x)$ and $H_i(t)$] in D and E. Finally, the fitted STRF (STRF$_m$, F) is modeled as the weighted sum of the statistically significant separable STRF components (from D and E).

Gabor STRF model accounts for receptive field structure of inferior colliculus neurons. Does the model adequately account for spectral and/or temporal receptive field structures? If so, how well does it account for joint spectro-temporal receptive field characteristics? We devised four metrics to independently quantify the spectral, temporal, and spectro-temporal goodness of fit of the model. Differences in receptive field shape between

the model and neural data were quantified individually for the SRF and TRF profiles as well as for the STRF. The spectral similarity index (SI$_s$), temporal similarity index (SI$_t$), and spectro-temporal similarity index (SI) each independently measure how well the model accounts for the structure of the SRF, TRF, and STRF, respectively. Each SI is equivalent to a correlation coefficient between the data and model, and, therefore, they assume numerical values between negative and positive one (DeAngelis et al. 1999; Escabí and Schreiner 2002; Miller et al. 2002). Errors due to energy differences between the model and data were characterized with an energy error metric—which we computed as a normalized mean square error (MSE; see METHODS) from the residual errors (difference between Gabor STRF model and the original STRF; Fig. 5, third column). This metric assumes values between zero and one, where zero indicates that the model provides a perfect fit and a value of one is indicative of a poor fit.

Figure 5 illustrates example fits of the STRF Gabor model of five ICC neurons and the residual errors between the model and data (third column). In most instances, the model accounts for the spectral, temporal, and spectro-temporal receptive field structure exceptionally well. For instance, the measured SI values (spectral SI = 0.992; temporal SI = 0.992; spectro-temporal SI = 0.967) and MSE (0.043) show that a strongly
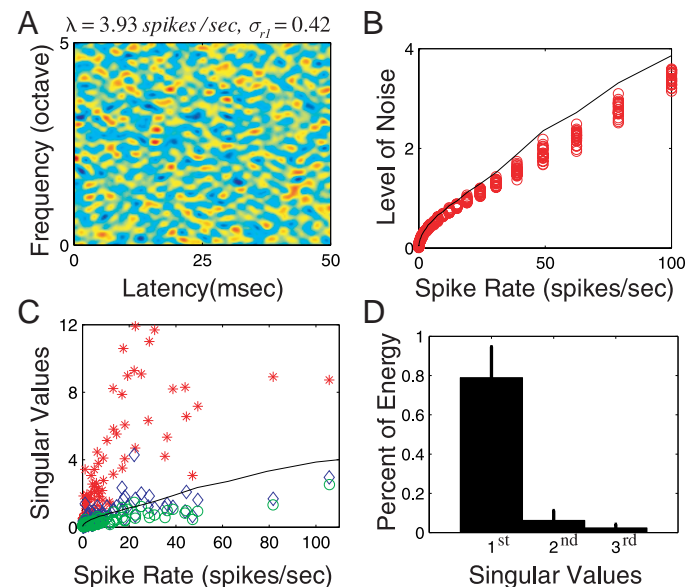

FIG. 4. Significance analysis of the Gabor STRF model. A random noise STRF, STRF$_r$ (A), is generated by reverse-correlating the dynamic moving ripple sound and a random, Poisson-distributed spike train at a specific spike rate ($\lambda = 3.93$ spikes/s for this example). The noise level is obtained by measuring the first singular value ($\sigma_{r1} = 0.42$ for the shown example) of the STRF$_r$ with the SVD method to break up the STRF into separable components (Fig. 3). For each spike rate this procedure is resimulated 25 times to estimate the distribution of noise-levels (vertical red circles in B). A resampling bootstrap technique is used to estimate the threshold-level required to achieve a significance of $P < 0.01$ at each spike rate (continuous line, B and C). The relationship between the noise-threshold level, the measured spike rate and the 1st, 2nd, and 3rd singular values obtained from the STRFs of all neurons is depicted in (C). All of the 1st-singular values (100%) exceed the noise threshold (red *), whereas only 39.7% of the 2nd (blue ◇), and 7.5% of the 3rd singular (green ○) values exceed significance ($P < 0.01$). Energy contribution of the separable STRF component (D). The 1st STRF component, (STRF$_1$) accounts for 78.9 ± 15.7% (mean ± SD) of the STRFs energy. The contributions of the 2nd (6.2 ± 5.0%) and 3rd (2.3 ± 1.8%) STRF components is significantly smaller.
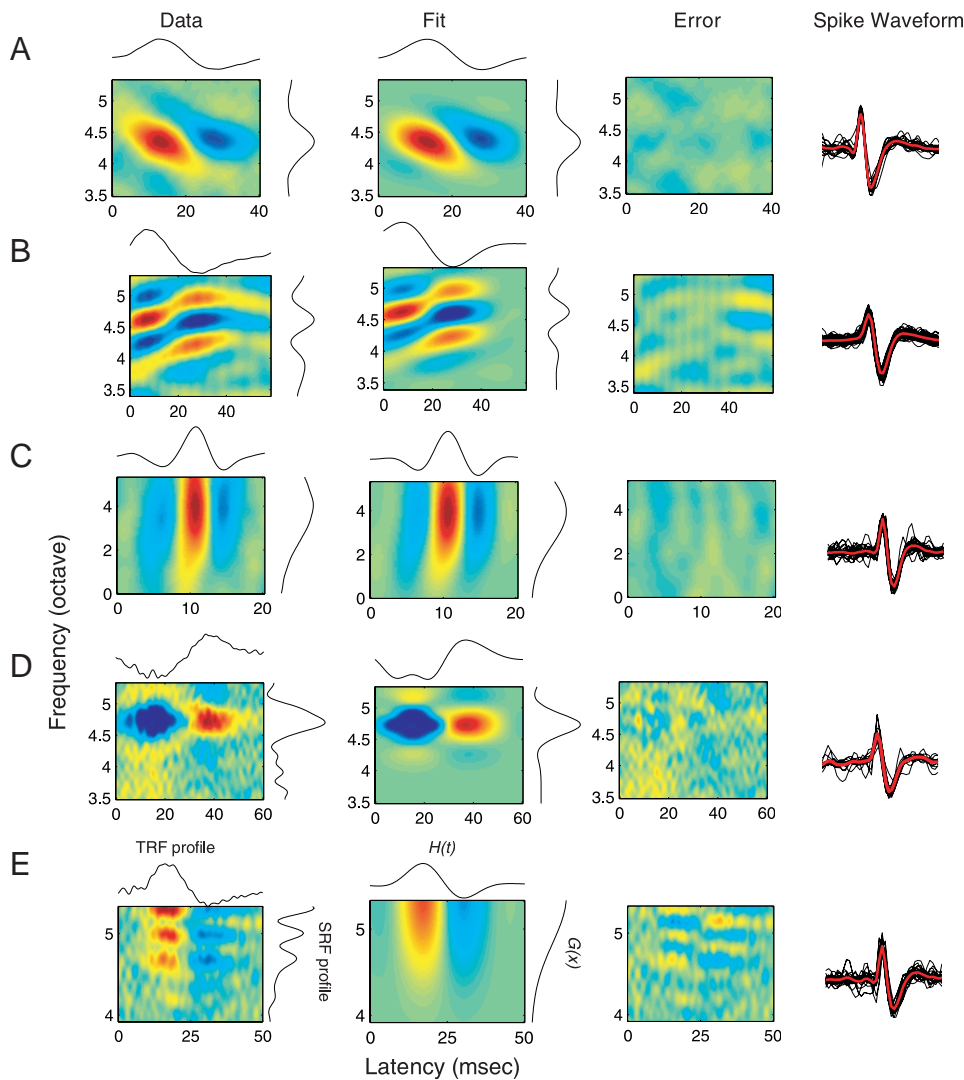
FIG. 5. Representative fits of the Gabor STRF model for 5 inferior colliculus neurons. Measured STRFs (*A–E, left*), fitted STRFs (STRF$_m$, *middle*), and error STRFs (*right*) are shown. The SRF and TRF profiles are shown on the *right* and *top* of measured and fitted STRFs. The measured, fitted, and error STRFs in each row are plotted using identical color scale. *A* and *B*: typical inseparable STRFs. *C*: typical separable STRF. *D*: typical inhibitory/separable STRF. *E*: poorly fitted STRF. Action potential traces are shown for reference at the *far right*.

nonseparable STRF (Fig. 5*A;* separability index = 0.692) can be adequately fit by the model. Not surprisingly, the structure of separable STRFs (Fig. 5*C*) is easily captured by the model (spectral SI = 0.993; temporal SI = 0.966; spectro-temporal SI = 0.976; MSE = 0.022); however, the number of STRF components required to fit a separable STRF is typically lower than for a nonseparable STRF (correlation between number of components and separability index: $r = -0.679 \pm 0.077$, $P < 0.001$).

The example STRFs of Fig. 5, *A–C,* were exceptionally clean with little additive noise. Other neurons had higher levels of noise (Fig. 5*D*), and yet, the model was able to account for their STRF structure (spectral SI = 0.955; temporal SI = 0.975; spectro-temporal SI = 0.941; MSE = 0.079).

Although the model was able to account for the structure of many neurons, it could not fit all receptive field structures. The neuron of Fig. 5*E,* for example, has multiple excitatory peaks that are displaced along the spectral axis. The measured SI values and MSE (spectral SI = 0.857; temporal SI = 0.970; spectro-temporal SI = 0.762; MSE = 0.434) indicate that the model accounts reasonably well for the temporal RF structure, which has a simple on-off TRF profile; however, the model can not fully account for the multiple excitatory spectral peaks

observed in the original SRF. This happens because the spectral oscillations of the STRF are strictly positive valued, whereas the Gabor model requires oscillatory components with negative and positive values. Accordingly, the model fails to account for the STRF structure because of its inability to model the SRF profile of the neuron.

The distribution for the three-similarity indices and the normalized MSE of all neurons are illustrated in Fig. 6. Overall the Gabor STRF model fully accounts for much of the spectral, temporal, and spectro-temporal structure of inferior colliculus neurons. In both instances, the mean spectral and temporal SIs (Fig. 6, *A* and *B*) are close to unity (0.938 ± 0.088 and 0.933 ± 0.075, respectively), suggesting that the shapes of the TRF and SRF profiles are readily accounted for by the Gabor model. Furthermore, the spectral and temporal SIs are not significantly different (paired *t*-test, $P > 0.57$), indicating that Gabor TRF and SRF models are equally well suited for describing the temporal and spectral receptive field profiles. The mean value of the spectro-temporal SI (0.846 ± 0.125; Fig. 6*C*) is lower than spectral and temporal SI (paired *t*-test; $P < 0.001$ and $P < 0.001$, respectively). This reduction in SI is accounted for by the fact that independent multiplicative errors are propagated from the SRF and TRF profiles to the STRF in the model,
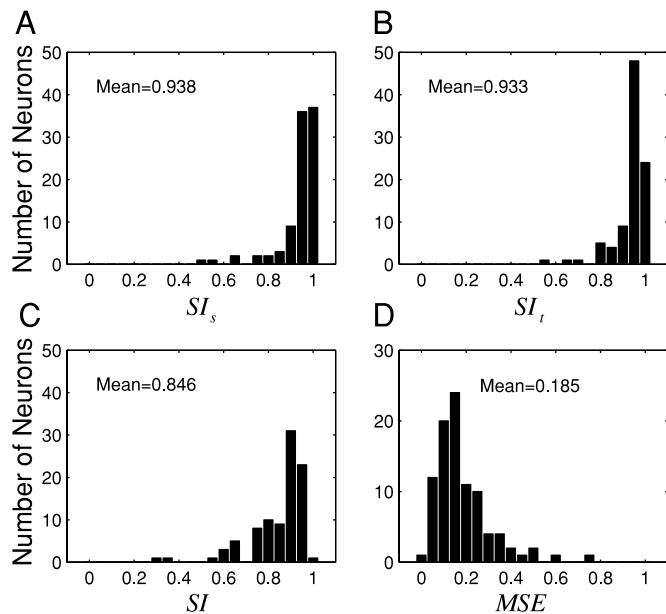
FIG. 6. Gabor STRF error analysis. Distribution of spectral similarity index (SI$_s$; A), temporal similarity index (SI$_t$; B), STRF similarity index (SI; C), and the energy error metric (MSE; D). The spectral and temporal SI quantify shape similarity between the measured and modeled SRF and TRF profiles, respectively. Both means are near unity suggesting that the Gabor model can adequately account for the shape of the SRF and TRF profiles. The STRF similarity index, assumes values that are slightly lower than for the SRF and TRF (C) because shape errors from the Gabor TRF and SRF models are propagated to the Gabor STRF model. The overall goodness of fit was measured with the energy error metric (lower values correspond to better fits), which typically assumed small values (D).

leading to a reduction in the spectro-temporal SI (using the spectral and temporal SI, the expected spectro-temporal SI assuming independent profiles is $0.938 \times 0.933 = 0.875$). Finally, the residual errors of the model (Fig. 6D) are typically small, as suggested by the MSE energy error metric (mean $\pm$ SD $= 0.185 \pm 0.126$), and were typically not significantly different from random noise (χ^2 test; $P < 0.01$ for 58 of 93 neurons; critical value, $\chi^2_{0.01,19} = 36.2$).

