Published in final edited form as:

Hear Res. 2009 October ; 256(1-2): 118–130. doi:10.1016/j.heares.2009.07.005.

Dynamics of Spectro-Temporal Tuning in Primary Auditory Cortex of the Awake Ferret

B. Shechter^{1,2}, H.D. Dobbins^{1,3}, P. Marvit⁴, and D.A. Depireux^{1,3}

¹ Department of Anatomy and Neurobiology, School of Medicine, University of Maryland, Baltimore, 20 Penn St. HSF II Rm. S251, Baltimore, MD 21201 USA

² Department of Biology, University of Maryland College Park, College Park, MD 20742 USA

³ Program in Neuroscience, School of Medicine, University of Maryland, Baltimore, Baltimore, MD 21201 USA

⁴ Program in Neuroscience and Cognitive Science, Department of Psychology, University of Maryland College Park, College Park, MD 20742 USA

Abstract

We previously characterized the steady-state spectro-temporal tuning properties of cortical cells with respect to broadband sounds by using sounds with sinusoidal spectro-temporal modulation envelope where spectral density and temporal periodicity were constant over several seconds. However, since speech and other natural sounds have spectro-temporal features that change substantially over milliseconds, we study the dynamics of tuning by using stimuli of constant overall intensity, but alternating between a flat spectro-temporal envelope and a modulated envelope with well defined spectral density and temporal periodicity. This allows us to define the tuning of cortical cells to speech-like and other rapid transitions, on the order of milliseconds, as well as the time evolution of this tuning in response to the appearance of new features in a sound. Responses of 92 cells in AI were analyzed based on the temporal evolution of the following measures of tuning after a rapid transition in the stimulus: center of mass and breadth of tuning; separability and direction selectivity; temporal and spectral asymmetry. We find that tuning center of mass increased in 70% of cells for spectral density and in 68% of cells for temporal periodicity, while roughly half of cells (47%) broadened their tuning, with the other half (53%) sharpening tuning. The majority of cells (73%) were initially not direction selective, as measured by an inseparability index, which had an initial low value that then increased to a higher steady state value. Most cells were characterized by temporal symmetry, while spectral symmetry was initially high and then progressed to low steady-state values (61%). We demonstrate that cortical neurons can be characterized by a lag-dependent modulation transfer function. This characterization, when measured through to steady-state, becomes equivalent to the classical spectro-temporal receptive field.

Keywords

Tuning dynamics; Auditory Cortex; AI; Temporal Symmetry; Inferior Colliculus; IC; Auditory grating; Spectro-Temporal Receptive Field; STRF

Contact Information: Didier A. Depireux, Department of Anatomy and Neurobiology, School of Medicine, University of Maryland, Baltimore, 20 Penn St. HSF II Rm. S251, Baltimore, MD 21201, depireux@gmail.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Auditory cortical neurons are tuned to specific aspects of the spectro-temporal content of the sound. Our previous work in the primary auditory cortex (AI) of the ferret (*Mustela Putorius furo*) characterized steady-state neural responses to ongoing broadband sounds with well-defined spectro-temporal content by constructing spectro-temporal receptive fields (STRFs), which measure how stimulus history (i.e. the recent content of each frequency channel as a function of time) influences neural activity. In these studies, auditory gratings with fixed spectro-temporal content were presented for several seconds, and STRFs were measured from the **steady-state** portion of the response. Using the STRFs, we successfully predicted steady-state responses in AI to new stationary sounds (Depireux et al. 2001; Fritz et al. 2003; Kowalski et al. 1996a; Schreiner and Calhoun 1994; Shechter and Depireux 2006; 2007). Other classes of broadband stimuli have been developed to measure STRFs, such as chord-like structured sounds (deCharms et al. 1998; Linden et al. 2003; Valentine and Eggermont 2004), natural sounds (Aertsen and Johannesma 1981; Schafer et al. 1992; Sen et al. 2001; Theunissen et al. 2001; Yeshurun et al. 1985), or slowly changing spectro-temporal content (Escabi and Schreiner 2002; Miller et al. 2001). Regardless of the method used, the features of most cortical receptive fields are similar in structure: an excitatory subfield occurring with a latency in the tens of milliseconds surrounded along the frequency axis and followed in time by inhibition.

Our long-term goal is to understand the coding of natural sounds and speech, which are broadband and typically have rapidly changing spectro-temporal content. To understand the coding of such sounds, which have complex statistics, we first study the encoding of simpler stimuli with rapidly changing but precisely defined structure.

One of the primary methods used to study auditory cortical neurons employs short duration pure tone stimuli of varying levels and frequencies. Neurons are characterized by a tuning curve measured over a defined period following stimulus onset. By counting action potentials over a fixed window, these studies implicitly assume that cortical tuning is instantaneous and static. In contrast, broadband sound studies typically use the neural response to an ongoing, long duration sound. Longer duration cortical responses obtained in awake preparations exhibit a large variety of response characteristics. As early as 1964, Evans and Whitfield (1964) demonstrated not only instantaneous tuning at the onset of the response, but also phasic and sustained responses to pure tones. More recently, Wang et al. (2005) showed that single neurons in AI exhibit an onset and sustained response to preferred stimuli, and onset-only response to non-preferred stimuli. This dichotomy of responses suggests complementary encoding schemes at different timescales during the response.

Sound *level* and spectro-temporal *content* are two different aspects of sound that are concurrently represented in cortical responses. We use the terms *level transient* to describe a change in level that occurs on a short time scale, and *feature transient* to describe a change in spectro-temporal content on a short time scale. The introduction of almost any broadband sound from silence induces a cortical response because of the level transient. As mentioned earlier, the STRF is useful both in characterizing the steady-state response of a neuron, and for predicting its steady-state response to novel stimuli. However, the onset response to an auditory grating is often not predicted well from the STRF derived using steady-state responses (Kowalski et al. 1996b). Classical STRFs are derived from the steady-state responses to stimuli with constant mean level. As a linear model, they describe and predict responses best in the context of stimulus deviations away from a constant mean level and, by design, are not expected to predict the response to sudden changes in level or spectro-temporal content.

The difficulty in characterizing the encoding of feature transients separately from that of level transients lies in dissociating the two components of the response. To address this issue, the sound stimuli in this study were constructed to have a constant mean level with a well-defined spectro-temporal envelope emerging from flat spectral noise (of the same level). Thus, an advantage is that our stimuli are derived from analytically defined spectro-temporal envelopes, with feature transients present independently of level transients. This is in contrast to the aforementioned auditory gratings and natural sounds, where level and feature transients occur simultaneously. Since our goal was to study the effects of feature transients independently of level transients, we have disambiguated as much as is possible two aspects of dynamic changes occurring in natural sounds.

Bredfeldt and Ringach (2002) studied dynamics of neural tuning in visual cortex in a similar way; they measured the dynamics of spatial frequency tuning using reverse correlation with respect to rapidly changing spatial luminance gratings. They found that tuning in most cells becomes more selective over the course of the response, and the preferred spatial frequency shifts from low to higher spatial frequencies. However, they only studied the dynamics of tuning in response to static gratings. Because natural sounds such as speech are inherently non-stationary, it is especially important for our investigation of the dynamics of tuning to employ stimuli having temporal modulations in addition to spectral modulations.

