J Neurophysiol 103: 887–903, 2010.
First published December 16, 2009; doi:10.1152/jn.00813.2009.

Spectral and Temporal Modulation Tradeoff in the Inferior Colliculus

Francisco A. Rodríguez,[1] Heather L. Read,[1,2] and Monty A. Escabí[1-3]
[1]*Biomedical Engineering,* [2]*Psychology,* [3]*Electrical and Computer Engineering, University of Connecticut, Storrs, Connecticut*

Submitted 9 September 2009; accepted in final form 11 December 2009

Rodríguez FA, Read HL, Escabí MA. Spectral and temporal modulation tradeoff in the inferior colliculus. *J Neurophysiol* 103: 887–903, 2010. First published December 16, 2009; doi:10.1152/jn.00813.2009. The cochlea encodes sounds through frequency-selective channels that exhibit low-pass modulation sensitivity. Unlike the cochlea, neurons in the auditory midbrain are tuned for spectral and temporal modulations found in natural sounds, yet the role of this transformation is not known. We report a distinct tradeoff in modulation sensitivity and tuning that is topographically ordered within the central nucleus of the inferior colliculus (CNIC). Spectrotemporal receptive fields (STRFs) were obtained with 16-channel electrodes inserted orthogonal to the isofrequency lamina. Surprisingly, temporal and spectral characteristics exhibited an opposing relationship along the tonotopic axis. For low best frequencies (BFs), units were selective for fast temporal and broad spectral modulations. A systematic progression was observed toward slower temporal and finer spectral modulation sensitivity at high BF. This tradeoff was strongly reflected in the arrangement of excitation and inhibition and, consequently, in the modulation tuning characteristics. Comparisons with auditory nerve fibers show that these trends oppose the pattern imposed by the peripheral filters. These results suggest that spectrotemporal preferences are reordered within the tonotopic axis of the CNIC. This topographic organization has profound implications for the coding of spectrotemporal features in natural sounds and could underlie a number of perceptual phenomena.

INTRODUCTION

The frequency decomposition performed by the cochlea is the principle that gives rise to tonotopic representation in the central auditory system. Beyond the cochlea, sounds are further decomposed into elementary spectral and temporal acoustic features. Spectral and temporal cues are critical for auditory percepts of pitch and timbre (Goldstein 1973; Irino and Patterson 1996; Plomp 1983; Schouten 1940; van-Veen and Houtgast 1985). Furthermore, natural sounds and speech can all be characterized by amplitude modulations of the signal over frequency and time (Chi et al. 1999; Singh and Theunissen 2003). Such spectrotemporal modulations are essential for speech recognition and for discriminating and detecting biologically relevant sounds (Chi et al. 1999). How neural ensembles represent such information in the central auditory system, however, is unknown.

The central nucleus of the inferior colliculus (CNIC) is well poised for processing spectrotemporal features in natural sounds. Its role as a key integrative center is well supported by its highly organized mosaic of laminar inputs (Malmierca et al. 1993; Oliver 2000). The CNIC receives convergent inputs from the cochlear nucleus, superior olivary complex, and the nucleus of the lateral lemniscus (Adams 1979; Malmierca et al. 2005; Oliver 2000), all of which have unique aural, temporal, and spectral response properties (Batra and Fitzpatrick 1999; Batra et al. 1989; Joris 1996; Kuwada et al. 2006; Rhode 1995; Spirou and Young 1991). Although numerous studies have characterized response properties within the CNIC, the relationship between the function and the underlying organization of this structure is poorly understood. A current hypothesis is that spectral and temporal receptive field preferences are systematically organized within the CNIC frequency lamina (Ehret et al. 2003; Langner et al. 2002; Schreiner and Langner 1988). Presently it is not clear whether receptive field preferences are systematically organized across other dimensions of the CNIC, including the principal tonotopic axis.

There is no concise model for how spectral and temporal modulations are processed and transformed within the CNIC. The auditory nerve (AN) is capable of following temporal modulations that exceed 1 kHz and yet the cortex can only reliably follow at 25 Hz (Joris et al. 2004). Within the AN, fibers are low-pass tuned for temporal modulations, and following rates tend to increase with characteristic frequency (Joris and Yin 1992). A simple hypothesis is that the CNIC inherits the response properties of AN fibers. However, this prediction is not supported by the finding that many CNIC neurons show band-pass functions for temporal modulations and temporal following rates are limited to ~300 Hz (Krishna and Semple 2000; Langner et al. 2002; Rees and Møller 1983; Rees and Palmer 1989). Langner and Schreiner (1988) found that best modulation frequencies increase with increasing best frequency (BF) analogous to the AN. However, others have not observed this relationship.

Neurons in the CNIC are also selective for spectral modulations that are present in natural sounds and speech (Singh and Theunissen 2003). The spectrum of natural sounds can contain peaks and notches such that the amplitude of the sound is modulated as a function of the sound frequency. Spectral modulation sensitivity has been tested in the IC with spectral ripple sounds (Escabí and Schreiner 2002) and a variety of broadband noises (Keller and Takahashi 2000; Lesica and Grothe 2008; Yu and Young 2000). CNIC units can exhibit tuned responses to spectral modulations with selectivity for spectral modulation extending up to ~3 cycle/octave (Qiu et al. 2003).

We tested whether spectrotemporal modulation preferences are ordered along the tonotopic axis of the CNIC. Based on the response pattern of AN fibers (Joris and Yin 1992; Kim and Young 1994) and CNIC neurons (Langner and Schreiner 1988) to AM signals, we hypothesized that temporal following rates would increase with increasing BF. Contrary to this expectation, we find that the Spectrotemporal receptive fields (STRFs) structure varied systematically from fast receptive fields for low BFs to substantially slower receptive fields for high BFs.

Address for reprint requests and other correspondence: M. A. Escabí, University of Connecticut Electrical, and Computer Engineering, 371 Fairfield Rd., Unit 2157, Storrs, CT 06269×1157 (E-mail: escabi@engr.uconn.edu).

Surprisingly, spectral modulation tuning preferences and the arrangement and strength of inhibition also covaried with these changes, suggesting a central basis for this tradeoff. Finally, we show that these trends differ substantially from AN fibers, indicating that spectral and temporal modulation preferences are dramatically transformed within the CNIC.

METHODS

Surgical procedure

Animals ($n = 6$) were housed and handled according to approved procedures by the University of Connecticut Animal Care and Use Committee and in accordance with National Institutes of Health and American Veterinary Medical Association guidelines. All efforts were made and alternatives were considered to minimize the number of animals in the study. Experiments were performed in an acute recording setting (48–72 h). The cat was initially anesthetized with a mixture of ketamine (10 mg/kg) and acepromazine (0.28 mg/kg im). A tracheotomy was performed to ensure adequate ventilation and reduce the nasal cavity acoustic noise. Afterward a craniotomy was performed under sodium pentobarbital (30 mg/kg) anesthesia. The overlying tissue in the occipital lobe and the bony tentorium were removed to expose the inferior colliculus (IC). Following surgery, the animal was maintained in a nonreflexive state of anesthesia by a continuous infusion of ketamine (2 $mg \cdot kg^{-1} \cdot h^{-1}$) and diazepam (3 $mg \cdot kg^{-1} \cdot h^{-1}$) in a lactated Ringer solution (4 $mg \cdot kg^{-1} \cdot h^{-1}$). Heart rate, temperature, breathing, and reflexes were monitored and used as physiological criteria for adjusting the anesthetic infusion rate.

Acoustic stimuli and delivery

Sounds were presented dichotically to the animal in a sound-shielded chamber (IAC, Bronx, NY) via a closed binaural speaker system attached to hollow ear-bars (Kopf Instruments, Tujunga, CA). The delivery system was calibrated for frequencies in the range of 1–47 kHz (±3 dB SPL) with a finite impulse response (FIR) inverse filter that was implemented on a TDT RX6 digital processor (TDT, Alchua, FL). Sounds were delivered with either a TDT RX6 or a RME DIGI 9652 sound card (Haimhausen, Germany) through electrostatic speaker drivers (TDT EC1).

At each recording location we first presented a sequence of pure tones to measure the frequency response areas (FRAs) of each recording site. FRAs were calculated during the experiment as a way to verify the CNIC tonotopic gradient (Merzenich and Reid 1974; Semple and Aitkin 1979). The stimuli consisted of a random sequence of pure tones (50-ms duration delivered every 300 ms with a 5-ms rise-time) spanning the range from 1.4 to 44.2 kHz (in 1/8-octave steps) and sound pressure levels from 5 to 85 dB SPL (in 10-dB steps).

Dynamic moving ripple (DMR) (Escabí and Schreiner 2002) was used to measure spectrotemporal preferences of CNIC neurons. Perceptually the DMR can quickly change its acoustic composition so as to elicit a number of perceptual qualities. The sound can undergo fast changes in spectral and temporal composition where the pitch and timbre of the sound is perceived to change dynamically with time. Similarly, the sound will also produce the percept of rhythmic patterns or FM sweeps and can quickly change its temporal properties to produce a more salient pitch quality. The DMR was generated digitally using a sampling rate of 96 kHz and 24-bit resolution. A continuous 10-min DMR was presented twice (20 min total) at 80 dB SPL (65 dB spectrum level per 1/3 octave). The DMR is a continuous time-varying broadband sound that covered a frequency range from 1 to 48 kHz and probed spectrotemporal preferences. At a fixed instant in time, the spectrum of the DMR is represented by a sinusoidal AM across the frequency axis were the density of peaks and troughs is determined by the spectral modulation frequency (SMF) parameter (cycle/octave). Temporally, the DMR sound contains amplitude modulations that are controlled by the temporal modulation frequency (TMF, Hz) parameter. The peak-to-peak amplitude of the DMR was set to 30 dB because this produces robust responses for the vast majority of IC neurons (Escabi et al. 2003). Both parameters (SMF and TMF) were varied randomly and continuously at a maximum rate of change of 0.25 Hz (SMF) and 1 Hz (TMF). SMF and TMF were varied independently of one another so that they fully probe the parameter space (0–4 cycle/octave, SMF, and 0–500 Hz, TMF) in an unbiased manner. During the 20-min presentation of the DMR, each ear was stimulated with an independent DMR. This allowed us to concurrently measure independent STRFs for the contralateral and ipsilateral ears (Qiu et al. 2003). For the purpose of this study, only the STRF for the contralateral ear are considered as it characterizes the dominant phase locked response of CNIC neurons (Qiu et al. 2003).

Electrophysiology

Customized acute 16-channel recording probes (NeuroNexus Technologies, Ann Arbor, MI) with 150-μm electrode separation and 177-μm^2 gold-plated contact area were used for the neural recordings. These electrodes were specifically selected for high impedance (3–5 MΩ at 1 kHz) across all of the 16 recording electrodes to enhance isolation characteristics. The probes were first positioned on the surface of the IC using a stereotaxic frame (Kopf Instruments). The probe shank was then inserted into the IC orthogonal to the isofrequency-band laminae (inserted at ~30° relative to the sagittal plane) (Schreiner and Langner 1997) with a LSS 6000 Inchiworm Microdive (Burleigh EXFO; Vanier, Quebec, Canada). Neuronal responses were digitized and recorded into a hard drive for off-line analysis with a RX5 Pentusa Base station (TDT). At each penetration site, a sequence of tones and a DMR (for STRF) were presented to the animal. Electrodes were repositioned by moving the probe ~300 μm along the IC surface either along the dorsomedial to ventrolateral axis or the rostrocaudal axis. Deviations from a prescribed stereotaxic grid were occasionally necessary to avoid the surface vasculature. The use of a 16-channel probe at each penetration allowed us to systematically record neural activity across principal frequency axis while uniformly covering the three-dimensional space of the CNIC in a grid like pattern.

STRF analysis

The neural responses to the DMR were used to compute a contralateral STRF at each recording location using a spike-triggered averaging procedure (Escabí and Schreiner 2002). Significance testing was performed against a noise STRF that was derived for a random firing neuron of identical spike rate. A two-tailed significance test was performed by requiring that positive and negative STRF values exceed 3.09 SD of the noise STRF. This strict criterion guarantees that we detect STRF components at a significance level of $P < 0.002$ ($P < 0.001$ for excitation and $P < 0.001$ for inhibition). As described in detail elsewhere, the color spectrum in all plots indicates spike rate relative to the mean such that blue and red denote decrease or increase below and above the mean, respectively (Escabí and Schreiner 2002). We refer to the blue domains in the STRF as inhibitory even though the reduction in spike rate could arise theoretically from synaptic inhibition or refractoriness at or below the level of the CNIC (see DISCUSSION).

Neural data selection

Neuronal recordings consisted of small multiunit clusters and single units from 4,656 recording locations in six animals. During the experiment recordings, we developed a conservative approach for selecting spike waveforms that would be representative of small multiunit clusters and would allow us to isolate single neurons for select recording sites. To select a "small" number of units from each

recording site, spike threshold discriminators were adjusted manually for each of the 16 channels prior to recording so as to achieve spike rates of ~5–20 spike/s. During this process, we had direct feedback from the discriminator for the instantaneous and averaged spike rate for each of the 16 sites (median spike rate: 7.5 spike/s). With the high-impedance probes used in this study, distinct action potential waveforms from multiunit clusters could readily be identified visually on roughly half of the recording sites. On average, we were able to isolate two single units across the 16 channels per recording session.

To assure that recording locations were confined to the CNIC proper, we first examined the frequency organization for each recording site. Recording sites that did not follow a consistent tonotopic gradient as in Fig. 1, *C* and *D,* were assumed to lie outside the CNIC and not included in the analysis (e.g., *F*). This initial pruning of the data resulted in 3,262 recording locations (2,917 multiunits; 345 single units). We next selected sites between 2 and 32 kHz (316 single units; 2,711 multiunits). The dataset was next truncated to assure that each sited exhibited a clean and well-defined STRF. We did this by computing the phase-locking index (PLI) for each unit and requiring that the PLI >0.08 (Escabí and Schreiner 2002). We have previously demonstrated that this criteria removes >99% of units with poor phase locking in the CNIC (Escabí and Schreiner 2002). This requirement assures that the parameter estimates of each unit and the overall population distribution was not corrupted by measurement noise. This final refinement of the data resulted in 2,249 multiunits and 285 single units.

