

Spectro-Temporal Response Field Characterization With Dynamic Ripples in Ferret Primary Auditory Cortex

DIDIER A. DEPIREUX,¹ JONATHAN Z. SIMON,¹ DAVID J. KLEIN,^{1,2} AND SHIHAB A. SHAMMA^{1,2}

¹*Institute for Systems Research and* ²*Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742-3311*

Received 15 November 1999; accepted in final form 12 October 2000

Depireux, Didier A., Jonathan Z. Simon, David J. Klein, and Shihab A. Shamma. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85: 1220–1234, 2001. To understand the neural representation of broadband, dynamic sounds in primary auditory cortex (AI), we characterize responses using the spectro-temporal response field (STRF). The STRF describes, predicts, and fully characterizes the linear dynamics of neurons in response to sounds with rich spectro-temporal envelopes. It is computed from the responses to elementary “ripples,” a family of sounds with drifting sinusoidal spectral envelopes. The collection of responses to all elementary ripples is the spectro-temporal transfer function. The complex spectro-temporal envelope of any broadband, dynamic sound can be expressed as the linear sum of individual ripples. Previous experiments using ripples with downward drifting spectra suggested that the transfer function is separable, i.e., it is reducible into a product of purely temporal and purely spectral functions. Here we measure the responses to upward and downward drifting ripples, assuming separability within each direction, to determine if the total bidirectional transfer function is fully separable. In general, the combined transfer function for two directions is not symmetric, and hence units in AI are not, in general, fully separable. Consequently, many AI units have complex response properties such as sensitivity to direction of motion, though most inseparable units are not strongly directionally selective. We show that for most neurons, the lack of full separability stems from differences between the upward and downward spectral cross-sections but not from the temporal cross-sections; this places strong constraints on the neural inputs of these AI units.

INTRODUCTION

Only a few general organizational features are known in primary auditory cortex (AI). They include a spatially ordered tonotopic axis (Evans et al. 1965), bands of alternating binaural response properties (Imig and Adrian 1977; Middlebrooks et al. 1980), and a variety of other response features that change systematically along the isofrequency planes such as thresholds (Heil et al. 1994; Schreiner et al. 1992), bandwidths (Schreiner and Sutter 1992), FM selectivity (Heil et al. 1992; Mendelson et al. 1993; Shamma et al. 1993), and asymmetry of response areas (RAs; the span of frequencies that influence, both through excitation and inhibition, the response of a cell) (Shamma et al. 1993). To derive a functionally coherent picture of these maps, it is necessary to integrate these features within a comprehensive descriptor of the unit responses; one that can

be quantitatively derived and employed to predict responses to novel stimuli.

Traditionally measured *response areas* are inadequate because they rarely include response dynamics and cannot be used to predict responses quantitatively. An alternative is the response field (RF) (Schreiner and Calhoun 1994; Shamma et al. 1995), a static, purely spectral function analogous to the RA except for the use of broadband sounds (but see Nelken et al. 1994; Sutter et al. 1996). A dynamic generalization of the RF is the spectro-temporal response field (STRF), a characteristic function of a neuron obtained using broadband sounds (Aertsen and Johannesma 1981; deCharms et al. 1998; Eggermont 1993 and references therein; Escabi and Schreiner 1999; Kowalski et al. 1996a; Kvale and Schreiner 1995; Theunissen et al. 2000). A schematic of an idealized STRF is illustrated in Fig. 1. Qualitatively, its spectral axis reflects the range of frequencies that influence the response or firing rate of the neuron being characterized, and its temporal axis reflects how this influence changes as a function of time. Positive-valued regions of the STRF describe excitatory influence, and negative regions describe inhibitory influence. The interplay between the spectral and temporal axes can give multiple interpretations to the STRF, e.g., as a time-evolving spectral response field or a family of impulse responses labeled by frequency band.

Over the last few years, we have developed new methods to derive the STRFs and characterize the responses of both single and multiple units in the ferret AI (Kowalski et al. 1996a,b). These methods use “moving ripples”: time-varying broadband sounds with sinusoidal spectral envelopes that drift a constant velocity along the logarithmic frequency axis. Figure 2 illustrates the spectrogram of such a stimulus. Neuronal responses are vigorous and well phase-locked to these spectral and temporal envelope modulations over a range of ripple velocities and densities. Measuring the amplitude and phase of the locked component of the response enables one to construct *transfer functions*. A transfer function can be inverse-Fourier transformed to obtain the STRF that characterizes a unit’s dynamics and selectivity along the tonotopic axis.

In developing these measurement and analysis methods, we use two fundamental assumptions. The first is that the responses are substantially linear with respect to the time-varying spectral envelope of stimuli. In particular, this implies that the

Address reprint requests to J. Z. Simon (E-mail: jzsimon@isr.umd.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

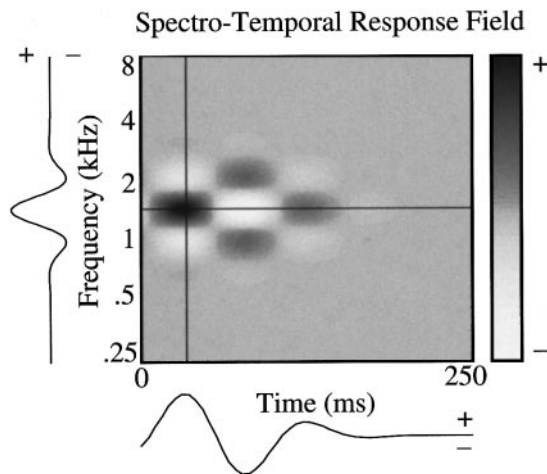


FIG. 1. An idealized spectro-temporal response field (STRF) with spectral and temporal 1-dimensional sections. The time axis is convolved with the time axis of the spectral envelope of a stimulus (as in Fig. 2) to predict the cell's response. For instance, a burst of energy between 1 and 2 kHz will produce a maximum firing rate after about 20 ms followed by inhibition.

response to the spectro-temporally rich stimulus—whose envelope can always be described as the sum of multiple moving ripples—will be the sum of its responses to the individual ripple components. This assumption was confirmed by successfully predicting responses to the superposition of multiple ripples (Kowalski et al. 1996b).

The second important assumption deals with the separability of the temporal and spectral aspects of the responses. Specifically we have demonstrated in other reports that temporal and spectral transfer functions can be measured independently of each other and then combined with a simple product to compute the total transfer function (Kowalski et al. 1996a). The importance of this finding stems from its experimental implications for measuring the STRFs and theoretical consequences for the biophysical and functional models of the STRFs. On the experimental side, separability makes it possible to infer responses to all ripple velocities and peak densities based on only a pair of temporal and spectral transfer functions. Without this assumption, measuring the two-dimensional transfer function is difficult because of the extended times needed to collect adequate spike counts. On the theoretical side, separability suggests that certain features of the STRF (as we shall discuss in detail in the following text) are formed by independent (and likely sequential) spectral and temporal processing stages.

In our earlier study (Kowalski et al. 1996a), separability was validated for ripples moving only in one direction (spectral envelope moving downward in frequency), a notion also known as “quadrant separability.” In this report, we compare the separable functions (spectral and temporal) across upward and downward quadrants. If the functions are the same across quadrants, the responses are “fully separable” (i.e., they are separable); otherwise they are quadrant separable, which is a (specialized) form of inseparability.

Like quadrant separability, full separability has experimental and theoretical implications. On the experimental side, fully separable STRFs can be measured with either upward or downward moving ripples. Theoretically, fully separable responses imply an STRF that is fully decomposable into the product of a purely temporal impulse response and a purely spectral

response field. It also implies a unit that responds equally well to upward and downward moving ripples and hence has necessarily a symmetric transfer function magnitude with respect to direction (Watson and Ahumada 1985). By contrast, cells that are only quadrant separable necessarily respond in asymmetric fashion with respect to direction, i.e., are direction sensitive.

We restrict our presentation in this paper to measurements with singly presented moving ripples in contrast to simultaneously presented ripples discussed in Klein et al. (2000).

There are several goals of this paper. We present a method of measuring the complete descriptor of the linear spectro-temporal properties of an auditory cell, the STRF. We describe examples of STRFs measured in AI and summarize the distribution of the STRF and transfer function parameters encountered. We show that there is a directional sensitivity in the response to the upward versus downward moving components of a sound's spectral envelope. This breaks the symmetry of full spectro-temporal separability and produces quadrant separability. We propose measures to quantify quadrant and full separability. Finally, we discuss the significance of the results and their relationship to results from similar auditory and analogous visual experimental paradigms.

METHODS

Surgery and animal preparation

Data were collected from a total of 11 domestic ferrets (*Mustela putorius*) supplied by Marshall Farms (Rochester, NY). The ferrets were anesthetized with pentobarbital sodium (40 mg/kg) and maintained under deep anesthesia during the surgery. Once the recording session started, a combination of ketamine ($8 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$), xylazine ($1.6 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$), atropine ($10 \mu\text{g} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$), and dexamethasone ($40 \mu\text{g} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$) was given throughout the experiment by continuous intravenous infusion, together with dextrose, 5% in Ringer solution, at a rate of $1 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$ to maintain metabolic stability. The ectosylvian gyrus, which includes the primary auditory cortex, was exposed by craniotomy and the dura was reflected. The contralateral ear canal was exposed and partly resected, and a cone-shaped speculum containing a miniature speaker (Sony MDR-E464) was sutured to the meatal stump. For more details on the surgery, see Shamma et al. (1993).

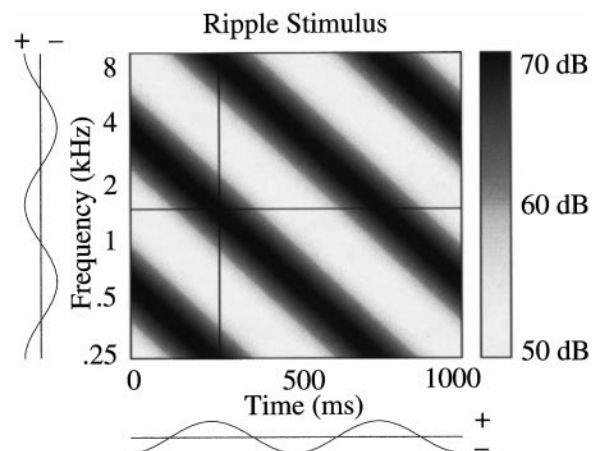


FIG. 2. Envelope of a moving ripple, $w = 2 \text{ Hz}$, $\Omega = 0.4 \text{ cycle/octave}$, $\Phi = -90^\circ$, with a 10 dB AM around a 60 dB base with spectral and temporal 1-dimensional sections. Ripple phase changes linearly with time and spectral position (in octaves).

Recordings

Action potentials from single units were recorded using glass-insulated tungsten microelectrodes with 5–7 MΩ tip impedance at 1 kHz. Neural signals were fed through a window discriminator, and the time of spike occurrence relative to stimulus delivery was stored using a computer. In each animal, electrode penetrations were made orthogonal to the cortical surface. In each penetration, cells were typically isolated at depths of 350–600 μm corresponding to cortical layers III and IV (Shamma et al. 1993). In many instances, it was difficult to isolate reliably a single unit for extended recordings, and hence several units were recorded instead. Such data were labeled “multiunit recordings” and are explicitly designated as such and separated from the single-unit records in all data presentations in the paper.