Spectral response preferences

Spectral response preferences of auditory neurons are typically determined with isolated pure-tones of varying frequency. The SRF is an extension of the methods used to study frequency response preferences using sound stimuli with spectral structure (Kowalski et al. 1996; Schreiner and Calhoun 1994; Versnel and Shamma 1998). This descriptor allows us to study spectral integration properties of single neurons to dynamic broadband sounds with a rich spectral structure. Spectral selectivity is captured by four parameters of the Gabor function SRF (*Eq. 11*)—center frequency (x_0), SRF bandwidth (BW), best ripple density (Ω_0), and spectral phase (P). The center frequency and bandwidth determine the central location and width of the SRF profile; the best ripple density determines the number of excitatory or inhibitory peaks in the SRF, and the spectral phase determines their alignment relative to the center frequency. Individually, each of these parameters reflects structural properties of the neuronal response area. The center frequency determines the central position of the SRF, whereas the bandwidth determines its spectral extent or selectivity. The

ripple density accounts for the interleaving pattern of excitation and inhibition observed in many neurons, whereas the spectral phase determines the exact position of the excitatory and inhibitory SRF subregions.

Due to some frequency bias in the sampling of ICC, the contralateral receptive field of the studied neurons covered a range of center frequencies from 1.47 to 5.3 oct. (between 1.393 and 20 kHz)—of which 64.5% were located in the range from 4 to 5 octaves (between 8 and 16 kHz; Fig. 7A). While the center frequency of the neuron determines the position along the primary sensory epithelium that preferentially activates the neuron, the spectral bandwidth accounts for the range of frequencies over which the neuron integrates spectral information, including both excitatory and inhibitory features. SRF bandwidths ranged from 0.14 to 4.8 octaves—although most neurons had bandwidths below ~2.0 octaves (93%). The SRF bandwidth follows a unimodal distribution with mean 0.988 octaves and median 0.654 octaves (Fig. 7C).

Auditory neurons can also respond selectively to oscillatory patterns of the stimulus spectrum (Kowalski et al. 1996; Schreiner and Calhoun 1994). Such selectivity arises via alternating excitatory and inhibitory subfields of the SRF profile. These excitatory and inhibitory RF features must overlap on and off features of the stimulus spectrum for the neuron to respond. Therefore such spectral selectivity is reflected in the SRF profile by alternating on and off subfields of the SRF profile, analogous to spatial grating selectivity in the visual system (Cai et al. 1997; DeAngelis et al. 1995, 1999). This form of spectral selectivity is captured by the Gabor model in the best ripple density parameter. The ripple density (units of cycles/octave) represents the number of spectral peaks in the stimulus spectrum existing over an octave range of frequencies. The best ripple density is defined as the number of stimulus spectral peaks that produces a maximal neural response. Alternately, it can also be thought of as the number of interleaved excitatory
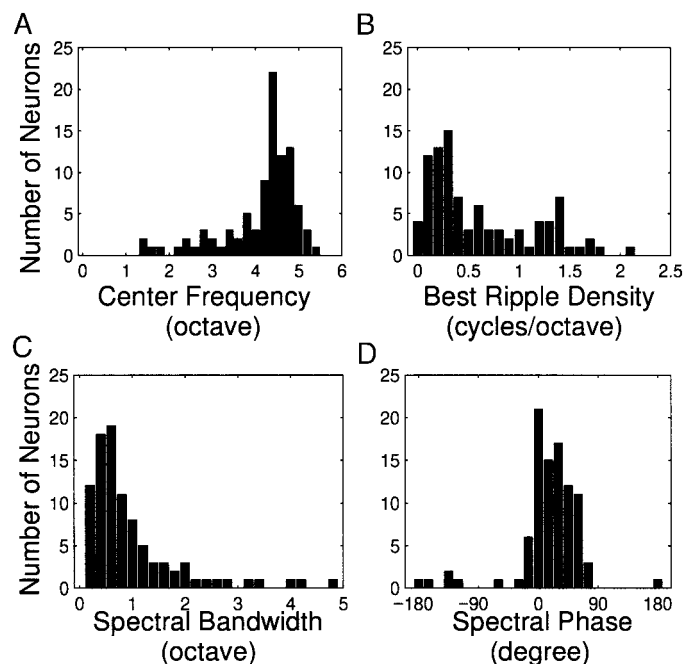

FIG. 7. Distributions of spectral STRF parameters. A: center frequency (x_0); B: the best ripple density (Ω_0); C: bandwidth of the SRF profile (BW), and D: the spectral phase (P) all assume unimodal distributions.

and inhibitory subunits of the SRF existing over a single octave (Escabí and Schreiner 2002; Klein et al. 2000; Miller et al. 2002; Schreiner and Calhoun 1994). Most neurons in our sample preferred low ripple densities (Fig. 7*B;* mean = 0.609 cycles/octave; median = 0.406 cycles/octave), indicating that they preferred broad spectral features of the dynamic moving ripple sound. The range of best ripple densities extended from nearly 0 (0.022 cycles/octave) to 2.113 cycles/octave although all neurons were tested up to 4 cycles/octave.

Finally, the spectral phase of the SRF profile determines the alignment of excitatory and inhibitory features relative to the center frequency of the neuron. Conceptually, a spectral phase shift corresponds to a frequency shift of the actual SRF maximum (not the envelope peak or center frequency). A positive phase value shifts the maximum of the spectral profile to lower frequencies; a negative phase shifts the SRF maximum to higher frequencies. Most of the STRFs (78.5%) have positive spectral phases, indicating that neurons favor lower frequencies than the center frequency (Fig. 7*D*).

The SRF profile allows us to study its arrangement in terms of spectral excitation and inhibition. The behavior of each neuron can also be interpreted directly in the ripple density or frequency domain (Kowalski et al. 1996; Miller et al. 2002; Schreiner and Calhoun 1994). To do this, the SRF is converted into a spectral modulation transfer function (sMTF). The sMTF measures the neurons response (spikes $\cdot$ s^{-1} $\cdot$ dB^{-1}) as a function of the applied ripple density. Using the Gabor model representation of the SRF profile (*Eq. 11*), the corresponding sMTF is obtained by applying a Fourier transform magnitude (FTM) to the SRF profile

$$G(x) = K \cdot e^{-[2(x-x_0)/\text{BW}]^2} \cdot \cos\left[2\pi \cdot \Omega_0 \cdot (x - x_0) + P\right]$$
$$\Downarrow \text{FTM} \qquad\qquad (14)$$
$$\text{sMTF}(\Omega) = A \cdot e^{-[\pi\text{BW}\cdot(\Omega - \Omega_0)/2]^2}$$

where all symbols are defined as in *Eq. 11*. The parameter *A,* determines the peak magnitude of the MTF or equivalently the gain of the neuron from stimulus to response (units spikes/s/dB). It is related to the magnitude of the SRF through the relationship: $A = K \cdot \text{BW} \cdot \sqrt{\pi}/4$. The sMTF acquires the structure of a Gaussian function with the center Ω_0 and standard deviation $\sqrt{2}/\pi/\text{BW}$. The bandwidth of the sMTF is defined as the width of the sMTF that accounts for 85% of the total energy under the Gaussian curve. This parameter determines the range of spectral oscillations (cycles/octave) in a stimulus that can potentially activate the neuron. According to this criterion, the tail points at the level of 1/*e* of the Gaussian sMTF peak value delineate the bandwidth of the sMTF. Compared to the bandwidth of the SRF profile, the bandwidth of the sMTF (4/π/BW) is inversely proportional to the bandwidth of the SRF profile (BW).

Figure 8, *A–C,* shows representative sMTFs of three single neurons in the ICC. To facilitate comparisons, each sMTF was normalized so that their total energy is equal to one; — shows the normalized sMTFs from *Eq. 14,* - - - corresponds to the normalized sMTFs obtained directly from measured SRF profiles. The Gabor sMTF model (*Eq. 14*) accounts for the structure and energy of the actual sMTFs quite well as depicted by the — and - - - in Fig. 8.

Neurons were individually classified according to their spectral filtering characteristics. These can, in theory, take the form of lowpass, bandpass, or highpass filtering response pattern.
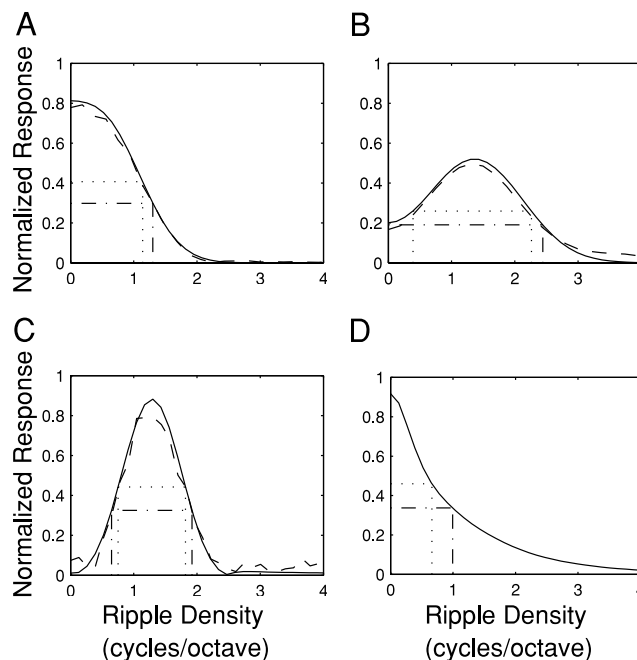

FIG. 8.   Representative spectral modulation transfer functions (sMTF). — and - - -, the fitted and measured sMTFs, respectively. All sMTFs are normalized for unit energy. *A*: a typical lowpass sMTF with the best ripple density (Ω_0 = 0 cycles/octave) and bandwidth (1.30 cycles/octave at upper 8.68-dB cutoff or 1.14 cycles/octave at upper 6-dB cutoff). *B* (best ripple density: 1.30 cycles/octave; bandwidth: 2.44 cycles/octave at upper 8.68-dB cutoff; 1.87 cycles/octave at upper 6-dB cutoff) and *C* (best ripple density: 1.30 cycles/octave, bandwidth: 1.27 cycles/octave at upper 8.68-dB cutoff; 1.07 cycles/octave at upper 6-dB cutoff) show typical sMTFs with bandpass filter characteristics. *D*: the composite population sMTF for the inferior colliculus (ICC) assumes a lowpass filter characteristic with a best ripple density of zero and bandwidth 0.995 cycles/octave (at upper 8.68-dB cutoff) or 0.662 cycles/octave (at upper 6-dB cutoff). $\cdots$ and $-\cdot-$, the upper 6- and 8.68-dB cutoff, respectively.