Ripple transfer function analysis

The modulation preferences of each unit were characterized directly from the ripple transfer function (RTF) (Escabí and Schreiner 2002). The RTF was obtained by performing a two-dimensional Fourier transform ($\Im_2\{\cdot\}$) on the significant STRF (significant probability $P < 0.002$) and subsequently computing the magnitude (Escabí and Schreiner 2002)

$$\mathrm{RTF}(F_m, \Omega) = |\Im_2\{\mathrm{STRF}(t, x)\}| \quad (1)$$

Here t is the time axis, $x = \log_2(f/f_0)$ is the frequency axis in octaves, $f_0 = 1$ kHz represents the minimum frequency of the DMR sound, F_m is the TMF, and Ω is the SMF. The spectral and temporal modulation transfer functions (sMTF and tMTF) were then obtained by collapsing the RTF along the spectral and temporal dimensions, respectively

$$\mathrm{sMTF}(\Omega) = \int \mathrm{RTF}(F_m,\Omega)\mathrm{d}F_m \quad (2A)$$

$$\mathrm{tMTF}(F_m) = \int \mathrm{RTF}(F_m,\Omega)\mathrm{d}\Omega \quad (2B)$$

and subsequently normalizing for a maximum gain of 1 (Miller et al. 2002). The peak responses from the sMTF and tMTF characterized the best spectral (bSMF) and the best TMF (bTMF), respectively. The spectral and temporal modulation upper cutoff frequencies were defined by the modulation frequency where the gain of the tMTF or sMTF were reduced by 3 dB or equivalently to 50% of the maximum power (Fig. 2, *B* and *D*). Neuronal responses were classified as either band-pass or low-pass tuned response pattern for both the spectral and temporal modulation dimensions (Qiu et al. 2003). To do this, the lower 3-dB cutoff was measured for the sMTF and tMTF independently. In both cases, the selectivity was classified as band-pass whenever the value of the lower 3-dB cutoff was greater than zero. Alternately, the sMTF or tMTF were classified as low-pass if the magnitude at zero TMF or SMF did not exceed the 3-dB criterion.

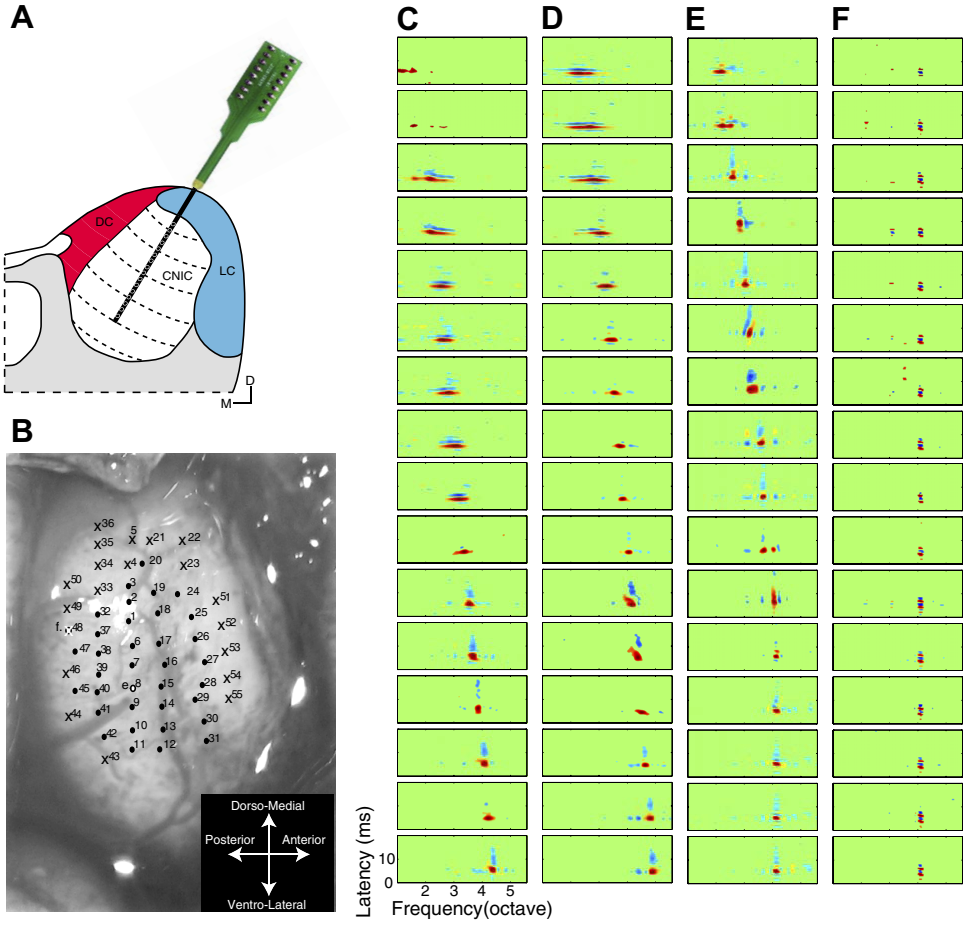

FIG. 1. Mapping spectrotemporal receptive field preferences within the central nucleus of the inferior colliculus (CNIC). *A*: illustration showing the placement of sixteen-channel recording probes into the CNIC. Probes were inserted at 30° relative to the sagittal plane and approximately orthogonal to the CNIC isofrequency lamina (---). The dorsal cortex (DC, red) and lateral cortex (LC, blue) of the IC are shown for reference. *B*: surface view of the IC from 1 experiment with the probe penetration sites superimposed. Penetration sites were selected to provide a uniform coverage of the IC in a grid-like pattern with ~300 μm along the anterior-posterior and mediolateral aspect. Sites that did not exhibit a definitive best frequency gradient are marked (×). These recording locations were presumed to be outside the central nucleus of the IC and were discarded from the analysis. *C–F*: example STRFs from penetrations within the IC of 4 separate experiments. Penetrations for *E* and *F* are from the experiment in *B* (sites 8 and 48, respectively). Recording sites within the CNIC exhibit a distinct tonotopic gradient with increasing penetration depth (*C–E*). Site *F* lacks this tonotopic organization. This site was presumed to lie outside the CNIC and was discarded from the analysis.

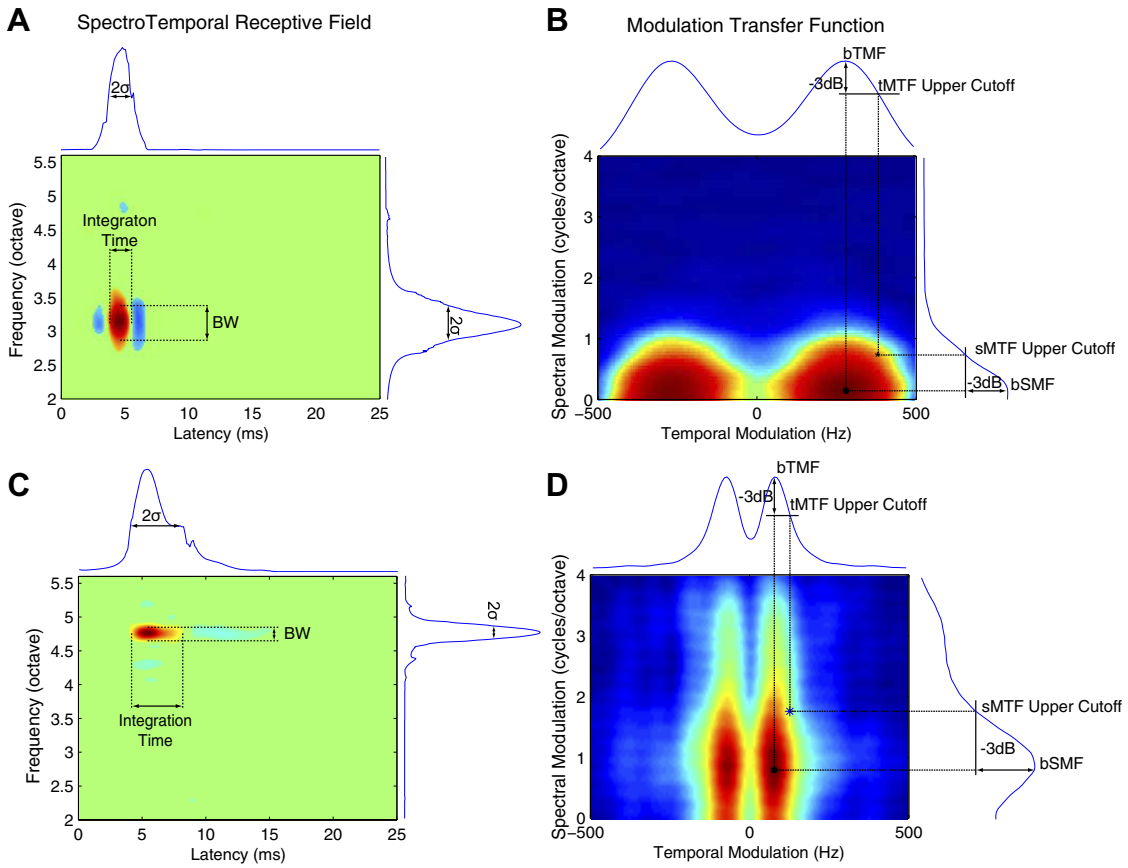

FIG. 2. Quantitative analysis of spectrotemporal and modulation tuning preferences. *A* and *C*: spectrotemporal receptive fields (STRFs) and the corresponding ripple transfer function (RTF, *B* and *D*) of 2 units. A unit with short STRF integration time that lacks sideband inhibition is shown in *A*. Temporally, this unit exhibits a fast OFF-ON-OFF pattern suggesting selectivity to fast temporal modulations. In contrast, the unit of *C* has longer integration time, narrower bandwidth, and significant spectral sideband and temporal inhibition. Only the significant portion of the STRF is shown ($P < 0.002$, see METHODS). The spectral and temporal marginals are obtained for each STRF by projecting the STRF along the spectral and temporal dimensions (blue lines, see METHODS). The receptive field bandwidth and integration time are defined as twice the SD of the spectral and temporal marginals, respectively. *B* and *D*: ripple transfer functions for the units of *A* and *C* are obtained by Fourier transforming the STRF (Escabí and Schreiner 2002). The latency axis of the STRF is now represented by the temporal modulation frequency (Hz), whereas the spectral axis is represented by the spectral modulation frequency (cycle/octave). The unit of *B*, same as *A*, is characterized by a band-pass tuned temporal response at ~265 Hz (bTMF) and low-pass spectral response with peak at 0.1 cycle/octave (bSMF; noted by black dot). The high bTMF and low bSMF for this unit are consistent with the broad but fast OFF-ON-OFF pattern observed in its STRF. The 3-dB spectral and temporal upper cutoffs (tMTF and sMTF upper cutoff, shown by asterisk) for this unit lie at 0.7 cycle/octave and 408 Hz (defined as the maximum frequencies for which RTF power exceeds 0.5 of the peak response power), consistent with the observed broad spectral bandwidth and fast temporal response for this unit. *D*: the RTF for the unit of *C* is characterized by a weakly tuned pattern about a bSMF of 1.6 cycle/octave, bTMF of 70.3 Hz, and 3-dB spectral and temporal upper cutoffs of 3.4 cycle/octave and 112.3 Hz. This is consistent with the slow temporal pattern, narrow bandwidth, and pronounced sideband inhibition as observed for this unit's STRF (*C*).

Finally, we also measured the spectral and temporal DC gains. These were defined as the sMTF and tMTF gains (in dB) at $\Omega = 0$ cycle/octave and $F_m = 0$ Hz, respectively.

STRF temporal and spectral resolution

We measured the temporal and spectral resolution of each unit by considering the integration area of the STRF filters. According to the uncertainty principle, the spectral and temporal resolution of a filter can be derived by considering the spectral and temporal power distributions of a filter (i.e., the power marginals) and measuring the average spread (i.e., the SD) across each of these dimensions (Cohen 1995; Gabor 1946). To implement this for our CNIC units, we first extracted the STRF envelope to define a time-frequency power distribution for the STRF filter. This power distribution was defined by the squared magnitude of the analytic STRF

$$P(t, x) = |\mathrm{STRF}(t, x) + j \cdot H\{\mathrm{STRF}(t, x)\}|^2 \quad (3)$$

where $H\{\cdot\}$ is the Hilbert transform and $j = \sqrt{-1}$. The spectral and temporal marginals were obtained by collapsing $P(t,x)$ along the temporal and spectral dimensions and normalizing for unit area, respectively

$$P_x(x) = \int P(t, x)\mathrm{d}t / \int\int P(t, x)\mathrm{d}t\mathrm{d}x \quad (4A)$$

$$p_t(t) = \int P(t, x)\mathrm{d}x / \int\int P(t, x)\mathrm{d}t\mathrm{d}x \quad (4B)$$

We have previously shown that the spectral and temporal marginals can be used to define the extent of the integration area of a CNIC unit as well as several functionally relevant parameters (Qiu et al. 2003). Note that the marginals do not contain inhibitory or suppressive STRF domains and have been normalized as a power distribution. The peaks of these distributions defined the STRF peak latency (PL) and BF. The STRF integration time (Δt) and bandwidth (Δf) were defined as twice the SD of the spectral and temporal marginals

$$\Delta t = 2 \times \sqrt{\int (t - \bar{t})^2 p_t(t)\mathrm{d}t} \quad (5A)$$

$$\Delta f = 2 \times \sqrt{\int (f - \bar{f})^2 p_f(f)\mathrm{d}f} \quad (5B)$$

where $\bar{f}$ and $\bar{t}$ represent the centroid values of the spectral and temporal marginals, respectively, and f the frequency variable in hertz. Note that spectral marginal of *Eq. 4A*, $p_x(x)$, was transformed into $p_f(f)$ using $x = \log_2(f/f_0)$ prior to computing the absolute bandwidth (in hertz). The uncertainty principle requires that the spectrotemporal resolution product obey: $\Delta t \cdot \Delta f > 1/\pi$ (Cohen 1995; Gabor 1946). Finally we also measured spectral bandwidths in octaves using an analogous equation

$$\Delta x = 2 \times \sqrt{\int (x - \bar{x})^2 p_x(x) \mathrm{d}x} \tag{5C}$$

Population frequency analysis

To determine how and if spectral and temporal preferences vary with BF, we first subdivided the data according the BF of each unit. The data were subdivided either into 1- or ½-octave intervals for BFs between 2 and 32 kHz. For each of the selected BF intervals, we computed the population RTF by averaging the RTFs of individual units that lie within each of the ½-octave intervals. Prior to averaging, the RTF of each unit was normalized for unit area so that spike rates do not bias the results. Statistical analysis was then performed as a function of BF (using 1-octave intervals) for both STRF (integration time and bandwidth) and RTF parameters (bTMF, bSMF, 3-dB temporal upper cutoff, and 3-dB spectral upper cutoff, etc.). All of the response parameters showed skewing and differed substantially from a normal distribution (χ^2 test, $P < 0.05$). We therefore employed the median as a statistical metric with Bonferroni correction for the number of comparisons.