Acoustic stimuli

All stimuli are computer synthesized. For each unit isolated, initial tests are carried out using tonal stimuli to measure the basic frequency response at several intensities to determine the best frequency (BF) and response threshold. All other stimuli used in these experiments have broadband spectra with a sinusoidally modulated (or rippled) envelope. We used the knowledge of the cell's BF to adjust the frequency range of the broadband sound so that the cell's excitatory and inhibitory regions lay well within the frequency range of the sounds.

In practice, it is hard to generate noise and then shape it with filters to a desired dynamic spectral envelope, so we generate ripples over a range of five octaves by taking logarithmically spaced pure tones with random (temporal) phases. The amplitude $S(t, x)$ of each tone is then

$$S(t, x) = L[1 + \Delta A \cdot \sin(2\pi \cdot w \cdot t + 2\pi \cdot \Omega \cdot x + \Phi)] \quad (1)$$

where $x = \log_2(f/f_0)$ is the number of octaves above the base frequency f_0 . The ripple envelope resembles a drifting one-dimensional grating as illustrated in Fig. 2. Five independent parameters characterize the ripple envelope: background level or loudness of the stimulus (L); AM of the ripple (ΔA) in percentage or decibels; ripple velocity (w) in units of cycles/s (or Hz); ripple density (Ω) in units of cycles/octave; and the initial phase of the ripple Φ . The spectra consist either of 20 or 100 tones per octave equally spaced along the logarithmic frequency axis or with a spacing of 1 tone/Hz with an amplitude decay producing equal power per octave. The spectra typically span five octaves (e.g., 0.25–8 kHz) with the range chosen such that the response area of the cell tested lay within the stimulus spectrum. The choice of a density of 20 or 100 tones per octave does not alter the cortical responses; hence we do not specify which density was used.

A single-ripple stimulus at overall level L dB SPL would typically be composed of N logarithmically spaced components, each at $L - 10 \log_{10}(N) \approx L - 20$ dB for $N = 101$. The overall stimulus level was chosen on the basis of threshold at BF; typically L was set 10–20 dB above threshold. High levels ($L > 70$ dB) were avoided to ensure the linearity of our stimulus delivery system. The amplitude of a single ripple was defined as the maximum percentage or logarithm change in the component amplitudes. Ripple amplitudes were either 90% (linear) or 10 dB (logarithmic) modulations.

The ripple velocities w and ripple densities Ω used were determined by the response properties of the neuron, but the typical range was $|w| < 25$ Hz (with some units requiring up to 100 Hz) and $|\Omega| < 1.6$ cycles/octaves (with some units requiring up to 4 cycles/octaves). Single ripples were always presented with $\Phi = 0$.

By the convention established in Eq. 1, a ripple whose spectral envelope is moving downward in frequency, as in Fig. 2, has positive w and positive Ω ; equivalently, it can be described by a ripple with negative w and negative Ω , and an added phase shift of π , by Eq. 1 and the identity $\sin(\alpha) = \sin(-\alpha + \pi)$. A ripple whose spectral peaks are moving upward in frequency has negative w and positive Ω ,

or by Eq. 1 and the same identity, positive w , negative Ω , and an added phase shift of π .

The stimulus bursts had an 8-ms rise/fall time and duration of 1.0 or 1.7 s, repeated every 3–4 s. All stimuli were gated and fed through an equalizer into the earphone. Calibration of the sound delivery system (to obtain a flat frequency response up to 20 kHz) was performed in situ with the use of a 1/8-in Brüel and Kjaer 4170 probe microphone. The microphone was inserted into the ear canal through the wall of the speculum to within 5 mm of the tympanic membrane. The speculum and microphone setup resembles closely that suggested by Evans (1979).

Theoretical considerations

DEFINING THE STRF. The fundamental tool to measure linearity and separability of primary cortical cell is to measure their STRF. The STRF is a spectro-temporal function $\text{STRF}(t, x)$. The linear response rate $y(t)$ of a cell is related to its $\text{STRF}(t, x)$ and the spectro-temporal envelope of the stimulus $S(t, x)$ by $y(t) = \int \int dt' dx' S(t' - t, x) \cdot \text{STRF}(t, x)$, i.e., convolution along the time dimension t and integration along the spectral dimension x .

The STRF is measured through its two-dimensional Fourier transform, or transfer function $T(w, \Omega) = \mathcal{F}_{w\Omega}[\text{STRF}(t - x)]$, and then inverse transformed to compute the STRF, where the coordinates dual to t and x are w and Ω , respectively (see Fig. 3). By measuring the sinusoidal component with temporal frequency w of the response $y_{w\Omega}(t)$ of a cell to a ripple of specific ripple velocity w and ripple density Ω , we can obtain the transfer function $T(w, \Omega)$ at one point in $w - \Omega$ space (Depireux et al. 1998)

$$\begin{aligned} y_{w\Omega}(t) &= \iint dt' dx' \text{STRF}(t', x') \sin 2\pi[w(t - t') + \Omega x'] \\ &= |T(w, \Omega)| \sin[2\pi wt + \Phi(w, \Omega)] \end{aligned} \quad (2)$$

This way, we derive the amplitude $|T(w, \Omega)|$ and phase $\Phi(w, \Omega)$ of the complex transfer function $T(w, \Omega)$ by measuring the amplitude and phase of the (real) response of the cell. Note that the use of complex numbers is not theoretically necessary, but it does simplify the calculations in the transfer function space considerably. By the definition of the transfer function, it follows that the inverse Fourier transform of $T(w, \Omega)$ is the STRF of the cell

$$\text{STRF}(t, x) = \mathcal{F}_{t-x}^{-1}[T_{w\Omega}] \quad (3)$$

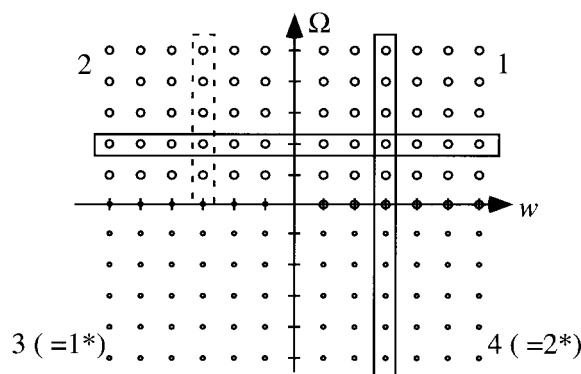


FIG. 3. To measure the complete ripple transfer function of an arbitrary STRF, we would need to measure the response of the cell to all the ripples represented by large circles. The small circles correspond to redundant ripples by complex conjugation. The value of the transfer function along the $w = 0$ axis is set to 0, because the modulation transfer function is not well defined there. Quadrant separability permits one to measure only the responses to ripples enclosed by the solid boxes. The transfer function in the dashed box is equal to the transfer function in the bottom half of the vertical box but with the opposite phase.

Because $\text{STRF}(t, x)$ is real but $T(w, \Omega)$ is complex, there is complex conjugate symmetry

$$T(-w, -\Omega) = T^*(w, \Omega) \quad (4)$$

which also holds for the Fourier transform of any real function of t and x .

DEFINING AND ASSESSING SEPARABILITY. Separability is an important property of the transfer functions. A fully separable transfer function is one that factorizes into a function of w and a function of Ω over all quadrants: $T(w, \Omega) = F(w) \cdot G(\Omega)$. This implies that $\text{STRF}(t, x)$ is time-spectrum separable: $\text{STRF}(t, x) = \text{IR}(t) \cdot \text{RF}(x)$. In this case, one needs only measure the transfer function for all Ω at a convenient w and for all w at a convenient Ω . $F(w)$ and $G(\Omega)$ are each complex-conjugate symmetric [$F(-w) = F^*(w)$, $G(-\Omega) = G^*(\Omega)$] because $\text{IR}(t)$ and $\text{RF}(x)$ are real, so one needs only consider the positive values of each. This dramatically decreases the number of measurements needed to characterize the STRF.

A transfer function may also be only partially separable in that it is separable only for ripples moving in a given direction (upward vs. downward). In this case, the transfer function is called quadrant separable and can be expressed as the product of two independent functions

$$T(w, \Omega) = \begin{cases} F_1(w)G_1(\Omega) & w > 0, \Omega > 0 \\ F_2(w)G_2(\Omega) & w < 0, \Omega > 0 \end{cases} \quad (5)$$

where the subscript 1 indicates the $w > 0, \Omega > 0$ quadrant, and the subscript 2 the $w < 0, \Omega > 0$ quadrant (see Fig. 3). Note that by reality of the STRF, the value of the transfer function in quadrants 3 ($w < 0, \Omega < 0$) and 4 ($w > 0, \Omega < 0$) is complex conjugate to the value in quadrants 1 and 2, respectively. In this case, the STRF is not separable in spectrum and time but is the linear superposition of two functions, one with support only in quadrant 1 (and 3) and one with support only in quadrant 2 (and 4).

Separability need not be an all-or-none property but rather can be assessed in a graded fashion. To do so, we apply singular value decomposition (SVD) of the matrix T of measured transfer-function values (Haykin 1996). T can be viewed as a matrix created by sampling the ideal transfer function at regularly spaced discrete values of w and Ω with random noise added to each sample. SVD decomposes T as

$$T = U \cdot \Lambda \cdot V^\dagger, \quad \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \quad \lambda_1 \geq \lambda_2 \geq \dots, \\ = \sum_{i=1}^n \lambda_i u_i \cdot v_i^\dagger \quad (6)$$

Here \dagger denotes the Hermitian transpose and U, V are matrices containing “singular” row vectors u_i and v_i corresponding to spectral and temporal cross-sections, respectively, of separable transfer functions. Thus the SVD can be viewed as decomposing T into a linear sum of n separable matrices, each weighted by its ability to approximate T as a weighted product of two vectors as in Eq. 6, as given by the “singular values” λ 's. Because of the presence of noise in the measurement, the λ 's are all expected to be nonzero with their values decreasing monotonically to a noise floor, which depends on the level of the noise.

With respect to this floor, the number of significant singular values depends on the nature of the measured transfer function T . The closer T 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. We have used this fact to define a single measure of the “distance” of the system from separability or alternatively the “degree of inseparability” α_{SVD}

$$\alpha_{\text{SVD}} = \left(1 - \lambda_1^2 / \left(\sum_i \lambda_i^2 \right) \right) \quad (7)$$

which is the proportion of T 's total power ($= \sum_i \lambda_i^2$), which is *not* accounted for by its best separable approximation. Values near zero indicate that only the first singular value has a large nonzero value (hence the STRF is separable). Values approaching 1 indicate an increasing dose of inseparability.

The handy measure of α_{SVD} brands inseparability by its strength but otherwise reveals nothing of its nature. Therefore we examine the origin of inseparability by other means. Specifically we shall analyze three factors that give rise to inseparability.

1) The relative power in the first and second quadrants

$$\alpha_d = \frac{P_2 - P_1}{P_2 + P_1} \quad (8)$$

where P_1 = power in quadrant 1 and P_2 = power in quadrant 2. Note that power is measured by summing the squared magnitudes of all transfer function values within the appropriate quadrant. An absolute value of α_d near one implies strong selectivity of the responses to the direction of ripple movement and hence strong inseparability.