Our study of the dynamics of tuning was initially motivated by the observations reported by Simon et. al. (2006): from their inference of the functional connectivity between subcortical and cortical circuits, we expect the cortical STRF to evolve in time. Simon et. al. (2006) showed that temporal symmetry of cortical receptive fields would be best explained by the existence of lagged cells earlier in the auditory pathway. Lagged cells are characterized by an inhibition followed by excitation, which is delayed by tens of milliseconds. Physiological evidence for such lagged cells can readily be found in the visual literature: Mastronarde (1987a; 1987b) and Saul and Humphrey (1990) showed the existence of lagged and non-lagged cells in cat LGN, each class having different temporal response; their staggered input contributions into cortex imply that the tuning of a cortical cell should evolve in time. We therefore hypothesized that coding in auditory cortex is best characterized by a modulation transfer function that evolves in time. Determining the exact timescales over which the symmetries reported by Simon et. al. (2006) emerge will provide a better understanding of the functional connectivity that gives rise to the cortical tuning and its evolution.

In this paper, we examine the dynamics of tuning to spectro-temporal content: how does cortical tuning evolve from purely spectral onset tuning to the complete steady state tuning?

Methods

Surgical preparation

All recordings were from awake, 3 to 12 month old domestic ferrets (*Mustela Putorius furo*) surgically implanted with chronic moveable multi-electrode arrays, custom made from a modification of the Neuralynx 12Drive-H (Neuralynx, Tucson AZ). For surgical preparation, ferrets were anesthetized with Halothane (3% induction, then 1.75% maintenance adjusted to keep heart rate, respiration, end-tidal CO₂ and SpO₂ within limits), and affixed within a stereotaxic frame. Body temperature was maintained at 37.5°C with a feedback heating pad. The skin on the skull was incised rostral-caudally along the midline from the nuchal crest to a line joining the eyes. The scalp was retracted and the temporalis muscles were resected bilaterally. A craniotomy was made unilaterally over the left visualized AI. To prevent re-growth and toughening of the dura, the mitotic inhibitor 5-fluorouracil was applied (Dobbins et al. 2007; Spinks et al. 2003). Stainless steel screws were inserted around the skull to anchor a head post and the multi-electrode microdrive to the skull with dental cement. The head post

was positioned rostrally. The microdrive was slowly lowered into the craniotomy, and the headpost and microdrive were mechanically bonded to the skull via screws and dental cement.

The electrode exit geometry was in a honeycomb pattern over AI, such that the minimum distance between adjacent electrodes was 225 μ m (Dobbins et al. 2007). After surgery, the ferrets were given Banamine (1mg/kg) and Baytril (0.2mg/kg) for three recovery days. All surgical and experimental procedures were approved by the University of Maryland Animal Care and Use Committee and were in accord with NIH Guidelines on the care and use of laboratory animals.

Neural Recordings

Recording sessions took place inside a double-walled sound booth (IAC, Bronx, NY, Noise Isolation Class of 70dB). The ferret was placed in a holder, with its head fixed using the implanted headpost to ensure the animal stayed within the calibrated sound field and to minimize movement noises during low level stimuli. The animal was monitored through a closed-circuit video. A low-pass filtered field potential from a low impedance electrode was used to monitor the emergence of slow wave EEG activity, taken to indicate drowsiness. Drowsiness was mitigated for a period of an hour or more by providing the ferret with treats (Ferretoone). Recording sessions typically lasted 3–4 hours. Activity of single neurons was simultaneously recorded from 6 to 12 parylene-coated tungsten microelectrodes (initial impedance 3.5–6M Ω at 1kHz, shaft diameter 76 μ m, Micro Probe, Inc, Gaithersburg, MD). Electrodes were individually advanced by manually turning a screw.

Neural activity was recorded and assigned to single neurons in two steps. During the recording, the electrode signal was band-pass filtered with low and high cutoff frequencies of 300Hz and 3kHz, respectively. Events were captured when the amplitude exceeded a threshold derived from the average power of the recorded signal; this threshold was set low enough to capture all spikes, but it also captured large excursions of the evoked potential. Event times were assigned by position of the peak. After recording, stored events were sorted into multiple classes, using a modification of the MClust package (Redish 2004), with the automated cutter KlustaKwik (Harris et al. 2000), based on each event's centroid of the Fourier transform, energy, and first two principal component projections. Our low threshold and conservative sorting typically yielded a large number of rejected events, which were included in a "miscellaneous" class and not considered neural spikes.

Stimulus Generation and Sound Presentation

All stimuli were generated in MATLAB (Natick, MA), then converted to an analog voltage (TDT RX6, Tucker David-Technologies, Alachua, FL) at 100kHz sampling rate, processed with an analog attenuator (TDT PA5), amplified (Crown DX-75) and presented from a speaker (Manger Transducer, Manger, Germany) located 1m at zenith relative to the animal's head. The sound field was calibrated so that the loudspeaker had a flat response (to within 1.5dB) from 250Hz to 32kHz at the position of the animal's head. The overall level of any given stimulus waveform was calibrated by adjusting its root mean square voltage with respect to a reference voltage obtained from a 1kHz tone played at 94 dB SPL.

Stimulus Set—Initial Characterization and Steady-State Response

Steady-state STRFs were measured with Temporally Orthogonal Ripple Combinations (TORC) stimuli (Klein et al. 2000). Briefly, the TORC stimuli are the sum of periodic ($T = 250$ msec) 7-octave auditory gratings, each having a spectro-temporal profile modulated sinusoidally in spectrum and in time. The modulation of a grating is characterized in spectrum by its spectral density (Ω , cycles/octave), in time by its temporal periodicity (w , Hz), and in amplitude by its excursions away from the mean level of the stimulus (modulation depth ΔA ,

% of mean). Each of the gratings comprising a TORC has the same spectral density and modulation depth, but differs in temporal periodicity, thus sampling a set of points in spectro-temporal parameter space with a single sound.

In the TORC stimulus, the amplitude $S(x,t)$ of each tone component is given by

$$S(x,t) = L[1 + \Delta A \cdot \sum_i \cos(2\pi(\Omega \cdot x + w_i \cdot t) + \varphi_i)], \quad (1)$$

which specifies a linear modulation. In the equation, the frequency of each tone component is given by x , where $x = \log_2(f/f_0)$ such that f_0 is the lower edge of the spectrum. L is related to the intensity of the stimulus and φ_i are the starting phases of each of the component gratings in the TORC. When both Ω and w are positive, the envelope drifts towards the low frequencies. The tones (f) that make up the gratings are logarithmically spaced, so as not to elicit a pitch percept. Typical TORC spectro-temporal envelopes are illustrated in Fig. 1.

Stimulus Set—Transient Tuning

To measure tuning dynamics, we used transient grating stimuli as illustrated in Fig. 2. These stimuli are broadband, spanning 5 octaves, and are 1.25 sec long. A typical stimulus spectro-temporal envelope is flat except for eight 50 msec intervals (transients) of modulation randomly distributed throughout the stimulus duration. Each transient consists of 50 msec of an auditory grating with specific spectral density, temporal periodicity and starting phase. In a given waveform, the 8 transients have the same density and temporal periodicity, but starting phases chosen from a random permutation of $\{2\pi \cdot x/8, x = 0, 1, \dots, 7\}$. Random inter-transient interval (ITI) durations are chosen from a normal distribution with a mean of 150 msec and a standard deviation of 50 msec, limited between 75 msec and 225 msec. The first transient begins 50 msec after the stimulus onset, and the ITIs are used to determine when subsequent transients begin. A 3 msec ramp is applied to the onset and offset of the spectro-temporal transient envelope, to avoid the perception of a click at the beginning and the end of the transients.

