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

Authors
Shechter, B.[1], Dobbins, H.D.[1,2], Marvit, P.[1], Depireux, D.A.[1,2]

Affiliations
[1] Department of Anatomy and Neurobiology, School of Medicine, University of Maryland, Baltimore, 20 Penn St. HSF II Rm. S251, Baltimore, MD 21201
[2] Program in Neuroscience, School of Medicine, University of Maryland, Baltimore, Baltimore, MD 21201

Running Head
Spectro-temporal Tuning Dynamics in Awake Ferret AI

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
ddepi001@umaryland.edu

Abstract

In this paper, we explore spectro-temporal tuning in primary auditory cortex following a sudden change in spectro-temporal content. We previously characterized the steady-state spectro-temporal tuning properties of cortical cells with respect to broadband sounds by presenting sounds with a spectro-temporal envelope of constant spectral density and angular frequency for several seconds. However, since speech and other natural sounds have spectro-temporal features that change substantially over milliseconds, we studied the dynamics of tuning by using stimuli of constant level but alternating between a flat spectro-temporal envelope and a modulated envelope with well defined spectral density and angular frequency. This allowed us to define the tuning of cortical cells to speech-like and other rapid transitions, at the order of milliseconds, as well as the time evolution of this tuning in response to the features of a new sound. We find that cortical neurons can be characterized by a lag-dependent modulation transfer function. This characterization, when measured through to steady-state, is equivalent to the classical spectro-temporal receptive field.

Keywords

Auditory gratings, ripple, AI, Auditory Cortex, Dynamics.

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,

Introduction

Auditory cortical neurons are tuned to specific aspects of the spectro-temporal content of the sounds an animal is exposed to. Is the first measurable tuning already at its steady state value, or does the tuning evolve in time? Our previous work in the primary auditory cortex (AI) of the ferret (*Mustela Putorius furo*) has characterized the steady-state neural responses to ongoing broadband sounds with well-defined spectro-temporal content. Using auditory gratings, or ripples, to construct an essentially linear model called the spectro-temporal receptive field (STRF), we characterized the responses of AI neurons and successfully predicted their steady-state responses to new stationary sounds (Depireux et al. 2001; Fritz et al. 2003; Kowalski et al. 1996a; Shechter and Depireux 2006; 2007). In these studies, stimuli with fixed spectro-temporal content were presented for several seconds, and only the steady-state response was used to construct an STRF. Other classes of broadband stimuli have been developed to obtain 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 in frequency and followed in time by inhibition. Our long-term goal is to understand the coding of natural sounds, and particularly speech, which is both broad-band and has rapidly changing spectro-temporal content. However, to understand the coding of such sounds which have complex statistics, we first study the encoding of stimuli with a known and precisely defined structure but are rapidly changing over short timescales.

One of the primary methods used to study auditory cortical neurons employs short duration pure tone stimuli of varying levels and frequencies. The 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, broad-band sound studies typically use the neural response to an on-going, long duration sound. However, 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 have an 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 *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. 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.

Investigators of primary visual cortex have studied dynamics of neural timing in similar ways. Bredfeldt and Ringach (2002) measured the dynamics of spatial frequency tuning using reverse

correlation with respect to rapidly changing spatial luminance gratings. They found that tuning in most V1 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, as opposed to moving 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 in a recent paper 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 showed that the separability of the cortical receptive fields is best explained by the existence of lagged cells earlier in the auditory pathway. The lagged cells are characterized by an excitation delayed by tens of milliseconds. Similar results are found in the visual literature: Mastronarde (1987a; 1987b) and Saul and Humphrey (1990) showed the existence of lagged and non-lagged cells in the A-layers of the cat LGN, each class having different temporal response and their staggered input contributions into cortex imply that the tuning of a cortical cell should evolve in time. We therefore hypothesized that the coding in auditory cortex is best characterized by a modulation transfer function evolving in time. Unpublished data from our lab on lagged cells in the inferior colliculus implies that convergence to the steady-state should occur over tens of milliseconds. The determination of the actual timescale over which the symmetries analyzed by Simon et al evolve over time will eventually provide a better understanding of how the functional connectivity that gives rise to cortical STRFs arises and evolves in time.

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). A detailed description can be found in (Dobbins et al. 2006). For surgical preparation, ferrets were fully anesthetized with Halothane (3% induction, then 1.75% maintenance adjusted to keep heart rate, respiration, end-tidal CO_2 and SpO_2 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 rostro-caudally along the midline from the nuchal crest to above a line joining the eyes. The scalp was retracted and the temporalis muscles were resected bilaterally. Stainless steel screws were inserted around the skull to anchor the subsequent head post, multi-electrode microdrive, and dental cement used to fix the experimental apparatus. The head post was positioned rostrally over the skull and affixed with a small amount of dental cement. A craniotomy was made unilaterally, on the left, over AI. To prevent re-growth and toughening of the dura, the mitotic inhibitor 5-flurouracil was applied (Dobbins et al. 2006; Spinks et al. 2003). The microdrive was slowly lowered into position, and mechanically bonded to the skull with dental cement. The electrode exit geometry was in a honeycomb pattern over AI, in such a way that the minimum distance between adjacent electrodes was 225µm. 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 head post 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 about an hour by providing the ferret with treats (e.g., Ferretone). Nevertheless, some level of drowsiness cannot be excluded in our recordings.

Recording sessions typically lasted 3-4 hours. Activity of single neurons was recorded with 6 to 12 parylene-coated tungsten microelectrodes simultaneously (initial impedance 3.5-6MΩ at 1kHz, shaft diameter 76μm, Micro Probe, Inc, Gaithersburg, MD). Electrodes were individually advanced by manually turning a screw. Because the electrode system is sealed, it is not possible to determine with certainty when the electrodes entered the brain. Therefore, the depth of recording for each electrode was estimated relative to the electrode position at which spikes were first detected.

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 and second 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 digitally 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 then 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—also called ripples—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 angular frequency (*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 angular frequency, 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

$$S(x,t) = L\left[\, 1 + \Delta A \cdot \Sigma_i \cos\left(2\pi\left(\Omega \cdot x + w_i \cdot t\right) + \phi_i\right) \right], \tag{1}$$

which specifies a linear modulation. In the equation, the frequency of each tone component is given by *x*, where $x = \log_2\left(f / f_0\right)$ 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 grating are logarithmically spaced, so that the pitch of the overall stimulus is indeterminate. Typical TORC spectro-temporal envelopes are illustrated in Fig. 1.