2) The asymmetry of the spectral transfer function around $\Omega = 0$ is

$$\alpha_s = 1 - \left| \frac{\sum_{\Omega > 0} G_1(\Omega) \cdot G_2^*(\Omega)}{\sqrt{\sum_{\Omega > 0} |G_1(\Omega)|^2} \cdot \sqrt{\sum_{\Omega > 0} |G_2(\Omega)|^2}} \right| \quad (9)$$

where the quantity inside the large absolute value bars is the (complex) correlation between $G_1(\Omega)$ and $G_2(\Omega)$. Index α_s values near one imply strong asymmetry (i.e., lack of correlation) in the transfer function to different directions and hence strong inseparability.

3) The asymmetry of the temporal transfer function around $w = 0$ is

$$\alpha_t = 1 - \left| \frac{\sum_{w > 0} F_1(w) \cdot F_2^*(-w)}{\sqrt{\sum_{w > 0} |F_1(w)|^2} \cdot \sqrt{\sum_{w > 0} |F_2(-w)|^2}} \right| \quad (10)$$

where the quantity inside the large absolute value bars is the (complex) correlation between $F_1(w)$ and $F_2^*(-w)$. Index α_t values near 1 imply strong asymmetry (i.e., lack of correlation) in the transfer function to different directions, and hence strong inseparability.

EFFECT OF FINITE SAMPLING. We measure the transfer function of cells by varying two parameters, ripple velocity and ripple density. For consistency's sake, we used the same range of parameters for a majority of cells. However, for some cells, the transfer function has not decreased significantly at the “edges” (for instance, in Fig. 9C, the temporal transfer function is clearly still strong at ± 64 Hz and above). This is equivalent to multiplying the true transfer function by a rectangular function which is zero everywhere except between -64 and 64 Hz, over which range it is 1. In the dual Fourier space of the transfer function space, that is, in the STRF space with coordinates t and x , this corresponds to convolving along each dimension the STRF with the Fourier transform of a rectangular pulse, that is, with $\sin(x)/x$. This leads to spurious oscillations in the *display* of the STRF as can be seen in Fig. 9C and others. These oscillations would disappear if we had measured the transfer functions all the way to their vanishing values.

Since all the characteristic parameters in this paper (see Table 1) are derived in transfer function space, it does not affect the analysis, but it may lead to misleading features in the STRFs.

DEVIATIONS FROM LINEARITY. Because the STRF is a measure of the linear part of the dynamics of a cell, we only consider effects that might modify the measurement of the first component of the Fourier transform of the period histograms. The most prominent nonlinearities are (approximate) half-wave rectification and compression. The half-wave rectification is primarily due to the positivity of spike rates (ordinarily the steady-state response to a flat spectrum is significantly less than half the peak firing rate of the unit); the distortion of a sinusoid due to half-wave rectification does not affect the phase of the response, and its effect on the amplitude of the first Fourier compo-

TABLE 1. Characteristic parameters of STRFs shown

STRF	f_m^1 , kHz	f_m^2 , kHz	τ_d^1 , ms	τ_d^2 , ms	ϕ_m , deg	θ , deg	δ , %	ϵ , %	Single/Multiple
Fig. 6	1.1	1.1	21	34	-24	-46	3	6	Mult
Fig. 8									
A	1.5	1.9	25	23	4	-69	3	14	Mult
B	1.4	1.6	32	23	-43	-58	8	29	Mult
C	1.1	2.0	20	35	-21	-105	5	25	Mult
Fig. 9									
A	3.8	4.5	29	29	-36	-110	3	4	Mult
B	1.5	1.8	21	20	7	-57	2	8	Mult
C	3.9	4.6	13	5	-68	-67	3	36	Sngl
Fig. 10									
A	0.56	0.68	21	12	-26	-63	2	6	Sngl
B	0.49	0.57	25	14	40	-35	9	24	Mult
C	1.2	1.14	47	43	4	140	4	11	Mult

STRF, spectro-temporal response field.

ment is a constant factor, independent of w and Ω . The distortion due to compression or saturation, similarly, does not affect the phase of the Fourier transform components of the response and similarly affects the amplitude only by an overall constant factor for stimuli of moderate level.

Nonlinearities of other types, such as static nonlinearities, if they exist, are quite small and have not shown up in our studies. Ultimately, the proof of linearity, and the relevance of the STRF, is found when one compares the predictions of the response of a cell to a new sound compared with the actual response. We have not found any evidence of systematic deviation between predicted and actual response that would indicate the presence of static nonlinearities.

Data reduction

Many of the data analysis methods described here are similar or straightforward extensions of those developed earlier in Kowalski et al. (1996a), and those will be only briefly reviewed here. Figures 4 and 5 illustrate the nature of the responses to the ripple stimuli and the analysis to extract the spectral (Fig. 4) and temporal (Fig. 5) transfer functions. In Fig. 4A, the ripples are presented at 8 Hz for ripple densities from -1.6 to 1.6 cycle/octave in steps of 0.2 cycle/octave. Each stimulus is presented 15 times.

For each ripple density, we compute at 16-bin period histogram based on the responses starting at 120 ms (to exclude the onset response; Fig. 4B). A 16-point Fourier transform (FFT) is then performed on the period histogram, and the amplitude and phase of the first component is taken to be the amplitude and phase of the transfer function. If the modulation of the response was that of a purely linear system, the higher FFT coefficients would be negligible, but because of half-wave rectification and compression, they sometimes are significant. In general $T_w(\Omega)$ can be written as

$$T_w(\Omega) = |T_w(\Omega)|e^{j\Phi_w(\Omega)} \quad (11)$$

where $j = \sqrt{-1}$. Figure 4C illustrates the magnitude $|T_w(\Omega)|$ and the unwrapped phase $\Phi_w(\Omega)$ of the transfer function $T_w(\Omega)$. The ripple density at which $|T_w(\Omega)|$ is a maximum is designated as Ω_m (= 0.0 octave/cycle in Fig. 4C).

Analogueous steps are followed in measuring the temporal transfer function as shown in Fig. 5 where ripples are presented at 0.2 cycle/octave for ripple velocities from -24 to 24 Hz in steps of 4 Hz.

Note that in the previous paper (Kowalski et al. 1996a), we weighted the measurement of the first component of the Fourier transforms of the period histograms by a weighted sum of the higher frequency components of the transform. This, however, is not compatible with the idea of a linear system so that the resultant STRF or equivalently the ripple transfer function T would not be expected to be the best possible predictor of the response to new sounds. Therefore

in this paper, the values of T correspond directly to the first component of the Fourier transform.

Once the ripple transfer function has been measured, it can be inverse Fourier transformed to display the STRF. Since the transfer function is typically measured over fewer than 8 points along each dimension in each quadrant, the resulting STRF as computed would look very jagged even if the underlying STRF was smooth. We therefore interpolate to a smooth STRF for display purposes, padding the transfer function with zeros to a size of 64×64 . All statistics and predictions use the measured unsmoothed STRF.

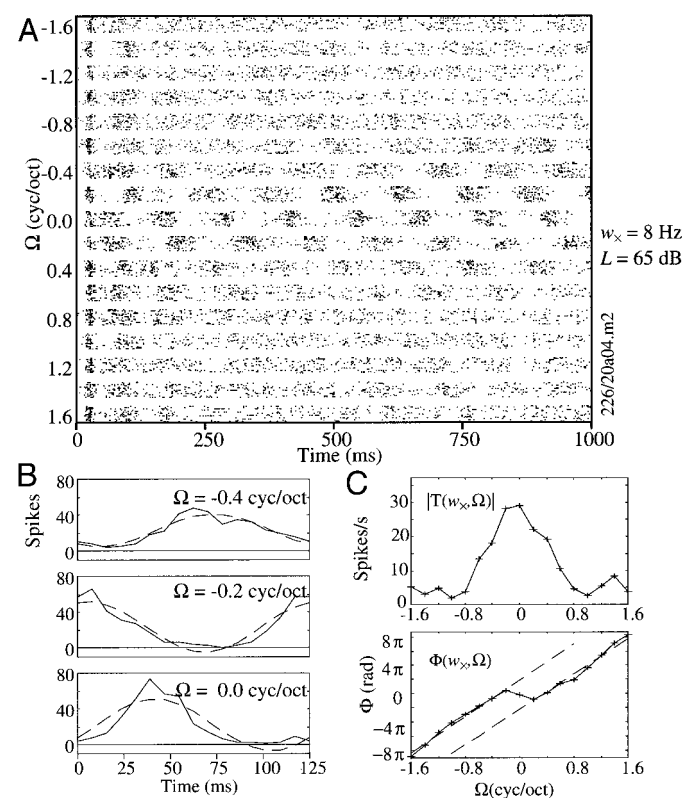


FIG. 4. Data analysis for ripples of fixed ripple velocity and varying ripple densities. A: raster plot of responses. Each point represents an action potential, and each ripple stimulus is presented 15 times. Note the position of the peaks changes linearly with ripple density. B: period histogram for 3 example ripple densities, with their sinusoidal fits. C: magnitude and phase of the period histogram fits. With the phase convention used for these stimuli, ripples with $\Omega < 0$ (quadrant 4) are equivalent to ripples with $w < 0$ (quadrant 2), using the conversion $(w, \Omega, \Phi) \rightarrow (-w, -\Omega, -\Phi + \pi)$.

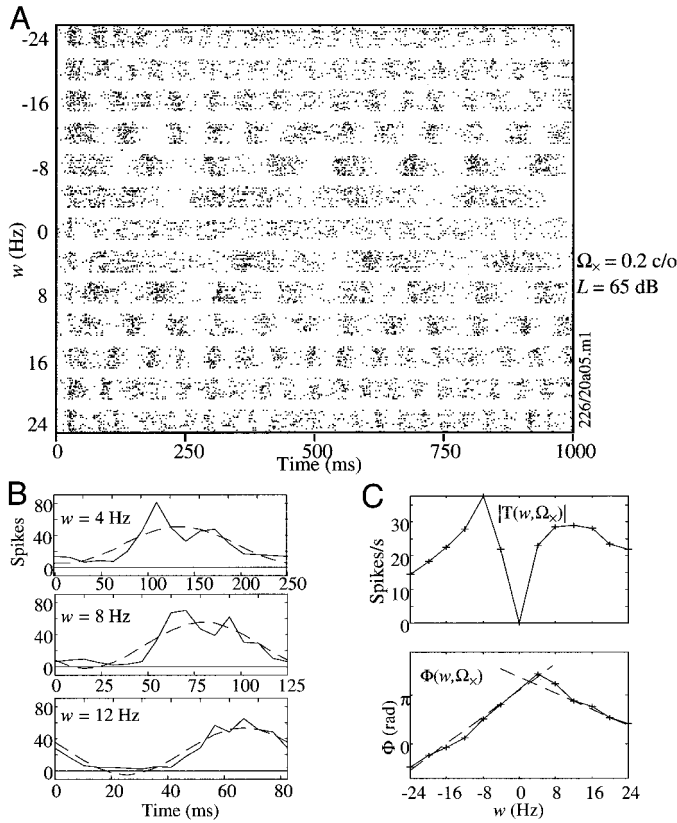


FIG. 5. Data analysis from ripples of fixed ripple density and varying ripple velocities. *A*: raster plot of responses. Each point represents an action potential, and each ripple stimulus is presented 15 times. *B*: period histogram for 3 example ripple velocities. Note how the position of the peak of the best fit changes linearly with ripple velocity. *C*: magnitude and phase of the period histogram fits.