These spectro-temporal envelopes are used to determine the amplitude of 100 tones per octave over 5 octaves as a function of the time. These carrier tones are in random temporal phase. The tones are added together to form a sound of almost constant power with no level onsets, but with a series of well-defined spectro-temporal feature transients. There are 63 transient grating stimuli: Spectral density of the transients ranged from -2 cyc/oct to 2 cyc/oct in steps of 0.5 cyc/oct and temporal periodicity from 0 Hz to 30 Hz in steps of 5 Hz, with 100% modulation depth.

Data Analysis

The derivation of the *steady-state* STRF from TORC stimuli is well established (Klein et al. 2000) and will not be repeated in full here; briefly, a reverse correlation method is used to obtain an STRF from the spike trains, as depicted in Fig. 1. Each spectro-temporal envelope is presented together with its inverse to compensate partially for non-linearities such as half-wave rectification. The first 250 msec of the response to each stimulus was omitted in order to analyze the response only after it reached a steady state. The STRF was then used to determine the appropriate frequency range of the transient grating stimuli.

In the case of the transient grating (the main focus of this paper), we are concerned with measuring the modulation transfer function—which is equivalent to measuring the receptive field—at a particular instant in time (as opposed to an average in steady-state over the entire stimulus duration). It is important to note that the method shown in Fig. 2 is equivalent to the

standard reverse correlation method for a long duration stimulus, as is illustrated in Fig. 3. the method of Fig. 2 could be applied to the steady-state regime of a neural response by presenting the same sound with 8 starting phases (for simplicity, the method is shown in Fig. 3 using only 4 starting phases). After a suitably long delay, the firing rate is measured over 1/8 of a cycle for each of the starting phases. The measurements are concatenated (Fig. 3, bottom-left) to obtain the response to a full cycle. The concatenated response would be equivalent to the response obtained from a single sound over a full cycle, as is done in a traditional reverse correlation (Fig. 3, bottom-right), once the response is in steady-state regime.

An example response to the feature transient stimuli is shown in Fig. 4. Following the transitions from a flat to modulated spectro-temporal envelope, the two exemplar cells shown in Fig. 4 respond in a graded manner dependent on 1) the spectro-temporal envelope statistics (Ω , w , and phase), 2) the relative position of the neuron on the tonotopic axis, and 3) as a function of time (or lag τ) after the change from flat to modulated spectrum. These 2 cells will be revisited in some of the following figures as examples of 2 broad categories of cells we found—those with dynamics in their tuning (cell 35) and those without dynamics (cell 47). The measure of time τ is defined as the lag, or time elapsed since the most recent transition from flat to modulated envelope. For the feature transient stimulus, we compute a transient modulation transfer function (tMTF), similar to the MTF obtained as a Fourier transform of the STRF (Kowalski et al, 1996a). Note that the MTF and the STRF are equivalent representations of the response characteristics of a neuron. The tMTF is measured for a set of chosen lags τ after the onset of feature transients (Fig. 2). For a stimulus of temporal periodicity w , our analysis window is $(8 \cdot w)^{-1}$ seconds in duration for each of the 8 transients. We compute the average spiking rate in a window starting at τ msec after each of the 8 transient onsets. For example, with $w = 25\text{Hz}$ and $\tau = 10$ msec, the analysis window is $(8 \cdot w)^{-1} = 5$ msec in duration: therefore, we measure the average number of spikes per second in a window starting 10 msec and ending 15 msec after the onset of each of the 8 phases. With the spike rate expressed as a function of transient onset phase (Fig. 2), we compute the Fourier transform and compensate for the phase shift in the stimulus due to the time from $\tau = 0$ elapsed to the center of the analysis window. The measurement is a best estimator of the spike rate at the center of the window. The phase compensation effectively ‘re-centers’ all analysis windows at τ . The amplitude and phase of the first Fourier component indicate the phase-locking strength and phase-delay of the response, respectively, with respect to the feature transient at τ msec after its onset. This is effectively the modulation of the neural response as a function of the initial phase of the transient and of the lag τ after the onset of the transient. Measuring this for all combinations of Ω and w , we obtain a tMTF as a function of lag. The tMTF is computed with overlapping sliding windows for each millisecond following the transient onsets. This millisecond resolution tMTF is used to compute various descriptive measures later in the analysis. In Fig. 5, we display the inverse Fourier transforms of the tMTFs at lags in multiples of 5 msec.

Once we obtain a set of tMTFs for a set of lags, we analyze how the tMTF evolves as a function of that lag. Our main interest in characterizing the tMTF is to study tuning dynamics and how the instantaneous tuning as it develops relates to the steady-state MTF—which is obtained by Fourier transform of the steady-state STRF. We found that the dynamics of several parameters (some developed in past studies to characterize steady-state STRFs (Depireux et al. 2001) and adapted to the present study) were especially useful. In particular, we considered the dynamics of the *center of mass* of tuning and the *breadth* of tuning around this center of mass in order to determine whether the average tuning changed and whether it broadened or sharpened. We also examined the dynamics of quadrant *separability* and symmetry of the *spectral* and *temporal* transfer functions, since our previous studies pointed to *a priori* unexpected results with respect to separability and temporal symmetry of STRFs.

Note that the tMTF has conjugate symmetry, and therefore (e.g. in Fig. 1) quadrants 1 and 2 are complex conjugates of quadrants 3 and 4, respectively. Specifically, calling $\Omega > 0$, $w > 0$ quadrant 1 and $\Omega < 0$, $w > 0$ quadrant 2, we use the following parameters:

1. **Center of mass of tuning** (Ω_{CM} , w_{CM}). This measure is a response-weighted mean spectral density and mean temporal periodicity. It is computed in the quadrant with the greater total modulation power.
2. **Breath of tuning** (α_b). For the quadrant in which the total modulation power is greater, this measure indicates how the tMTF power is spread around its center of mass. It is defined in one quadrant by a normalized distance from the center of mass of the modulation transfer function as:

$$\alpha_b = \frac{1}{\sum_{\Omega, w} P_{\Omega, w}} \cdot \sum_{\Omega, w} P_{\Omega, w} \cdot \sqrt{\left(\frac{\Omega - \Omega_{CM}}{\Omega_{max}}\right)^2 + \left(\frac{w - w_{CM}}{w_{max}}\right)^2} \quad (2)$$

over all measured Ω and w in that quadrant. $P_{\Omega, w}$ is the power of modulation in the response at (Ω, w) , or in other words the square of the amplitude of the (Ω, w) component of the tMTF. (Ω_{CM}, w_{CM}) , is the tMTF center of mass in that quadrant, and (Ω_{max}, w_{max}) are the maximum spectral density and temporal periodicity tested, respectively. If the cell's tuning sharpens or broadens with increasing lag, α_b will decrease or increase, respectively.