Neurons in our sample only exhibited lowpass (Fig. 8*A*) and bandpass (Fig. 8, *B* and *C*) spectral selectivity. The criterion for classifying each neuron from the sMTF consisted of comparing the sMTF bandwidth of each neuron in relation to its best ripple density. Specifically, we required that the measured best ripple density (Ω_0) be greater than half the sMTF bandwidth for bandpass neurons. This requirement guarantees that bandpass neurons have a residual DC level response of less than half the sMTF peak magnitude; whereas lowpass neurons will have a significant DC response with >50% of the peak response magnitude. Figure 8*A* illustrates this procedure for a typical sMTF with lowpass selectivity (same as Fig. 5*A*), which shows a nonoscillatory on-spectral response pattern. Its sMTF indicates that the structure of the STRF along the spectral dimension is dominantly excitatory or inhibitory. A neuron with bandpass filter characteristics is illustrated by the examples of Fig. 8*C* (same as Fig. 5*B*). This neuron has an SRF with strong alternating excitatory and inhibitory subfields. An intermediate scenario occurs for the neuron of Fig. 8*B* (same as Fig. 2*A*), which shows a significant DC level response in the sMTF; however, the neuron exhibits weak inhibitory sidebands and, consequently, a best ripple density that is offset from zero. In the STRF domain, this neurons shows a strong pattern of excitation and a significant, but subtle, inhibitory subregion. According to our criterion, we found that 80 of 93 neurons exhibited lowpass response preferences; 83 neurons (13 band-

pass and 70 lowpass) had best ripple densities offset from zero (as for Fig. 8*B*) and 69 had best ripple densities <1 cycle/octave. Thirteen neurons exhibited bandpass selectivity, and no neurons had highpass response preferences.

Each individual sMTF tells us about the spectral selectivity of individual neurons and tells us little about the overall spectral filtering capabilities of the inferior colliculus. Therefore, we determined the overall spectral selectivity of the inferior colliculus by computing a population sMTF. The population sMTF of the inferior colliculus (Fig. 8*D*) was obtained by averaging the amplitude-normalized sMTFs of all single neurons. Using the criterion defined for single unit sMTFs, we find that the spectral selectivity of the ICC (in the sampled frequency range) is lowpass with a bandwidth of 0.995 cycles/octave (at upper 8.68 dB cutoff; according to the 1/*e* bandwidth criterion) or 0.662 cycles/octave (at upper 6 dB cutoff) and centered about a best ripple density of zero cycles/octave. Thus the ICC as a whole has a significant preference for broadband stimuli.

Temporal response preferences

Neurons in the ICC show a diverse range of response preferences to temporally modulated stimuli (e.g., Krishna and Semple 2000; Langner and Schreiner 1988; Ramachandran et al. 1999; Rees and Møller 1983). While numerous studies have identified the output-response characteristics of ICC neurons to simple time-varying stimuli, the receptive field structure leading to these response preferences has previously not been studied. Temporal response characteristics of ICC neurons can be interpreted by four parameters of the temporal Gabor model (*Eq. 13*)—the best temporal modulation frequency (F_{m_0}), the peak latency (T_0), the response duration (D), and the temporal phase (Q). Together, the peak latency and response duration determine the locality and width of the TRF profile, respectively; the best temporal modulation frequency and temporal phase determine the rate and alignment of the temporal oscillation of the TRF profile.

Figure 9 illustrates distributions for these parameters for the contralateral receptive field. The absolute value of the best temporal modulation frequency ranged from 0 to 255.5 Hz and the distribution peaks at 30 Hz (Fig. 9*A*). Thus although numerous neurons can respond selectively to exceedingly fast temporal modulations of the dynamic moving ripple, most neurons preferred low modulation rates.

The peak latency is defined as the time of maximal neural response (excitation or inhibition) following the onset of stimulation, whereas the response duration determines the time period over which the neurons integrate acoustic information. From the distributions in Fig. 9*B*, the peak latency was usually <20 ms (range: 3.5–27.4 ms; mean: 10.1 ms; median: 8.5 ms) and is consistent with previous observations using pure tone and noise stimuli (Krishna and Semple 2000; Langner and Schreiner 1988). The response durations extended over a broad range (observed range: 1.8–82.6 ms), although most neurons typically had short response durations (mean: 12.1 ms, median = 6.2 ms).

Finally, the temporal phase determines the arrangement of excitation and inhibition of the TRF profile, relative to the peak latency or centroid position—which is determined from the TRF envelope. Positive temporal phases shift the TRF profile to the left of the peak latency; negative values shift the TRF
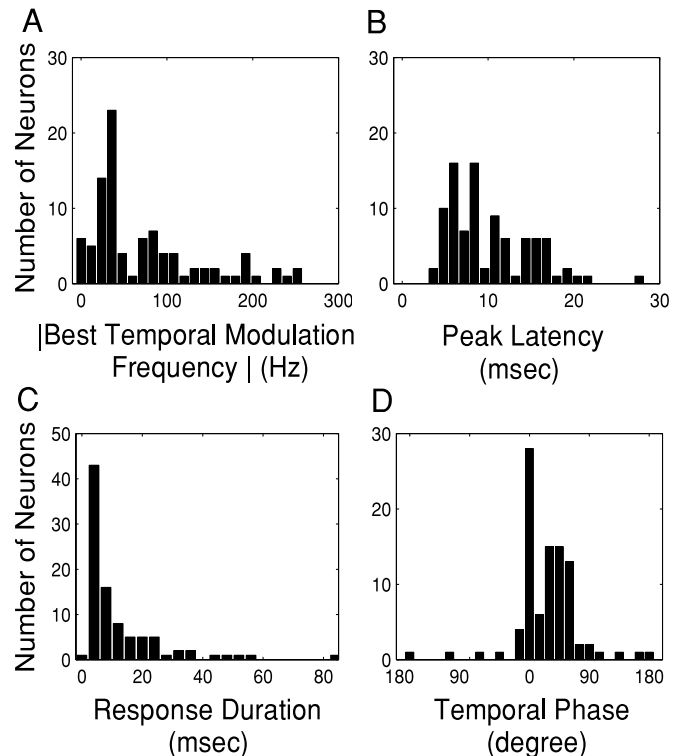

FIG. 9. Distributions for temporal STRF parameters. *A–D*: the best temporal modulation frequency (F_{m_0}), the peak latency (T_0), the response duration (D), and the temporal phase (Q), respectively.

profile to longer latencies. The temporal phase distribution (Fig. 9*D*) shows that 78.5% of temporal phases are positive, thus indicating that the peaks of the TRF profiles are typically shifted to the left of the peak derived from the temporal envelope. Therefore excitation typically precedes inhibition.

The TRF profile allows us to study the timing of the neural response and the temporal arrangement of excitation and inhibition. The behavior of each neuron can also be interpreted and studied directly in the frequency domain. By converting the TRF profile (measured at the center frequency) into the Fourier domain, we can obtain the temporal modulation transfer function (tMTF) of each neuron. The tMTF characterizes the time-locked response of the neuron as a function of the temporal modulation frequency. Using the Gabor function TRF profile (*Eq. 13*), the tMTF can be represented by a Gaussian function of the form

$$\text{tMTF}(F_m) = A \cdot e^{-[\pi \cdot D \cdot (F_m - F_{m_0})/2]^2} \qquad (15)$$

where F_{m_0} and D are as in *Eq. 13* and the tMTF is expressed in units of spikes/sec/dB. The parameter A corresponds to response strength. To facilitate comparisons, each tMTF was normalized for unit energy. The criterion for choosing the bandwidth of the tMTF and for classifying them according to lowpass and bandpass selectivity follows the same procedure as for the sMTF (see previous section). Thus the duration of the TRF profile (D) is inversely proportional to the bandwidth of the tMTF ($4/\pi/D$).

Figure 10 shows three representative inferior colliculus tMTFs. The examples of Fig. 10, *A* and *B*, have a significant DC level response and are therefore classified as having lowpass sensitivity to the temporal modulation frequency. While
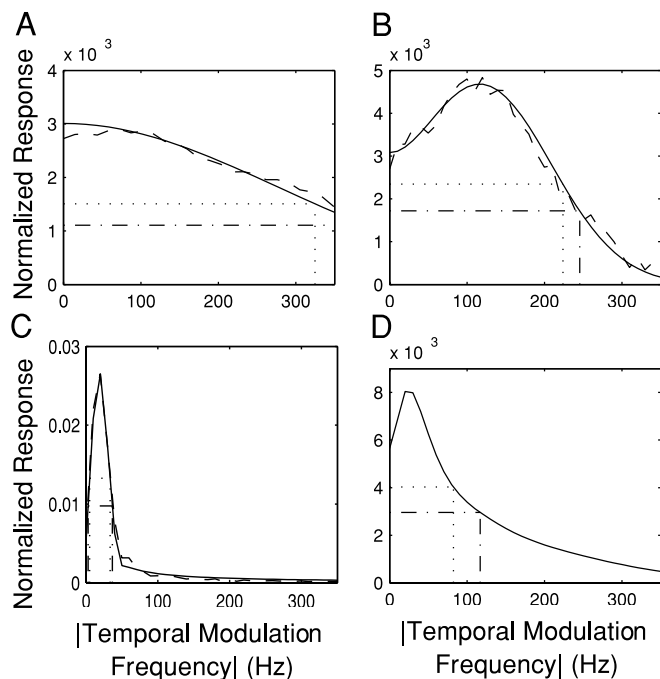
FIG. 10. Representative temporal modulation transfer functions (tMTFs). — and - - -, the fitted and measured tMTFs (normalized for unit energy), respectively. *A* and *B*: typical lowpass tMTFs with 0 (*A*) and non-0 (*B*) best temporal modulation frequencies (bandwidths, *A*: 350 Hz at upper 8.68 dB cutoff; 324.7 Hz at 6 dB upper cutoff; *B*: 245.4 Hz at upper 8.68 dB cutoff; 223.8 Hz at upper 6 dB cutoff). *C*: a typical bandpass tMTF (best temporal modulation frequency: 20.0 Hz; bandwidth: 34.0 Hz at upper 8.68 dB cutoff; 28.5 Hz at upper 6 dB cutoff). *D*: the composite population tMTF for the ICC is lowpass in character with non-0 best temporal modulation rate (30.0 Hz; bandwidth: 117.0 Hz at upper 8.68 dB cutoff; 82.5 Hz at upper 6 dB cutoff). ··· and – · –, the upper 6 dB and 8.68 dB cutoff, respectively.

the first neuron has its strongest response at zero frequency, the latter neuron has a best temporal modulation frequency of 130.3 Hz. Both neurons responded over a large range of modulation frequencies as suggested by their response bandwidths. The bandwidths of the tMTF for Fig. 10, *A* and *B,* are 350.0 Hz (at upper 8.68 dB cutoff or 324.7 Hz at upper 6 dB cutoff) and 245.4 Hz (at upper 8.68 dB cutoff or 223.8 Hz at upper 6 dB cutoff), respectively.

The timing pattern of the STRF is critical for determining the behavior of the tMTF and its classification as lowpass or bandpass sensitivity—this behavior, in turn, depends strongly on the patterning of temporal excitation and inhibition of the STRF. Typical STRFs that show lowpass tMTFs with zero best temporal modulation frequency contain purely excitatory or inhibitory features in the temporal cross-section of the STRF (e.g., Fig. 10*A;* same as contra in Fig. 13*D*); alternately, if the neuron has a lowpass tMTF with non-zero best temporal modulation frequency, its STRF will show an interleaved arrangement of excitation and inhibition—although typically not of the same strength (Fig. 10*B*). A tMTFs with bandpass sensitivity is depicted in Fig. 10*C* (same neuron as Fig. 5*B*). This neuron has a best temporal modulation frequency and bandwidth of 20.0 and 34.0 Hz at upper 8.68 dB cutoff (or bandwidth of 28.5 Hz at upper 6 dB cutoff), respectively. Such STRFs have an alternating arrangement of excitation and inhibition along the temporal axis of the TRF profile. Across the entire population, 51 neurons show lowpass temporal sensitivity—of which *n* = 4 had best temporal modulation frequency of exactly zero.