To construct the two-dimensional transfer function, we assume quadrant separability, measure the transfer function along the cross-sections shown in Fig. 3, to combine these spectral and temporal cross sections as illustrated in Fig. 6. For each quadrant, the transfer function is the outer product of the cross-section, divided by the (complex) value of the transfer function at the crossover (\times) point. In Fig. 6, the point is $(w_{\times 1}, \Omega_{\times 1}) = (8 \text{ Hz}, 0.2 \text{ cycle/octave})$ in quadrant 1 and $(w_{\times 2}, \Omega_{\times 2}) = (-8 \text{ Hz}, 0.2 \text{ cycles/octave})$ in quadrant 2.

$$T(w, \Omega) = T(w_{\times q}, \Omega) \cdot T(w, \Omega_{\times q}) / T(w_{\times q}, \Omega_{\times q}) \quad (12)$$

where $q = 1$ and $q = 2$ are the independent quadrants 1 and 2. In practice, the value of the transfer function along the two cross-sections was measured at two different times, giving two measurements of the transfer function at each crossover point $T(w_{\times q}, \Omega_{\times q})$. The results of the two measurements may differ, and so we use the (complex) geometric mean of the two measured values as the divisor in Eq. 12, $T_{\text{eff}}(w_{\times q}, \Omega_{\times q}) = [T_{1\text{st}}(w_{\times q}, \Omega_{\times q}) T_{2\text{nd}}(w_{\times q}, \Omega_{\times q})]^{1/2}$.

The ratio $T_{1\text{st}}(w_{\times q}, \Omega_{\times q}) / T_{2\text{nd}}(w_{\times q}, \Omega_{\times q})$, which should be unity, reflects noise in the system and is used to estimate reliability in the following text.

The value of the transfer function along the $w = 0$ axis is set to zero because the modulation transfer function is not well defined there, i.e., there is no modulation of firing rate around the DC (average) rate with a frequency of 0 Hz. The value of the transfer function along the $\Omega = 0$ axis is not measured directly, so the value used is the mean of the value inferred from being the boundary of quadrant 1 and that inferred from being the boundary of quadrant 2.

Once the values of transfer functions for quadrants 1 and 2 and their boundaries are measured, the values for quadrants 3 and 4 are given

by Eq. 4 (see also Fig. 3). The STRF is then computed by an inverse Fourier transform (as in Eq. 3) and is illustrated in Fig. 6*B* (left). This interpolated version of the STRF (used for display) is obtained by using Eq. 3 on the transfer function padded with zeros at high $|w|$ and $|\Omega|$ (see Fig. 6*A*).

Deriving STRF parameters from the phase functions

Numerous parameters can be derived from the STRF (or equivalently the transfer function) that are analogous to traditional response measures such as BF, tuning curve bandwidth, and latency. Most of these parameters are best derived from analysis of the phase of the transfer functions (Fig. 7).

We model the phase of the transfer function within each quadrant $\Phi^q(w, \Omega)$, $q = 1, 2$ (see Eq. 2) as a linear function of w and Ω

$$\Phi^q(\Omega, w) = -2\pi w \tau_d^q + 2\pi \Omega x_m^q + \chi^q \quad (13)$$

where τ_d^q is the mean or group delay of the STRF (a portion of which comes from the response latency), $x_m^q = \log(f_m^q/f_0)$ is the mean frequency (in octaves above the base frequency of the ripple, see Eq. 1) around which the STRF is centered (putting it near the BF), and χ^q is a constant phase angle, for each quadrant q . The complex-conjugate symmetry of the transfer function means that these six independent parameters describe the phase everywhere in the $w - \Omega$ plane. The convention of the minus sign before τ_d allows the time-dependent responses to be functions of $(t - \tau_d)$ as is appropriate for a delay.

The justification for assuming linear fits of the phase functions has been discussed in detail earlier in (Depireux et al. 1998) and is strongly motivated by the data (Kowalski et al. 1996a). Note, however, that the assumption of *phase* linearity is used only for parameter estimation and is not assumed in computing the STRF. The first linear term in Eq. 13 stems from the fact that auditory units differing in their mean neural delays will exhibit linear phase dependence on w with different slope depending on delay. Analogous arguments apply for units that are located at different places along the tonotopic axis: the response phase of different units (with otherwise identical STRFs) changes linearly with Ω at different rates, depending on the relative center frequency locations. In both cases, the slopes of the linear phase function indicate the absolute shift of the STRF relative to the origin, i.e., the mean time delay τ_d^q relative to the start of the stimulus, and the center frequency x_m^q relative to the low frequency edge of the ripple spectrum. The linear phase model does not assume that the linear phase shifts, τ_d^q and x_m^q , are equal across quadrants, but tonotopy suggests that f_m^1 and f_m^2 should be approximately equal and $\tau_d^1 \approx \tau_d^2$ since the temporal delays of the neural inputs are not segregated by quadrant. This is shown experimentally in the following text.

An interpretation of τ_d , for each quadrant, is that it is the sum of the pure response latency and (roughly) half the temporal width of the STRF. This is in contrast to the STRF's peak delay, τ_{STRF} , defined to be the delay for which the STRF achieves its maximum value, which may lead or lag τ_d , depending on the constant temporal phase shift, θ , defined in the following text. Similarly, f_m for each quadrant may or may not fall on the STRF's best frequency, BF_{STRF} , defined to be the frequency at which the STRF achieves its maximum value, depending on the constant spectral phase shift, ϕ , defined in the following text.

A convenient convention for interpreting the constant component of the phase is to break up the constant phase angle χ^q into two parts

$$\chi^1 = -\theta + \phi, \quad \chi^2 = \theta + \phi \quad (14)$$

θ and ϕ are, respectively, the temporal polarity and spectral asymmetry of the STRF. Spectral asymmetry parameterizes the balance of the STRF along the spectral axis about its center. For example, a unit with $\phi = 0$ would have its BF_{STRF} in the center of the spectral envelope of the STRF, possibly surrounded by inhibitory regions. A unit with $\phi > 0$ would have its BF_{STRF} at a lower frequency than the center of the STRF with an inhibitory sideband above BF_{STRF} . A unit with $\phi < 0$

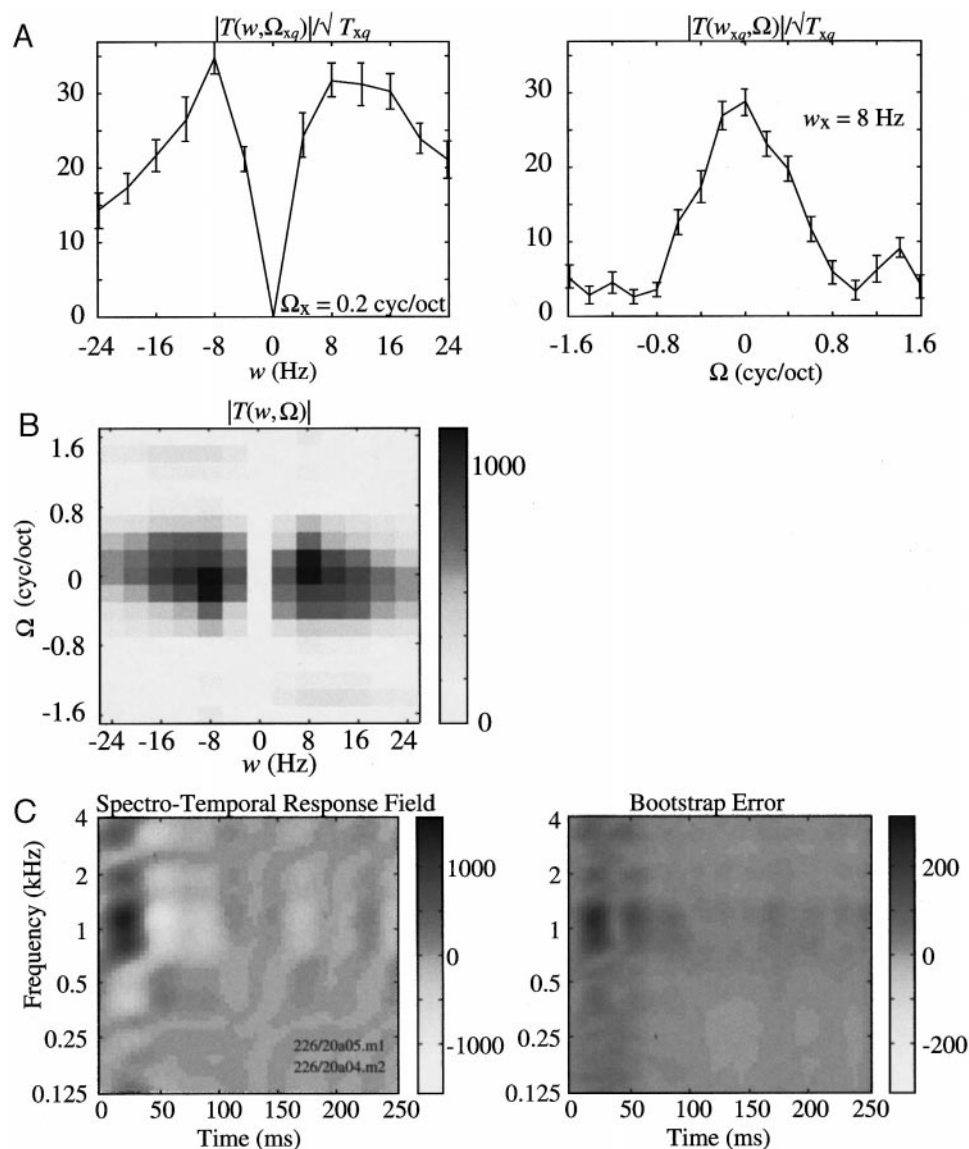


FIG. 6. Deriving the spectro-temporal transfer function, STRF, and related parameters. A: magnitude of the temporal (left) and spectral (right) transfer function cross-sections, normalized by the values at the cross-over points (Eq. 12). The error bars are computed by the bootstrap method, explained below. B: the magnitude of the full transfer function, resulting from the outer product of the functions in A. C: the STRF of the cell computed by an inverse Fourier transform of the complex transfer functions. To the right is the error estimate of the STRF, using the same scale multiplied by a factor of 5 (for legibility), resulting in error parameters of $\delta = 0.03$ and $\epsilon = 0.06$. See Table 1 for details.

would have its BF_{STRF} at a higher frequency than the center of the STRF, with an inhibitory sideband below BF_{STRF} (see example in Fig. 4C of Shamma et al. 1995). Similarly the temporal polarity parameterizes the balance of the STRF along the temporal axis about its center: whether the peak response occurs before or after regions of inhibition, respectively, $\theta < 0$ ("onset response at BF") or $\theta > 0$ ("offset response at BF"). There is an ambiguity in fixing θ and ϕ that we remove by restricting ϕ to lie between -90 and $+90^\circ$, while θ ranges the full -180 to $+180^\circ$. See Fig. 7 as an illustration of the phase behavior in the different quadrants.

In past reports (Kowalski et al. 1996a), θ and ϕ could be measured without measuring the transfer function in the upward moving quadrant 2 by measuring the constant component of the phase in quadrant 1 ($\chi^1 = -\theta + \phi$) and along the w axis, where the constant component of the phase is expected to be the mean across the quadrants [$(\chi^1 - \chi^2)/2 = -\theta$; note the change in convention of $\theta \rightarrow -\theta$ between the present work and Kowalski et al. (1996a)].