[Fig. 1 to be inserted around here]

Stimulus Set—Transient Tuning

To measure the cells' tuning dynamics, we used transient grating stimuli as illustrated in Fig. 2. These stimuli are broadband, spanning 5 octaves, and they 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, angular frequency and starting phase.

In a given waveform, the 8 transients have the same density and angular frequency, but the starting phases are chosen from a random permutation of $\{2\pi \cdot x/8,\ x = 0,1,\ldots 7\}$. 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, but hard-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.

Once generated, the spectro-temporal envelope is then used to determine the amplitude of 100 tones per octave over 5 octaves as a function of the time. The 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 spectro-temporally well-defined feature transients. There are 63 transient grating stimuli: Spectral densities of the transients ranged from -2 cyc/oct to 2 cyc/oct in steps of 0.5 cyc/oct and temporal angular frequencies from 0 Hz to 30 Hz in steps of 5 Hz, with a 100% modulation depth.

[Fig. 2 to be inserted around here]

Data Analysis

The derivation of the *steady-state* STRF from the TORCs is now standard and will not be repeated here (Klein et al. 2000). 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 the cell's response 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, which is the focus of this paper, we are concerned with measuring the modulation transfer function, or equivalently the receptive field, at an instant in time. 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 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 (4 are shown) 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 8 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), assuming the response was in a steady-state regime.

[Fig. 3 to be inserted around here]

An example of the response to the feature transient stimuli is shown in Fig. 4. Following the transitions from a flat spectro-temporal envelope to a modulated one, the two exemplar cells shown in Fig. 4 respond to some of the transient changes in a way that depends on the spectro-temporal envelope of the stimulus (Ω and w), and the corresponding relative position of the neuron on the tonotopic axis, and as a function of time or lag τ since 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). This measure of time τ is defined as the lag, or the duration since the last change to a 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 angular frequency 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 $\tau = 10\,\text{msec}$, 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 of the transients. 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 elapsed to the center of the analysis window. The

measurement is a best estimator of the spike rate at the center of the window, which changes based on *w* for given τ. 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 tMTF for each lag relates to the steady-state MTF, obtained as a Fourier transform of the standard 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 to study. In particular, we will consider the dynamics of the *center of mass* of tuning and the *breadth* of tuning around this center of mass, to determine whether the average tuning changed and whether it broadened or sharpened. We will also examine the dynamics of quadrant *separability* and symmetry of the *spectral* and *temporal* transfer functions, since our previous studies pointed to unexpected results with respect to separability and temporal symmetry of STRFs.

Note that the tMTF has conjugate symmetry, and therefore 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 (see Fig. 1). With this in mind, 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 angular frequency. It is computed in the quadrant with the greater total modulation power.

2. **Sharpness 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 a normalized distance from the center of mass of the modulation transfer function as:

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

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 (defined below), and $(\Omega_{\max}, w_{\max})$ are the maximum spectral density and angular frequency 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})**: Separability is an important property of transfer functions. 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 (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

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

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 receptive field will not be directionally selective at short lags, and this selectivity will only manifest at 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

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

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

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 (resp. F_1 and F_2) are, the closer the absolute value in the equations above 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 explore the time evolution of α_t as it reaches the symmetric steady-state value.

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:

$$\mathrm{SNR}_{\Omega,w} = \frac{\left| \sum_{bootstraps} \psi(\Omega, w) \right|^2}{N_{boot} \cdot \sigma_\psi^2} \quad (7)$$

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

Only receptive fields which had an $SNR > 0.5$ were included in the later analyses.

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.

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

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 $0.1\ spikes^2/sec^2$, 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 183 single-unit recordings in 3 ferrets. Of these cells, 92 (50%) showed reliable steady-state phase-locking to modulations in the stimulus, as measured by an SNR larger than 0.5; in other words, the rejected units were either poorly phase-locked or were inconsistent

from trial to trial. Furthermore, we found that cells with low SNR (< 0.5) also did not exhibit a significant response to the transient gratings.

With respect to the transient gratings, cells were considered to have a significant response if they met the criterion (see Methods) that the total power in the transient modulation transfer function (tMTF) exceeded threshold continuously for at least 30 msec. We found that 57 cells (62%) out of the 92 that had a reliable steady-state characterization met all criteria for significant response to the transient gratings. The reasons for this yield vary. In the response to transients, 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 ($< 0.1\ spikes^2/sec^2$).

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 a lot of variability from cell to cell and these numbers only serve to indicate that on average, 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 illustrates its preference to the spectro-temporal content of ongoing sounds. In this paper we develop a method of analyzing responses with respect to the onset of novel spectro-temporal features. Here, we expose the dynamics of a neuron's receptive field with respect to the onset of feature transients. Fig. 5 shows the evolution of tuning for two cells at multiple lags after the onset of a feature transient. As mentioned earlier, these 2 cells are representative of two broad categories of cells we found—those with dynamics in their tuning (cell 35) and those without dynamics (cell 47). We compare these evolutions to their steady-state counterparts. The inverse Fourier transform of the tMTF for **Cell #35** (Fig. 5A) exhibits an excitatory region (which corresponds 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). It is interesting to note that 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 dynamical 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 a steady-state starting near a lag of 25 ms, this cell develops direction selectivity. The direction selectivity is evident from the excitatory and inhibitory regions assuming an oblique orientation. For longer lags, the tSTRF becomes increasingly similar to the steady-state (Fig. 5E).

We show the total modulation power of the tMTFs for these two cells in Fig. 5C,F, which was used to determine the lags for which the cells were responding to the feature transients. The significance threshold for each is depicted by the dashed line, and the total modulation power for those lags τ where the threshold was exceeded are stressed in bold. The 16-msec analysis windows in which we computed the trends for the α parameters used to characterize the tMTFs are indicated by the gray bars at the bottom of the graphs.

[Fig. 5 to be inserted around here]

Center of Mass

Among the ideas put forth in this paper, we hypothesized that the envelope is encoded in cortex dynamically, which we characterize with a lag dependent modulation transfer function. In other words, 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 angular frequency, 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 angular frequency) 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 angular frequency (*w*) (see Fig. 6A).