Forty-two ICC neurons were classified as having bandpass tMTFs—all of which had non-zero best temporal modulation frequencies.

The overall temporal selectivity of the ICC was determined by averaging all normalized tMTFs to approximate the composite tMTF for the population. The population tMTF shows lowpass selectivity to the dynamic moving ripple stimulus (Fig. 10*D*), although the best temporal modulation rate is offset from zero (peak: 30.0 Hz; bandwidth: 117.0 Hz at upper 8.68 dB cutoff or 82.5 Hz at upper 6 dB cutoff).

Time-frequency separability

Central auditory neurons can exhibit time-frequency interactions in response to sounds with spectral and temporal structure as observed for the coding of frequency-modulated stimuli (Kowalski et al. 1996; Rees and Møller 1983). Such neural interactions may be used for encoding of time-frequency conjunctions, although the neural basis for such selectivity is unknown. Speech and other vocalization signals exhibit directionally oriented time-frequency sweeps and time-dependent frequency modulations in the signal spectrum. Neuronal selectivity to oriented stimulus features may arise through spectro-temporal filters that are selectively oriented to the direction of a frequency sweep—analogous to the motion selective neurons in the visual system (DeAngelis et al. 1993b). Alternately, it is also possible that directionally oriented stimulus features interact with excitatory and inhibitory RF subregions of *unoriented* spectro-temporal receptive fields; and the saliency for oriented stimulus information would instead be explained by the population response of unoriented spectro-temporal filters. We can address this issue in the ICC by analyzing the detailed structure of the STRF, TRF, and SRF. Specifically, we are interested in determining how the TRF profile changes with frequency or the SRF profile changes with time and how each of the model parameters contributes to the STRF structure. Are the spectral and temporal dimensions of the stimulus integrated independently at the colliculus level? To address these questions, we can initially slice through the STRF at different latencies (e.g., Fig. 1, *B* and *C*) or at different frequencies (e.g., Fig. 2*B*) to study the time-frequency interactions of neuronal responses.

Figure 11*B* shows a typical time-frequency inseparable STRF. To examine how the structure of the SRF profile changes with time, we use the spectral Gabor function (*Eq. 11*) to fit several cross-sections of this STRF at different latencies and to extract physiologically relevant information of the SRF profiles. The black lines with open circles in Fig. 11, *C–F,* illustrate how four parameters of the Gabor function vary with latency. The center frequency (x_0), the bandwidth of the SRF (BW), and the best ripple density (Ω_0) do not change substantially with latency (*C–E,* respectively). However, the phase (*P*) gradually changes with latency by roughly 180°, accounting for the obliquely oriented transition from excitation to inhibition with increasing latency. This example illustrates how the time-varying spectral phase of the SRF profile accounts for much of the structure of the inseparable STRF.

In contrast to the STRF of Fig. 11*B,* the STRF of Fig. 11*A* has a time-frequency separable structure. For this neuron, the center frequency (x_0), the bandwidth (BW), and the best ripple density (Ω_0) are not uniquely specified for all latencies (dotted red lines in Fig. 11, *C–E*). The spectral phase (*P*) alternates by ~180° with latency in a manner that is directly correlated with
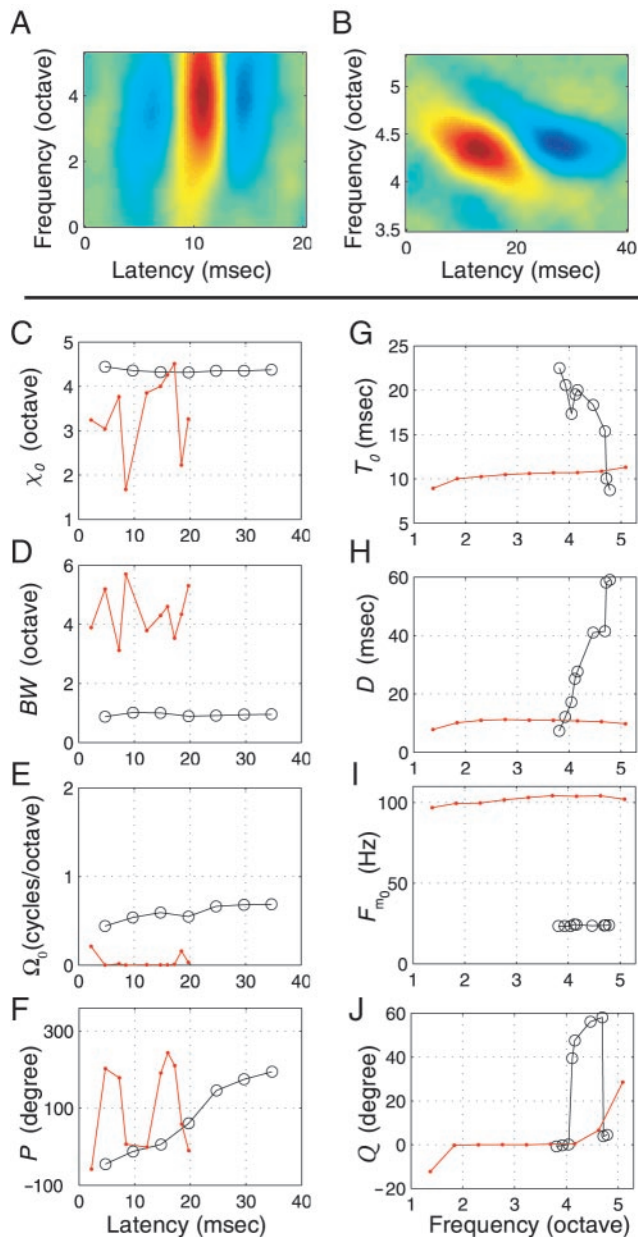
FIG. 11. Separability analysis. *A* and *B*: typical neurons with separable and inseparable STRFs, respectively. *C–F* illustrate how spectral parameters vary over latency: center frequency (x_0), the bandwidth of the SRF profile (BW), the best ripple density (Ω_0), and the spectral phase (*P*), respectively. *G–J* illustrate how temporal parameters change over frequency: peak latency (T_0), the response duration (*D*), the best temporal modulation frequency (F_{m_0}), and the temporal phase (*Q*), respectively. Black lines with open circles represent spectral and temporal parameters for neuron *B*. Red lines with solid circles indicate spectral and temporal parameters for neuron *A*.

the excitatory and inhibitory subregions of the STRF. In the excitatory subregion, the measured phase of ~10° extends over the entire duration of the excitation (between 8 and 12 ms); but in the inhibitory regions, the phase increases sharply to ~200° (between 5–8 and 12–18 ms). From these examples, it is clear that the spectral phase determines the sign of the neuron's SRF profile and, therefore, accounts for the alignment of neural excitation and/or inhibition observed in the STRF.

We can use the same technique as for the SRF profile to investigate how the TRF profile change as a function of fre-

quency. Temporal cross-sections of the STRF obtained at different frequencies are individually fitted by Gabor functions (*Eq. 13;* Fig. 2*D*). The changes of four temporal parameters in the Gabor function are illustrated in the Fig. 11, *G–J,* for neurons *A* and *B* of Fig. 11. Neuron *B* has a peak latency (T_0) and response duration (*D*) that vary with frequency (black lines with open circles in Fig. 11, *G* and *H,* respectively); however, its best temporal modulation frequency (F_{m_0}) (black line with open circle in Fig. 11*I*) is constant. The temporal phase (*Q*) of this neuron changes gradually from ~0 to ~60° with frequency within the response region (between 4 and 5 octaves) (black line with open circle in Fig. 11*J*). Alternately for neuron *A,* the peak latency (T_0), response duration (*D*), and best temporal modulation frequency (F_{m_0}) do not vary substantially over frequency (red lines with solid circle in *G–I,* respectively). Because the temporal pattern of the excitation and inhibition is similar at all frequencies, the temporal phase is roughly constant throughout the extent of the STRF (red line with solid circle in *J*).

The preceding analysis demonstrates that inseparable STRFs do not have unique spectral phase over latency. Furthermore it shows that the peak latency, duration, and temporal phase are not necessarily constant with changing frequency. Separable STRFs, alternately, have unique spectral phase (±180° increment), peak latency, response duration, and temporal phase over frequency within the specified response region.

The Gabor STRF model is built up as sum of STRF components, each of which is a time-frequency separable STRF. Therefore a measure of separability can be obtained by considering the energy of the first-singular value in relationship to the total energy of the higher-order singular values of the fitted Gabor model. The separability index (α_d; see METHODS) assumes values between 0 and 1. If the measured STRF is perfectly separable, α_d assumes a values of 1; alternately, an STRF with highly inseparable time-frequency features has a separability index near zero. As an example, the STRF of Fig. 11*A* is approximately time-frequency separable and, consequently, its separability index is high (0.934). Neurons with non-separable oblique features typically have lower separability indices (e.g., Fig. 11*B*, 0.692).

Most neurons in the inferior colliculus have time-frequency separable structure and, therefore, independently integrate spectral and temporal stimulus attributes. The separability index distribution of all neurons (Fig. 12) contains a sharp peak near $\alpha_d = 1$ (observed range: 0.292–1). Measured separability index values are skewed toward one as suggested by the mean and median values (mean = 0.919, median = 1). Of those neurons (40%) that exhibit time-frequency inseparable structure ($\alpha_d < 1$), only a few neurons exhibited highly inseparable receptive field arrangements (as in Figs. 5, *A* and *B,* and 13*C*) and many more had separability indices near one. Thus in contrast to motion selectivity in the visual system—where a large proportion of visual cortex neurons exhibit highly inseparable receptive fields (DeAngelis et al. 1993a,b, 1995)—most ICC STRFs are either purely separable or only weakly inseparable. This finding supports the hypothesis that the majority of selectivity to FM stimuli in the auditory system arises through stimulus interactions with excitatory and inhibitory RF subregions and not through strongly oriented neural receptive fields. Furthermore, the high proportion of separable STRFs may be important for encoding comodulated components in natural
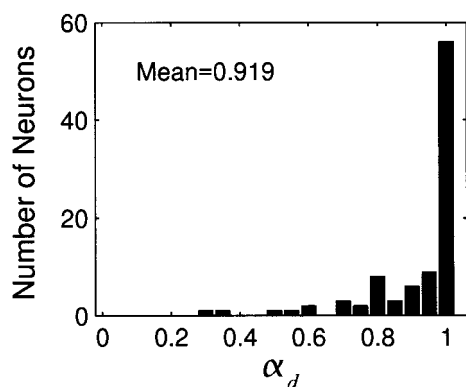
FIG. 12. Distribution of separability index (α_d). Most neurons (56/93) have perfectly separable time-frequency structure ($\alpha_d = 1$). The mean separability index (0.919) is exceptionally high, indicating that most collicular STRFs are well approximated by the product of the TRF and SRF profile.

signals that are time-frequency separable (Nelken et al. 1999), whereas the small proportion of highly inseparable receptive fields may play a specific role in the coding of strongly oriented frequency sweeps, which appear to be less prevalent in natural signals.

Binaurality

Binaural interactions are well described in the central auditory system (Goldberg and Brown 1969; Irvine and Gago 1990; Kuwada et al. 1997; Schnupp et al. 2001). Most binaural studies use structurally simple stimuli that are simultaneously presented to each ear to identify neural mechanisms of sound localization. Although a great deal is known about the response characteristics to such stimulus combinations, little is known about the general receptive field arrangements underlying binaural interactions. For this reason, we apply our Gabor model to compare the arrangements of neural receptive fields for contralateral and ipsilateral inputs to the ICC.