3. **The degree of inseparability** (α_{SVD}): Although there is no reason to expect it, *a priori*, separability turns out to be an important property of cortical MTFs. A fully separable transfer function is one that can be factorized into a product of functions of Ω and w : $MTF(\Omega, w) = G(\Omega) \cdot F(w)$, or equivalently the STRF(x, t) is time-spectrum separable: $STRF(x, t) = RF(x) \cdot IR(t)$. Separability need not be an all-or-none property but rather can be assessed in a graded fashion by using a singular value decomposition (SVD). This method decomposes a function into a sum of fully separable functions; a detailed explanation is available in Abdi (2007). Briefly, SVD decomposes a matrix into the product of a *diagonal* matrix Λ and two unitary matrices U and V so that $U \times \Lambda \times V^T$ is the original matrix. Λ has the same dimensions as the original matrix with nonnegative decreasing diagonal elements (λ_i). SVD therefore decomposes the tMTF into a weighted sum of fully separable components, where each component is the product of a spectral and a temporal transfer function weighted by a diagonal element λ_i of Λ . These spectral and temporal transfer functions are the columns of U and V , respectively, and are ordered in decreasing contribution to the overall sum. Using SVD, we want to measure how much of the total tMTF power is accounted for by its first singular. We define

$$\alpha_{SVD} = \left(1 - \lambda_1^2 / \left(\sum_i \lambda_i^2 \right) \right) \quad (4)$$

α_{SVD} therefore defines a single measure of the “distance” of the system from separability or alternatively the “degree of inseparability”. An α_{SVD} value of 0 means the tMTF is fully separable (i.e., it is a product of a spectral transfer function and a temporal transfer function), whereas values approaching 1 correspond with inseparability (the closer the tMTF is to being separable, the more dominant the first singular value λ_1 will be over its counterparts, which share the residual error in a

manner that depends on the precise nature of the inseparability). Separability implies the absence of direction selectivity. Since the directionality of the envelope of a sound is indeterminate at short lags, we hypothesize that a neuron's response will not be direction selective at short lags, and this selectivity will only manifest with increasing lag.

4. **The spectral and temporal asymmetry (α_s and α_t).** These measures indicate how asymmetric the tMTF is around $\Omega = 0$ and $w = 0$, respectively, in terms of the absolute values of normalized complex cross-correlations between the principal spectral and temporal sections in quadrants 1 and 2. Taken together, these two indices α_s and α_t afford another way of analyzing the time-dependent build-up of direction selectivity towards the steady-state receptive field by quantifying how asymmetric the transfer functions are with respect to the down-moving (quadrant 1) versus the up-moving (quadrant 2) components of the spectro-temporal envelope. We define

$$\alpha_s = 1 - \left| \frac{\sum_{\Omega > 0} G_1(\Omega) \cdot G_2^*(\Omega)}{\sqrt{\sum_{\Omega > 0} |G_1(\Omega)|^2 \cdot \sum_{\Omega > 0} |G_2(\Omega)|^2}} \right| \quad (5)$$

$$\alpha_t = 1 - \left| \frac{\sum_{w > 0} F_1(w) \cdot F_2^*(-w)}{\sqrt{\sum_{w > 0} |F_1(w)|^2 \cdot \sum_{w > 0} |F_2(-w)|^2}} \right| \quad (6)$$

where G and F are the spectral and temporal transfer functions of the tMTF quadrants respectively, and the subscripts 1,2 indicate the quadrant for which they are computed. These functions (G , F) are the first columns of U and V from a singular value decomposition in each quadrant. The more similar G_1 and G_2 (respectively, F_1 and F_2) are, the closer the absolute value in Eqs. 5,6 will be to 1. Therefore, α values near 0 correspond to symmetric transfer functions, whereas values near 1 correspond to more asymmetric transfer functions. It has previously been shown that steady-state STRFs in AI of the ferret are by and large quadrant separable and temporally symmetric (Simon et al. 2006). We also explore the time evolution of α_t as it reaches symmetry in the steady-state.

Response variability and bootstrap

To determine the reliability of the steady-state STRF, we computed a signal-to-noise ratio (SNR) of its modulation transfer function (MTF, the 2-dimensional Fourier transform of the STRF). N_{boot} (here, 100) bootstrap estimates $\psi(\Omega, w)$ were generated for each point of the MTF, where the response to each period of the stimulus waveform was treated as an independent measurement. The SNR of each point was defined as the average power divided by the variance of the estimates. The total SNR of the STRF was computed as the power-weighted mean of the SNR at each (Ω, w) point:

$$SNR_{\Omega, w} = \frac{|\sum_{bootstraps} \psi(\Omega, w)|^2}{N_{boot} \cdot \sigma_{\psi}^2} \quad (7)$$

$$\text{SNR} = \frac{\sum_{\Omega, w} P_{\Omega, w} \cdot \text{SNR}_{\Omega, w}}{\sum_{\Omega, w} P_{\Omega, w}} \quad (8)$$

Transient MTF Response Thresholding

In order to determine the lags at which responses to the transient features were significant, we compared the total tMTF power P_{tMTF} (Eq. 9) at each lag to a baseline modulation power.

$$P_{tMTF}(\tau) = \sum_{\Omega, w} P_{\Omega, w} \quad (9)$$

The baseline modulation power was defined as the average total tMTF power from lags 0 msec to 8 msec. These lags occur before the 10 msec minimal expected response latency of a cortical neuron, so that this is an average measure of power in the absence of a response. The significance threshold for each cell was defined as 10% of the maximum modulation power above baseline. If this threshold was below an absolute threshold of $0.1 \text{ spikes}^2/\text{sec}^2$, then the absolute threshold was used instead. Only cells for which the modulation power exceeded threshold for at least 30 msec continuously were further analyzed. A cell was considered to have a significant response only for those lags at which the power exceeded threshold.

Based on the modulation power, we selected a 16 msec time interval on which to extract the trends of the dynamics of tuning. This window was chosen centered at the first lag for which there was a significant peak in the total power in order to normalize for the different response latencies and durations observed. The window's duration was set long enough to allow for the measurement of trends, but short enough so that the dynamics were not averaged out.

Results

General characteristics of responses

The responses to transient and continuous spectro-temporal modulations of broadband noise were collected from 92 single-unit recordings in 3 ferrets, which showed reliable steady-state phase-locking to modulations in the stimulus, as measured by an SNR larger than 0.5.

With respect to the transient gratings, cells were considered to have a significant response if the total power in the transient modulation transfer function (tMTF) exceeded threshold continuously for at least 30 msec (see Methods). We found that 57 cells (62%) had a reliable steady-state characterization and met all criteria for transient response significance. In response to transient gratings, phase-locking from a few units was poor. The criteria used in classifying the presence of a response were strict in that they excluded some cells which, by visual examination, were deemed to phase-lock to the transient gratings. These six cells were of very short duration transient response (< 25 msec in response to both pure tones and broad-band sounds), high latency (> 60 msec), or low modulation power (< $0.1 \text{ spikes}^2/\text{sec}^2$).

The average spontaneous spike rate for all cells in the study, measured between sound presentations (at least one second after a sound was off) was 15.7 spikes/sec. During the sustained sounds (flat noise and transient gratings), the average evoked spike rate was 18.7 spikes/sec. There was considerable variability from cell to cell; these numbers serve only to indicate that, with 8 feature transients lasting 50 msec each, we had on average 7–8 spikes per (Ω, w) combination per sweep.