[Fig. 6 to be inserted around here]

Sharpness of Tuning

To analyze sharpness of tuning in spectral density and temporal periodicity space, we computed α_b in the quadrant of the tMTF with the greater modulation power for every lag τ. This parameter quantifies the spread of power around its center of mass—effectively a weighted measure of variance. A transfer function with a large spread will be broadly tuned (in Ω-*w* space), and therefore have a large value of α_b, and 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 Ω-*w* space, i.e. in the tMTF, will correspond to a broadening of the corresponding tuning in x-t space, i.e. the STRF. 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.

Since the analysis window is of finite duration, the response measure is an average, as opposed to an instantaneous value. Therefore, there is a transition period from noise to signal when the cell starts responding. 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 one of the former category while Cell #47 represents one of the latter category.

[Fig. 7 to be inserted around here]

Separability

Direction selectivity in a cells' steady-state response does not imply direction selectivity in the response close to the onset. Depending on the spectral density and angular velocity 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. 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).

[Fig. 8 to be inserted around here]

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. At small lags, it is not possible for a cell to distinguish between upward and downward moving gratings, because there is an ambiguity in measuring the grating direction from a sample of the stimulus that is too short. Therefore, the tMTF is highly separable at small lags, corresponding to the initial absence of direction selectivity. The separability of the transfer function is measured by α_{SVD}, with values near zero corresponding to higher degrees of separability. For lags before the total modulation power becomes significant, the cell is responding to white noise (flat spectro-temporal envelope). The value of α_{SVD} for this initial period was random with a high mean. The majority of cells (73%) had an α_{SVD} which initially decreased from the initial random value, since there cannot be directionality in the first few milliseconds of the response. Then, α_{SVD} increased for a short period and finally 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 terms 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 acquired some directionality with increasing lag. The absence of highly negative values of τ^2 confirms that no cells had a progression, indicated by α_{SVD}, that started separable before feature onset, progressed to inseparable during the first few milliseconds after

feature onset, and then returned to separable at long lags; the absence of such behavior confirms our physical intuition that inseparability cannot appear at short lags.

Spectral and temporal symmetry, α_s and α_t

Direction selectivity is a property of the auditory response which 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 to analyze the stimulus, and thus direction selectivity should progress towards its steady-state preference of one direction over the other. 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 function and temporal impulse response function.

In the steady-state, we (Depireux et al. 2001; Simon et al. 2006) showed that most cells in AI demonstrate a high degree of temporal symmetry (small values of α_t). In the transient response we found that with increasing lag, α_t quickly drops to a near-zero value for most cells, and then rebounds within about 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 and $1/\mathrm{SNR}$ (slope = 0.5).

Since most cells initially had an 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, the dynamics of α_s would determine the degree of separability, and thus indirectly, its direction selectivity. Most cells

(61%) exhibited this type of behavior, where 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, measured by α_s as a function of lag.

[Fig. 9 to be inserted around here]

Discussion

In this study, we took a first step towards measuring how spectral density and angular frequency 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 a transient modulation transfer function at a set of chosen lags τ, by measuring the modulation of the neural response to gratings as a function of the initial phase and lag. With a set of tMTFs for a set of lags, we analyzed how the tMTF dynamically evolves as a function of lag towards the steady-state MTF measured by steady-state linear methods. These dynamics were characterized with the following statistics:

- The center of mass (CM) and breadth α_b of the tuning, that is the range of densities and velocities within 1 σ of CM,
- α_{SVD}, which describes how separable, i.e. how close to a simple product of a function of time and a function of density the tMTF is
- α_s, the asymmetry of the response to the spectral components of the up-moving *vs.* down-moving components of the envelope, and
- α_t, the asymmetry of the response to the temporal components 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 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 densities to higher ones, and from lower periodicities or angular frequencies to higher ones, as the cells' tuning 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 onset 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 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 temporal oscillations of temporal impulse functions under an 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 cell 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 (i.e., 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 the 30 ms or so of our analysis window, the STRF tuning gets sharper confirming 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 are ruled out. 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 α_t. Rather, temporal symmetry models predict that the thalamic inputs can be almost unrestricted in their spectral support, but they 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 significant response. Quadrant separability (but not full separability), as well as temporal symmetry, emerges thereafter in AI. This output is sent back to the IC (Bajo et al. 2006) and may contribute to the emergence of hypothesized lagged cells in IC. These 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.

Acknowledgment

The authors thank Yadong "KK" Ji for extensive help in animal care and data acquisition, and Sridhar Kalluri for improving the manuscript considerably. This research was funded by NIH/NIDCD 1 RO01 DC005937 awarded to DAD. PM also received support from training grant NIH/NINDS 2T32NS007375-11

References

Abdi H. Singular Value Decomposition (SVD) and Generalized Singular Value Decomposition (GSVD). In: *Encyclopedia of Measurement and Statistics*, edited by Salkind NJ. Thousand Oaks, CA: Sage, 2007.

Aertsen AM, and Johannesma PI. A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biological Cybernetics* 42: 145-156, 1981.

Bajo VM, Nodal FR, Bizley JK, Moore DR, and King AJ. The Ferret Auditory Cortex: Descending Projections to the Inferior Colliculus. *Cereb Cortex* 2006.

Bredfeldt CE, and Ringach DL. Dynamics of spatial frequency tuning in macaque V1. *J Neurosci* 22: 1976-1984, 2002.

deCharms RC, Blake DT, and Merzenich MM. Optimizing sound features for cortical neurons.[see comment]. *Science* 280: 1439-1443, 1998.

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

Dobbins HD, Marvit P, Ji Y, and Depireux DA. A chronic multi-electrode implant for recording in an awake ferret cortex. *J Neurosc Methods* accepted: 2006.

Escabi MA, and Schreiner CE. Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *Journal of Neuroscience* 22: 4114-4131, 2002.

Evans EF, and Whitfield IC. Classification of Unit Responses in the Auditory Cortex of the Unanaesthetized and Unrestrained Cat. *The Journal of physiology* 171: 476-493, 1964.

Fritz J, Shamma S, Elhilali M, and Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6: 1216-1223, 2003.

Galazyuk AV, Lin W, Llano D, and Feng AS. Leading inhibition to neural oscillation is important for time-domain processing in the auditory midbrain. *Journal of Neurophysiology* 94: 314-326, 2005.