Because of response variability, we only fit to those points of the transfer function that have more than half of the response power in the first component of the Fourier transform. Then the fit is done across the entire two-dimensional phase plane for each quadrant. Ultimately our unwrapping method is less than ideal, and estimates of θ and ϕ especially reflect that (Ghiglia and Pritt 1998).

Estimating response variability: the bootstrap method

Variability in our experiments originates from multiple sources, including internal neural mechanisms (e.g., Poisson-like distributions of spike times), extracellular recording/identifying methods, and equipment noise. Quantitative estimates of the reliability of our measurements is crucial to its analysis and subsequent interpretation. A method of variability estimation that is especially appropriate to these measurements is the bootstrap method (Efron and Tibshirani 1993; Politis 1998).

The essence of this method is to use "resamples," in which N samples of bootstrap data are drawn *with replacement* from the N original samples of data. Repeating this procedure a large number of times creates a population of bootstrap resamples whose probability distribution is a good estimator of the probability distribution from which the original data were drawn.

To illustrate this procedure, consider measuring the transfer function at a point (w, Ω) . This is done by presenting the same (w, Ω) stimulus N times and constructing a period histogram based on all N sweeps. The amplitude and phase of the first Fourier component of the period histogram are assigned to the amplitude and phase of the transfer function. A single bootstrap resampling of the responses will have N sweeps, where, because they are drawn from the original

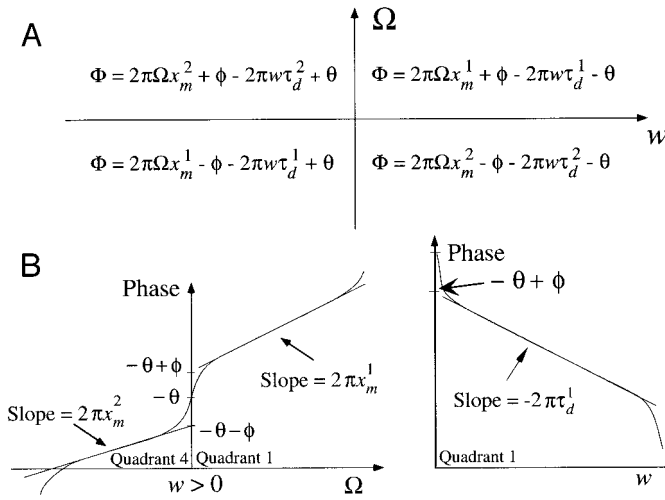


FIG. 7. A: the phase of the transfer function can be well described by a linear fit containing 6 parameters over most of the relevant regions of the w - Ω plane. B: in this cartoon, the slope is constant for most of the curves, after $(left) -2\pi w \tau_d^q$ has been subtracted in each quadrant, corresponding to a center frequency that is independent of the ripple density, and $(right) 2\pi\Omega x_m^1$ has been subtracted, corresponding to a delay that is independent of ripple velocity. At very small ripple densities (long ripple periodicity), center frequency is less meaningful, and similarly for small ripple velocity and delay, respectively. At large ripple velocity the slope asymptotes to the signal front delay, but when this occurs, the small amplitude of the transfer function makes it difficult to measure the phase. See Dong and Atick (1995) and Papoulis (1962).

responses with replacement, some will be duplicated and some will be unused. Nevertheless a period of histogram is constructed, and the bootstrap estimate of the transfer function is assigned to its first Fourier component. Performing a large number of bootstrap resamples results in a population of estimates for the transfer function. This population has a mean, variance, and higher-order moments. These moments are estimators of the moments of the original population (of all transfer functions of all allowable neuronal responses to the stimulus). For example, the standard deviation of all bootstrap estimates of the transfer function is an estimator of the standard deviation of measurements of the transfer function. This allows us to put error bars on our transfer functions and STRFs.

Effects of crossover point errors

Another significant source of error is the difference between the responses of repeated measurements at the transfer function crossover points. The ratio of these independent measurements, $T_{1st}(w_{\times}^q, \Omega_{\times}^q) / T_{2nd}(w_{\times}^q, \Omega_{\times}^q)$ should be unity. When not unity, it reflects the same variability measured by the bootstrap method but also additional systematic error from having measured the two transfer function cross-sections at different times. To account for this disparity, the total squared error of the STRF is set to the sum of the bootstrap STRF variance and the square of the crossover error σ_{\times}

$$\sigma_{STRF}^2 = \sigma_{Bootstrap}^2 + \sigma_{\times}^2 \quad (15)$$

where $\sigma_{\times}(t, x)$ captures the systematic error from not having taken all data at the same time and is given by

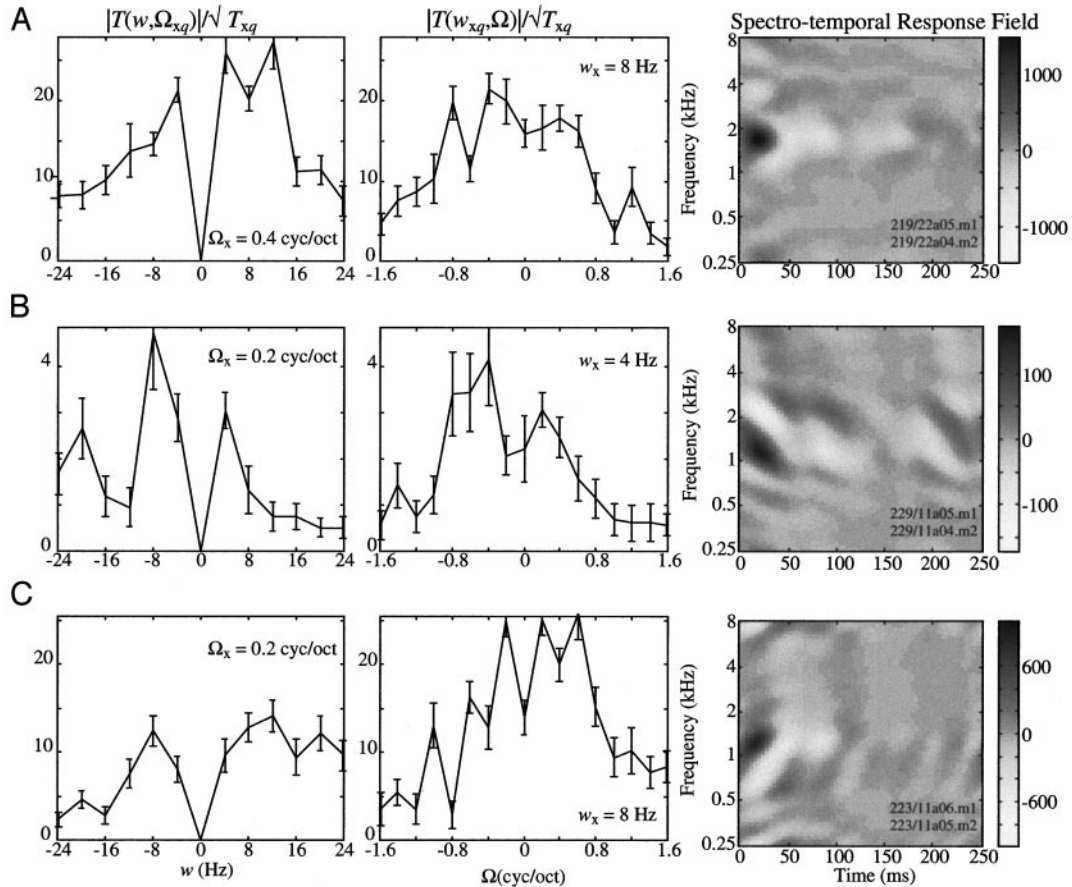


FIG. 8. Three examples of spectro-temporal transfer function sections, corresponding STRFs. For each row A–C: magnitude of the temporal (left) and spectral (middle) transfer functions. All other details are as in Fig. 6; (right), STRFs.

$$\sigma_{\times}(t, x) = \left[\max_q \left(\frac{(\max(|T^q(w_{\times}^q, \Omega_{\times}^q)|, |T^q(w_{\times}^q, \Omega_{\times}^q)|)}{|T_{\text{eff}}^q(w_{\times}^q, \Omega_{\times}^q)|)} - 1 \right) \right] \times |\text{STRF}(t, x)| \quad (16)$$

Finally, we collapse the error over the entire (t, x) plane into two dimensionless terms δ and ϵ

$$\delta = \frac{1}{\Delta T \Delta X} \int \int dt dx \sigma_{\text{STRF}}(t, x) \frac{1}{\max(|\text{STRF}(t, x)|)} \quad (17)$$

$$\epsilon = \int \int dt dx [\sigma_{\text{STRF}}(t, x)]^2 / \int \int dt dx [\text{STRF}(t, x)]^2 \quad (18)$$

where ΔT and ΔX are the length of time and number of octaves over which the STRF was measured.

δ is a measure of the average standard deviation in units of the maximum of the STRF. ϵ is a measure of the variance in units of power. If noise is additive, then $\epsilon = P_{\epsilon}/(P + P_{\epsilon}) = 1/(\text{SNR} + 1)$, with P = power, P_{ϵ} = noise power, and SNR = signal-to-noise ratio. ϵ should go down with the number of recordings, assuming the system can be described as the time-invariant random process.

RESULTS

Data presented here were collected from 22 single-unit and 54 multiunit recordings in 11 ferrets. In the summary histograms, both single units and multiunit are included but are distinguished from each other.

Most units encountered in AI respond well to moving rip-

ples. Responses are typically phase-locked to the moving envelope of the ripple over a range of ripple velocities and densities. However, of a total of 172 recordings made, only 76 cases provided adequate quality and quantity of responses. The reasons for this low yield vary. For example, we have encountered responses from a few units that were either poorly phase-locked or were inconsistent from trial to trial; such units were abandoned since our analysis methods are unsuitable for their characterization. Also because of extended recording times, typically over an hour, units were sometimes lost before sufficient data could be collected to carry out a full analysis. In other cases, the unit or animal changed state during the recording session, rendering the data unreliable. The reason for the extended recording time is to present ripple sounds and other sounds consisting of combinations of ripples, so we can verify linearity by using the STRFs to predict the response of the cell to new sounds. We found empirically that about 10,000 spikes are typically needed to obtain an STRF with well-defined features in response to single ripples, which with our sound paradigm usually corresponds to a 20-min presentation per cross-section. To eliminate data corresponding to unreliable cells, as described in the preceding text, we use units only with values of $\delta \leq 0.12$ and $\epsilon \leq 0.7$ (see METHODS) as the threshold for rejecting the data. These reliability statistics takes into account most of the preceding sources of error. The values of 0.12 and 0.7 are somewhat arbitrary, though we found that cells tended to separate themselves into two populations above