Dynamics

The steady-state measurement of a neuron's receptive field quantifies its preference for the spectro-temporal content of ongoing sounds. In this paper, we expose the dynamics of a neuron's receptive field with respect to the onset of feature transients. In this regard, we develop a method of analyzing responses with respect to the onset of novel spectro-temporal features. Fig. 5 shows the evolution of tuning for two average cells at multiple lags after the onset of a feature transient. These 2 cells are representative of two broad categories of cells we found—those with dynamics in their tuning over the first 50 ms of an unchanging sound, (cell 35) and those without dynamics (cell 47). We compare these evolutions to their steady-state counterparts by computing the inverse Fourier transform of the tMTF (i.e., the tSTRF). **Cell #35** (Fig. 5A) exhibits an excitatory region (corresponding to the excitatory region of the steady-state response) at short lags. Sideband inhibitory regions develop at intermediate lags (from $\tau = 20$ msec to $\tau = 40$ msec). Finally, an inhibitory region follows the main excitatory region (from $\tau = 40$ msec). While the STRF (Fig. 5B) captures both the main excitatory and inhibitory regions, it fails to capture any significant spectral sideband inhibitory regions that appear in the dynamic characterization.

Fig. 5D shows the same characterization for **Cell #47**. Here the size and location of the inhibitory and excitatory regions of the receptive field do not change significantly with increasing lag. However, as the receptive field stabilizes toward steady-state (starting at $\tau = 25$ msec), the cell develops direction selectivity, measured as the asymmetry of power in one quadrant of the MTF versus the other. The direction selectivity is evident from the excitatory and inhibitory regions of the tSTRFs assuming an oblique orientation. For longer lags, the tSTRF becomes increasingly similar to the steady-state (Fig. 5E).

In Fig. 5C,F, we show for these two cells the total modulation power of the tMTFs, which was used to determine the lags for which the cells were responding to the feature transients.

We further characterized these dynamics with a number of descriptive measures (see Methods).

Center of Mass

We hypothesized that the spectro-temporal envelope is encoded in cortex dynamically, which we demonstrate by measuring the lag dependent modulation transfer function. Such a transfer function shows how neural tuning changes as a function of lag. We define the preferred stimulus as the center of mass of the transfer function, and track the dynamics of the best spectral density and temporal periodicity, as shown in Fig. 6B and C.

We extracted the slope of the best linear fit to the center of mass (both in spectral density and temporal periodicity) around the first significant peak in the modulation power for the quadrant with the greater modulation power. Center of mass increased with lag in 70% of cells for spectral density (Ω) and in 68% of cells for temporal periodicity (w) (see Fig. 6A). This overall increase in the center of mass of tuning in MTF space can correspond to a sharpening of the features in spectro-temporal space if the breadth of tuning (about the center of mass) of the MTF does not change.

Breadth of Tuning

To analyze breadth of tuning in spectral density and temporal periodicity space, we computed α_b , which quantifies the spread of power around its center of mass—effectively a weighted measure of variance (Equation: see Methods), for every lag τ in the tMTF quadrant with the greater modulation power. A transfer function with a large spread will be broadly tuned (in Ω - w space), and therefore have a large value of α_b . Conversely, a sharply tuned transfer function concentrated mostly around its center of mass will have a comparatively small value of α_b .

Because of the duality between the compactness of a function and that of its Fourier transform, a reduction in tuning breadth in the MTF will correspond to a broadening of the corresponding tuning as measured by the STRF.

The analysis window is of finite duration, which makes the response measure an average value throughout that window as opposed to an instantaneous value. Therefore, the neural response transitions from noise to signal as the cell starts responding. Since our analysis starts at the onset of the change from flat noise to a specific spectro-temporal content—i.e., before the neural conduction time—the initial tMTF is composed of random noise, with uniform power. Therefore, values of α_b for small τ are random with a high mean.

We extracted the slope of the best linear fit to $\alpha_b(\tau)$ around the first significant peak in the modulation power. Approximately half of cells (47%) broadened their tuning and the other half (53%) sharpened tuning as a function of lag (Fig. 7). Cell #35 represents a member of the former category while Cell #47 represents one of the latter category. Combined with the overall increase in center of mass we showed earlier, this result implies a gradual sharpening of the features in spectro-temporal space after the onset of a transient.

Separability

Direction selectivity in a cell's steady-state response does not imply direction selectivity in the response close to the onset. Depending on the spectral density and temporal periodicity of the grating, we expect that the direction of the grating, although well defined from its spectro-temporal envelope, cannot be initially determined with confidence given *only* the sound waveform. Note that there is an inherent time-frequency compromise when determining the spectro-temporal content from a sound pressure waveform; this contributes to the uncertainty in the spectro-temporal analysis given only a short segment of the sound. However, even in the steady-state, neurons in AI are not wholly selective for a single direction, but rather show a relative preference for upwards versus downwards moving gratings (or vice versa).

The STRF can be viewed as a series of temporal profiles arranged along the spectral axis. Direction selectivity requires a precise organization of these profiles along that spectral axis, such that their exact arrangement determines the preferred direction for stimulus frequency content. Spectral and temporal processing are interdependent in direction selective cells; their STRFs are thus inseparable and cannot be represented simply as the product of spectral and temporal functions. Conversely, an *STRF* (x, t) which is fully separable into a product of spectral and temporal functions, $STRF(x, t) = RF(x) \cdot IR(t)$ cannot be direction selective. The separability of the transfer function is measured by α_{SVD} , with values near zero corresponding to a high degree of separability. Initially, the cell is responding only to the white noise segment of the stimulus (flat spectro-temporal envelope). For the same reasons presented earlier, the value of α_{SVD} for this initial period will be random with a high mean. At the beginning of the neural response to the modulated stimulus (measured by the modulation power of the tMTF), it is not possible for a cell to distinguish between upward and downward moving gratings, because of the ambiguity in measuring grating direction from a short sample of the stimulus. Therefore, the tMTF is highly separable at small lags with correspondingly small values of α_{SVD} . The majority of cells (73%) had α_{SVD} which decreased from its initial random value, which corresponds well with the inability to determine direction during the first few milliseconds of the response. Following this initial decrease, α_{SVD} then increased over a short period and quickly leveled off.

This trend is characterized by a concavity in the α_{SVD} curve (see Fig. 8B,C). When fit to a second order polynomial, this concavity is described by a positive coefficient for the second order term τ^2 of the polynomial. In Fig. 8, we show the distribution of second-order coefficients for the $\alpha_{SVD}(\tau)$ fits and individual traces for the two example cells shown in Fig. 5. The temporal

progression of inseparability implied by the τ^2 coefficient shows that the majority of cells (73%) acquired some direction selectivity with increasing lag. The absence of highly negative values of τ^2 confirms that no cells had a progression from separability before feature onset, to inseparability during the first few milliseconds after feature onset, and then a return to separability at long lags; the absence of such behavior confirms our physical intuition that inseparability does not appear at short lags.

Spectral and temporal symmetry, α_s and α_t

Direction selectivity is a property of the auditory response that is inherently expected to be dynamic. Even if a neuron is direction selective in its steady-state response, it initially should be unable to determine the direction of the spectro-temporal sweep (and therefore should not be direction selective). With increasing lag, the neuron has more time and a larger sample over which to analyze the stimulus, and thus direction selectivity should progress towards its steady-state value. α_{SVD} allows us to indirectly analyze direction selectivity through inseparability of the transfer function. However, we can further characterize separability by considering the symmetry between quadrants 1 and 2 of the spectral transfer functions and the temporal impulse response functions separately—namely through α_s and α_t , respectively. In addition to a high α_{SVD} , corresponding to a large number of significant singular values and singular vectors, lack of separability can also arise from asymmetry in either or both of the corresponding spectral transfer functions and temporal impulse response functions.