Harris KD, Henze DA, Csicsvari J, Hirase H, and Buzsaki G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J Neurophysiol* 84: 401-414, 2000.

Klein DJ, Depireux DA, Simon JZ, and Shamma SA. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *Journal of Computational Neuroscience* 9: 85-111, 2000.

Klein DJ, Simon JZ, Depireux DA, and Shamma SA. Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex. *J Comput Neurosci* 20: 111-136, 2006.

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

Kowalski N, Depireux DA, and Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *Journal of Neurophysiology* 76: 3524-3534, 1996b.

Linden JF, Liu RC, Sahani M, Schreiner CE, and Merzenich MM. Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *Journal of Neurophysiology* 90: 2660-2675, 2003.

Mastronarde DN. Two classes of single-input X-cells in cat lateral geniculate nucleus. I. Receptive-field properties and classification of cells. *J Neurophysiol* 57: 357-380, 1987a.
Mastronarde 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* 57: 381-413, 1987b.
Miller LM, Escabi MA, and Schreiner CE. Feature selectivity and interneuronal cooperation in the thalamocortical system. *Journal of Neuroscience* 21: 8136-8144, 2001.
Redish AD. MClust: a spike-sorting toolbox, freely available software http://www.cbc.umn.edu/~redish/mclust.
Saul AB, and Humphrey AL. Spatial and temporal response properties of lagged and nonlagged cells in cat lateral geniculate nucleus. *Journal of Neurophysiology* 64: 206-224, 1990.
Schafer M, Rubsamen R, Dorrscheidt GJ, and 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* 57: 231-244, 1992.
Sen K, Theunissen FE, and Doupe AJ. Feature analysis of natural sounds in the songbird auditory forebrain. *Journal of Neurophysiology* 86: 1445-1458, 2001.
Shechter B, and Depireux DA. Response adaptation to broadband sounds in primary auditory cortex of the awake ferret. *Hear Res* 221: 91-103, 2006.
Shechter B, and Depireux DA. Stability of spectro-temporal tuning over several seconds in primary auditory cortex of the awake ferret. *Neuroscience* 148: 806-814, 2007.
Simon JZ, Depireux DA, Klein DJ, Fritz JB, and Shamma SA. Temporal Symmetry in Primary Auditory Cortex: Implications for Cortical Connectivity. *Neural Computation* Accepted: 2006.
Spinks RL, Baker SN, Jackson A, Khaw PT, and Lemon RN. Problem of dural scarring in recording from awake, behaving monkeys: a solution using 5-fluorouracil. *J Neurophysiol* 90: 1324-1332, 2003.
Sullivan WE, 3rd. Possible neural mechanisms of target distance coding in auditory system of the echolocating bat Myotis lucifugus. *J Neurophysiol* 48: 1033-1047, 1982.
Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, and Gallant JL. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network-Computation in Neural Systems* 12: 289-316, 2001.
Valentine PA, and Eggermont JJ. Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex. *Hear Res* 196: 119-133, 2004.
Wang X, Lu T, Snider RK, and Liang L. Sustained firing in auditory cortex evoked by preferred stimuli. *Nature* 435: 341-346, 2005.
Yeshurun Y, Wollberg Z, Dyn N, and Allon N. Identification of MGB cells by Volterra kernels. I. Prediction of responses to species specific vocalizations. *Biol Cybern* 51: 383-390, 1985.

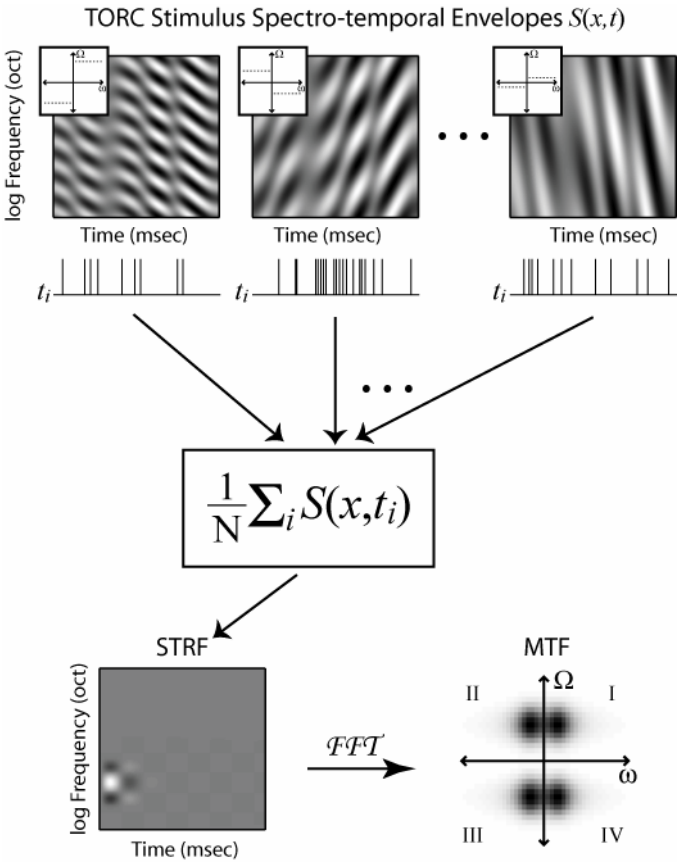

Fig. 1. Spectro-temporal reverse correlation with TORC stimuli. (Top) Three typical TORC spectro-temporal envelopes. They are the sums of gratings with spectral densities 1 cyc/oct (left), 3/7 cyc/oct (middle), and 1/7 cyc/oct (right). Each stimulus depicted contained temporal angular frequencies from 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.