Analysis of excitation and inhibition

To characterize how functional inhibition contributes to modulation tuning and the observed trends, we quantitatively examined the excitatory and inhibitory STRF components. The STRF of each unit was first decomposed into its excitatory (STRFe) and inhibitory (STRFi) components (Fig. 8) by selecting the positive or negative STRF values, respectively. Following this decomposition, we re-estimated all of the previously described STRF parameters, although these were now computed separately for the STRFe and STRFi. These parameters included the excitatory and inhibitory BF (BFe, BFi), bandwidths (BWe and BWi), and integration time (Δt_e and Δt_i). The relationship between all of the STRF parameters was then examined as a function of strength of inhibition and BF.

To quantify the amount of inhibition present in the STRF of a particular unit, we first computed the STRF power (Escabí and Schreiner 2002) associated with the STRFe and STRFi components. The excitatory power was derived from STRFe as

$$\sigma_e^2 = \int\int \mathrm{STRF}_e(\tau, x)^2 \mathrm{d}\tau \mathrm{d}x \tag{6A}$$

and the inhibitory power was derived from the STRFi as

$$\sigma_i^2 = \int\int \mathrm{STRF}_i(\tau, x)^2 \mathrm{d}\tau \mathrm{d}x \tag{6B}$$

The fractional inhibitory power in the STRF was then quantified as the inhibitory-to-excitatory power ratio (IER)

$$\mathrm{IER} = \begin{cases} \dfrac{\sigma_i^2}{\sigma_e^2} & \sigma_e^2 \geq \sigma_i^2 \\ \dfrac{\sigma_e^2}{\sigma_i^2} & \sigma_e^2 < \sigma_i^2 \end{cases} \tag{7}$$

This metric quantifies the relative strength between excitation and inhibition in the STRF. A value of 0 indicates that the STRF was either purely excitatory or purely inhibitory. In contrast, a value of 1 indicates that the strength of inhibition is equal to the strength of excitation. Although the IER metric cannot distinguish whether STRF is dominated by excitatory ($\sigma_e^2 > \sigma_i^2$) or inhibitory ($\sigma_e^2 < \sigma_i^2$) power, for the vast majority of units the excitatory power dominated ($\sigma_e^2 > \sigma_i^2$, 2,522 of 2,534 units; $\sigma_e^2 < \sigma_i^2$,12 units).

STRF Gabor-alpha model

To further characterize the relationship between functional inhibition and modulation tuning, we developed a mathematical model of the STRF in which the strength and shape of excitation and inhibition could be systematically and independently adjusted along the temporal and spectral dimensions of the STRF. This model is analogous to the separable spectrotemporal Gabor model we have employed previously to characterize CNIC STRFs (Qiu et al. 2003). The primary difference is that the envelope of the temporal receptive consists of an alpha function, whereas previously we employed a Gaussian temporal envelope. The separable model STRF is expressed as

$$\mathrm{STRF}_m(t, x) = W(t) \cdot G(x) \tag{8A}$$

where

$$G(x) = e^{-[2(x - x_0)/\mathrm{BW}]^2} \cos[2\pi \cdot \Omega_0 \cdot (x - x_0) + P] \tag{8B}$$

is the Gabor spectral receptive field profile. The spectral parameters of the model include the BF in octaves (x_0), the receptive field bandwidth (BW, octave), the best SMF (Ω_0, cycle/octave), and the spectral phase (P). The temporal component of the STRF is expressed as the product of an alpha function and a cosine waveform

$$W(t) = E(t) \cdot \cos[2\pi \cdot F_{m0} \cdot (t - t_0) + Q] \tag{8C}$$

where the temporal envelope is given by the piecewise alpha function

$$E(t) = \begin{cases} 0 & 0 < t \leq t_0 \\ K \cdot \dfrac{t - t_0}{\tau_1 \cdot e^{-1}} \cdot e^{-(t - t_0)/\tau_1} & t > t_0 \;\&\; t \leq \tau_1 + t, \\ K \cdot \dfrac{t - t_0 + \tau_2 - \tau_1}{\tau_2 \cdot e^{-1}} \cdot e^{-(t - t_0 + \tau_2 - \tau_1)/\tau_2} & t > \tau_1 + t_0 \end{cases} \tag{8D}$$

Here the temporal parameters of the model include the response latency (t_0), the receptive field rise (τ_1) and decay times (τ_2), the best TMF (F_{m0}, Hz), and the temporal phase (Q). Two separate simulations were performed to examine the effects of temporal and spectral inhibition in the STRF.

First we varied the amount of temporal inhibition in the STRF independently of the spectral receptive field characteristics (Fig. 7, *A–C*). The model simulation was performed by fixing all parameters with the exception of the receptive field decay time (τ_2), which was varied between 0.5 and 7 ms. For short decay times, the STRF composition for this model consists strictly of a brief excitatory response. As the decay time increases, a STRFi component becomes apparent, and this component becomes more prominent with increasing decay time (Fig. 7*A*). The remaining spectral ($x_0 = 0$ octave, $\Omega_0 = 0$ cycle/octave, BW = 1 octave, $Q = 0$) and temporal ($t_0 = 5$ ms, $F_{m0} = 100$ Hz, $\tau_1 = 2$ ms, $q = \pi/6$) parameters were fixed and chosen to lie within the observed physiological range for a typical neuron with 100-Hz bTMF (Qiu et al. 2003).

A second simulation was performed in which we characterized the role of spectral inhibition (Fig. 7, *D–F*). For this simulation, the proportion of spectral inhibition was varied independently of the STRF temporal characteristics by varying the receptive field bandwidth (BW = 0.3 – 1.5 octave). For short bandwidths, the STRF model compositions consists of a temporally brief and spectrally narrow excitatory response with no temporal or spectral inhibition. As the bandwidth is increased, sideband inhibition is recruited and becomes more pronounced, although the temporal characteristics remain unchanged (Fig. 7*D*). For this simu-

lation, the remaining spectral ($x_0 = 0$ octave, $\Omega_0 = 1$ cycle/octave, $Q = 0$) and temporal ($t_0 = 5$ ms, $f_{m0} = 0$ Hz, $\tau_1 = 1$ ms, $\tau_2 = 2$ ms, $q = \pi/6$) parameters were fixed and chosen to lie within the observed physiological range for a typical temporal low-pass neuron with 0-Hz bTMF (Qiu et al. 2003).

Population STRFe and STRFi

To characterize the structure of the excitatory and inhibitory receptive field domains, we computed the population averaged STRFe and STRFi for different BF intervals (BF range = 2–4, 4–8, 8–16, and 16–32 kHz). To do this, it was necessary to account for differences in latency and BF of each unit otherwise the population STRFe and STRFi would exhibit a diffused pattern that does not properly reflect the sound integration dynamics. Prior to averaging, we therefore aligned the STRFe and STRFi by subtracting the peak latency and BF from each unit (aligned to 0 octave and 0 ms). This procedure effectively aligned the STRFs to the region of maximum amplitude prior to averaging.

AN data analysis

Spectrotemporal resolution estimates were derived for data from the AN provided by Joris and Yin (1992) and Kim and Young (1994) to compare with our CNIC results. In the data from Joris and Yin, AN filter impulse responses were not available that would allow us to directly estimate Δf and Δt. Estimates were therefore derived from the FRAs and MTFs to sinusoidal amplitude modulated tones. We used the FRA bandwidths measured 10 dB above threshold to approximate Δf. The temporal resolution, Δt, was approximated as half the period of the fastest temporal modulation that the unit could follow (3 dB upper cutoff frequency from the MTFs). The Kim and Young data consisted of STRFs measured using Gaussian noise. In this case, estimates of the temporal (Δt) and spectral resolution (Δf) were obtained directly by applying *Eqs. 4* and *5* as for the CNIC data.

RESULTS

STRFs were measured in the cat CNIC with a 16-channel probe inserted orthogonal to the isofrequency lamina (Fig. 1*A*). The location of each penetration on the dorsal surface of the IC is shown for one experiment in Fig. 1*B*. In Fig. 1, *C–E,* are the STRFs from representative penetrations. In all cases, a defining low- to high-frequency gradient is observed, consistent with the placement of the multichannel electrode within the CNIC at an orientation parallel to the tonotopic axis (Merzenich and Reid 1974; Semple and Aitkin 1979) and orthogonal to the isofrequency lamina (Schreiner and Langner 1997). The responses in Fig. 1*F* are from a penetration site along the posterior edge of the IC in which all of the STRFs exhibited similar BFs. Such constant BF tuning is consistent with the placement of the probes in the lateral cortex (LC) of the IC where the lamina have been shown to course orthogonal to the CNIC (Loftus et al. 2008). To limit our recordings to the CNIC proper, all recording tracks that did not exhibit a consistent BF gradient as displayed for Fig. 1, *C–E,* were discarded from the analysis (sites marked by *x* in Fig. 1*B*). Thus the selected recording locations all exhibit a monotonic low to high BF gradient with electrode position as expected for the CNIC.

Spectral and temporal modulation preferences vary with tonotopic position

Spectral and temporal response properties in the CNIC appeared to vary systematically with tonotopic position (Fig. 1, *C–E*). At first glance, it appeared that low-frequency sites were often characterized by short integration times and spectrally broad STRFs, whereas those at higher frequencies often had longer integration times and were spectrally narrower (Fig. 1, *C–E*). We therefore sought to quantify the relationship between spectral and temporal preferences along the depth axis of the CNIC.

To characterize the relationship between spectral and temporal response preferences and BF, we extracted several parameters directly from the STRF and the corresponding (RTF; Figs. 2 and 3) (Miller et al. 2002). The receptive field integration resolution was characterized by measuring the STRF integration time and bandwidth (in octaves) as illustrated for two units in Fig. 2, *A* and *C*. These were defined as the average widths of the spectral and temporal marginals (Fig. 2, *A* and *C,* blue contours; see METHODS). For each unit, we computed the temporal and spectral MTF (tMTF and sMTF; Fig. 2, *B* and *D,* blue contours) and estimated the best spectral (bSMF) and best temporal (bTMF) modulation frequencies, and spectral and temporal 3-dB cutoff frequencies (see METHODS). The bTMF and bSMF define the stimulus modulations that produce maximal phase-locked response, whereas the 3-dB cutoff frequencies provide an upper response limit where the phase-locked response power has been reduced by 50% of the maximum response. Based on the derived functional parameters, the unit of Fig. 2*A* exhibits relatively broad spectral (bandwidth = 0.51 octaves) and high temporal resolution (STRF integration time = 1.69 ms). Spectrally this unit has a broad "ON" receptive field and brief "OFF-ON-OFF" temporal receptive field pattern. Accordingly, this unit exhibits low-pass selectivity for spectral and band-pass selectivity for temporal modulations, respectively (Fig. 2*B,* bSMF = 0.1 cycle/octave; bTMF = 256 Hz). In all instances, low-pass modulation tuning is observed for STRFs that lacked spectral or temporal inhibition (Miller et al. 2002; Qiu et al. 2003). The unit of Fig. 2*C,* by comparison, exhibits narrower spectral (bandwidth = 0.17 octave) and lower temporal resolution (4.97-ms integration time). This unit exhibits a noticeable amount of sideband inhibition around its BF (14.88 kHz), and subsequently its sMTF is tuned with bSMF = 0.8 cycle/octave. Similarly, this unit exhibits a strong inhibitory/suppressive temporal pattern and is therefore tuned along the temporal modulation dimension (bTMF = 70.3 Hz). A variety of complementary STRF arrangements and spectrally and/or temporally tuned RTF patterns are observed across the population of recording sites (Fig. 3).

The relationship between spectral and temporal modulation preferences and resolution formed a continuum from high to low resolution in each domain for the population of CNIC sites examined (Fig. 3*A*). At one extreme, units exhibit narrow spectral tuning, strong sideband inhibition, and band-pass sMTF (Fig. 3, *B–D*). At the opposite extreme, units could exhibit broad spectral tuning, lack of sideband inhibition and low-pass sMTF (Fig. 3, *H* and *I*). Units with strong temporal inhibition exhibit band-pass tMTF (Fig. 3, *D–I*), whereas those with weak or absent temporal inhibition are characterized by low-pass tMTF (Fig. 3, *B* and *C*). Such spectral and temporal preferences systematically traded off and formed a continuum (Fig. 3*A*). Figure 3*A* illustrates the relationship between bSMF and bTMF (285 single units, ◍; 2,249 multiunits, ●). At the extremes, units with fast temporal responses and thus high bTMF typically have broad spectral bandwidths and coarse

spectral resolution as indicated by lower bSMF (Fig. 3, *H* and *I*). Similarly, units with slow temporal responses and thus low bTMF can have narrow spectral bandwidths (high bSMF) and consequently stronger sideband inhibition along the frequency dimension (e.g., Fig. 3, *B–D*). A large proportion of neurons is also observed with low bTMF (<100 Hz) and bSMF (<1 cycle/octave).