In the steady-state, Depireux et al (2001) and Simon et al (2006) showed that most cells in AI demonstrate a high degree of temporal symmetry (small values of α_t). In the current study of the response to transients, we found that with increasing lag, α_t quickly drops to a near-zero value for most cells, and then rebounds within on average 30 msec to a slightly higher value. Note that those cells in which α_t converged quickly to a minimal steady-state value had a higher quality of recording, as measured by steady-state SNR and tMTF modulation power. We found a positive correlation between the median value of α_t measured over the period of time of significant response in the transient response and $1/\text{SNR}$ (slope = 0.5).

Since most cells initially had a largely separable transient receptive field (low α_{SVD}), we expected α_s to have a low value for small lags. Note that for fully separable receptive fields, both α_t and α_s are small. Given the low steady-state value of α_t already observed in previous papers, the dynamics of α_s would determine the degree of separability, and thus indirectly, direction selectivity. Most cells (61%) exhibited this type of behavior, whereby α_s was initially low and then climbed to a non-zero steady-state value (see Fig. 9). As was the case for α_t , the estimation of α_s was limited by the quality of the recording. The onset of the spectro-temporal feature should produce a low α_s indicating spectral symmetry, since the spectrum of incoming sounds is almost instantaneously represented in the cochlea using the unique mammalian time-frequency representation. Spectral asymmetry should therefore take a certain integration time before direction selectivity could manifest itself, as measured by α_s as a function of lag.

Discussion

In this study, we took a first step towards measuring how spectral density and temporal periodicity tunings arise and evolve as a function of lag after a sudden change in the spectro-temporal content of the envelope of a broadband sound. We measured transient modulation transfer functions at a set of chosen lags τ , by measuring the modulation of the neural response to auditory gratings as a function of the initial phase and lag. With a set of tMTFs at given lags, we analyzed how tuning dynamically evolves towards the steady-state MTF (which was measured by established steady-state linear methods). We characterized tuning dynamics with the following statistics:

- The center of mass (CM) and breadth α_b of tuning, which measures the range of densities and velocities within 1σ of CM,
- α_{SVD} , which measures the separability of the tMTF, i.e. the degree to which the tMTF can be represented as the product of a temporal and a spectral function,
- α_s , the asymmetry of the response to the spectral component(s) of the up-moving vs. down-moving components of the envelope, and
- α_t , the asymmetry of the response to the temporal component(s) of the up-moving vs. down-moving components of the envelope.

We characterized neural dynamics by the time evolution of these measures in the response to auditory transients, on a millisecond timescale. Most cells demonstrated a change in these parameters, but also convergence to a steady-state value after a remarkably short period of time.

The convergence was particularly interesting regarding the temporal symmetry parameter α_t , which for most cells evolved from a random value to a near-zero value within 30 msec after the significant response onset. While the convergence to zero is expected from the modeling in Simon et al (2006), the timescale involved in this convergence is novel. Equally relevant was the finding that the center of mass of tuning progressed from lower spectral densities to higher ones, and from lower temporal periodicities to higher ones, while the breadth of tuning around this center of mass did not change. This increase in center of mass with no change in breadth of tuning corresponds to a sharpening of the features in the corresponding dynamically changing STRF. The cells' tunings sharpened to encode the spectral envelope content more accurately.

In addition to the dynamics of spectro-temporal tuning, the short-time evolution of other properties of cortical neurons can be derived from our measures of dynamics. One such property is direction selectivity, which is not expected to emerge until a neuron has had sufficient time to detect the direction of drift of a spectral envelope. The majority (73%) of cells exhibited such behavior by virtue of their initial high degree of separability and subsequent increasing inseparability. This was seen as a concavity in the time progression of α_{SVD} , which upon response onset decreased from the noise value, and then climbed to its steady state value.

Transient evolution of spectro-temporal tuning

When using sounds with rapidly changing spectro-temporal content, the current study shows that the classical receptive field can be modified so that neural dynamics of the response to a sudden change in spectral profile can be modeled without the explicit introduction of a new timescale or new construct, as would happen if, for instance, we had uncovered the existence of two distinct and separate tunings, one onset and one sustained, and their corresponding STRFs. Even in earlier studies, we documented our inability to predict the response to sudden transitions in spectro-temporal content with the classical receptive field (Klein et al. 2006; Kowalski et al. 1996b). The current study extends the standard linear STRF model by introducing and measuring a lag-dependent modulation transfer function. The new cortical model enables us to quantitatively describe 1) how tuning evolves from the comparatively simpler collicular representation to the cortical one, and 2) how tuning to the spectral envelope continuously evolves from the onset of the envelope to the steady-state tuning. Simon et al (2006) studied temporal symmetry of steady-state cortical STRFs. The simplest model that accounts for temporal symmetry predicts the existence of two populations of cells earlier in the pathway, called *lagged* and *non-lagged*, where lagged cells have temporally phase shifted STRFs compared to canonical, non-lagged cells. This is not just a delayed response, but rather a change of phase of the oscillations of temporal impulse functions under its envelope which

is roughly the same for all cells. The effect of this change of phase would produce an initial inhibition followed by a delayed excitation; this has been observed in bat IC (Galazyuk et al. 2005; Sullivan 1982) and might be induced by the existence of modulating cortico-collicular projections (Bajo et al. 2006). On the one hand, the existence of lagged cells was postulated to explain the presence of several properties we had observed in steady-state STRFs, and on the other, the delayed input of these lagged cells into the cortical circuitry is likely to account for some of the aspects of the dynamics of tuning reported in this paper.

Dynamic Filters

In classical terms of system modeling, we note that the goals of the auditory system are several and not compatible *prima facie*. In particular, the dual goals of sound detection and sound identification have opposite requirements: detection is most easily accomplished with filters that integrate power over as wide a bandwidth as possible, whereas identification is usually accomplished through narrowly tuned filters. Therefore, it is reasonable that cortical cells might perform a continuously changing, dynamically adapting filtering capable of accomplishing both goals of detection and identification. On the other hand, tuning to spectro-temporal content need not be dynamic; it is conceivable that one neural population would detect *changes* in the spectral profile, while a different distinct population would encode the *content* of the spectral envelope. Still other coding schemes are possible, of course. Our findings indicate that at the level of AI, at least, the same cells are coding for both the detection of change in spectro-temporal content and for its coding. We found that as a population, the center of mass of tuning increases towards higher $|\Omega|$ and w , while α_b does not change. This indicates that over 30 ms or so, the STRF tuning gets sharper supporting the hypothesis stated above.