Hypothetically, binaural interactions to simple stimuli should be reflected in the structure and/or energy of the contra- and ipsi-STRFs. One possibility is that binaural receptive fields have identical spectro-temporal structure. Under such a model, differences in average input drive (e.g., STRF energy) from each ear could potentially account for binaural sensitivities, although each neuron would encode for identical spectro-temporal stimulus features in both ears. Alternately, it is also possible that the contra- and ipsi-STRFs are distinctly different and systematic differences in the converging receptive field structures account for binaural sensitivities. Figure 13 illustrates typical receptive fields obtained with simultaneous binaural stimulation with statistically independent contra and ipsi dynamic moving ripple stimuli (Escabí and Schreiner 2002). In the previous sections, we examined only the structure of the dominant contralateral STRFs. We find that 36/99 ICC neurons also exhibit significant ipsilateral STRFs. In terms of the dominant excitatory or dominant inhibitory interactions (Goldberg and Brown 1969), neurons with binaural sensitivity can be classified as principally excitatory-excitatory (EE), excitatory-inhibitory (EI), excitatory-unresponsive (EO), etc. Although most neurons exhibit no discernable STRF structure for the ipsilateral ear ($P < 0.002$; EO; 62/99; Fig. 13, E and C), 23 neurons exhibited dominant excitatory binaural interactions

(EE; Fig. 13A); six neurons responded exclusively to the ipsilateral ear (OE; Fig. 13F); 4 had a dominant ipsilateral inhibitory subregion (EI; Fig. 13B); 3 exhibited dominant contralateral inhibition (IE; Fig. 13D); and one neuron had a dominant inhibitory contralateral subregion (IO; Fig. 13E).

The preceding examples illustrate the diversity of binaural STRF composition observed in the ICC. Differences between the contra- and ipsi-STRFs can, in theory, manifest solely along the temporal dimension of the TRF profile, the spectral dimension of the SRF profile, or along both—the spectral and temporal dimension of the STRF. Therefore we compared the spectral and temporal composition of the contra- and ipsi-STRFs to determine which dimensions and parameters contribute to binaural sensitivities.
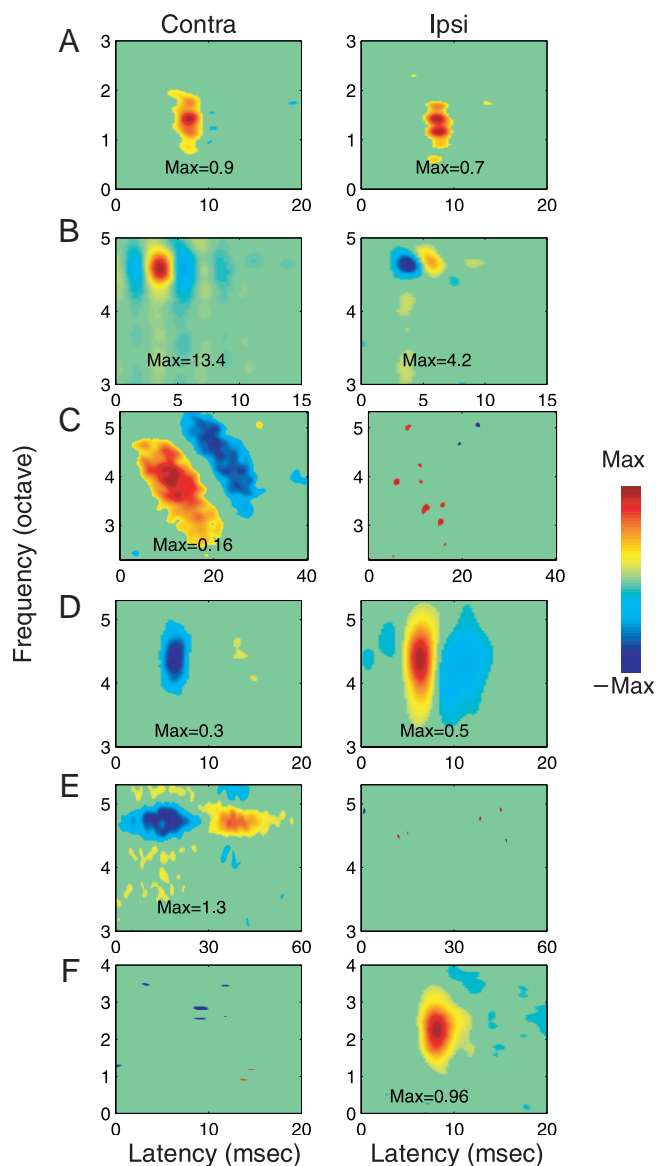

FIG. 13.   Representative binaural STRFs. Contralateral (left) and ipsilateral (right) STRFs exhibited a variety of arrangements of spectral and/or temporal excitation and/or inhibition, although all binaural receptive fields showed a large degree of spectro-temporal overlap. Only STRF subregions that exceeded significance ($P < 0.002$) are shown. Neurons could exhibit purely excitatory interactions (A), excitatory and inhibitory interactions (D and B), or could respond exclusively to the contra- or ipsi-ear (C, E, and F).
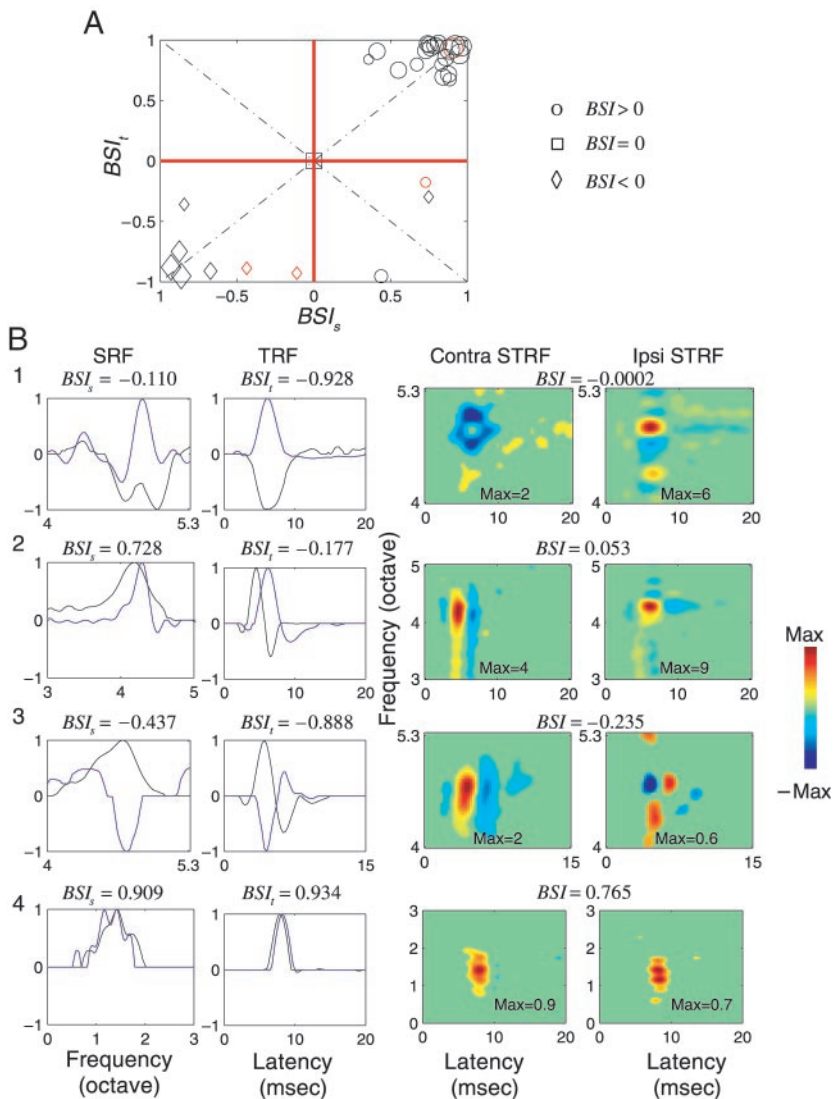
FIG. 14. Analysis of binaural receptive field disparity. *A*: the joint distribution of spectral (BSI_s), temporal (BSI_t), and spectro-temporal (BSI) binaural similarity indexes. Spectral and temporal BSIs are designated along the abscissa and ordinate, respectively, whereas the magnitude and sign of the spectro-temporal BSI are represented symbolically by symbols of different sizes (open diamonds and open circles indicate BSI < 0 and BSI > 0, respectively; larger symbols corresponds to larger BSI magnitudes). Neurons with 0 BSI showed only monaural preferences (open squares symbol). *B*: example neurons showing various spectral, temporal, and spectro-temporal receptive field arrangements (denoted by red symbols in *A*). Neurons can exhibit correlated (e.g., neuron 4), uncorrelated (e.g., neuron 1, 2), or anticorrelated (e.g., neuron 1–3) spectral and/or temporal receptive field structures. Receptive field data is shown at a significance level of *P* < 0.002.

The spectral, temporal, and spectro-temporal arrangement of binaural receptive fields was first analyzed by considering the structural similarity between the contra- and ipsi-STRF. Three metrics were devised to quantify the relative degree of structural aural similarity for TRF profiles, SRF profiles, and the entire STRF (see METHODS; *Eqs. 4–6*). The binaural similarity index (BSI) is analogous to the correlation coefficient between the contralateral and ipsilateral STRF. The spectral BSI (BSI_s) and the temporal BSI (BSI_t) are analogous to a correlation coefficient between the contra- and ipsi-SRF profiles and the TRF profiles, respectively.

Example binaural response profiles along with the respective TRF and SRF profiles are shown in Fig. 14*B*. Some neurons exhibited temporally orthogonal receptive field arrangements (Fig. 14*B;* neuron 2; $BSI_t = -0.177$) whereas others had anticorrelated TRF profiles (Fig. 14*B;* neuron 1, $BSI_t = -0.928$; neuron 3, $BSI_t = -0.888$). Spectral profiles could also exhibit correlated (Fig. 14*B;* neuron 2, $BSI_s = 0.728$; neuron 4, $BSI_s = 0.909$), anticorrelated (Fig. 14*B;* neuron 3; $BSI_s = -0.437$), or uncorrelated (Fig. 14*B;* neuron 1; $BSI_s = -0.110$) arrangements between the contra- and ipsi-STRFs. Such differences either occurred simultaneously in time and

frequency (Fig. 14*B;* neuron 3) or independently for each dimension (Fig. 14*B;* neuron 2). For instance, neuron 2 of Fig. 14*B* has correlated SRF profiles and a temporally misaligned (uncorrelated) TRF profiles, whereas neuron 3 has misaligned (anticorrelated) SRF and TRF profiles. Other neurons had perfectly aligned receptive field structure with similar SRF and TRF profiles (Fig. 14*B;* neuron 4).