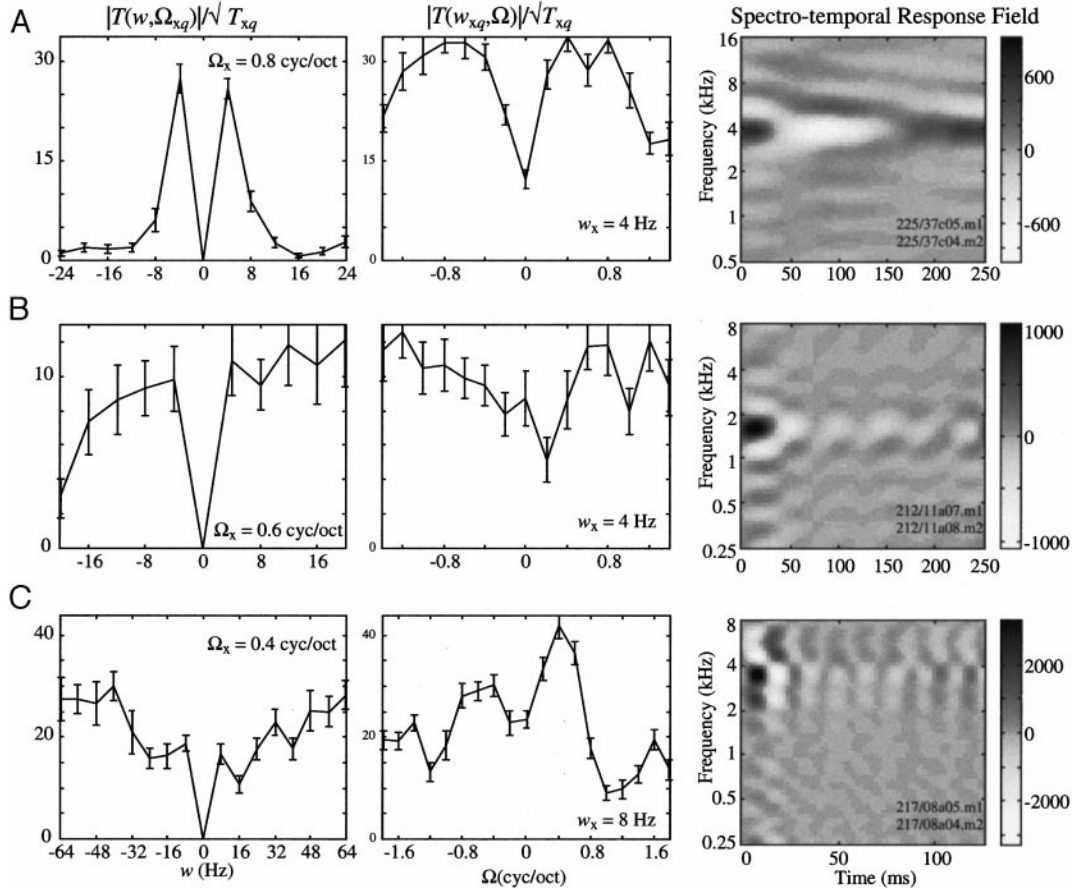


FIG. 9. Further examples of spectro-temporal transfer function sections, corresponding STRFs. Conventions as in Fig. 8. A: with narrow ripple velocity bandwidth. B: with broad ripple velocity bandwidth. C: a spectrally asymmetric unit.

and below these thresholds, respectively, and that the mathematical criteria of reliable versus noisy cell corresponded well with our intuitive perception based on visual inspection.

Responses to moving ripples

On average, AI units synchronize their responses to upward and downward moving ripples equally effectively with ripple velocities ranging from 2 to over 100 Hz, and ripple densities up to 4 cycle/octave. Examples of several temporal and spectral transfer function magnitudes are shown in Figs. 8–10, each with its corresponding STRF. In all cases, units respond well only over a specific range of ripple velocities and ripple densities, but the detailed shape and extent of the transfer functions vary from one unit to another. For instance, the unit in Fig. 9A responds well only to ripple velocities of ± 4 Hz, whereas the unit in Fig. 9C responds well at least up to ± 64 Hz. The unit in Fig. 6 responds well to ripple densities within ± 0.4 cycle/octave, whereas the unit in Fig. 10A responds over a wider range of densities but poorly at 0 cycle/octave.

As described in the preceding text, the transfer function at $w = 0$ is set to 0 since it is not well defined (and so has 0 contribution to the STRF). Additionally, for 12 cells (not shown), the transfer function was measured from ± 8 to ± 1 Hz in 1-Hz steps, and in all cases, the transfer function was negligible at the slowest ripple velocities (in contrast to the average firing rates, which remained significant).

Units also vary significantly in the asymmetry of their transfer functions with respect to the direction of the moving ripple. For example, responses to the two directions are relatively equal (transfer functions are roughly symmetric) in Figs. 6 and 9A. By comparison, the temporal transfer functions in Fig. 8, A–C, are asymmetric. The unit in Fig. 8B responds better to upward moving ripples; the unit in Fig. 8C responds over a wider range to downward moving ripples. These asymmetries are discussed in depth later in the context of transfer function separability.

The STRFs derived from these transfer functions commonly exhibit alternating significant regions of positive peaks and negative basins, interpreted here as excitatory and inhibitory regions, respectively. The four STRFs illustrated in Figs. 6 and 8 are of units that are tuned between 1 and 2 kHz. However, the shapes of the surrounding inhibitory regions vary considerably reflecting the different temporal and spectral transfer functions (see Fig. 11). For instance, STRFs may be relatively symmetric (Fig. 8A) or asymmetric (Fig. 9C). They can be clearly directional, i.e., tilted one way (Fig. 8B) or the other (Fig. 8C) on the spectro-temporal surface.

STRFs display a wide variety of shapes that are briefly described in the following text. The majority of AI cells exhibit STRFs with a simple excitatory field and varying amounts of inhibitory surround. The first peak of the excitatory portion indicates the BF_{STRF} of the unit, while its extent reflects its tuning curve at a given level.

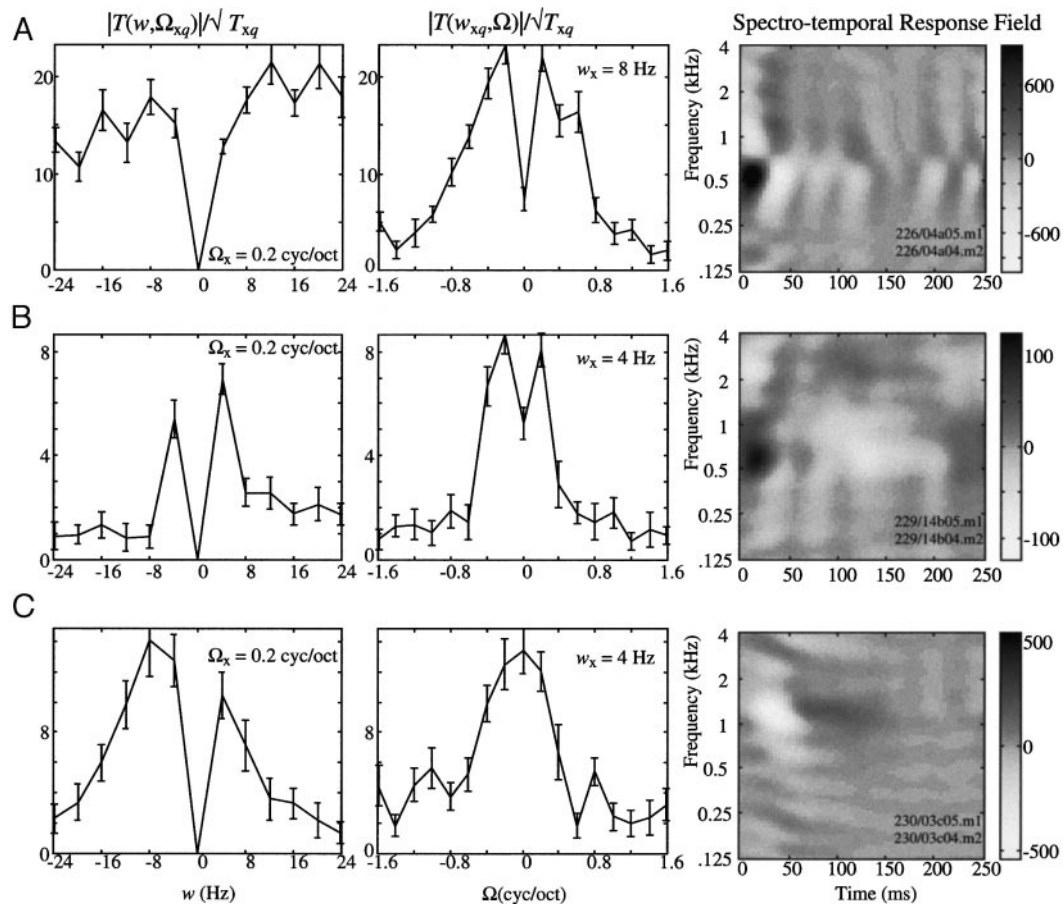


FIG. 10. Further examples of spectro-temporal transfer function sections, corresponding STRFs. Conventions as in Fig. 8. A: fast dynamics. B: slow dynamics. C: offset cell.

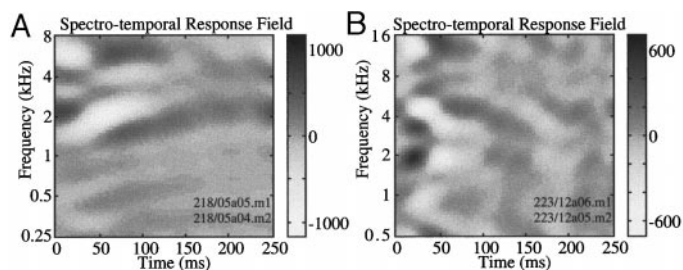


FIG. 11. Two cells with unusual receptive fields.

In many cases, the inhibitory surround is spectrally asymmetric around the BF_{STRF} (Fig. 9C); such asymmetry is effectively captured by the parameter ϕ (Eq. 14), where ϕ values near zero indicate roughly symmetric STRFs, while $\phi \approx 90^\circ$ indicate strong inhibition below the BF_{STRF} , and $\phi \approx -90^\circ$ indicates strong

inhibition above the BF_{STRF} . The ϕ distribution in our sample is summarized in Fig. 12C. It closely resembles that seen earlier with downward moving and stationary ripples (Kowalski et al. 1996a; Schreiner and Calhoun 1994; Versnel et al. 1995).