Models of cortical tuning

The model of spectral envelope feature extraction presented in this paper has augmented the basic STRF model. This extension allows better modeling of how the STRF complexity (temporal symmetry and direction selectivity, for instance) is built up in AI from the comparatively simpler thalamic or collicular receptive fields of cells that eventually project to AI. Simon et al (2006) show that neurons in ferret AI are well described in the steady state by their STRF (or equivalently their MTF), but also have a distinctive property called temporal symmetry: Every temporal cross-section of the STRF (impulse response) is the same function of time, but for an overall scaling and Hilbert rotation (a shift of phase under an fixed envelope). Temporal symmetry is highly constraining regarding possible models of functional neural connectivity within and into AI. The simplest models of the thalamocortical functional connectivity are those in which the only constraint is that thalamic inputs to an AI cell have a best frequency to within half an octave of the AI cell. Such models are ruled out because they yield STRFs that are incompatible with the constraints of the observed cortical temporal symmetry (measured by α_t). Rather, temporally symmetric models predict that the thalamic inputs can be almost unrestricted in their spectral support, but must have the same low frequency temporal structure (e.g. approximately constant amplitude and phase linearity for a few tens of Hz). The majority of cortical cells display responses that are fully separable during the first 20 msec or so of the response. Quadrant separability (but not full separability), as well as temporal symmetry, emerges thereafter in AI. This output is likely sent back to the IC (Bajo et al. 2006) and may contribute to the emergence of the hypothesized lagged cells in IC. Lagged cells would provide additional input to AI cells, contributing to the dynamics of tuning we observed. This would form the basis for the cortical lag-dependent tMTF introduced in this study, which reaches a steady-state tuning that remains relatively stable over hundreds of milliseconds to several seconds (Shechter and Depireux 2007). These findings reinforce the idea that temporal symmetry without full separability is not an automatic property of the network, but rather arises from the contribution of several populations of cells.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Yadong “KK” Ji for extensive help in animal care and data acquisition, and Sridhar Kalluri and Asaf Keller for help in preparation of this manuscript. This research was funded by NIH/NIDCD 1 RO01 DC005937 awarded to DAD. PM also received support from training grant NIH/NINDS 2T32NS007375-11

Abbreviations

AI	Primary auditory cortex
IC	inferior colliculus
MTF	modulation transfer function
tMTF	transient modulation transfer function
STRF	spectro-temporal receptive field
TORC	temporally orthogonal ripple combination

References

- Abdi, H. Singular Value Decomposition (SVD) and Generalized Singular Value Decomposition (GSVD). In: Salkind, NJ., editor. Encyclopedia of Measurement and Statistics. Thousand Oaks, CA: Sage; 2007.
- Aertsen AM, Johannesma PI. A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biological Cybernetics* 1981;42:145–156. [PubMed: 6976799]
- Bajo VM, Nodal FR, Bizley JK, Moore DR, King AJ. The Ferret Auditory Cortex: Descending Projections to the Inferior Colliculus. *Cereb Cortex*. 2006
- Bredfeldt CE, Ringach DL. Dynamics of spatial frequency tuning in macaque V1. *J Neurosci* 2002;22:1976–1984. [PubMed: 11880528]
- deCharms RC, Blake DT, Merzenich MM. Optimizing sound features for cortical neurons.[see comment]. *Science* 1998;280:1439–1443. [PubMed: 9603734]
- Depireux DA, Simon JZ, Klein DJ, Shamma SA. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophys* 2001;85:1220–1234.
- Dobbins HD, Marvit P, Ji YD, Depireux DA. Chronically Recording with a Multi-Electrode Array Device in the Auditory Cortex of an Awake Ferret. *J Neurosci Meth* 2007;159
- Escabi MA, Schreiner CE. Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *Journal of Neuroscience* 2002;22:4114–4131. [PubMed: 12019330]
- Evans EF, Whitfield IC. Classification of Unit Responses in the Auditory Cortex of the Unanaesthetized and Unrestrained Cat. *The Journal of physiology* 1964;171:476–493. [PubMed: 14193935]
- Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 2003;6:1216–1223. [PubMed: 14583754]
- Galazyuk AV, Lin W, Llano D, Feng AS. Leading inhibition to neural oscillation is important for time-domain processing in the auditory midbrain. *Journal of Neurophysiology* 2005;94:314–326. [PubMed: 15772243]
- Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsaki G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J Neurophysiol* 2000;84:401–414. [PubMed: 10899214]
- Klein DJ, Depireux DA, Simon JZ, Shamma SA. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *Journal of Computational Neuroscience* 2000;9:85–111. [PubMed: 10946994]

- Klein DJ, Simon JZ, Depireux DA, Shamma SA. Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex. *J Comput Neurosci* 2006;20:111–136. [PubMed: 16518572]
- Kowalski N, Depireux DA, Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *Journal of Neurophysiology* 1996a;76:3503–3523. [PubMed: 8930289]
- Kowalski N, Depireux DA, Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *Journal of Neurophysiology* 1996b; 76:3524–3534. [PubMed: 8930290]
- Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM. Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *Journal of Neurophysiology* 2003;90:2660–2675. [PubMed: 12815016]
- Mastroratte DN. Two classes of single-input X-cells in cat lateral geniculate nucleus. I. Receptive-field properties and classification of cells. *J Neurophysiol* 1987a;57:357–380. [PubMed: 3559684]
- Mastroratte DN. Two classes of single-input X-cells in cat lateral geniculate nucleus. II. Retinal inputs and the generation of receptive-field properties. *J Neurophysiol* 1987b;57:381–413. [PubMed: 3559685]
- Miller LM, Escabi MA, Schreiner CE. Feature selectivity and interneuronal cooperation in the thalamocortical system. *Journal of Neuroscience* 2001;21:8136–8144. [PubMed: 11588186]
- Redish, AD. MClust: a spike-sorting toolbox, freely available software .
<http://www.cbc.umn.edu/~redish/mclust>
- Saul AB, Humphrey AL. Spatial and temporal response properties of lagged and nonlagged cells in cat lateral geniculate nucleus. *Journal of Neurophysiology* 1990;64:206–224. [PubMed: 2388066]
- Schafer M, Rubsamen R, Dorrscheidt GJ, Knipschild M. Setting complex tasks to single units in the avian auditory forebrain. II. Do we really need natural stimuli to describe neuronal response characteristics? *Hear Res* 1992;57:231–244. [PubMed: 1733915]
- Schreiner CE, Calhoun BM. Spectral envelope coding in cat primary auditory cortex: Properties of ripple transfer functions. *Aud Neurosci* 1994;1:39–61.
- Sen K, Theunissen FE, Doupe AJ. Feature analysis of natural sounds in the songbird auditory forebrain. *Journal of Neurophysiology* 2001;86:1445–1458. [PubMed: 11535690]
- Shechter B, Depireux DA. Response adaptation to broadband sounds in primary auditory cortex of the awake ferret. *Hear Res* 2006;221:91–103. [PubMed: 16982164]
- Shechter B, Depireux DA. Stability of spectro-temporal tuning over several seconds in primary auditory cortex of the awake ferret. *Neuroscience* 2007;148:806–814. [PubMed: 17693032]
- Simon JZ, Depireux DA, Klein DJ, Fritz JB, Shamma SA. Temporal Symmetry in Primary Auditory Cortex: Implications for Cortical Connectivity. *Neural Computation*. 2006 Accepted.
- Spinks RL, Baker SN, Jackson A, Khaw PT, Lemon RN. Problem of dural scarring in recording from awake, behaving monkeys: a solution using 5-fluorouracil. *J Neurophysiol* 2003;90:1324–1332. [PubMed: 12904511]
- Sullivan WE 3rd. Possible neural mechanisms of target distance coding in auditory system of the echolocating bat *Myotis lucifugus*. *J Neurophysiol* 1982;48:1033–1047. [PubMed: 7143031]
- Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network-Computation in Neural Systems* 2001;12:289–316.
- Valentine PA, Eggermont JJ. Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex. *Hear Res* 2004;196:119–133. [PubMed: 15464309]
- Wang X, Lu T, Snider RK, Liang L. Sustained firing in auditory cortex evoked by preferred stimuli. *Nature* 2005;435:341–346. [PubMed: 15902257]
- Yeshurun Y, Wollberg Z, Dyn N, Allon N. Identification of MGB cells by Volterra kernels. I. Prediction of responses to species specific vocalizations. *Biol Cybern* 1985;51:383–390. [PubMed: 3995095]

Fig. 1.