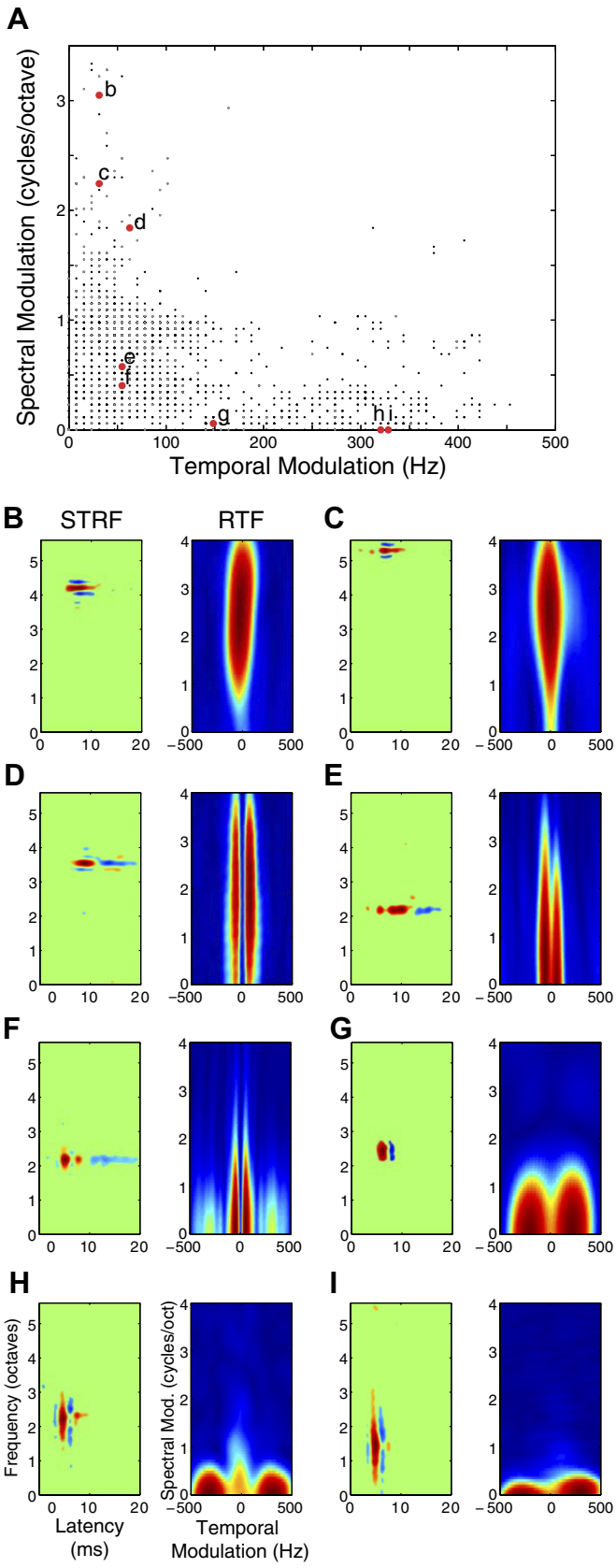

FIG. 3. Spectrotemporal tradeoff in modulation selection. Best temporal and spectral modulation frequencies covary. Sites with fast temporal preferences (high bTMF) have coarse spectral resolution (low bSMF), whereas sites with high spectral resolution (high bSMF) prefer slow temporal modulations (low bTMF). *B—I*: STRFs (*A–H, left*) and RTFs (*B–I, right*) from examples sites are selected to cover the CNIC response space (red dots in *A*). *B—D*: recording sites that exhibit narrow spectral tuning and strong sideband inhibition are characterized by high bSMF. These same units are tuned to slow temporal modulations. *E* and *F*: recording sites with narrow bandwidths and lack of sideband inhibition are tuned for low bSMF. These 2 units have long integration times and respond optimally to low bTMF. *G*: a fast neuron that lacks sideband inhibition but has a fast ON-OFF STRF pattern. *H* and *I*: units with broadband spectral selectivity that lack sideband inhibition are tuned to low spectral modulations. Both of these units exhibit fast temporal integration with interleaving patterns of excitation and inhibition and are tuned for fast temporal modulations.

Spectral and temporal modulation preferences in the CNIC covaried with BF in a manner distinct from that observed in the periphery where temporal modulation upper cutoffs increase with BF and units exhibit low-pass modulation sensitivity (Joris and Yin 1992). A qualitative comparison of the example STRFs reveals how modulation preferences vary with BF (Figs. 1 and 3). Sites with low BFs are often characterized by broad STRFs with temporally succinct ON-OFF temporal response pattern (e.g., Fig. 3, *H* and *I*). Accordingly, these sites have a high bTMF (309 and 281 Hz) and low bSMF (0.40 and 0.23 cycle/octave). Recording sites with higher BFs (e.g., Fig. 3, *B* and *C*) tend to have shorter STRF integration times, narrower spectral tuning, and stronger sideband inhibition. The latter examples have high bSMF (3.04 and 2.53 cycle/octave) and low bTMF (31.25 and 31.26 Hz). Sites with intermediate frequencies (e.g., Fig. 3, *E–G*), by comparison, tend to fall between these two extremes.

Quantitative analysis of the spectrotemporal modulation preferences confirms the described trends for the selected examples. We assessed the BF dependence by generating population RTFs grouped according to BF (in ½-octave frequency intervals, see METHODS). Figure 4 shows the population averaged RTF for each ½-octave frequency interval along with the corresponding bTMF and bSMF values from individual units (Fig. 4, superimposed dots). At low BF (2.0–2.8 kHz), the dominant response from the population RTF is confined to a lobe that extended to several hundred hertz TMF (population peak bTMF = 226.56 Hz) and SMF <1 cycle/octave (population peak bSMF = 0.11 cycle/octave). Individual bTMF and bSMF values for each unit (●, multiunit; ◎, single unit; Fig. 4) largely overlap the population RTF for this frequency range. At the opposite extreme, the population RTF of sites with high BFs (22.6–32 kHz, Fig. 4*H*) is dominated by energy at high bSMF (population peak bSMF = 0.86 cycle/octave) and low bTMF values (population peak bTMF = 39.06 Hz), with a larger proportion of individual unit's bTMF and bSMF values falling within this dominant response as indicated by the cluster of dots. A smooth transition between the low and high BF sites is observed suggesting that tuning characteristic change systematically with BF.

To quantify the observed relationships, we performed a statistical analysis of the spectral and temporal modulation preferences as a function of BF (Fig. 5). The median bTMF and

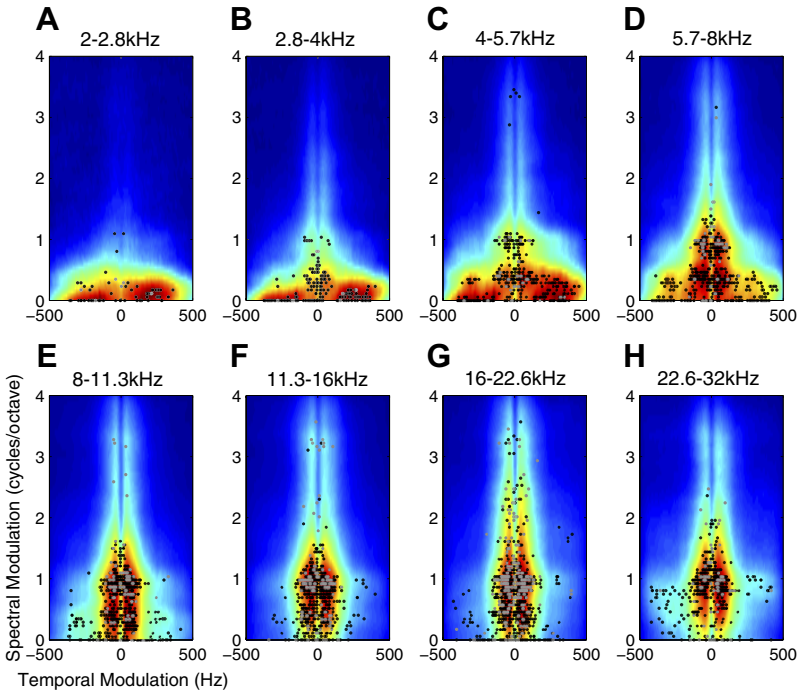

FIG. 4. Population spectrotemporal tuning varies systematically with best frequencies. Ripple transfer functions were obtained for each unit and averaged for different best frequency (BF) ranges. BFs were first partitioned into ½-octave bands and the population RTF was obtained for each ½-octave range (frequency range is noted above each panel). Best temporal and spectral modulation frequencies are shown for each unit as dots on each panel (●, multiunits; ◎, single units). Low-frequency units were concentrated about low spectral resolution (low bSMF) and fast temporal modulations (high bTMF). Accordingly, the population RTFs for low frequencies consisted of a dominant response lobe that extended to high bTMF values and had little energy for bSMF >0.5 cycle/octave (e.g., *A* and *B*). In contrast high-frequency sites (e.g., *F–H*) prefer finer spectral (high bSMF) and slower temporal modulations (low bTMF). This was evident both in the density of units as well as the energy distribution for the population RTF for each frequency band. A smooth continuum was observed such that spectrotemporal preferences varied systematically with increasing BF.

tMTF upper cutoffs (Fig. 5, *A* and *C*) decreases significantly with increasing BF for both single and multiunits (Wilcoxon rank sum, $P < 0.05$), suggesting that temporal modulation resolution decreases with BF. For frequencies between 2 and 4 kHz, the median bTMF and upper cutoffs (single units: bTMF = 218 Hz, median upper cutoff = 275 Hz; multiunits: median bTMF = 148 Hz, median upper cutoff = 250 Hz) are more than doubled compared with 16–32 kHz (single units: bTMF = 62 Hz, upper cutoff = 123 Hz; multiunits: bTMF = 63 Hz, upper cutoff = 126 Hz). An opposing trend is observed for both single and multiunits along the spectral dimension (Fig. 5, *B* and *D*). The median bSMF and sMTF upper cutoff

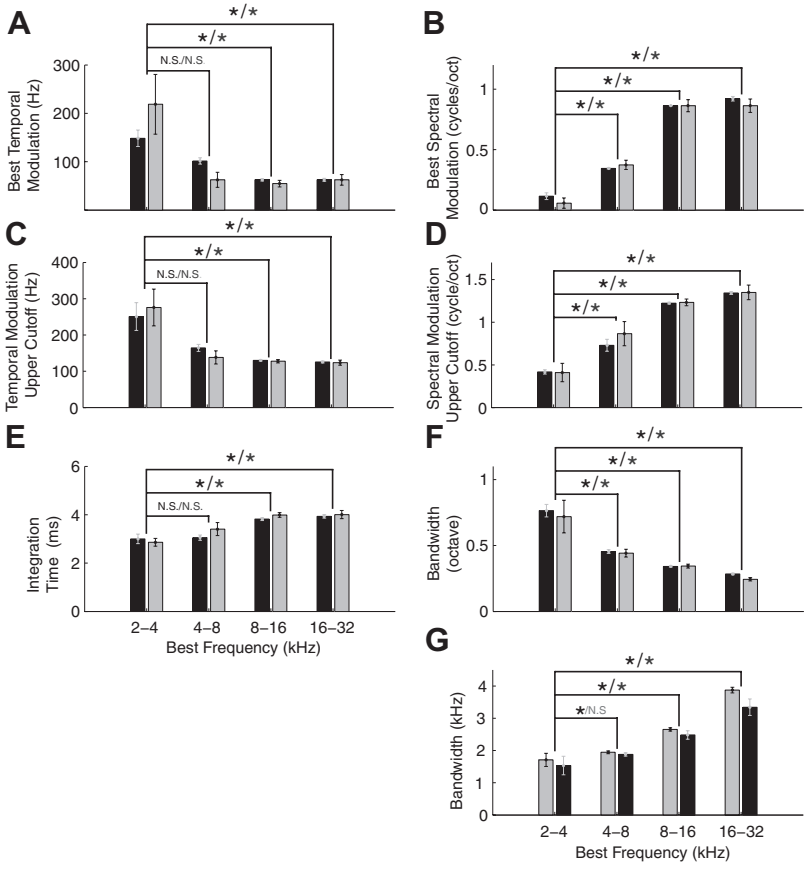

FIG. 5. Spectrotemporal resolution and modulation tuning characteristics vary systematically with recording location BF. Quantitative analysis of the STRF preferences reveals that temporal and spectral modulation selectivity exhibit opposing trends with increasing BF. ▩, median values for single units; ■, multi units. *A* and *C*: median best temporal modulation and upper cutoffs decrease, while median best spectral modulation and upper cutoffs increase with increasing BF (*B* and *D*). A related trend is also observed for the STRF integration time and proportional bandwidth (in octaves). Median integration times (*E*) and proportional bandwidths (*F*) exhibit an opposing trend with increasing BF. Temporal integration times are shorter and spectral bandwidths (in octaves) are broader for low-frequency sites (2–4 kHz). *G*: in contrast, absolute bandwidths (Hz) increase with increasing BF. *, designate significant comparisons (Wilcoxon rank sum, $P < 0.05$).

increases with BF, demonstrating that spectral resolution increases with BF (Wilcoxon rank sum, $P < 0.05$). Median bSMFs (single units: 0.86 cycle/octave for 16–32 kHz; 0.05 cycle/octave for 2–4 kHz; multiunits: 0.92 cycle/octave for 16–32 kHz; 0.11 cycle/octave for 2–4 kHz) and spectral modulation upper cutoffs (single units: 1.35 cycle/octave for 16–32 kHz; 0.41 cycle/octave for 2–4 kHz; multiunits: 1.34 cycle/octave for 16–32 kHz; 0.42 cycle/octave for 2–4 kHz) were considerably larger for high-frequency sites. In the STRF domain, a similar but weaker pattern is observed. The median STRF integration times (Fig. 5*E*) is longer for sites with high BFs (single units: 2.9 ms for 2- to 4-kHz range; 4.0 ms for 16- to 32-kHz range; multiunits: 3.0 for 2- to 4-kHz range, 3.9 for 16- to 32-kHz range; Wilcoxon rank sum, $P < 0.05$) while the median proportional bandwidth (i.e., in octaves; Fig. 5*F*) decreases considerably with increasing BF (single units: 0.72 octaves for 2- to 4-kHz range; 0.24 octaves for 16- to 32-kHz range; multiunits: 0.76 octaves for 2- to 4-kHz range; 0.28 octaves for 16- to 32-kHz range; Wilcoxon rank sum, $P < 0.05$). In contrast, the STRF absolute bandwidth (in Hz; Fig. 5*G*) increases with increasing BF (single units: 1.5 kHz for 2- to 4-kHz range; 3.3 kHz for 16- to 32-kHz range; multiunits: 1.7 kHz for 2- to 4-kHz range; 3.9 kHz for 16- to 32-kHz range; Wilcoxon rank sum, $P < 0.05$).

Relationship to the uncertainty principle and cochlear frequency resolution

The systematic relationship between BFs and CNIC receptive field properties may be related to the constraints imposed by the uncertainty principle (Gabor 1946) and the frequency resolution of AN fibers (Liberman and Kiang 1978). According to the uncertainty principle, the minimum allowed temporal and spectral resolutions of a filter are inversely related such that: $\Delta t \cdot \Delta f \geq 1/\pi$. Here Δt corresponds to the average temporal width of the filter, whereas Δf corresponds to the average filter bandwidth (in hertz). As a consequence, the bandwidth of a cochlear filter constrain its temporal resolution, such that the fastest temporal modulations encoded through the filter are partly limited by the frequency interaction products within the filter. Thus in theory, the fastest temporal modulation that can be faithfully represented at the output of a cochlear filter is approximately half the filter bandwidth.