Population data for the spectral, temporal, and spectro-temporal BSI are shown in Fig. 14*A*. For the vast majority of binaural neurons, the spectral and temporal BSIs are clustered near high negative and positive values (Fig. 14*A*), thus indicating that the contra- and ipsi-SRF and TRF profiles can assume a correlated or anticorrelated structure. The absolute magnitude of the spectral and temporal BSIs (spectral, 0.723 ± 0.199; temporal, 0.760 ± 0.244; mean ± SD) are reasonably high, whereas the absolute magnitude of the joint spectro-temporal BSI is significantly lower (0.513 ± 0.2352; mean ± SD; paired *t*-test, $P < 0.001$). This finding suggests that, individually, the temporal and spectral dimensions of the contra- and ipsi-STRF share some common features in the TRF and SRF profiles; however, the spectro-temporal arrangements of the contra- and ipsi-STRFs appear to be less matched.
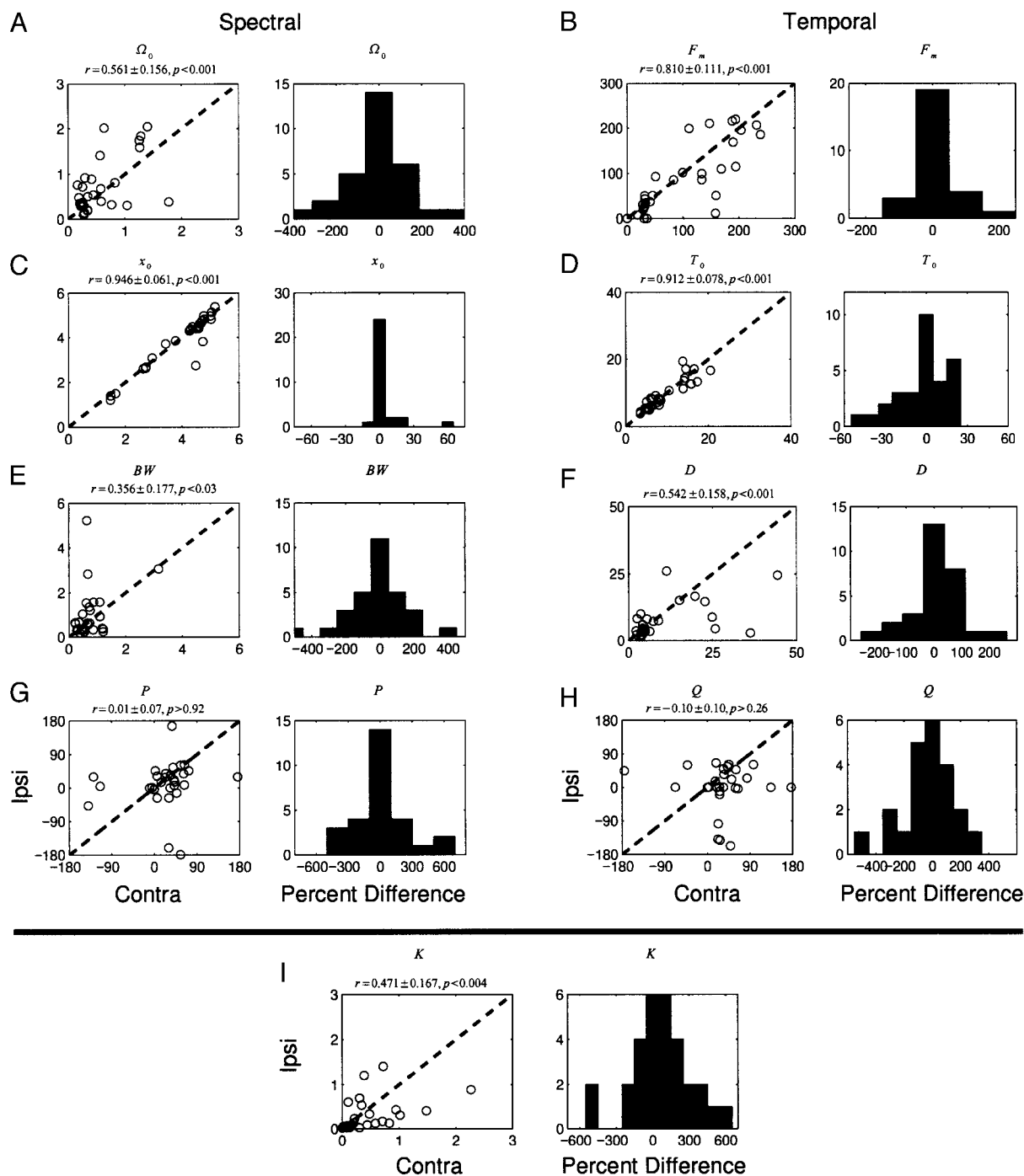
FIG. 15. Contralateral and ipsilateral receptive field parameter distributions. *A–I*: scatter plots and disparity distributions (percent difference between contra and ipsi) showing spectral (*left*) and temporal (*right*) parameters of the contra and ipsi receptive field: *A*: best ripple density (Ω_0); *B*: the best temporal modulation frequency (F_{m_0}); *C*: the center frequency (x_0); *D*: the peak latency (T_0); *E*: the bandwidth of the SRF profile (BW); *F*: the response duration (D); *G*: the spectral phase (P); *H*: the temporal phase (Q); and *I*: response strength (K), respectively. Correlation coefficients and the corresponding significance levels are indicated on *top* of each scatter plot.

Systematic differences in contra- and ipsilateral STRF structure can potentially account for some aspects of binaural sensitivities in the ICC. Which receptive field dimensions (temporal or spectral) and neural parameters contribute to the observed binaural receptive field mismatch? To identify the source of this mismatch, we first fitted the contra- and ipsi-STRFs to the Gabor STRF model. Contralateral and ipsilateral

parameters for each receptive field were then individually compared. Figure 15 illustrates scatter plots for the spectral and temporal parameters derived from the contra- and ipsi-STRFs. Some spectral and temporal parameters, including the peak latency (T_0, Fig. 15*D*; $r = 0.912 \pm 0.078$, *t*-test, $P < 0.001$) and center frequency (x_0, Fig. 15*C*; $r = 0.946 \pm 0.061$, *t*-test, $P < 0.001$), were highly conserved; other parameters showed

lower correlation values although statistically significant. Comparing temporal (F_{m_0}, D) and spectral parameters (Ω_0, BW), we find that the temporal receptive field dimensions are more highly matched for the two inputs (F_{m_0}: $r = 0.810 \pm 0.111$, t-test, $P < 0.001$; D: $r = 0.542 \pm 0.158$, t-test, $P < 0.001$; Ω_0: $r = 0.561 \pm 0.156$, t-test, $P < 0.001$; BW: $r = 0.356 \pm 0.177$, t-test, $P < 0.03$). All spectral and temporal parameters were statistically correlated, with the exception of the spectral and temporal phases (circular correlation analysis; P: $r = 0.01 \pm 0.07$, bootstrap, $P > 0.92$; Q: $r = -0.10 \pm 0.10$, bootstrap, $P > 0.26$). Thus although numerous STRF parameters collectively contributed to the mismatch of ipsi- and contra-receptive fields, the spectral and temporal phases contributed the most to the binaural receptive field misalignments. Together, this suggests that the overall extent and centers of the spectral and temporal receptive field integration area are typically closely matched binaurally. However, the degree of binaural alignment of excitation and inhibition can vary widely among neurons, thus providing a currently little appreciated binaural integration condition beyond intra-aural time and level differences.

As proposed in the visual system, systematic differences in the binocular receptive field properties may be used to detect the depth of a visual object. In the studies by Anzai et al. (1999), visual cortex neurons show systematic differences retinotopic position and spatial phase between the left and right inputs that are consistent with models of binocular depth perception. Similarly, our analysis of the binaural composition of the auditory STRF suggests that differences in the binaural alignment of excitatory and inhibitory RF features may provide a mechanism for encoding differences in the converging binaural spectrum; which, in turn, can be used to determine the position of a sound source in space. Unlike visual RFs, we find that the central position of colliculus STRFs is conserved binaurally, and therefore positional cues do not appear to contribute to binaural detection as for the visual system. Significant disparities in the spectro-temporal phase, however, lead to interleaved patterns of excitation and inhibition binaurally. Such aural differences may be important for analyzing spectral notches in the spectrum of a sound source, which vary significantly as a function of spatial position (Hartmann and Witternberg 1996; Kulkarini and Colburn 1998).

DISCUSSION

We have studied the monaural and binaural spectro-temporal receptive field structure of 99 phase-locking neurons in the cat ICC (Escabí and Schreiner 2002). A time-frequency Gabor STRF model is presented that allows us to quantify the receptive field structure of auditory STRFs. This model can be used to remove measurement noise in the STRF and to extract physiologically meaningful information of the receptive field structure. Our results provide the following new insights: *1* the Gabor function is an adequate descriptor of the SRF and TRF profile (Figs. 1 and 2). Using the described singular value decomposition method, we can extend the fitting procedure to the entire STRF. The STRF can be described by the weighted sum of independent separable STRF components, which are the product of a spectral waveform and a temporal waveform (Figs. 3 and 5). These can in turn be fitted with the time-frequency Gabor STRF model. *2* From the analysis of the

contralateral sMTF and tMTF, ICC neurons exhibited lowpass and/or bandpass spectral and temporal selectivity. *3* The separability index (α_d) measures the degree of time-frequency separability of the STRF. Most neurons (60.2%) exhibited time-frequency separable receptive field structure and, therefore, independently process spectral and temporal stimulus attributes. *4* Finally, we used the model to study differences in the converging ipsi- and contralateral receptive field structure. Our results indicate that for neurons exhibiting binaural convergence most STRF properties for the two inputs are highly correlated. However, subtle spectro-temporal differences in the alignment of excitation and inhibition contribute significantly to binaural processing in the ICC. Together, the model provides a uniform description of the receptive field structure that allows us to jointly evaluate spectral, temporal, spectro-temporal, and binaural aspects of the stimulus-response relationship.

Gabor STRF model

The STRF is an approximation of the neural receptive field obtained by the spike-triggered average method using finite experimental data (Miller et al. 2002; Escabí and Schreiner 2002). A time-frequency Gabor model was used to remove measurement noise and to quantitatively evaluate the receptive field structure of ICC neurons. Both the spectral RF and temporal RF profiles are equally well described by a unidimensional Gabor function, as indicated by the high temporal (mean = 0.933) and spectral (mean = 0.938) similarity indices of the fits to the raw data. The structure of the entire STRF showed a subtle reduction in the spectro-temporal SI (mean = 0.846) that can be accounted for by multiplicative errors that are propagated independently when the STRF is built up as a product of SRF and TRF profiles. Differences in the entire STRF structure were evaluated by measuring the normalized MSE between the model and measured STRF. Most neurons had low MSE values (mean ± SD = 0.185 ± 0.126; Fig. 6D), indicating that the receptive field structures were well accounted for both in shape and energy.

By analyzing the statistical structure of the receptive field measurement noise (Fig. 4), we were able to determine the number of independent receptive field dimensions required to properly fit collicular STRFs. Typically, we find that one or two STRF components are sufficient to capture the structure of inferior colliculus receptive fields. Only 39.7 and 7.5% of the neurons had significant second and third components each accounting, respectively, for only 6.2 ± 5.0 and 2.3 ± 1.8% of the total receptive field energy. Because each Gabor function requires 9 independent parameters, ICC STRFs therefore typically require 9 or 18 independent parameters to fully account for the entire receptive field structure.

Spectro-temporal receptive field structure

The spectral modulation transfer function (sMTF) was used to quantify the spectral selectivity of the SRF profile. Most ICC neurons exhibited lowpass sMTF (86%, $n = 80$; 14% bandpass, $n = 13$) although in most of those cases (70 of 83 lowpass neurons), a non-zero best ripple density (a peak in the filter function) could be identified (ranging from 0.022 to 2.113 cycles/octave). By comparing the distribution of best ripple

density in the ICC to those in the thalamus and the cortex, we find that spectral preferences are highly conserved between the inferior colliculus and auditory thalamus (Miller et al. 2001, 2002) (Wilcoxon rank test, $P > 0.33$). Compared to the primary auditory cortex, the distribution of ripple densities was significantly different for the ICC (Wilcoxon rank test, $P < 0.001$) although both were grossly overlapped. When we recomputed the population sMTF according to the energy normalization procedure of Miller et al. (2002), we found that the collicular, thalamic, and cortical population sMTFs were closely matched, with similar upper 6-dB cutoff (upper 6-dB cutoff: ICC, 1.46 cycles/octave; thalamus, 1.30 cycles/octave; cortex, 1.37 cycles/octave; sMTF correlation coefficient: thalamus vs. ICC, $r = 0.99 \pm 0.01$; cortex vs. ICC, $r = 0.99 \pm 0.01$, mean $\pm$ SD). Furthermore, the observed range of sMTF bandwidths was comparable to those found in cortex with static ripple stimuli (Calhoun and Schreiner 1998; Schreiner and Calhoun 1994) and in the thalamocortical system with dynamic moving ripple (Miller et al. 2002). Together, the data indicate that the range of spectral selectivity, as determined with ripple spectra, is highly conserved in the colliculus and throughout the thalamocortical network (Miller et al. 2001).

The best ripple density reflects the periodicity pattern of spectral excitation and inhibition of the SRF profile while the spectral phase contributes to their spectral alignment (i.e., the dominant SRF profile peak position relative to the peak of the SRF envelope). Most STRFs have positive spectral phases distributed between 0 and 90°. Therefore, the frequency of the dominant excitatory SRF peak is typically below the neuron's center frequency (i.e., the peak of the SRF envelope), while the dominant inhibitory mode is typically above the center frequency.