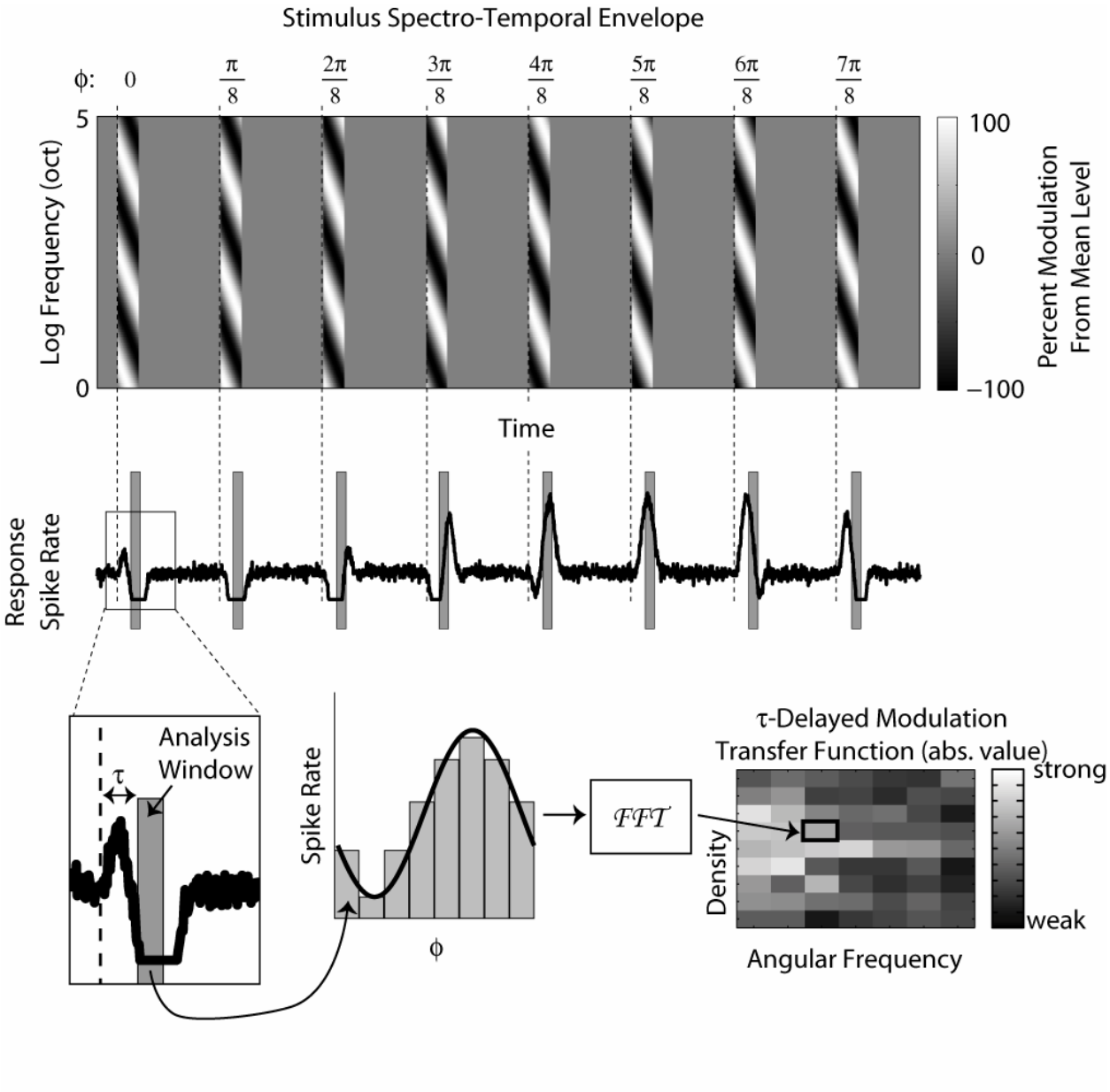

Fig. 2. (Top) Spectro-temporal envelope of a transient grating stimulus. The stimulus has a flat envelope with eight 50 msec transient segments of grating modulation interspersed. Each transient has the same density (Ω) and angular frequency (w), but starts at a different phase (ϕ). The actual stimulus has random interstimulus 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.

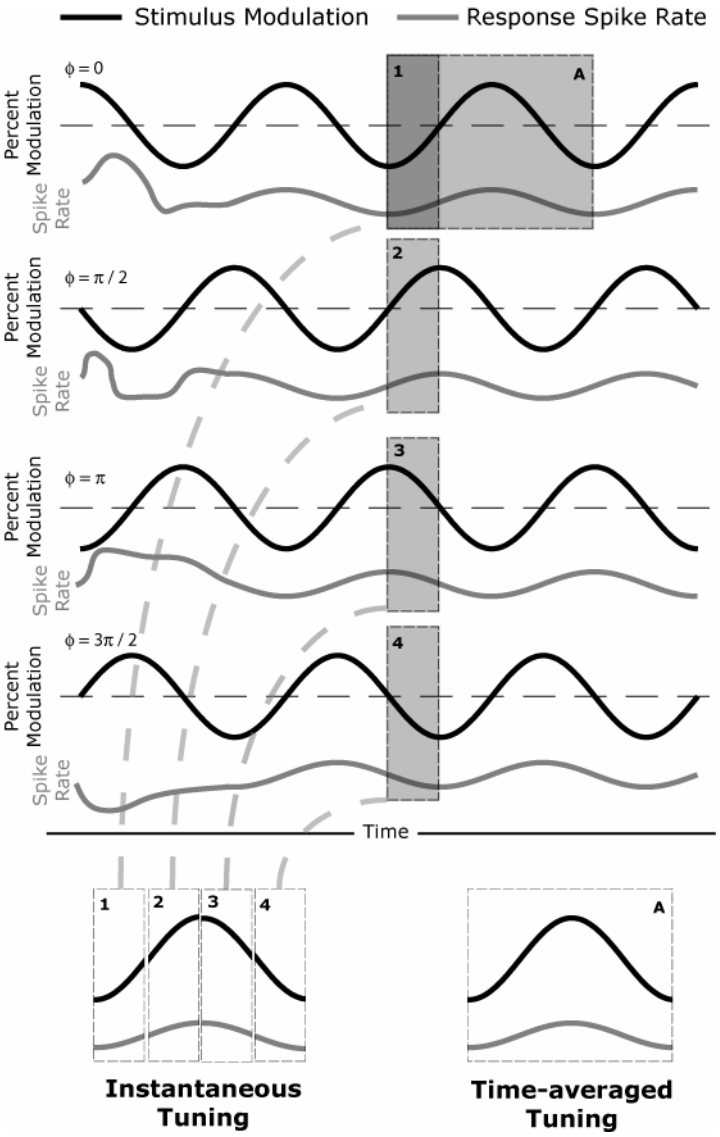

Fig. 3. Equivalence of the standard method of deriving STRFs with the instantaneous method used in this paper, for steady-state sounds. In the standard method (bottom right), we measure the phase and amplitude of the neural response for a full period of the stimulus. Alternatively, for an instantaneous 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. 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.