The observed spectrotemporal tradeoff in the CNIC was not strictly limited by the receptive field bandwidth. In the AN, a strong dependence among temporal modulation upper cutoff frequencies, BFs, and absolute bandwidths (in hertz) is observed (Joris and Yin 1992). Temporal modulation upper cutoffs increase with BF for frequencies up to ~10 kHz; beyond 10 kHz temporal modulation upper cutoffs tend to saturate at ~1 kHz (Fig. 14 of Joris and Yin 1992). A trend that mirrors the uncertainty principle tradeoff is observed for peripheral AN fibers [Fig. 6, *A* and *B;* red data are from Joris and Yin (1992); blue data are from Kim and Young (1994); dots, fiber recordings that fall within the CNIC BF range of 2–32 kHz; plus sign, fiber recordings outside the 2- to 32-kHz range; dotted lines, predictions from the uncertainty principle]. There is a strong negative correlation between AN fiber spectral and temporal resolution estimates obtained with SAM tones and

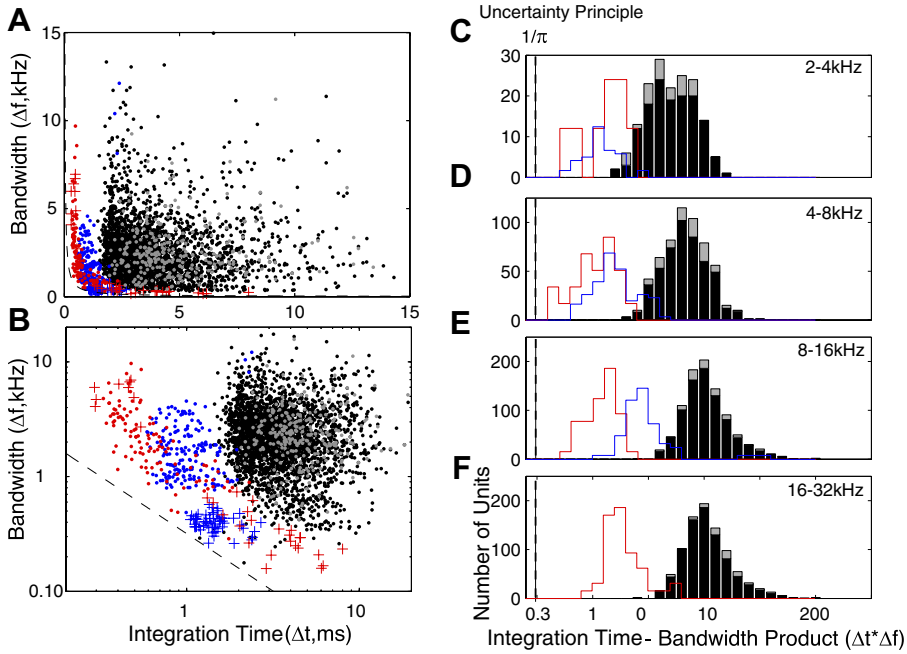

FIG. 6. Cochlear filter bandwidths and the uncertainty principle do not account for the observed spectrotemporal resolution tradeoff. *A*: the time-frequency resolution (Δt vs. Δf) of AN fibers (red: data from Joris and Yin 1992; blue: data from Kim and Young 1994; dots: AN fibers with $2 < BF < 32$ kHz; plus sign: AN Fibers with BF outside this range) roughly follow the inverse relationship for the uncertainty principle: $\Delta t \cdot \Delta f \geq 1/\pi$ (dotted line). By comparison, the time-frequency resolution (Δt vs. Δf) of CNIC units deviates substantially from this theoretical bound (a; black dots = multiunits; gray dots = single units). Units with $\Delta t > 15$ ms are not shown ($n = 5$). *B*: the same data as in *A* are shown using a double-logarithmic axis. Note that the time and frequency resolution of AN fibers is strongly negatively correlated and is within an order of magnitude of the uncertainty principle. *C–F*: the distribution of time-frequency resolution products, $\Delta t \cdot \Delta f$, from all CNIC recording sites (gray stacked histogram corresponds to single units; black stacked histogram corresponds to multiunits) and AN fibers (red histogram = Joris and Yin 1992; blue histogram = Kim and Young 1994) is shown as a function of best frequency (2–4, 4–8, 8–16, 16–32 kHz). For reference, the theoretical lower bound of $1/\pi$ imposed by the uncertainty principle is noted by dashed lines. For the vast majority of CNIC units, the time-frequency resolution product is greater than for AN fibers and more than an order of magnitude from the theoretical limit. Furthermore, the time-frequency resolution increases systematically with increasing BF.

tone pips [$r = -0.89 \pm 0.02$, t-test $P < 0.05$; $\log(\Delta t)$ vs. $\log(\Delta f)$] (Joris and Yin 1992). A weaker trend is observed for AN fiber STRFs obtained with Gaussian white noise [$r = -0.40 \pm 0.08$, t-test $P < 0.05$; $\log(\Delta t)$ vs. $\log(\Delta f)$] (Kim and Young 1994) although this dataset did not include frequencies beyond 15 kHz. In contrast, CNIC neurons exhibit a weaker trend ($r = -0.12 \pm 0.02$, t-test $P < 0.05$; Fig. 6, *A* and *B;* black circles, multiunits; gray circles, single units) and are substantially further away from the theoretical limit (Fig. 6, *A* and *B*, dotted curve). Viewed in a different way, the uncertainty principle predicts that temporal modulation upper cutoffs increase with increasing bandwidth and consequently BF. As expected CNIC absolute bandwidths (in kHz) increase with increasing BF (Fig. 5*G;* $r = 0.75 \pm 0.01$, mean ± SE; t-test $P < 0.05$). Mc Laughlin et al. also found this relationship in the IC and AN fibers (Mc Laughlin et al. 2007). However, in opposition to the uncertainty principle is our finding that temporal modulation upper cutoff frequencies decrease with increasing BF (Fig. 5C*)*. This trend opposes the relationship observed for AN fibers where temporal modulation upper cutoff frequencies tend to increase with BF (Joris and Yin 1992).

To compare AN fibers and CNIC neurons with the predictions of the uncertainty principle, we measured the spectrotemporal resolution product, $\Delta t \cdot \Delta f$. For the uncertainty principle, this product is constant ($1/\pi$) regardless of the filter BF. In the AN, this product increases with BF for spectrotemporal resolution estimates derived from the responses to SAM tones and tone pips (Fig. 6, *C–F,* red histograms) (Joris and Yin 1992). A similar trend is also observed for AN fiber STRFs obtained with Gaussian white noise, although these tend to deviate more from the uncertainty principle for high BF sites (Fig. 6, *C–F,* blue histograms) (Kim and Young 1994). Unlike AN fibers, which are within an order of magnitude from the uncertainty principle, this product is on average more than one order of magnitude away from the theoretical limit for CNIC neurons (single units: $\Delta t \cdot \Delta f = 9.3 \pm 0.4$; multiunits: $\Delta t \cdot \Delta f = 10.2 \pm 0.2$) and is significantly larger than for AN fibers ($P < 0.05$, Wilcoxon rank sum; Fig. 6, *C–F;* red, AN fibers data from Joris and Yin; blue, AN fiber data from Kim and Young; black, CNIC). This quantity increased with increasing BF in the CNIC for both single (median $\Delta t \cdot \Delta f$ for 2- to 4-kHz range = 4.1 ± 0.6; 4- to 8-kHz range = 7.4 ± 0.4; 8- to 16-kHz range = 9.9 ± 0.6; 16- to 32-kHz range = 14.7 ± 0.8; Wilcoxon rank sum, $P < 0.001$) and multiunits (median $\Delta t \cdot \Delta t$ for 2- to 4-kHz range = 5.0 ± 0.4; 4- to 8-kHz range = 6.7 ± 0.2; 8- to 16-kHz range = 9.9 ± 0.2; 16- to 32-kHz range = 16.0 ± 0.4; Wilcoxon rank sum, $P < 0.001$), implying that the spectrotemporal resolution was poorer for high BF sites.

Role of functional inhibition in modulation tuning

It is possible that the observed filtering characteristics are inherited directly from the periphery. However, this is not the case because best and upper temporal modulation frequencies opposes the trends between BFs and temporal modulation upper cutoffs observed in the AN (Joris and Yin 1992). Furthermore, the presence of band-pass-tuned sMTFs and tMTFs in the CNIC differs substantially from the AN where fibers exhibit low-pass temporal modulation tuning (Joris and Yin 1992). Thus the modulation sensitivity of the peripheral filters in the cochlea alone cannot account for the observed trends. We therefore tested whether inhibition and/or suppression potentially play a role in the observed spectrotemporal modulation tradeoff. Throughout we will refer to the negative STRF domains as "inhibitory," although these strictly reflect lack of a sound energy at a particular time and frequency prior to the initiation of an action potential, and could arise through several mechanisms such as neural inhibition or suppression (see DISCUSSION).

As previously proposed (Miller et al. 2002; Qiu et al. 2003), neurons with strong temporal inhibition in the STRF exhibit a band-pass-tuned tMTF (e.g., Fig. 3, *F–I*) ,whereas units that lack such temporal inhibition are characterized by low-pass tMTF (e.g., Fig. 3, *B* and *C*). Similarly, units that exhibit spectral sideband inhibition will exhibit band-pass sMTF (Fig. 3, *B–D*). Figure 7 illustrates this concept with a model STRF (see METHODS) in which the balance between excitation and inhibition is systematically varied along the temporal (*A–C*) or spectral (*D–F*) dimensions. When no inhibition is present (Fig. 7, *A* and *D, left*), the tMTF and sMTF both exhibit a low-pass pattern (i.e., untuned; Fig. 7, *B* and *E, left*). Increasing the amount of temporal inhibition in the STRF model leads to a more pronounced band-pass tMTF; however, the spectral tuning remains unchanged (Fig. 7, *A* and *B, middle* and *right*). Similarly, an increase in sideband inhibition strictly enhances the spectral selectivity such that the sMTF is band-pass tuned (Fig. 7, *D* and *E, middle* and *right*). Figure 7, *C* and *F,* plots the IER from the STRF versus the amount of spectral or temporal tuning as measured by the spectral and temporal DC gains (see METHODS). Increasing the IER by adding temporal inhibition to the model STRF decreases the temporal DC gain (Fig. 7*C*), thus indicating stronger tMTF tuning. Similarly, an increase in the amount of spectral inhibition leads to more negative spectral DC gains and thus more pronounced spectral modulation tuning (Fig. 7*F*). The model results demonstrate the intricate interaction between STRFe and STRFi domains and their role in generating modulation tuning. A purely excitatory STRF exhibits low-pass selectivity for both spectral and temporal modulations, whereas a STRF with excitation and inhibition exhibits band-pass modulation tuning.

Neural responses from the CNIC mirror the model predictions, suggesting that the pattern of inhibition shapes the spectrotemporal tuning. For each unit, we segmented the STRF into its STRFe and STRFi components (Fig. 8, *A–C*) and determined how the balance between excitation and inhibition in the STRF alters modulation tuning. Figure 8*A* shows a unit with a purely excitatory STRF (IER = 0) that exhibits a low-pass function along both temporal and spectral modulation dimensions (temporal and spectral DC gains = 0 dB). The unit of Fig. 8*B* has inhibitory subregions flanking an excitatory region in the time domain (IER = 0.16). Consistent with the model, this unit exhibits a band-pass tuned tMTF (temporal DC gain = −5.3 dB) although it is untuned for spectral modulations (spectral DC gain = 0 dB). Finally, the unit of Fig. 8*C* has a complex primarily spectral inhibitory component (IER = 0.21) and is therefore band-pass tuned primarily for spectral modulations (spectral DC gain = −8.2 dB; temporal DC gain = −1.1 dB). Figure 8, *D–F,* shows histograms for the relationship between IER and the DC gains across all units. The selected example units of *A–C* are marked by ×, +, and * on the histograms, respectively. Similar to the model findings,

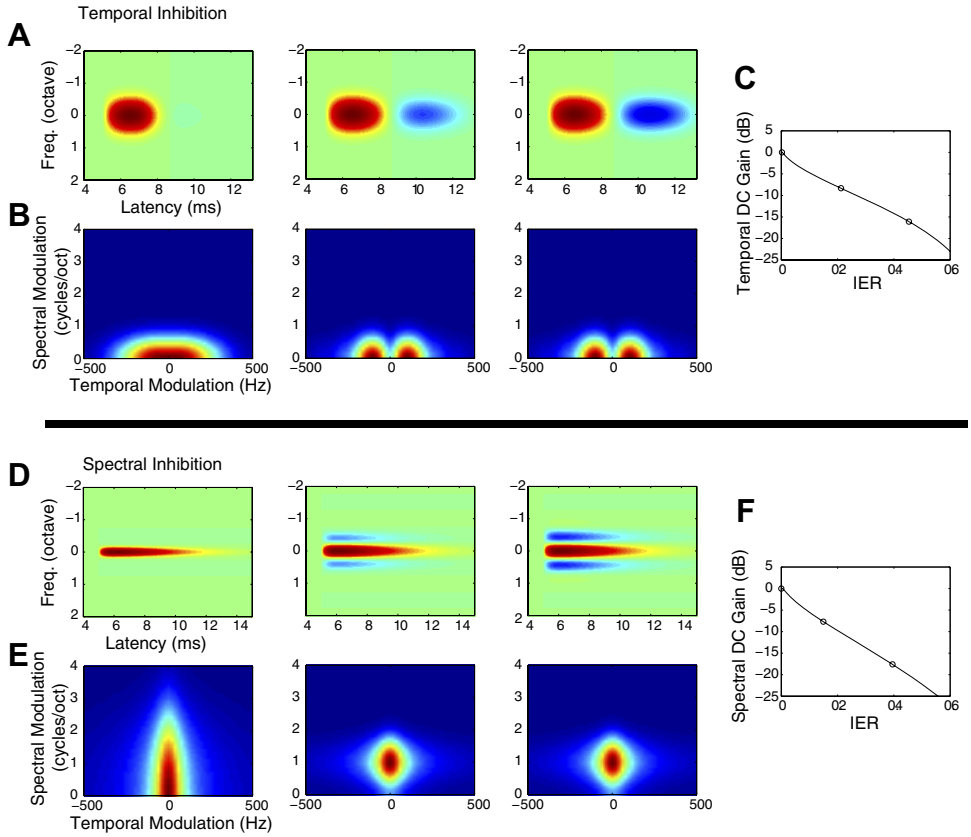

FIG. 7. The role of excitation and inhibition in modulation tuning. *A*: increasing the amount of temporal inhibition in a model STRF (*A, left* to *right*) changes the temporal tuning in the RTF from low- to band-pass (*B, left* to *right*). The strength of inhibition in the STRF was quantified by the inhibitory-to-excitatory power ratio (IER) while the strength of modulation tuning is characterized by the DC gain (see METHODS). *C*: temporal DC gains from *B* decrease with increasing IER indicating that stronger temporal inhibition leads to stronger temporal tuning. *D, left* to *right*: increasing the amount of sideband inhibition in a model STRF leads to more pronounced spectral tuning (*E, left* to *right*). *F*: spectral DC gains from the corresponding RTFs (*E*) decrease with increasing IER, indicating that stronger sideband inhibition leads to stronger spectral modulation tuning.

the DC gains decrease as the strength of inhibition (IER) increases. These parameters exhibit a significant negative correlation (IER vs. temporal DC gain: $r = -0.52 \pm 0.02$, $P < 0.001$; IER vs. spectral DC gain: $r = -0.32 \pm 0.02$, $P < 0.001$; IER vs. maximum DC gain: $r = -0.60 \pm 0.02$, $P < 0.001$), suggesting that stronger STRF inhibition leads to more pronounced tuning.