Spectro-temporal reverse correlation with TORC stimuli. (Top) Three typical TORC spectro-temporal envelopes. They are the sums of gratings with spectral densities 1 cyc/oct (left), 3/7 cyc/oct (middle), and 1/7 cyc/oct (right). Each stimulus depicted contained temporal periodicities from 4 to 32 Hz in steps of 4 Hz. Presentation of each stimulus lasted for 6 seconds (24 periods of 250 msec each). Inset are the two dimensional Fourier transforms of the stimuli. Below each stimulus is a cartoon representation of the spike trains they elicit. (Middle) Reverse correlation of the spike trains with the stimulus spectro-temporal envelopes. For each spike event, we average a full period of the stimulus envelope that preceded it to obtain the cell's STRF (Bottom). The cell's modulation transfer function (MTF) is the two dimensional Fourier

transform of its STRF. Note that the MTF is complex-conjugate symmetric, where quadrants I and III and quadrants II and IV are equal in amplitude, but opposite in phase.

Fig. 2.

(Top) Spectro-temporal envelope of a transient grating stimulus. The stimulus has a flat envelope with eight 50 msec transient segments of grating modulation interspersed. Each transient has the same density (Ω) and temporal periodicity (w), but starts at a different phase (ϕ). The actual stimulus has random intertransient intervals. (Middle) The response of a neuron to the stimulus shown at top, aligned to the stimulus onset time. We compute transient modulation transfer functions (tMTF) for each lag, τ . For a given τ and for each starting phase, we compute the average spiking rate in a window of duration $(8 \cdot w)^{-1}$ sec. These windows for a specific choice of τ are depicted by the gray boxes overlying the response curve. (Bottom) For the same τ , the spiking rates evoked by each phase of the modulation are Fourier transformed and the amplitude and phase of the first harmonic extracted. This is a measure of phase-locking to modulations in the stimulus at τ msec after the transient onset for the (Ω , w) pair presented.

Fig. 3.

Equivalence for steady-state sounds of the standard method of deriving STRFs with the instantaneous method used in this paper. In the standard method (bottom right), we measure the phase and amplitude of the neural response for a full period of the stimulus (denoted by 'A'). Alternatively, for an instantaneous measurement (bottom left), we present our continuous or long duration sound 4 times, each with a different starting phase and at a given time, measure the cell's response over 1/4 of a cycle (denoted by '1'–'4'). Concatenating these 1/4 cycle responses yields the cell's tuning for a full cycle of the response, but at a specific moment in time. This method is used in the paper with 8 starting phases instead of 4.

Fig. 4.

Raster plot of responses to transient gratings. Each waveform is presented 10 times, with the grating transients presented at 8 starting phases (phase values shown above the rasters). Each transient is 50 msec in duration. Each dot represents an action potential. The gray line is a PSTH, measured with 25 msec bins. The vertical dotted lines indicate the times at which each feature transient began. The solid vertical lines at 0 msec and 1250 msec indicate the times at which the sound was turned on and off, respectively. The 2 cells shown here and used in the following figures are examples of the 2 broad categories of cells we have found—those with dynamics in their tuning (cell 35) and those without dynamics (cell 47).

Fig. 5.

(A,D) Lag Dependent Transient Receptive Fields are shown for two representative cells in auditory cortex. Each frame shown is the inverse Fourier transform of the tMTF at 5 msec interval lags, which was computed using the method depicted in Fig. 2. Cell 35 (A) shows tuning with sideband inhibitory regions at intermediate lags (from $\tau = 20$ msec to $\tau = 40$ msec), but these regions are not seen in the steady-state. Cell 47 (D) has tuning which exhibits an accumulation of direction selectivity with increasing lag. (B,E) The steady-state STRFs obtained through reverse correlation with TORC stimuli for the same cells in A and D, respectively. (C,F) The total power in the transient modulation transfer functions is plotted as a function of lag for the two cells. This value is used to determine whether there is a significant response to the transient gratings. The thresholds are plotted by the dashed lines at 10% of the maximum response from baseline and points above threshold are stressed in **bold**. Horizontal gray bars indicate the 16-msec analysis windows used to compute the trends for the α parameters. Data for an additional 4 cells are shown in supplemental figures.

Fig. 6.

The center of mass was computed for the tMTF at each lag in the quadrant which had the greater power throughout the response. We fit the temporal progression linearly (both in spectral density Ω and in temporal periodicity ω) at the lag corresponding to the first significant peak in the modulation power (gray bars in B and C). A) The distributions of fit slopes describing the change in center of mass with increasing lag are shown in the histograms (left: Ω , right: ω). Both quantities increased for most cells (70% for Ω and 68% for ω). B,C) Center of mass (left: Ω , right: ω) as a function of lag for Cells 35 and 47 (Fig. 5). Lags at which the response was significant (determined from the total modulation power) are indicated in bold.

Fig. 7.

α_b was computed for the tMTF at each lag. The best linear fit to $\alpha_b(\tau)$ was found at the first significant peak in the modulation power (indicated by the gray bars in B and C). A) The distribution of fit slopes for all cells analyzed in this study are depicted. 47% broadened their tuning and 53% sharpened tuning as a function of lag. B,C) Example traces of $\alpha_b(\tau)$ for Cells 35 and 47 (see Fig. 5). Lags for which the response was significant are indicated in bold.

Fig. 8.

α_{SVD} was computed for the tMTF at each lag, and its temporal progression, $\alpha_{SVD}(\tau)$ was fit with a second-order polynomial. A) The distribution of the second order coefficients τ^2 for all cells analyzed in this study are shown in the histogram. The majority of cells (73%) had a positive second order coefficient, corresponding to a concave behavior, in which α_{SVD} first decreases and then rises back to either a steady-state or continues to increase. B,C) Example traces of α_{SVD} for Cells 35 and 47 (Fig. 5). Both cells exhibit the behavior described above. The significant part of the response is indicated in bold. The dashed line indicates the value of α_{SVD} obtained from the steady-state STRF. The first significant peaks of the modulation powers for the two cells are indicated by the horizontal gray bars.

Fig. 9.

A) α_t (left) and α_s (right) for the tMTF at each lag for Cell 35. Both α_t and α_s quickly decay to a near-zero value, and α_s gradually increases to the steady-state value. B) Same as in A for Cell 47. Both α_t and α_s decay to a near-zero value again, but with a longer latency. In contrast to Cell 35, α_s reaches a higher steady-state value at a shorter latency. Significant responses are indicated in bold, and the dashed line indicates the steady-state value.