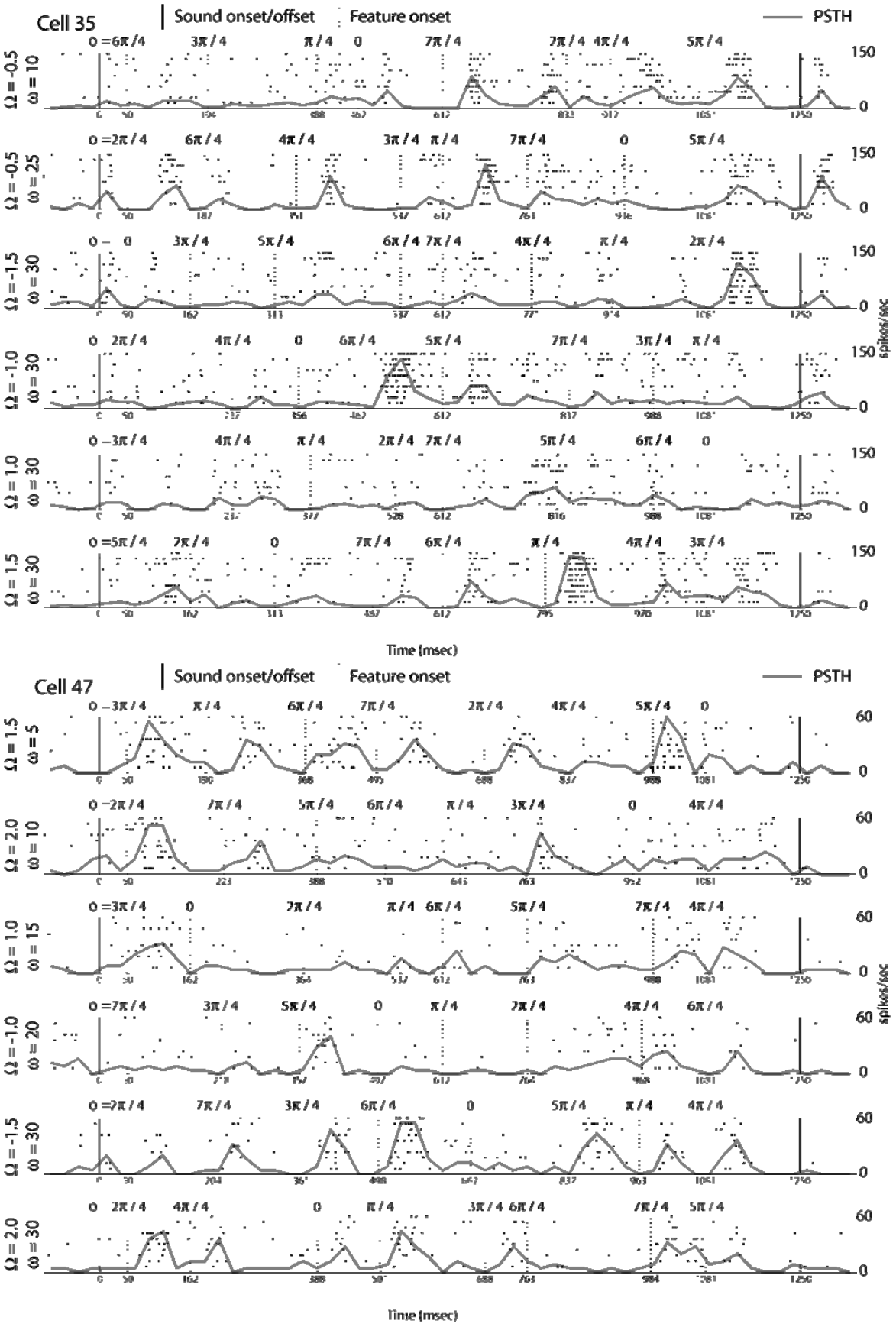
Cell 35
Sound onset/offset
Feature onset
PSTH
Ω = -0.5 ω = 10
Ω = -0.5 ω = 25
Ω = -1.5 ω = 30
Ω = -1.0 ω = 30
Ω = 1.0 ω = 30
Ω = 1.5 ω = 30
150
0
spikes/sec
Time (msec)
Cell 47
Sound onset/offset
Feature onset
PSTH
Ω = 1.5 ω = 5
Ω = 2.0 ω = 10
Ω = 1.0 ω = 15
Ω = -1.0 ω = 20
Ω = -1.5 ω = 30
Ω = 2.0 ω = 30
60
0
spikes/sec
Time (msec)

Fig. 4. Raster plot of responses to a few transient gratings. Each waveform is presented 10 times, with the grating transients presented at 8 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).