The excitatory and inhibitory components of the STRF exhibit marked changes in their spectral and temporal pattern with increasing BF. Figure 9 shows the population averaged (single and multiunits combined) excitatory (*top*) and inhibitory (*bottom*) STRF for different BF ranges (see METHODS; 2–4, 4–8, 8–16, and 16–32 kHz). For the lowest BFs (2–4 kHz), the STRFe exhibits a spectrally broad (~1.3 octave bandwidth) and temporally succinct pattern (~3-ms integration time), whereas for the highest frequency range the STRFe exhibits narrower bandwidth (~0.25 octave) of similar integration time. The population STRFi (Fig. 9, *bottom*) also exhibits systematic changes with BF, although the observed patterns and changes were generally more complex. For the lowest BFs (2–4 kHz), the dominant component of the STRFi consists of a spectrally broad and temporally succinct component that lagged the excitatory component by ~1 ms. A weaker although temporally broad inhibitory component is also observed that lags the excitatory component (starting at ~5 and ending at ~18 ms), whereas a weak and temporally succinct inhibitory component precedes the excitation (~2 ms lead). For high BFs sites (16–32 kHz), the STRFi consists of a prominent inhibitory component that surrounds the excitation. Like the low BF sites, a small, but noticeable amount of inhibition precedes the excitatory response. However, the high BF sites exhibit prominent sideband inhibition, whereas the low BF sites did not. A prominent and temporally broad secondary inhibitory component resembles the lagging component in the 2- to 4-kHz range; however, this component becomes more pronounced with increasing BF. These changes in the STRFe and STRFi structure are quantified in Fig. 10. As can be seen, the median excitatory and inhibitory integration times (Fig. 10, *A* and *D*) increase systematically with BF while the excitatory and inhibitory bandwidths (*B* and *E*) tend to decrease (Wilcoxon rank sum, $P < 0.05$). There are significant differences between single and multiunit inhibitory bandwidths for the low-frequency range (2–8 kHz; Fig. 10, *E* and *F*). While multiunits exhibit a significant decreasing trend for inhibitory bandwidths (in octaves), their inhibitory bandwidths in hertz does not vary systematically with BF (Wilcoxon rank sum, N.S across all BF comparisons). The alternative was true for single units, which exhibit a systematic trend for inhibitory bandwidths in kilohertz but not in octaves (Wilcoxon rank sum, N.S. across all BF comparisons). It is possible that these differences are due to the fact that there are few single-unit samples for the low and high BF range and the STRFi component is generally weaker and noisier than the excitatory component making it difficult to obtain reliable measurements. Overall, the changes with BF in the excitatory and inhibitory patterns are consistent with the hypothesis that excitation and inhibition provide a basis for changes in modulation tuning.

The results of Figs. 9 and 10 demonstrate that the shape of the STRFe and STRFi components changes systematically with BF, and these changes lead to changes in spectrotemporal modulation tuning as observed in Figs. 4 and 5. It is also possible that higher BF sites exhibit stronger inhibition, which

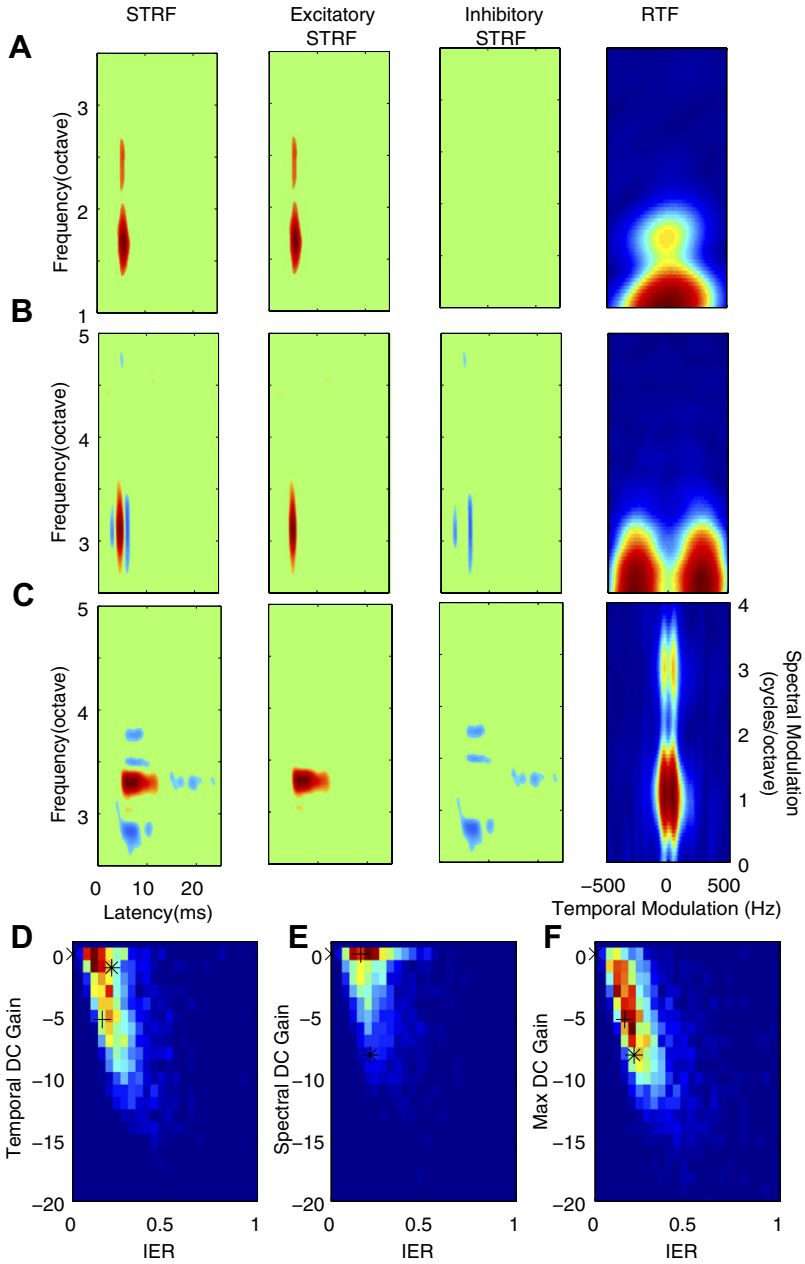

FIG. 8. Strength of modulation tuning covaries with the strength of STRF inhibition. *A–C*: quantitative analysis of excitatory and inhibitory STRF components. Each STRF was broken up into an inhibitory and excitatory component. The unit of *A* exhibits a strictly excitatory STRF. The corresponding RTF for this unit (*far right*) is low-pass for both spectral and temporal modulations. *B*: this unit exhibits a strong temporal excitatory and inhibitory pattern. Accordingly its RTF (*far right*) is strongly tuned for temporal modulations. The unit of *C* exhibits strong sideband inhibition and a weak temporal inhibitory component. This is reflected in its RTF as weak temporal tuning and strong spectral tuning (*far right*). *D–F*: histograms showing that the strength of STRF inhibition covaries with the amount of modulation tuning across all CNIC units. The IER is negatively correlated with the temporal DC gains (*D*), spectral DC gains (*E*), and the maximum of the temporal or spectral DC gains (*F*). The units of *A–C* are represented by ×, +, and *, respectively.

would further refine spectrotemporal selectivity for high-frequency sites. Figure 11*E* plots the average IER versus BF and demonstrates that the strength of inhibition did not increase appreciably with increasing BF (Wilcoxon rank sum, N.S.). Despite this, there is a subtle but significant increase in the proportion of temporal band-pass single and multiunits (*t*-test, $P < 0.05$; Fig. 11*A*) and a significant reduction in the temporal DC gain (Wilcoxon rank sum, $P < 0.05$; *C*) for BFs between 16 and 32 kHz compared with 2–4 kHz; however, these trends are not systematic along the low to high BF continuum (Wilcoxon rank sum, N.S. across BF comparisons). The corresponding trends are considerably stronger for the spectral dimension where single and multiunits exhibit substantially more band-pass tuning (*t*-test, $P < 0.05$; Fig. 11*B*) and a marked reduction in the spectral DC gains for high BFs (Wilcoxon rank sum, $P < 0.05$; Fig. 11*D*). Only 4.5% (multiunits = 4.5%; single units = 4.3%) of low-frequency units with BFs between 2 and 4 kHz exhibit spectrally band-pass RTFs while 58.7% (multiunits = 58.8%; single units = 57.5%) of the units between 16 and 32 kHz exhibit spectrally band-pass selectivity.

DISCUSSION

This study provides a new surprising view of how spectrotemporal features are transformed in the auditory midbrain. Spectrotemporal filtering resolution exhibits a systematic tradeoff along the tonotopic axis of the CNIC that cannot be accounted by the spectrotemporal resolution of the peripheral auditory filters. At the extremes, low BF sites are characterized by high temporal (short integration times and high bTMF values) and poor spectral resolution (broad proportional bandwidths and low bSMF), whereas high BF sites exhibit poorer temporal (longer integration times and

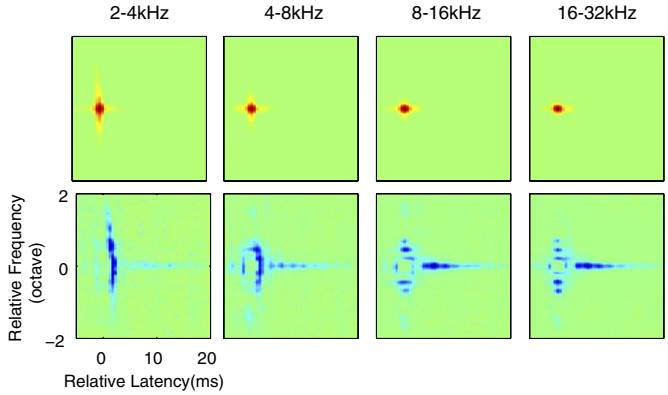

FIG. 9. Excitatory and inhibitory STRF preferences vary with BF. The population averaged excitatory (*top*) and inhibitory (*bottom*) STRF is shown for four frequency ranges (2–4, 4–8, 8–16, 16–32 kHz). At low frequencies (2–4 kHz), the excitatory STRF exhibits substantially broader bandwidths compared with high-frequency sites (16–32 kHz). In contrast, the structure of the inhibitory STRF is much more complex. For low frequencies (2–4 kHz), the STRFi is dominated by a brief (~2 ms) but strong temporal inhibitory component that spans ~1.8 octave. For this range of frequencies (2–4 kHz), sideband inhibition is not present. For high BFs (16–32 kHz), sideband inhibition becomes stronger and the STRF temporal inhibition is substantially narrower and longer.

low bTMF) and significantly better spectral resolution (narrower proportional bandwidths and higher bSMF). Together, the results suggest a dramatic transformation in the organization of spectrotemporal preferences at the level of the CNIC.

Comparison to the cochlear filters

Spectrotemporal filtering characteristics of CNIC neurons differ in a number of ways from AN fibers where modulation upper cutoffs tend to increase as a function of the fiber BF (Joris and Yin 1992). First, CNIC temporal tuning properties shift systematically with BF although the resulting trends oppose the observed pattern in the AN. Temporal upper cutoff frequencies and bTMFs decrease with increasing BF (Fig. 5, *A* and *C*). Furthermore, unlike the peripheral auditory filters, which strictly exhibit low-pass temporal modulation sensitivity, a substantial proportion of CNIC units also exhibit band-pass temporal modulation tuning.

Most CNIC studies have not identified a consistent relationship between BF and temporal modulation preferences in the CNIC (Krishna and Semple 2000; Rees and Møller 1983; Rees and Palmer 1989); however, Langner and colleagues have argued that BFs and best temporal modulation frequencies follow a trend consistent with the AN (Langner and Schreiner 1988; Langner et al. 1987, 2002). The main factor leading to this disparity has to do with the fact that Langner et al. sampled neurons with very low BFs (down to ~200 Hz), whereas most other studies focused exclusively on BFs greater than ~1 kHz. Note that to properly sample the SAM tones used in these studies, it is required that the carrier frequency be at least twice the frequency of the sinusoidal modulation. Consequently, there is a strong dependence between the neuron's BF and the maximum modulation frequency that can be tested. For low BF units (e.g., 250 Hz), CNIC neurons can readily phase lock to the fastest temporal modulations (e.g., 125 Hz) that are allowed by the sound carrier (250 Hz). Thus a direct relationship between BF and modulation sensitivity as observed by Langner and colleagues is expected a priori for the low-frequency range (i.e., <1 kHz). In our study, we were unable to sample the low-frequency range (<2 kHz) because the dynamic ripple sound has temporal modulation ≤500 Hz, and this requires carrier frequencies >1 kHz. Practically, to assure that the STRF was fully within the stimulus frequency range, we limited our recordings to BFs >2 kHz.