STRFs also vary considerably in their temporal dynamics, best seen in the $t-x$ domain. Some are fast with envelopes that decay relatively rapidly (Figs. 9C and 10A). Others are slow, taking over 150 ms to decay (as in Figs. 9A and 10B). These response dynamics reflect details of the temporal transfer function such as the ripple velocity at which it peaks (characteristic ripple velocity) and its width (ripple velocity bandwidth). STRFs also exhibit an onset delay (or latency) that is captured by the τ_d values, derived from the phase function (Eq. 13). The distribution of this delay tends to be well clustered around 25 ms as seen in Fig. 12B. Finally, unit STRFs can be generally classified as either *onset* (Figs. 9, A–C, and 10,

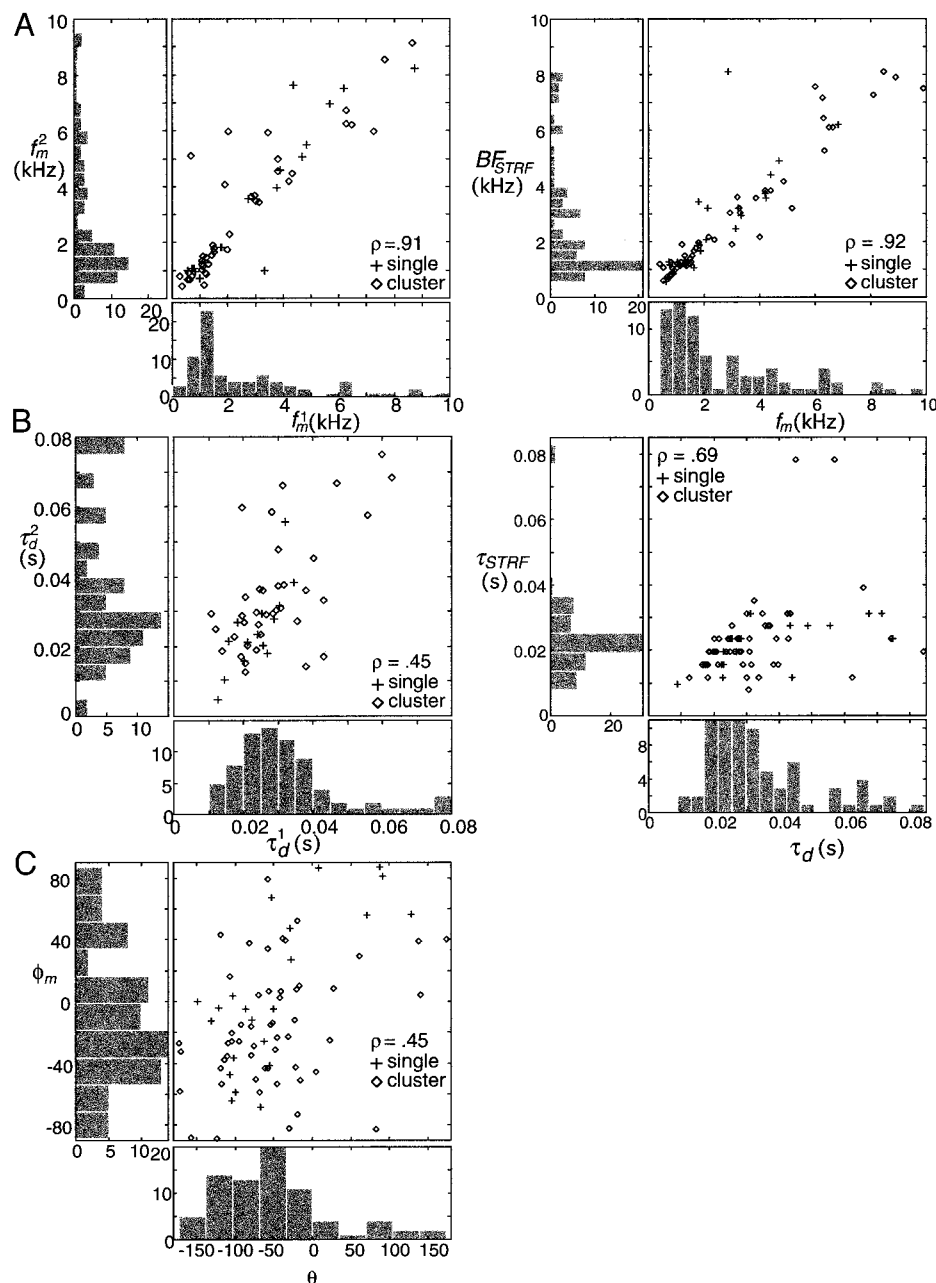


FIG. 12. The statistical distribution of parameters. *A, left:* f_m^1 vs. f_m^2 , i.e., center frequency f_m as determined by quadrant 1 vs. quadrant 2. *Right:* f_m vs. BF_{STRF} , i.e., quadrant averaged center frequency f_m vs. frequency giving highest STRF peak. *B, left:* τ_d^1 vs. τ_d^2 , i.e., mean delay τ_d as determined by quadrant 1 vs. quadrant 2. *Right:* τ_d vs. τ_{STRF} , i.e., quadrant averaged mean delay τ_d vs. delay giving highest STRF peak. *C:* temporal phase θ vs. spectral phase ϕ_m . Note that some outliers are not included in the figure; however, all data points are used in statistics.

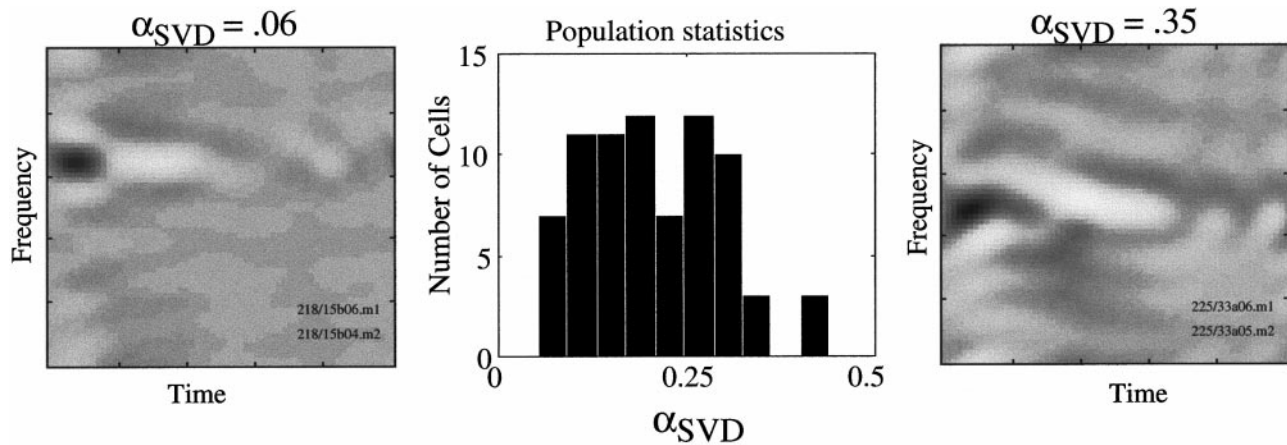


FIG. 13. Small and a large α_{SVD} cell, with the distribution of α_{SVD} in the *middle*. There is no obvious separation of cells into 2 populations, fully separable and inseparable, but rather a continuum.

A and B, most cells) or *offset* (Fig. 10C), a property that corresponds, respectively, to the negative or positive sign of the parameter θ . Onset STRFs are far more common in our sample as seen in the θ distribution in Fig. 12C.

Finally, STRFs may display very complex dynamics and spectro-temporal selectivity that are not easily captured by simple parameters. Two examples of such STRFs are shown in Fig. 11. One might be tempted to dismiss such STRFs as mere

aberration or noise except that they are derived from repeatable responses ($\delta = 0.10$ and $\epsilon = 0.49$ for Fig. 11A and $\delta = 0.03$ and $\epsilon = 0.04$ for Fig. 11B).

Separability and its relation to STRF shape

Separability is an important property of the transfer functions that has significant experimental and theoretical implica-

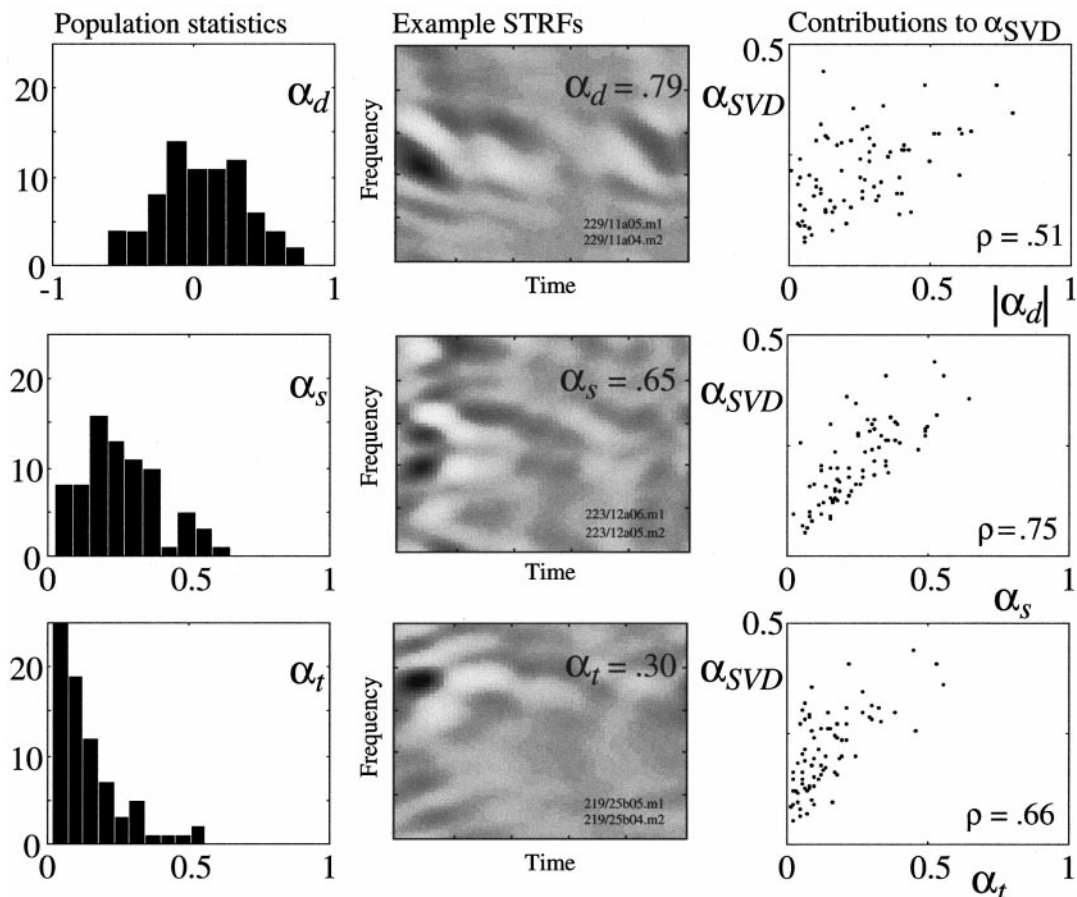


FIG. 14. The distribution of 3 insepability indicators, α_d , α_s , and α_t . *Middle*: examples of STRFs with extreme values of the corresponding insepability indicator. *Right*: distributions of each insepability indicator, plotted against total insepability, α_{SVD} . Because there is always some level of noise, no cell has an α_s or α_t exactly equal to 0.

tions. In this paper, we assume quadrant separability and ask whether responses are fully separable, the degree of inseparability, and the origin of the inseparability. Each of these indicators has a potentially useful interpretation for the shape of the STRF and the underlying structure of processes that give rise to it.

The simplest and most general way to examine full separability is to compute the SVD matrix α_{SVD} (Eq. 6). Figure 13 illustrates the distribution of α_{SVD} , Eq. 7, computed from all the cells used. Values near 0 indicate that only the first singular value has a large nonzero value and hence that the STRF is fully separable. Increasing values indicate increasing degree of inseparability. A significant fraction of cells deviate from full separability.

It can be shown that fully separable transfer functions must have magnitudes that are symmetric about the (w, Ω) origin. Alone, α_{SVD} offers no insight into the specific nature of these departures from the symmetric, separable case. However, it will be shown that there are three parameters (Eqs. 8–10) that in combination form α_{SVD} and that each corresponds to a specific distortion of a separable transfer function:

1) α_d , the response directionality, or the imbalance in the overall strength of the responses to the upward and downward moving ripples;

2) α_t , the asymmetry in the temporal transfer function $F(w)$;

3) α_s , the asymmetry in the spectral transfer function $G(\Omega)$.