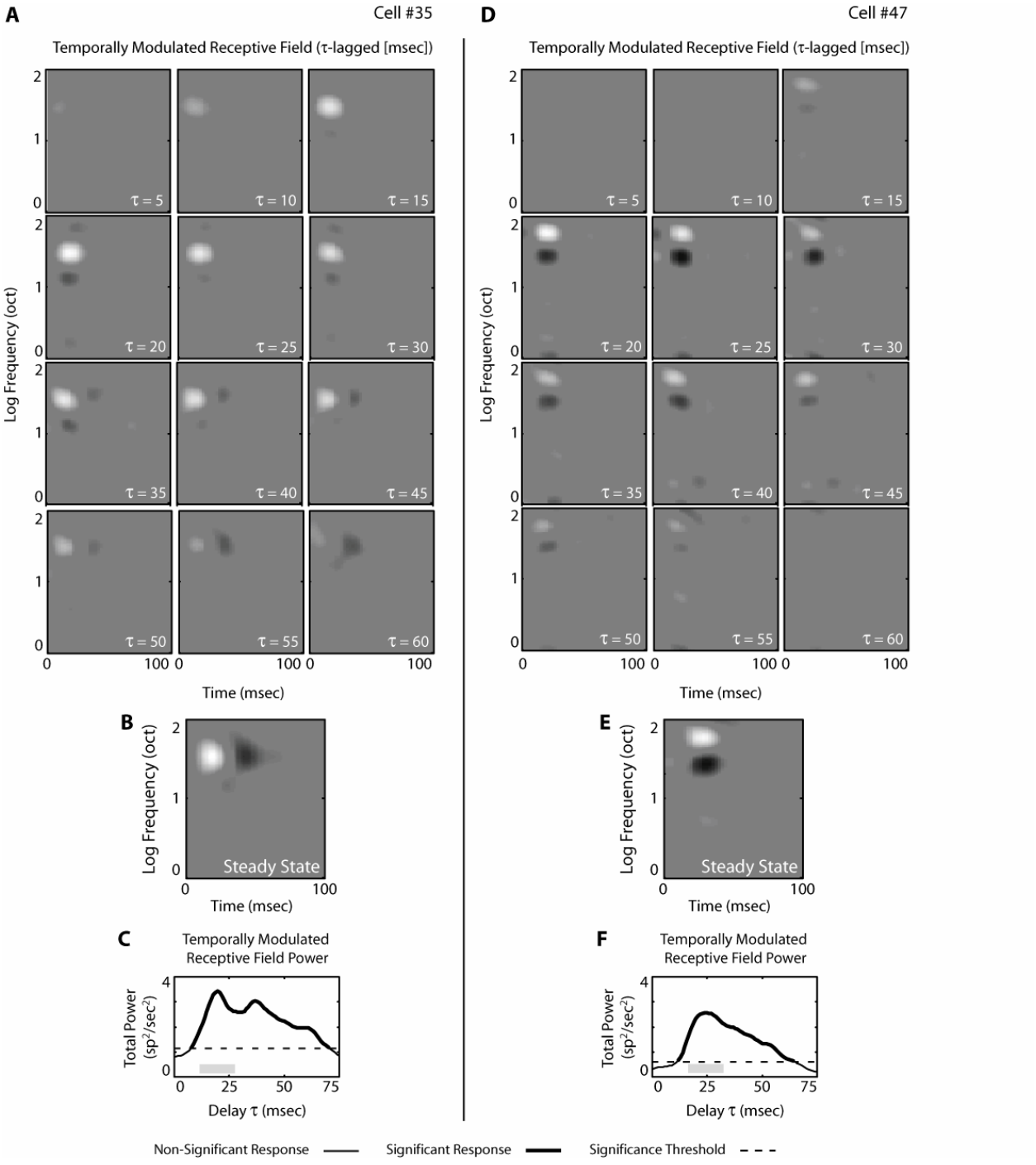

Fig. 5. (A,D) Lag Dependent Transient Receptive Fields are shown for two representative cells in auditory cortex. Each frame shown is the inverse Fourier transform of the tMTF at 5 msec interval lags, which was computed using the method depicted in Fig. 2. Cell 35 (A) shows tuning with sideband inhibitory 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.

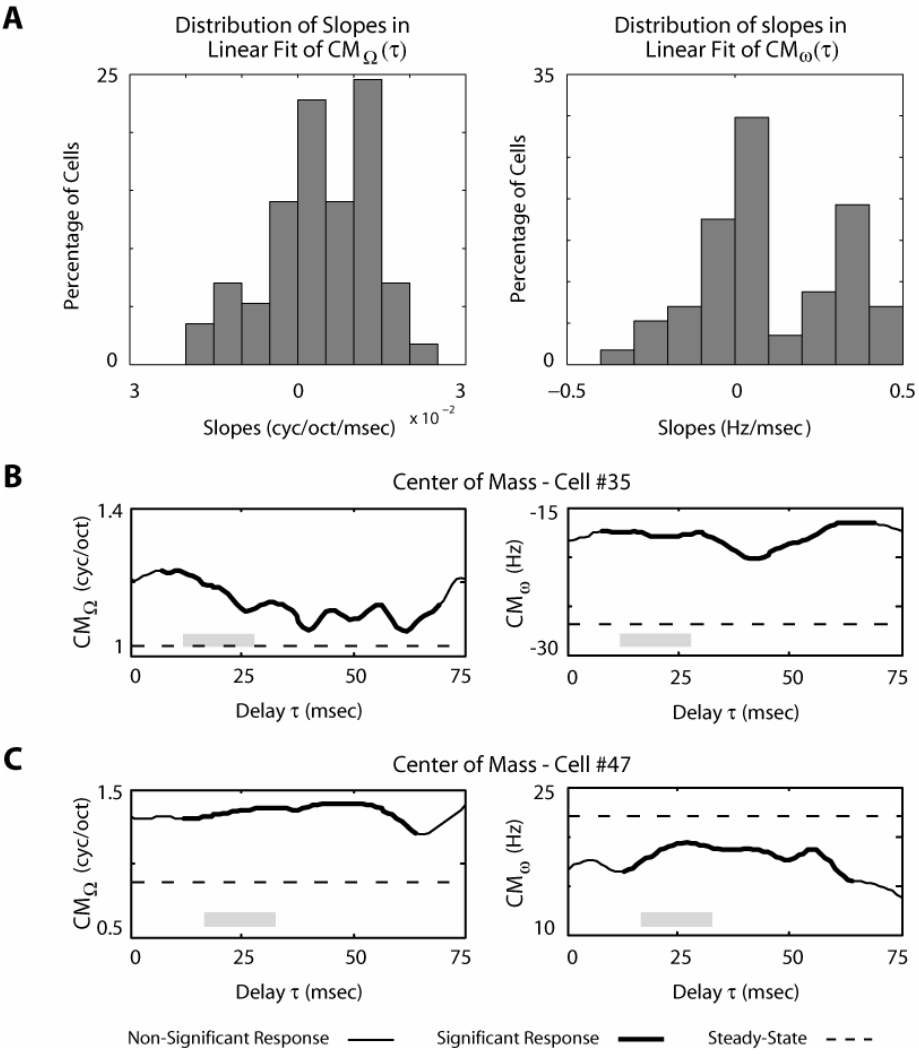

Fig. 6. The center of mass was computed for the tMTF at each lag in the quadrant which had the greater power throughout the response. We fit the temporal progression linearly (both in spectral density Ω and in angular frequency w) at the lag corresponding to the first significant peak in the modulation 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: w). Both quantities were biased to increase (70% for Ω and 68% for w). B,C) Center of mass (left: Ω, right: w) 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.