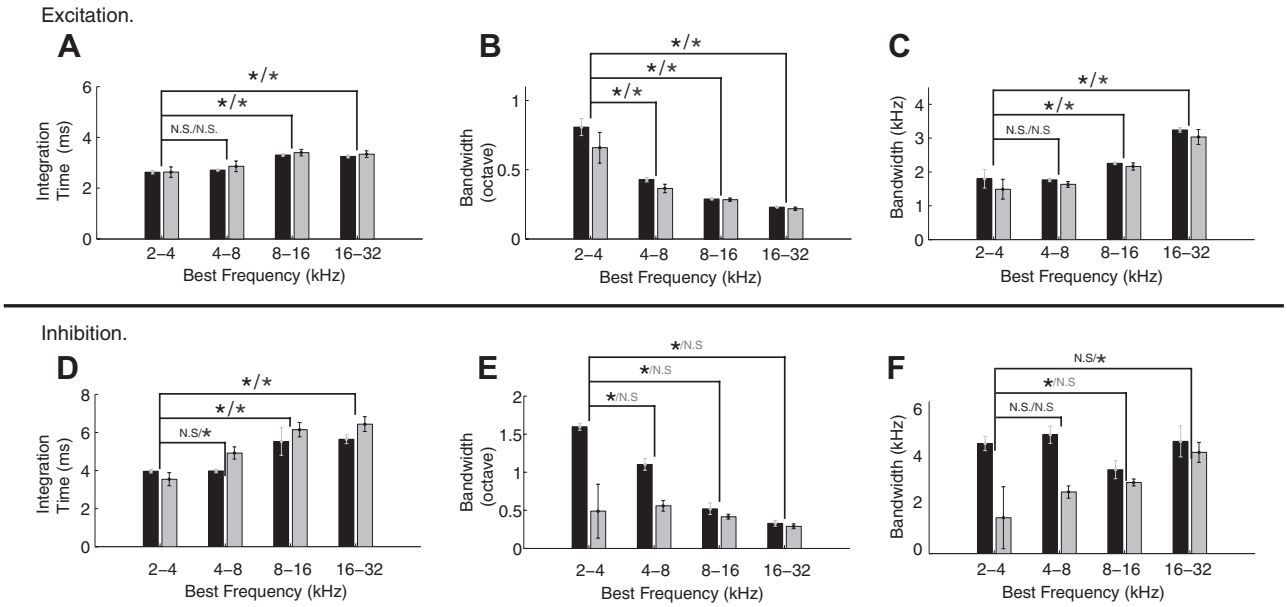

FIG. 10. Excitatory and inhibitory STRF preferences vary with recording location BF. Quantitative analysis of the excitatory and inhibitory STRF demonstrates that temporal and spectral preferences shift in opposing directions with increasing BF. ▨, median values to single units; ■, multiunits. The excitatory (*A*) and inhibitory (*D*) integration time both increase with BF while STRF excitatory bandwidths (in octaves) decrease (*B*). In contrast, absolute excitatory bandwidths increase with BF (*C*). The corresponding bandwidth trends for the inhibitory STRF component (*E* and *F*) were less systematic and differed somewhat between single and multiunits. *, significant comparisons (Wilcoxon rank sum, $P < 0.05$).

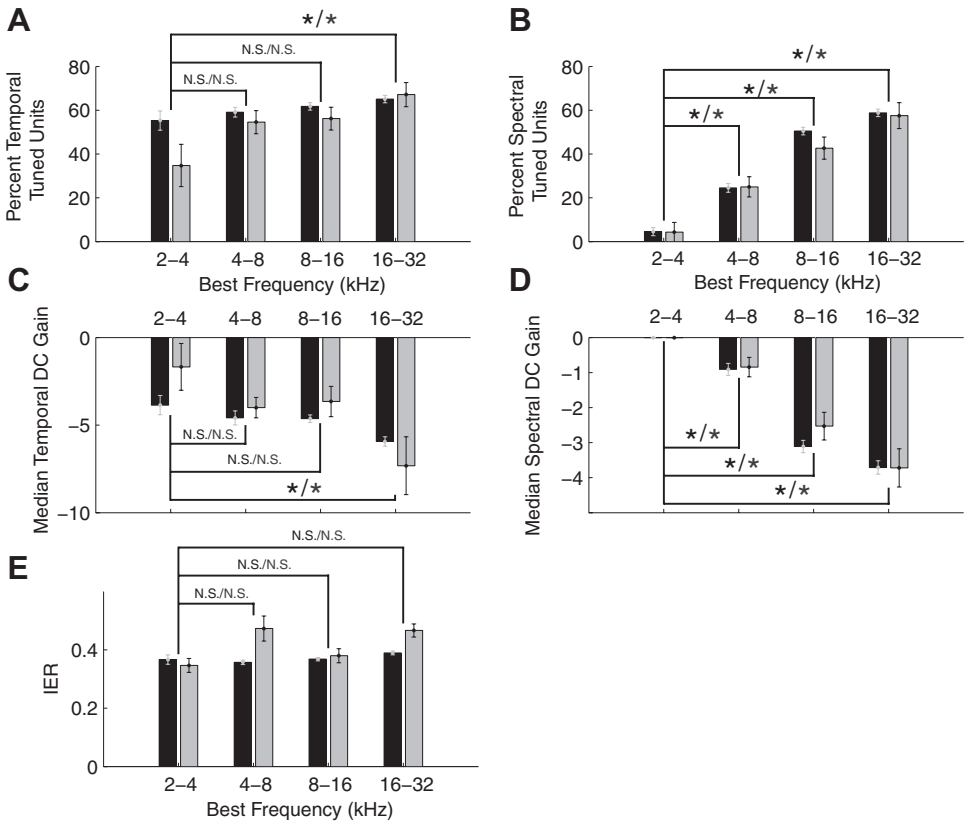

FIG. 11. Strength of modulation tuning varies with BF. The proportion of single and multiunits that exhibit temporal (*A*) or spectral (*B*) band-pass modulation tuning (3-dB criterion, see METHODS) increases with BF (*t*-test, $P < 0.05$). ▫, single units; ■, multiunits. Median temporal (*C*) and spectral (*D*) DC gains decrease with BF (Wilcoxon rank sum, $P < 0.05$) consistent with stronger temporal and spectral band-pass modulation tuning for higher frequency sites. The inhibitory-to-excitatory power ratio (IER) does not exhibit a significant increase for higher frequency sites (Wilcoxon rank sum, N.S.). *, significant comparisons ($P < 0.05$).

Spectral modulation sensitivity also varies in an unexpected manner with BF although spectral resolution mirrored AN fibers. In the AN (Liberman and Kiang 1978) and in the CNIC, absolute bandwidths (in hertz) increase with BF, whereas proportional bandwidth (in octaves) decrease with increasing BF (Fig. 5, *F* and *G*). Unlike the AN, however, strong sideband inhibition is present in many CNIC units. Notably, the presence of sideband inhibition serves to sharpen spectral modulation selectivity by creating tuned response (Fig. 8). Our data suggest that the amount and strength of spectral inhibition and consequently spectral modulation tuning increases with increasing BF (Fig. 11, *B* and *D*). This in turn produces a systematic increase in the preferred spectral modulation and upper cutoffs with increasing BF (Fig. 5, *B* and *D*). Interestingly, best spectral modulations extended up to ~3 cycle/octave so that the spectral resolution of the most acute CNIC neurons is in line with 1/3 octave critical band perceptual resolution (Fletcher 1940).

The spectrotemporal resolution of CNIC units also differs substantially from AN fibers. The comparison of our data to AN fibers has to be approached within the context of the stimulus measurements employed and analysis paradigm. Clearly there are differences between AN fiber measurements obtained for broadband noise (Kim and Young 1994) and narrowband signals (Joris and Yin 1992) that likely result because of differences in the stimulus employed. As can be seen in Fig. 6, measurements from AN fiber STRFs are overlapped with data obtained for narrowband signals. By comparison, CNIC neurons are more than an order of magnitude from the uncertainty principle and substantially more dispersed than both sets of AN fiber measurements (Fig. 6). AN fibers roughly follow the inverse relationship for bandwidth and integration time and are within one order of magnitude of the uncertainty principle. In particular, the inverse relationship between bandwidth and integration time is most pronounced for narrowband signals. By comparison, the spectro-temporal resolution of AN fiber STRFs is restricted to a narrower region of the allowed space. For CNIC, the spectrotemporal resolution product increases substantially with BF (Fig. 6, *C–F*). A similar but weaker trend is also present for AN fibers. In the AN this likely results from hair cell membrane and synaptic filtering that limit temporal phase-locking to ~1 kHz (Palmer and Russell 1986). Within the CNIC, this systematic change is seen in the temporal receptive field structure, which opposes the expected trend from the uncertainty principle. Specifically, absolute bandwidths increase with BF while temporal upper cutoffs decrease (Fig. 5, *C* and *G*). This trend leads to an increase in the spectrotemporal resolution product for higher BF sites (Fig. 6, *C–F*).

Functional implications

One plausible explanation for the observed tradeoff is that CNIC receptive field patterns may reflect specialized integration and resolving properties of the subcollicular inputs (Ramachandran et al. 1999). In particular, this tradeoff could reflect the complementary and overlapping pattern of medial (MSO) and lateral superior olivary (LSO) nuclei inputs. The excitatory inputs from the MSO and LSO to the CNIC are graded from low to high frequencies, such that MSO afferents dominate the low-frequency portion of the CNIC and LSO afferents are much more prevalent at high frequencies (Oliver 2000). As such, the observed trends could arise because of MSO and LSO specific receptive field differences, which could

be important for sound localization phenomena. Although we have not probed binaural preferences in this study, it is nonetheless possible that the graded patterns of MSO and LSO input contribute to the observed trends.

The combined effects of excitation and inhibition likely play a role in the generation of the observed STRF patterns and the corresponding frequency-dependent modulation trends (Figs. 9–11). Inhibitory circuits are pervasive within the auditory brain stem and midbrain and shape spectral and temporal sensitivity of units. Within the CNIC, local inhibition has a profound effect on the spectrotemporal modulation preferences of neurons (Andoni et al. 2007; Casseday et al. 2000; Kuwada et al. 1997; LeBeau et al. 2001; Nataraj and Wenstrup 2006; Zhang and Kelly 2003). In principle, the negative component of the STRF reflects lack of sound energy at a particular time and frequency prior to the initiation of an action potential (Escabi et al. 2005), which could theoretically arise through inhibition, neural adaptation or refractoriness. Several factors, however, point toward inhibition as the most likely mechanism. First, the time scales of the negative components of CNIC STRFs are much too short to be accounted by neural adaptation since adaptation in the midbrain takes place over hundredths of milliseconds (Kvale and Schreiner 2004; Nagel and Doupe 2006) to tens of seconds (Malmierca et al. 2009). This is consistent with our data because average spike rates to the continuous DMR adapt to ~50% of the initial firing rate with an average time constant of 18 s (data not shown). In a similar manner, although refractoriness could in theory contribute for some very fast inhibitory regions (e.g., 1- to 2-ms inhibitory integration time), it cannot account for the wide range of inhibitory integration times observed in our STRFs, which can extend up to tens of milliseconds. For many CNIC STRFs, the structure of the inhibitory pattern is actually inconsistent with adaptation because the inhibitory patters actually precede (e.g., Figs. 1, 3, *H* and *I,* and 8*B*) or occur concurrently with excitation as for sideband inhibitory regions (e.g., Figs. 1, 2*C,* and 3, *B–D*). Sideband inhibitory regions could arise through basilar membrane suppression (Ruggero et al. 1992); however, sideband suppression is substantially weaker and spectrally narrower in AN fibers STRFs (Kim and Young 1994) compared with the patterns we observe. A limitation of our approach is the use of anesthesia for the neural recordings. Although anesthesia is known to alter spectral and temporal response preferences, its overall effects are substantially less pronounced in the CNIC compared with auditory cortex (Ter-Mikaelian et al. 2007). Despite this, the anesthetic condition cannot account for our frequency dependent trends because anesthesia would indiscriminately target all frequencies. Future studies will need to consider the origins and mechanisms that lead to the observed sensitivity and trends.

Changes in the temporal arrangement of neural excitation, inhibition, and/or suppression are also consistent with the changes in temporal modulation sensitivity observed with BF. Intriguingly, the proportion of STRF inhibitory power does not change appreciably with BF (Fig. 11) although the spectrotemporal arrangement of excitation and inhibition varies considerably (Fig. 9) with more prevalent sideband inhibition present for high BF sites. This results in a dramatic increase in the strength and number of spectrally tuned sites at high frequencies. By comparison, the strength and amount of temporal modulation tuning also increases slightly with BF, and this is accompanied by a dramatic reduction of bTMF and temporal upper cutoffs as seen in Fig. 5. These changes in the modulation tuning are consistent with the observed changes in the spectrotemporal pattern of excitation and inhibition (Fig. 9). STRF integration time measurements are in line with measurements of CNIC cell membrane time constants (~5 ms) (Geis and Borst 2009). Specifically, the systematic trend toward slower STRFe's and STRFi's (Figs. 9 and 10) and slower modulation sensitivity for high BFs (Fig. 5, *A* and *C*) is consistent with the fact that cell membrane time constants tend to increase with BF in the CNIC (Geis and Borst 2009). Furthermore, these results resemble a comparable trend observed in the CNIC during electrical stimulation of the AN where higher limiting stimulation rates and shorter latencies were observed for low BF CNIC neurons (Middlebrooks and Snyder 2008).

The described spectrotemporal modulation trends have not been observed in the CNIC with other types of sounds and measurement techniques. One notable difference between this study and others is the large number of recording sites obtained with the 16-channel acute recording probes. Previous mapping studies have exclusively used single electrodes and were unable to achieve high-density sampling (e.g., >400 site/animal recorded on average for this study) within a large volume of the CNIC. A second difference is that the STRF strictly measures phase-locked neural activity and provides a description of the spectrotemporal sound pattern that is most likely to produce a single phase-locked spike (Escabi et al. 2005). In contrast, previous studies that have looked for an organization for amplitude modulations have focused primarily on rate MTF measurements (Langner et al. 2002; Schreiner and Langner 1988). Recently we have demonstrated that synchrony- and rate-tuning patterns are not complementary and can lead to distinctly different neural selectivity (Zheng and Escabi 2008). Thus it is likely that our findings would not be reproducible with rate based MTF approaches. Furthermore, the DMR sound employed in this study dynamically activates the entire sensory epithelium and simultaneously engages excitatory and inhibitory inputs across a broad frequency range. Given that the observed receptive field trends depend strongly on the arrangement of the receptive field inhibition (Figs. 9 and 10), it is likely that observed CNIC results would not be observed for simpler narrowband sounds (e.g., SAM tones, tone pips) because these cannot efficiently probe sideband inhibitory inputs.