In contrast to the spectral response, the temporal response pattern is more intricate. First, the structure of the temporal receptive profile is not symmetric about its peak point, and, therefore, it is necessary to skew the time axis to account for the sharp onsets response observed for the temporal envelopes of nearly all neurons (as determined from the positive asymmetry index). This property of the temporal receptive field likely accounts for the phasic nature of onset responses observed at the colliculus level for pure tones and throughout the auditory pathway (Heil and Irvine 1997). Furthermore, the temporal receptive field asymmetry may explain the perceptual saliency for asymmetrically ramped auditory stimuli (Neuhoff 2000; Patterson 1994).

Temporal response parameters that quantify the timing of ICC response were derived from the Gabor STRF model and the population tMTFs. The relative alignment of excitation and inhibition was determined from the temporal phase of the TRF profile. As for the SRF profile, we find that most STRFs have positive temporal phases between 0 and 90°, and therefore, the TRF profile of most neurons show an initial excitatory receptive field domain that is followed by an inhibitory/suppressive period. Latency values measured directly from the peak of the TRF profile are consistent with those reported previously for simpler stimuli (Krishna and Semple 2000; Langner and Schreiner 1988). The median value of peak latency (8.5 ms) is shorter than those in the thalamus and cortex (10.5 and 13.0 ms); (Miller et al. 2002). However, the distributions of the peak latencies for these three stations grossly overlap, and, therefore, all three stations are substantially coactivated.

The main temporal modulation preferences observed in this study largely match the ranges observed in previous studies with amplitude modulated tones or noise (e.g., Krishna and Semple 2000; Langner and Schreiner 1988; Rees and Møller, 1983). By comparing the tMTF of ICC, thalamus, and cortex (Miller et al. 2002) we confirm that temporal modulation preferences systematically deteriorate from the ICC to the primary auditory cortex (Schreiner and Langner, 1988a). The range of the best temporal modulation preferences in the ICC is broader than those in the thalamus and cortex (Miller et al. 2002), but narrower than for auditory nerve (AN) fibers (Joris and Yin 1992). There is a significant reduction in the population tMTF upper 6-dB cutoff (ICC, 82.5 Hz; thalamus, 62.9 Hz; cortex, 37.4 Hz) as well as the peak modulation following rate (ICC, 30 Hz; thalamus, 21.9 Hz; cortex, 12.8 Hz). Thus in contrast to the spectral selectivity, which is highly preserved, temporal response preferences degrade dramatically across these three stations. More than 50% of ICC neurons prefer best temporal modulation frequencies below the measured population mean (73.6 Hz); therefore suggesting that the population tMTF selectivity is biased toward low-modulation frequencies in the ICC.

According to our bandwidth criterion, we find that ~55% of ICC neurons exhibited lowpass sensitivity although the majority of lowpass neurons have tMTF peaks away from 0 Hz despite a significant DC level response; bandpass neurons, by comparison, had no evident DC component. The dramatic increase of bandpass behavior and response selectivity in the ICC compared to the auditory nerve (Joris and Yin 1992) is likely due to the interleaved patterns of temporal excitation and inhibition that is evident in nearly all ICC STRFs.

Analysis of the combined spectro-temporal receptive field structure reveals that the vast majority of ICC neurons are time-frequency separable (separability index: range, 0.292–1; mean, 0.919; median, 1) although some neurons exhibit obliquely oriented excitatory and inhibitory STRF subregions, or spectro-temporally misaligned excitatory/inhibitory components. This finding suggests that the majority of ICC neurons independently process temporal and spectral stimulus information. This is consistent with the fact that the first STRF component obtained from the SVD accounts for most of the STRF energy.

Spectro-temporal selectivity can also be evaluated by comparing the spectral and temporal parameters of the Gabor STRF model. Although the separability index indicates that the structure of the STRF can be built up from the TRF and SRF profiles, it is nonetheless possible that the parameters of the SRF and TRF profiles covary. By comparing the spectral bandwidth and temporal duration of the Gabor STRF model, we find that there is an evident time-frequency resolution tradeoff in the receptive field size (Fig. 16C). Furthermore, the best ripple density and best temporal modulation rate also showed a significant negative correlation ($r = -0.452 \pm 0.094$; $P < 0.001$; Fig. 16D)—indicative of a time-frequency tradeoff in the modulation filtering resolution (Escabí and Schreiner 2002).

Larger receptive fields can potentially accommodate a larger number of inhibitory/excitatory receptive field components as observed for feature selectivity in the songbird system (Sen et al. 2001). By analyzing the structure of the SRF and TRF profiles, we find a distinct trend between the receptive field size
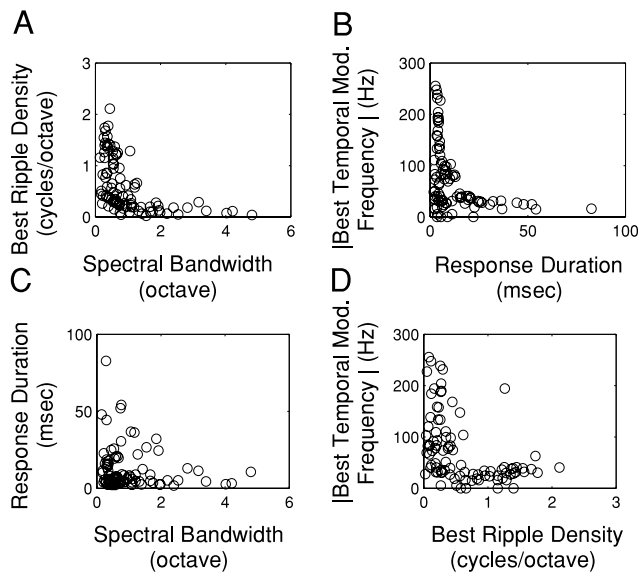
FIG. 16. Relationship between spectral end temporal receptive field parameters shows a tradeoff in time-frequency and modulation filtering resolution. *A*: spectral bandwidth vs. best ripple density; *B*: response duration vs. best temporal modulation frequency; *C*: spectral bandwidth vs. response duration; *D*: best ripple density vs. best temporal modulation frequency.

and the observed modulation preference (Fig. 16*A*). Neurons with broad spectral bandwidths (>1.5 octaves) responded only to low ripple densities (<0.5 cycles/octave), whereas neurons that responded to a limited range of frequencies (<1.5 octaves) responded over the entire range of measured best ripple densities (~0–2.1 cycles/octave). Likewise, the response duration also determined the number of temporal oscillations of the temporal receptive field profile (Fig. 16*B*). STRFs with short durations responded over the entire range of measured temporal modulation rates (~0–255 Hz) whereas neurons that had long-lasting temporal response profiles only exhibited slow temporal modulation rates (<50 Hz). This trend suggests that the number of excitatory and inhibitory subregions of the STRF is constrained by the receptive field bandwidth and duration, respectively. Such spectro-temporal tradeoffs in receptive field resolution and modulation filtering are consistent with a topographically distributed spectro-temporal tradeoff observed across the extent of the ICC isofrequency band lamina (Schreiner and Langner 1988b). Furthermore, such a tradeoffs may be important for the coding of natural sounds, which show a similar time-frequency tradeoff (Lewicki 2002; Theunissen et al. 2000).

Structure of visual versus auditory STRFs

Recent studies in the auditory system indicate that the structure of the auditory and visual STRFs exhibit similar time-varying structure (de Charms et al. 1998; Shamma 2001). These inferences are largely drawn from qualitative features of the auditory STRF, although the fine structure of auditory and visual STRFs has not been quantitatively compared. The Gabor STRF model provides a basis for comparing the structure of auditory STRFs directly with those obtained in the visual system using a set of nearly identical analytic equations (Adelson and Bergen 1985; Cai et al. 1997; DeAngelis et al. 1999; Jones and Palmer 1987; Watson and Ahumada 1985).

Comparing our results with those in the visual system reveals that auditory and visual STRFs are reasonably well described by a sum of time-frequency or time-space separable Gabor functions. As observed in the visual system (DeAngelis et al. 1999), error estimates (Fig. 6*D*) and similarity index (Fig. 6*C*) measurements confirm that most of the structure of auditory STRF is captured with as little as two independent time-frequency Gabor components. Furthermore, comparable percent errors observed for both visual (DeAngelis et al. 1999) and auditory STRFs indicate that the Gabor STRF model is equally well suited for describing auditory and visual receptive fields.

Aside from the faster temporal modulation preferences in the ICC, both visual and auditory temporal receptive field share several structural properties. Similar to visual receptive fields (Cai et al. 1997; DeAngelis et al. 1993a,b, 1999), the timing profile of auditory midbrain STRFs exhibit a distinct temporal asymmetry that is typified by a short rise time and long-lasting decay and requires time-warping function to achieve symmetry.

The spectral dimension of the auditory STRF is analogous to the spatial dimension of the visual STRF; however, the retinal sensory epithelium is a two-dimensional surface, whereas the primary sensory epithelium in the cochlea is unidimensional. When the spatial dimension of visual STRFs is collapsed along the direction of preferred orientation, visual and auditory STRF can be described by a nearly identical two-dimensional Gabor function (DeAngelis et al. 1999). Using this convention, the structure of auditory and visual STRFs is remarkably similar although the extents of their spectral and spatial structure are substantially different. In the visual system, the width of the Gabor-function defines the spatial extent over which the visual neurons integrate visual information, whereas the SRF bandwidth describes the extent of frequencies over which auditory neurons integrate sound information. In the auditory system, 1 octave corresponds to ~0.279 mm of receptor surface in the cochlea (Greenwood 1990). Therefore the observed range of bandwidths (0.14–4.8 octaves; mean ± SD = 0.987 ± 0.915 octaves) extended over 0.04–1.34 mm (mean ± SD = 0.275 ± 0.255 mm) of cochlear epithelium, which is broader than the range of spatial extents in VI receptive fields in the cat (~0.035–0.4 mm of retinal receptor surface); (Bishop et al. 1962; Tusa et al. 1978). Interestingly, the minimum sensory epithelium distance covered by both auditory and visual RFs is comparable in its extent (~0.04 vs. 0.035 mm).

Finally, the spectral phase of collicular neurons is largely limited to the range from 0 to 90°. Therefore the arrangement of excitation and inhibition appears to show similar relationships for the visual and auditory STRFs, in which excitation and inhibition can exhibit a variety of spectral alignments with respect to the center of the receptive field (Anzai et al. 1999). This structural property may enable ICC neurons to decipher spectral information about sounds with uniquely aligned spectral notches or resonances.

Binaural response preferences

Most binaural studies in the inferior colliculus focus on the analysis of interaural timing (ITD) and level (ILD) differences cues (e.g., Goldberg and Brown 1969; Irvine and Gago 1990; Kuwada et al. 1997). While such cues clearly contribute to

binaural phenomena, little is known about the converging spectro-temporal receptive field arrangements that contribute to binaural response integration and sound localization in the ICC.

By comparing the ipsilateral and contralateral receptive fields derived from simultaneously presented but statistically independent DMR stimuli to the two ears, we were able to characterize the structural properties of the converging spectro-temporal information. In $\sim \frac{1}{3}$ of the recorded neurons, STRFs for both ears could be obtained. Individually, the magnitude of the spectral and temporal similarity indices can be quite high (mean, 0.738 and 0.816, respectively), whereas the magnitude of the combined spectro-temporal binaural similarity index is typically much lower (mean = 0.513; paired t-test, $P < 0.001$). This disparity is partly accounted for by subtle spectral and temporal phase differences between the SRF or TRF profiles, thus resulting in STRF structures where the contra and ipsi excitatory and inhibitory subfields are spectro-temporally mismatched. Although, some of the reduction in the BSI is also caused by other STRF parameters that only showed a weak correlation (e.g., spectral bandwidth and response duration), the spectral and temporal phases likely provide the greatest contribution to this reduction (statistically uncorrelated aurally, $P > 0.92$ and $P > 0.26$, respectively). Other receptive field parameters, including the center frequency, peak latency, best ripple density, and the temporal modulation rate are significantly correlated. Thus although excitatory and inhibitory inputs to the ICC are aurally mismatched, their receptive fields are centrally overlapped with similar modulation preferences.