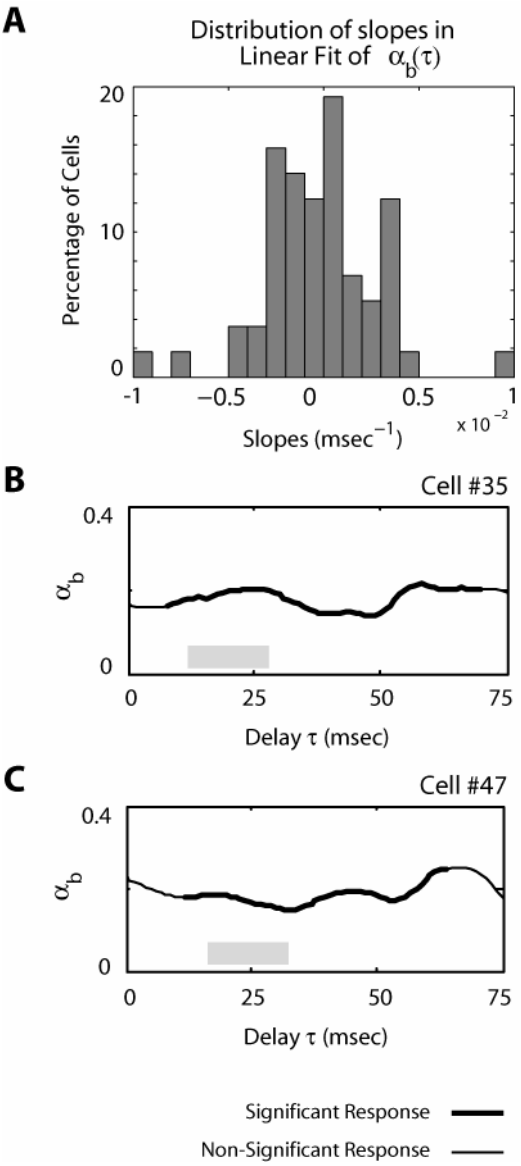

Fig. 7. α_b was computed for the tMTF at each lag. The best linear fit to $\alpha_b(\tau)$ was found at the first significant peak in the modulation power (indicated by the gray bars in B and C). A) The distribution of fit slopes for all cells analyzed in this study are depicted. 47% broadened their 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.

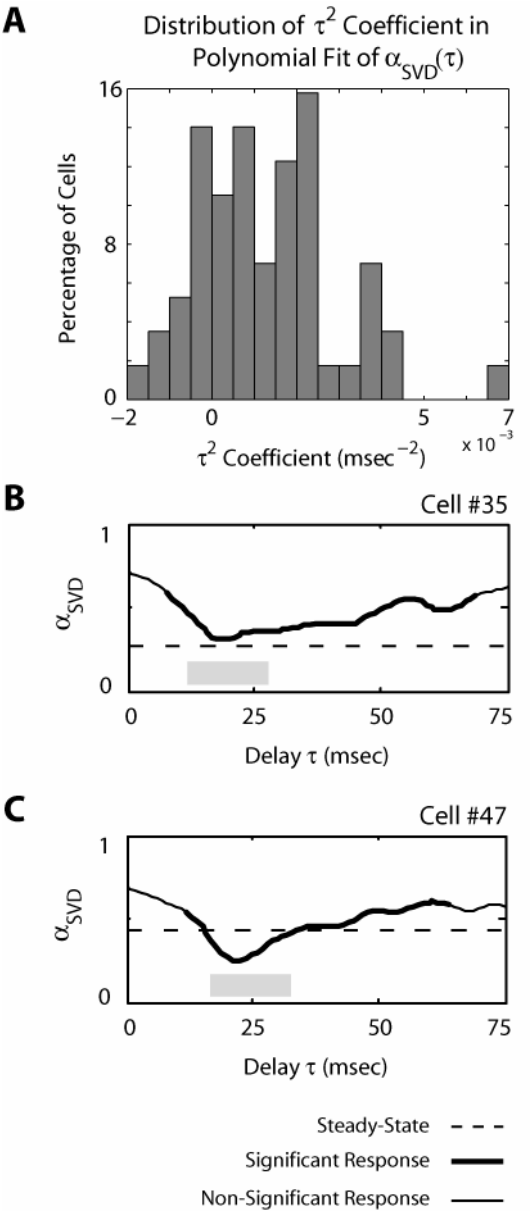

Fig. 8. α_{SVD} was computed for the tMTF at each lag, and its temporal progression, $\alpha_{SVD}(\tau)$ was fit with a second-order polynomial. A) The distribution of the second order coefficients τ^2 for all cells analyzed in this study are shown in the histogram. The majority of cells (73%) had a 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.

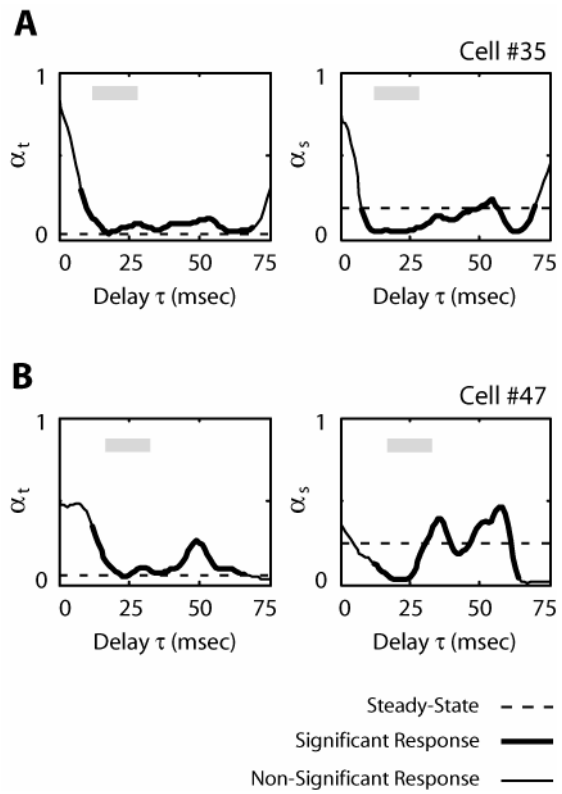

Fig. 9. A) α_t (left) and α_s (right) for the tMTF at each lag for Cell 35. Both α_t and α_s quickly decay to a near-zero value, and α_s gradually increases to the steady-state value. B) Same as in A for Cell 47. Both α_t and α_s decay to a near-zero value again, but with a longer latency. In contrast 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.