Implication for the analysis of natural sounds

This study raises several new questions regarding the cost of encoding spectral and temporal features in natural sounds. CNIC neurons operate suboptimally when one considers the theoretical resolution limit set by the uncertainty principle. This is in contrast to the cochlear filters, which deviate slightly but appear to operate much closer to the theoretical resolution limit (Fig. 6).

One explanation for this dichotomy is that the cochlea serves to efficiently encode and preserve information in the incoming sound waveform while CNIC neurons operate to extract spectrotemporal features that are commonly found in natural sounds. The spectrotemporal modulations in speech and other natural sounds exhibit 1/*f* modulation spectrum (Attias and

Schreiner 1998a; Singh and Theunissen 2003; Voss and Clarke 1975) and log-normal amplitude statistics (Attias and Schreiner 1998a; Escabi et al. 2003) and exhibit a spectrotemporal resolution tradeoff that mirrors our neuronal data (Singh and Theunissen 2003). In particular, Singh and Theunissen demonstrated that spectral and temporal modulation spectra of animal vocalization sounds and speech are not independent. Sounds with fast temporal modulations tend to have coarse spectral modulations while sounds with fine spectral modulations tend to have slower temporal modulations, analogous to the observed neural sensitivity in Fig. 3*A*. Neurons in the CNIC have been shown to operate optimally with respect to several low- and higher-order statistics of the modulation envelope of natural sounds (Attias and Schreiner 1998b; Escabi et al. 2003). Thus it is possible that the observed frequency-dependent trends are important for encoding features that are represented across the low- to high-frequency continuum of hearing. Specifically, most of the relevant information for animal communication sounds and speech is derived from the mid- and low-frequency channels (French and Steinberg 1947; Pavlovic 1987). Within this frequency range precise temporal information is essential for speech intelligibility although fine spectral information is not as important (Shannon et al. 1995). High temporal resolution is also essential for the analysis of interaural timing cues, which are important for sound localization and are dominant for the low-frequency range of hearing (Macpherson and Middlebrooks 2002). Interestingly, the low-frequency units in our dataset exhibit the coarsest spectral resolution (broad bandwidth, low bSMF) and fastest temporal resolutions (short integration times, high bTMF).

In contrast, the observed spectral selectivity for high-frequency CNIC units may be necessary to properly resolve spectral details in head-related transfer function cues that are prominent in the high-frequency range of hearing and do not require fine temporal resolution (Kulkarni and Colburn 1998; Macpherson and Middlebrooks 2003; Tollin and Yin 2003). In contrast to AN fibers, CNIC units exhibit sideband inhibitory responses that enhance spectral modulation selectivity (Figs. 8 and 9). The fact that spectral modulation tuning is substantially more prominent for our high-frequency sites indicates that selectivity for spectral cues is enhanced for high frequencies. This is consistent with the hypothesis that enhanced spectral modulation selectivity at high frequencies could contribute to the analysis of head related spectral cues.

Thus unlike the peripheral receptors, which efficiently preserve information in natural sounds (Lewicki 2002), our results suggest that spectrotemporal sound modulations are selectively filtered and reordered within the CNIC. Future studies will need to address whether frequency-dependent differences in the statistical structure of natural sounds can be identified that correspond to the observed filtering characteristics.

ACKNOWLEDGMENTS

We thank D. L. Oliver, S. Kuwada, and C. E. Schreiner for reviewing the manuscript and providing thoughtful feedback and D. Kim for several insightful discussions. We also thank P. X. Joris, T.C.T. Yin, and E. D. Young for providing auditory nerve fiber data.

GRANTS

This work was supported by National Institute of Deafness and Other Communication Disorders Grant DC-006397.

REFERENCES

Adams JC. Ascending projections to the inferior colliculus. *J Comp Neurol* 183: 519–538, 1979.

Andoni S, Li N, Pollak GD. Spectrotemporal receptive fields in the inferior colliculus revealing selectivity for spectral motion in conspecific vocalizations. *J Neurosci* 27: 4882–4893, 2007.

Attias H, Schreiner C. Low-order temporal statistics of natural sounds. *Adv Neural Inform Process Syst* 9: 27–33, 1998a.

Attias H, Schreiner C. Coding of naturalistic stimuli by auditory midbrain neurons. *Adv Neural Inform Process Syst* 10: 103–109, 1998b.

Batra R, Fitzpatrick DC. Discharge patterns of neurons in the ventral nucleus of the lateral lemniscus of the unanesthetized rabbit. *J Neurophysiol* 82: 1097–1113, 1999.

Batra R, Kuwada S, Stanford TR. Temporal coding of envelopes and their interaural delays in the inferior colliculus of the unanesthetized rabbit. *J Neurophysiol* 61: 257–268, 1989.

Casseday JH, Ehrlich D, Covey E. Neural measurement of sound duration: control by excitatory-inhibitory interactions in the inferior colliculus. *J Neurophysiol* 84: 1475–1487, 2000.

Chi T, Gao Y, Guyton MC, Ru P, Shamma S. Spectro-temporal modulation transfer functions and speech intelligibility. *J Acoust Soc Am* 106: 2719–2732, 1999.

Cohen L. *Time-Frequency Analysis.* Englewood Cliffs, NJ: Prentice Hall, 1995.

Ehret G, Egorova M, Hage SR, Muller BA. Spatial map of frequency tuning-curve shapes in the mouse inferior colliculus. *Neuroreport* 14: 1365–1369, 2003.

Escabí MA, Miller LM, Read HL, Schreiner CE. Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. *J Neurosci* 23: 11489–11504, 2003.

Escabí MA, Nassiri R, Miller LM, Schreiner CE, Read HL. The contribution of spike threshold to acoustic feature selectivity, spike information content, and information throughput. *J Neurosci* 25: 9524–9534, 2005.

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

Fletcher H. Auditory patterns. *Rev Mod Phys* 12: 47–65, 1940.

French NR, Steinberg JC. Factors governing the intelligibility of speech sounds. *J Acoust Soc Am* 19: 90–119, 1947.

Gabor D. Theory of communication. *J Inst Elec Engr* 93: 429–457, 1946.

Geis HR, Borst JG. Intracellular responses of neurons in the mouse inferior colliculus to sinusoidal amplitude-modulated tones. *J Neurophysiol* 101: 2002–2016, 2009.

Goldstein JL. An optimum processor theory for the central formation of the pitch of complex tones. *J Acoust Soc Am* 54: 1496–1516, 1973.

Irino T, Patterson RD. Temporal asymmetry in the auditory system. *J Acoust Soc Am* 99: 2316–2331, 1996.

Joris PX. Envelope coding in the lateral superior olive. II. Characteristic delays and comparison with responses in the medial superior olive. *J Neurophysiol* 76: 2137–2156, 1996.

Joris PX, Yin TC. Responses to amplitude-modulated tones in the auditory nerve of the cat. *J Acoust Soc Am* 91: 215–232, 1992.

Joris PX, Schreiner CE, Rees A. Neural processing of amplitude-modulated sounds. *Physiol Rev* 84: 541–577, 2004.

Keller CH, Takahashi TT. Representation of temporal features of complex sounds by the discharge patterns of neurons in the owl's inferior colliculus. *J Neurophysiol* 84: 2638–2650, 2000.

Kim PJ, Young ED. Comparative analysis of spectro-temporal receptive fields, reverse correlation functions, and frequency tuning curves of auditory-nerve fibers. *J Acoust Soc Am* 95: 410–422, 1994.

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

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

Kuwada S, Fitzpatrick DC, Batra R, Ostapoff EM. Sensitivity to interaural time differences in the dorsal nucleus of the lateral lemniscus of the unanesthetized rabbit: comparison with other structures. *J Neurophysiol* 95: 1309–1322, 2006.

Kuwada S, Batra R, Yin TC, Oliver DL, Haberly LB, Stanford TR. Intracellular recordings in response to monaural and binaural stimulation of neurons in the inferior colliculus of the cat. *J Neurosci* 17: 7565–7581, 1997.

Kvale MN, Schreiner CE. Short-term adaptation of auditory receptive fields to dynamic stimuli. *J Neurophysiol* 91: 604–612, 2004.

Langner G, Albert M, Briede T. Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (*Chinchilla lanigger*). *Hear Res* 168: 110–130, 2002.

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

Langner G, Schreiner C, Merzenich MM. Covariation of latency and temporal resolution in the inferior colliculus of the cat. *Hear Res* 31: 197–201, 1987.

LeBeau FE, Malmierca MS, Rees A. Iontophoresis in vivo demonstrates a key role for GABA(A) and glycinergic inhibition in shaping frequency response areas in the inferior colliculus of guinea pig. *J Neurosci* 21: 7303–7312, 2001.

Lesica NA, Grothe B. Dynamic spectrotemporal feature selectivity in the auditory midbrain. *J Neurosci* 28: 5412–5421, 2008.

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

Liberman MC, Kiang NY. Acoustic trauma in cats. Cochlear pathology and auditory-nerve activity. *Acta Otolaryngol Suppl* 358: 1–63, 1978.

Loftus WC, Malmierca MS, Bishop DC, Oliver DL The cytoarchitecture of the inferior colliculus revisited: a common organization of the lateral cortex in rat and cat. *Neuroscience* 154: 196–205, 2008.

Macpherson EA, Middlebrooks JC. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *J Acoust Soc Am* 111: 2219–2236, 2002.

Macpherson EA, Middlebrooks JC. Vertical-plane sound localization probed with ripple-spectrum noise. *J Acoust Soc Am* 114: 430–445, 2003.

Malmierca MS, Blackstad TW, Osen KK, Karagulle T, Molowny RL. The central nucleus of the inferior colliculus in rat: a Golgi and computer reconstruction study of neuronal and laminar structure. *J Comp Neurol* 333: 1–27, 1993.

Malmierca MS, Cristaudo S, Perez-Gonzalez D, Covey E. Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *J Neurosci* 29: 5483–5493, 2009.

Malmierca MS, Saint Marie RL, Merchan MA, Oliver DL. Laminar inputs from dorsal cochlear nucleus and ventral cochlear nucleus to the central nucleus of the inferior colliculus: two patterns of convergence. *Neuroscience* 136: 883–894, 2005.

Mc Laughlin M, Van de Sande B, van der Heijden M, Joris PX. Comparison of bandwidths in the inferior colliculus and the auditory nerve. I. Measurement using a spectrally manipulated stimulus. *J Neurophysiol* 98: 2566–2579, 2007.

Merzenich MM, Reid MD. Representation of the cochlea within the inferior colliculus of the cat. *Brain Res* 77: 397–415, 1974.

Middlebrooks JC, Snyder RL. Intraneural stimulation for auditory prosthesis: modiolar trunk and intracranial stimulation sites. *Hear Res* 242: 52–63, 2008.

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

Nagel KI, Doupe AJ. Temporal processing and adaptation in the songbird auditory forebrain. *Neuron* 51: 845–859, 2006.

Nataraj K, Wenstrup JJ. Roles of inhibition in complex auditory responses in the inferior colliculus: inhibited combination-sensitive neurons. *J Neurophysiol* 95: 2179–2192, 2006.

Oliver DL. Ascending efferent projections of the superior olivary complex. *Microsc Res Tech* 51: 355–363, 2000.

Palmer AR, Russell IJ. Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair-cells. *Hear Res* 24: 1–15, 1986.

Pavlovic CV. Derivation of primary parameters and procedures for use in speech intelligibility predictions. *J Acoust Soc Am* 82: 413–422, 1987.

Plomp R The role of modulations in hearing. In: *Hearing: Physiological Bases and Psychophysics,* edited by Klinke R, Hartmann R. New York: Springer Verlag, 1983, p. 270–275.

Qiu A, Schreiner CE, Escabí MA Gabor analysis of auditory midbrain receptive fields: spectro-temporal and binaural composition. *J Neurophysiol* 90: 456–476, 2003.

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

Rees A, Palmer AR. Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise. *J Acoust Soc Am* 85: 1978–1994, 1989.

Rhode WS. Interspike intervals as a correlate of periodicity pitch in cat cochlear nucleus. *J Acoust Soc Am* 97: 2414–2429, 1995.

Ruggero MA, Robles L, Rich NC. Two-tone suppression in the basilar membrane of the cochlea: mechanical basis of auditory-nerve rate suppression. *J Neurophysiol* 68: 1087–1099, 1992.

Schouten JF. The residue and the mechanisms of hearing. *Proc K Ned Akad Wet* 43: 991–999, 1940.

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

Schreiner CE, Langner G. Laminar fine structure of frequency organization in auditory midbrain. *Nature* 388: 383–386, 1997.

Semple MN, Aitkin LM. Representation of sound frequency and laterality by units in central nucleus of cat inferior colliculus. *J Neurophysiol* 42: 1626–1639, 1979.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science* 270: 303–304, 1995.

Singh NC, Theunissen FE. Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 114: 3394–3411, 2003.

Spirou GA, Young ED. Organization of dorsal cochlear nucleus type IV unit response maps and their relationship to activation by bandlimited noise. *J Neurophysiol* 66: 1750–1768, 1991.

Ter-Mikaelian M, Sanes DH, Semple MN. Transformation of temporal properties between auditory midbrain and cortex in the awake Mongolian gerbil. *J Neurosci* 27: 6091–6102, 2007.

Tollin DJ, Yin TC. Spectral cues explain illusory elevation effects with stereo sounds in cats. *J Neurophysiol* 90: 525–530, 2003.

van-Veen TM, Houtgast T. Spectral sharpness and vowel dissimilarity. *J Acoust Soc Am* 77: 628–634, 1985.

Voss RF, Clarke J. "1/f noise" in music and speech. *Nature* 258: 317–318, 1975.

Yu JJ, Young ED. Linear and nonlinear pathways of spectral information transmission in the cochlear nucleus. *Proc Natl Acad Sci USA* 97: 11780–11786, 2000.

Zhang H, Kelly JB. Glutamatergic and GABAergic regulation of neural responses in inferior colliculus to amplitude-modulated sounds. *J Neurophysiol* 90: 477–490, 2003.

Zheng Y, Escabi MA. Distinct roles for onset and sustained activity in the neuronal code for temporal periodicity and acoustic envelope shape. *J Neurosci* 28: 14230–14244, 2008.