Although the magnitude of the spectral and temporal BSI determine the correspondence in shape of the contra- and ipsi-TRF and -SRF profiles, the sign of the BSI determines the relative alignment of excitation and inhibition. BSI values are clustered for negative and positive values, indicating that SRF and TRF profiles either exhibited a partly correlated or anti-correlated arrangement. The sign of the spectral, temporal, and spectro-temporal BSIs was conserved across all three metrics (Fig. 14A), and therefore, the specific relationship observed for the STRF (correlated/anticorrelated) was mutually preserved for the SRF and TRF profiles (spectral vs. spectro-temporal: $r = 0.915 \pm 0.076$, $P < 0.001$; temporal vs. spectro-temporal: $r = 0.853 \pm 0.099$, $P < 0.001$). In contrast, the magnitude of the spectral and temporal BSIs show no specific correlation (spectral vs. temporal: $r = -0.089 \pm 0.188$; $P > 0.5$), although the magnitude of the spectral and temporal BSIs individually contributed to the spectro-temporal BSI (spectral vs. spectro-temporal: $r = 0.670 \pm 0.140$, $P < 0.001$; temporal vs. spectro-temporal: $r = 0.531 \pm 0.160$, $P < 0.003$).

The binaural receptive field structure should, in theory, account for binaural response preferences of auditory neurons; however, the exact role of the binaural STRF needs to be more fully investigated. Specifically, how does the binaural receptive field structure contribute to sound localization and binaural phenomena? Because of the slow time course of the TRF profile (Fig. 9C), it is unlikely that STRF arrangements contribute to ITD sensitivities in the ICC (usually in the hundredths of microseconds range). Instead, the described receptive field arrangements likely contribute to ILD sensitivities and location-specific spectral filtering of broadband sound. The diversity and complexity of observed binaural STRF arrangements (e.g., Fig. 13) indicate that simple classification schemes based on the dominant excitatory or inhibitory receptive field contribution (Goldberg and Brown 1969) are too simplistic to fully account for the binaural preferences to dynamic broadband stimuli. Differences in the phase, bandwidth, and ripple density of the SRF structure could potentially be used to localize broadband sound sources that are highly susceptible to differentially filtered spectrum (Hartmann and Witternberg 1996; Kulkarini and Colburn 1998). Thus it is possible that interaural receptive field disparities are integrated at the colliculus and beyond to compute the spatial position of a sound source, analogous to the integration of binocular disparities in the primary visual cortex (Anzai et al. 1999).

As observed for visual cortex neurons we find that ICC STRFs share similar structural parameters binaurally although their spectral and temporal phases appear to be misaligned (Anzai et al. 1999); however, unlike visual receptive fields, we find no disparities in the central position of the STRF. The relevance of this finding for sound localization can be understood by noting that the binaural detection problem is fundamentally different from binocular fusion. In the visual system, external visual stimuli can project onto different spatial positions of the retinal epithelium. Deciphering the distance to a visual object requires that visual neurons analyze positional shifts in the contra and ipsi projecting images and subtle phase disparities in the local image structure. Sound localization, however, arises via differential filtering of the incoming signal spectrum by the listener's head and pinnae (Hartmann and Witternberg 1996; Kulkarini and Colburn 1998). This differential filtering modifies the frequency content of the incoming sound by superimposing binaurally misaligned spectral notches; yet, unlike the visual system, the sound's spectral content is never displaced along the cochlear epithelium. Binaural cues are, in this manner, interwoven with the frequency spectrum of the sound, which is relevant for determining the sound source content. Thus the observed similarities in the contra and ipsi STRFs (e.g., center frequency, ripple density, duration etc.) may be important for extracting information about the sound source content, whereas the misaligned receptive field phases may be necessary to decipher interaural disparities arising from the sound source position.

Recent studies have demonstrated that binaural STRFs account for much of the structure in spatial selectivity profiles of cortical neurons (Schnupp et al. 2001), and it is likely that the proposed interaural filtering mechanisms account for the observed spatial preferences. The wide assortments of binaural receptive field arrangements in the colliculus, thalamus, and primary auditory cortex (Miller et al. 2002) may therefore be necessary for the brain to efficiently compute and decipher differences in the incident spectrum, which arise through head shadowing and pinnae filtering and which depend on the sound source position. Furthermore, temporal differences in the contra- and ipsi-STRF structure may be necessary to dynamically track changes in the spectrum of a moving sound source. Such interaural filtering, along with the observed receptive field arrangements, may provide a basis for encoding binaural disparities in the source spectrum independently of contextual information in complex environmental stimuli.

REFERENCES

Adelson EH and Bergen JR. Spatiotemporal energy models for the perception of motion. *J Opt Soc Am* 2: 284–299, 1985.

Aersten AMHJ, Olders JHJ, and Johannesma PIM. 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, 1980.

Anzai A, Ohzawa I, and Freeman RD. Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *J Neurophysiol* 82: 874–890, 1999.

Bishop PO, Kozak W, and Vakkur GJ. Some quantitative aspects of the cat's eye: axis and plane of reference, visual field co-ordinates and optics. *J Physiol* 163: 466–502, 1962.

Bliss CI. *Statistics in Biology* New York: McGraw Hill, 1967.

Cai DQ, DeAngelis GC, and Freeman RD. Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. *J Neurophysiol* 78: 1045–1061, 1997.

Calhoun B and Schreiner CE. Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics. *J Euro Neurosci* 10: 926–940, 1998.

Daugman JG. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J Opt Soc Am A* 2: 1160–1169, 1985.

DeAngelis GC, Ghose GM, Ohzawa I, and Freeman RD. Functional microorganization of primary visual cortex: receptive field analysis of nearby neurons. *J Neurosci* 19: 4046–4064, 1999.

DeAngelis GC, Ohzawa I, and Freeman RD. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *J Neurophysiol* 69: 1091–1117, 1993a.

DeAngelis GC, Ohzawa I, and Freeman RD. 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, 1993b.

DeAngelis GC, Ohzawa I, and Freeman RD. Receptive-field dynamics in the central visual pathways. *Trends Neurosci* 18: 451–458, 1995.

deCharms RC, Blake DT, and Merzenich MM. Optimizing sound features for cortical neurons. *Science* 280: 1439–1443, 1998.

Depireux DA, Simon JZ, Klein DJ, and Shamma SA. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85: 1220–1234, 2001.

De Valois RL and Cottaris NP. Inputs to directionally selective simple cells in macaque striate cortex. *Proc Natl Acad Sci USA* 95: 14488–14493, 1998.

Efron B and Tibshirani RJ. *An Introduction to the Bootstrap.* New York: Chapman & Hall, 1993.

Escabí MA and Schreiner CE. Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *J Neurosci* 22: 4114–4131, 2002.

Goldberg JM and Brown PB. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J Neurophysiol* 32: 613–636, 1969.

Greenwood D. A cochlear frequency-position function for several species—29 years later. *J Acoust Soc Am* 87: 2592–2605, 1990.

Hartmann WM and Wittenberg A. On the externalization of sound images. *J Acoust Soc Am* 99: 3678–3688, 1996.

Heil P and Irvine DRF. First-spike timing of auditory-nerve fibers and comparison with auditory cortex. *J Neurophysiol* 78: 2438–2454, 1997.

Irvine DRF and Gago G. Binaural interaction in high-frequency neurons 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, 1990.

Jones JP and Palmer LA. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol* 58: 1233–1258, 1987a.

Jones JP and Palmer LA. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J Neurophysiol* 58: 1187–1211, 1987b.

Jones JP, Stepnoski A, and Palmer LA. The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *J Neurophysiol* 59: 1212–1232, 1987.

Joris PX and Yin TCT. Response to amplitude-modulated tones in the auditory nerve. *J Acoust Soc Am* 91: 215–232, 1992.

Klein DJ, Depireux DA, Simon JZ, and Shamma SA. Robust spectro-temporal reverse correlation for the auditory system: optimizing stimulus design. *J Comp Neurosci* 9: 85–111, 2000.

Kowalski N, Depireux DA, and Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J Neurophysiol* 76: 3503–3523, 1996.

Krishna BS and Semple MN. Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol* 84: 255–273, 2000.

Kulkarni A and Colburn HS. Role of spectral detail in sound-source localization. *Nature* 396: 747–749, 1998.

Kuwada S, Batra R, Yin TCT, Oliver DL, Haberly LB, and Stanford TR. Intracellular recordings in response to monaural and binaural stimulation of neurons in the inferior colliculus of the cat. *J Neurosci* 17: 1565–7581, 1997.

Langner G and Schreiner CE. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol* 60: 1799–1822, 1988.

Lewicki MS. Bayesian modeling and classification of neural signals. *Neural Comput* 6: 1005–1029, 1994.

Lewicki MS. Efficient coding of natural sounds. *Nat Neurosci* 5: 356–363, 2002.

Lu T, Liang L, and Wang X. Neural representation of temporally asymmetric stimuli in the auditory cortex of awake primates. *J Neurophysiol* 85: 2364–2380, 2001.

Marcelja S. Mathematical description of the response of simple cortical cells. *J Opt Soc Am A* 70: 1297–1300, 1980.

Miller LM, Escabí MA, Read HL, and Schreiner CE. Functional convergence of response properties in the auditory thalamocortical system. *Neuron* 32: 151–160, 2001.

Miller LM, Escabí MA, Read HL, and Schreiner CE. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol* 87: 516–527, 2002.

Nelken I, Kim PJ, and Young ED. 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, 1997.

Nelken I, Rotman Y, and Yosef OB. Responses of auditory-cortex neurons to structural features of natural sounds. *Nature* 37: 154–157, 1999.

Neuhoff JG. Perceptual bias for rising tones. *Nature* 395: 123–124, 1998.

Patterson RD. The sound of a sinusoid: spectral models. *J Acoust Soc Am* 96: 1409–1418, 1994.

Press WH, Teukolsky SA, Vetterling WT, and Flannery BP. *Numerical Recipes in C* (2nd ed.). Cambridge, UK: Cambridge University Press, 1995.

Ramachandran R, Davis KA, and May BJ. Single-unit responses in the inferior colliculus of decerebrate cats. I. Classification based on frequency response maps. *J Neurophysiol* 82: 152–163, 1999.

Rees A and Møller AR. Responses of neurons in the inferior colliculus of the rat to AM and FM tones. *Hear Res* 10: 301–330, 1983.

Reid RC, Soodak RE, and Shapley RM. Directional selectivity and spatio-temporal structure of receptive fields of simple cells in cat striate cortex. *J Neurophysiol* 66: 505–529, 1991.

Sen K, Theunissen FE, and Doupe AJ. Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol* 86: 1445–1458, 2001.

Shamma S. On the role of space and time in auditory processing. *Trends Cogn Sci* 5: 340–348, 2001.

Schnupp JWH, Mrsic-Flogel TD, and King AJ. Linear processing of spatial cues in primary auditory cortex. *Nature* 414: 200–204, 2001.

Schreiner CE and Calhoun BM. Spectral envelope coding in cat primary auditory cortex. *Aud Neurosci* 1: 39–61, 1994.

Schreiner CE and Langner G. Coding of temporal patterns in the central auditory system. In: *Auditory Function: Neurobiological Bases of Hearing.* New York: Wiley, 337–362, 1988a.

Schreiner CE and Langner G. Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* 60: 1823–1840, 1988b.

Theunissen FE, Sen K, and Doupe AJ. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315–2331, 2000.

Tusa RJ, Palmer LA, and Rosenquist AC. The retinotopic organization of Area 17 (striate cortex) in the cat. *J Comp Neurol* 177: 213–236, 1978.

Versnel H and Shamma SA. Spectral-ripple representation of steady-state vowels in primary auditory cortex. *J Acoust Soc Am* 103: 2502–2514, 1998.

Watson AB and Ahumada AJ. Model of human visual-motion sensing. *J Opt Soc Am A* 2: 322–342, 1985.