The distribution of these three parameters is shown in Fig. 14. The directionality parameter α_d is distributed approximately normally between negative and positive values. This parameter is closely related to the directional selectivity of the STRF. STRFs with large $|\alpha_d|$ values exhibit obvious directional shapes such as seen in Fig. 14 (top, middle). A significant proportion of units (37%) also have spectral dissimilarity values (α_s) exceeding 0.3. An STRF with especially large α_s is shown in Fig. 14 (middle). Note that these STRFs may not necessarily exhibit obvious directionally selective shapes.

A strikingly different finding is the dearth of units (12%) with significant temporal dissimilarity ($\alpha > 0.3$) as seen in the distribution in Fig. 14 (bottom, left). An STRF with $\alpha = 0.30$ is displayed in Fig. 14 (bottom, middle): it is difficult to detect simple correlates of the large α_t values in the shape of the STRF. Note that this is not due to measuring the temporal transfer function at six points and the spectral transfer function at eight points in each quadrant: when the last two points of the spectral cross section are removed, the same results are obtained.

The three inseparability indicators do not appear to be significantly correlated, based on the pairwise scatter plots in Fig. 15, suggesting that independent mechanisms underlie the expression of each factor. By contrast, each factor (as expected) is well correlated with the total SVD index as seen in Fig. 14 (right).

We can define a composite measure of inseparability, the

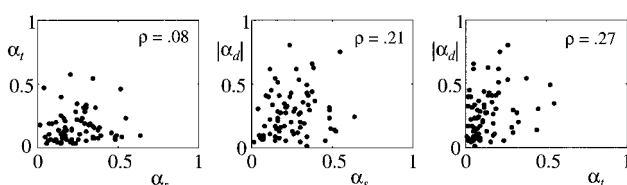


FIG. 15. The correlation of 3 inseparability indicators, α_t , α_s , and α_d .

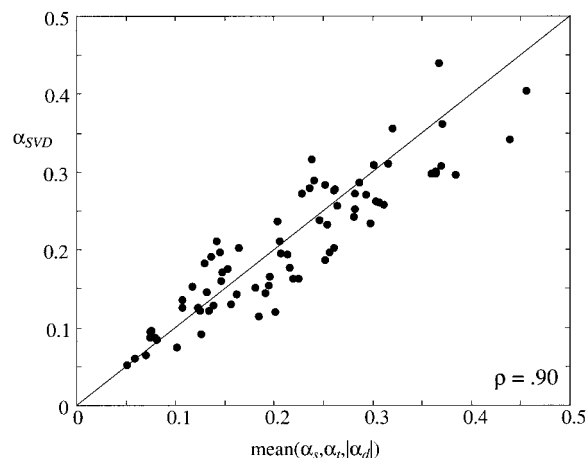


FIG. 16. The correlation between the mean of α_t , α_s , and $|\alpha_d|$ and α_{SVD} .

mean of α_t , α_s , and $|\alpha_d|$. Figure 16 illustrates that this measure is highly correlated to α_{SVD} and hence is an equally valid measure of inseparability.

There is no sharp threshold for inseparability. In Fig. 13, for instance, $\alpha_{\text{SVD}} \approx 0.35$ clearly corresponds to an inseparable cell. However, because of the continuum of values for α_{SVD} , there is no obvious cutoff.

DISCUSSION

Summary of results

The emphasis of this work has been on presenting a technique to describe neural response patterns of units in the cortex. More precisely, we use moving ripples to characterize the spectral and temporal properties of responses of auditory cortical neurons, although this is a general method that can be used for any population of neurons for which responses are shown to be substantially linear for broadband stimuli.

We have examined the nature of AI responses to rippled spectra moving in both upward and downward directions and incorporated these responses into the STRF. A summary of the main results follows.

1) We confirm earlier findings (Kowalski et al. 1996a) that AI units respond in a phase-locked fashion to the moving ripples over a range of velocities and directions that depend on the ripple density of the spectrum. In particular, responses are usually tuned around a specific ripple velocity and density. In the ferret, responses are commonly best in the 4- to 16-Hz range and densities lower than 2 cycle/octave. These findings are roughly consistent with those found in different species using different experimental paradigms: experiments with dynamic spectra (e.g., narrowband such as AM and FM tones or broadband such as modulated noise and click trains) have found similar maximum rates of synchronized responses in AI (Eggermont 1994; Schreiner and Urbas 1988).

2) We demonstrate a similarity between responses to upward and downward moving ripples. Specifically, the response parameter values and distributions to either direction are comparable (even if unequal), and hence reflect general dynamic response properties, not direction specific properties per se.

3) Complete spectro-temporal transfer functions are measured that exhibit a rich variety of shapes and cover a wide range of stimulus parameters. The STRF describes the way AI

units integrate stimulus power along the spectro-temporal dimensions.

4) We illustrate a variety of STRFs with a broad range of BFs, bandwidths, asymmetrical inhibition, temporal dynamics, and direction selectivity. We have assessed the prevalence of these features over all sampled units by examining the distribution of specific parameters that reflect each of these features.

5) The degree and origin of inseparability of the unit transfer functions is assessed using two methods. In the first, SVD analysis is applied to the entire transfer function to determine the number and ratio of the resulting singular values. The results indicate that AI units span a relatively uniform distribution between full separability to moderate inseparability. In the second method, we examine the origin of inseparability and find that it is primarily due to two factors: imbalance in the response power and an asymmetry in the spectral transfer function relative to the direction of ripple motion. Interestingly, we find that temporal (but not spectral) transfer functions are relatively symmetric and hence contribute little to overall transfer function inseparability.

In Kowalski et al. (1996a,b), pentobarbital was used for anesthesia; in the present study, a ketamine/xylazine combination was used. In Kohn et al. (1996), the effect of different anesthetics on the tuning properties of auditory cortical cells as a whole was presented. Under ketamine, a wider variety of responses was found, tuning to ripple density was slightly lower (from 1.05 cycle/octave under pentobarbital to 0.8 cycle/octave under ketamine), and no significant change in temporal tuning was observed. Other properties, though, such as linearity of the STRF for downward moving ripples, were unchanged. These results can be accounted for by assuming that overall, response fields measured with ripples have less inhibition under ketamine than under pentobarbital.

Separability and its implications

An important property of the responses is that for ripples moving in only one direction, the spectral and temporal functions are separable: within each quadrant they can be measured independently of each other. The property of quadrant separability makes it possible to measure the overall spectro-temporal transfer function in reasonable times using only single ripples since only a few velocity and spectral density combinations need to be measured. We have established (Kowalski et al. 1996a) that all recorded transfer functions in AI exhibit quadrant separability. In the experiments reported here, we assumed quadrant separability (Kowalski et al. 1996a,b) and proceeded to examine whether the resulting two-dimensional transfer functions are fully separable. Our findings indicate that AI responses fall uniformly on a continuum between moderately to fully separable.

A fully separable cell cannot be directionally selective in its responses. Inseparability is a necessary condition for the formation of more complex STRFs; direction selectivity is one possible consequence of inseparability. A directionally selective STRF usually has a distinctive elongated form along a spectro-temporal direction that matches that of its most sensitive ripple stimulus. For example, the STRF illustrated in Fig. 8B is most responsive to a ripple $\Omega = -0.4$ cycle/octave, $w = -8$ Hz, whose spectrogram matches well the outline of the STRF spacing and orientation. Direction selectivity implies

that a unit is differentially responsive to one direction of ripple movement and hence must have a significant nonzero directionality index. Therefore direction selectivity necessarily implies an inseparable STRF. The opposite is not true: an inseparable STRF might reflect other factors such as asymmetric temporal and/or spectral transfer functions (α_t or $\alpha_s \neq 0$), which do not manifest themselves in an obvious elongated form or preferential responses to one direction or another (as shown in Fig. 14, *center column, middle and bottom*).

Separability also places strong constraints on the underlying biological processes that give rise to the STRF shapes. For example, full separability suggests that the STRF is constituted of independent temporal and spectral processing stages. By contrast, inseparability (or just quadrant separability) implies spectrally and temporally intertwined stages of processing with the specific form of the model being entirely dependent on the details of the transfer functions. Quadrant separability in particular is a very strong constraint on both the neural inputs and the processing of the unit: almost all neural networks (whether linear or nonlinear) with multiple fully separable STRFs as inputs will in general produce a totally inseparable STRF. In particular, the naive procedure of constructing a directionally sensitive STRF by taking the simple sum of two fully separable STRFs with differing f_m and τ_d will produce a totally inseparable STRF which is not quadrant separable. To produce a quadrant separable STRF requires special inputs and/or special processing.

It can be shown that a quadrant separable, temporally symmetric (i.e., $\alpha_t \ll 1$), cortical neuron can be easily constructed by taking inputs from (potentially) many units with (potentially) different spectral response fields and even with (potentially) different temporal impulse response properties as long as the temporal dynamics of the inputs to the cortical cell are fast compared with the temporal dynamics of the cortical cell itself (Simon et al. 2000). Quadrant separability then occurs when the inputs are temporally phase-lagged relative to each other [though not necessarily 90° as in Saul and Humphrey (1990) and Dong and Atick (1995)].

This is consistent with the input neural connectivity one expects from layer IV cortical neurons, which receive input from thalamic medial geniculate body (MGB). MGB neurons may have fully separable STRF [as is the case for typical inferior colliculus central (ICC) neurons (Escabi and Schreiner 1999)] with different spectral response fields (differing in width, extent/location of inhibitory bands, and to a lesser extent, best frequency). MGB temporal cross-sections of transfer functions are essentially constant when low-passed at a cutoff frequency appropriate to cortical behavior (e.g., typically well below 100 Hz) (Yeshurun et al. 1985). Furthermore some MGB neurons may have a temporal phase lag, as in the visual system's lateral geniculate's "lagged cells" (Saul and Humphrey 1990).

Significantly, the property of quadrant separability with temporal symmetry does not allow for any cortical inputs unless those inputs have the same temporal behavior as the neuron studied. If, for instance, all neurons in the same cortical column have similar temporal properties, including similar neural delays, this would be consistent with quadrant separability. Otherwise, cortical inputs would break quadrant separability and create a totally inseparable neuron. Total inseparability would be expected for cortical neurons in layers that receive signifi-

cant input from other cortical columns or from any other neural source with significantly different temporal processing, including (but not limited to) any significant delays.

It is possible that this extremely constraining result is an anesthesia-induced effect. If not, the result is a fascinating constraint on the neural network providing input to a given cortical cell.

We particularly thank A. Saul. We also thank J. Eggermont, I. Ohzawa, and M. Slaney for very helpful and illuminating discussions.

This work was supported by Office of Naval Research Multidisciplinary University Research Initiative Grant N00014-97-1-0501, National Institute on Deafness and Other Communication Disorders Training Grant T32 DC-00046-01, and National Science Foundation Grant NSFD CD8803012.

REFERENCES

- AERTSEN AMHJ AND JOHANNESMA PIM. The spectro-temporal receptive field: a functional characteristic of auditory neurons. *Biol Cybern* 42: 133–143, 1981.
- DECHARMS RC, BLAKE DT, AND MERZENICH MM. Optimizing sound features for cortical neurons. *Science* 280: 1439–1443, 1998.
- DEPIREUX DA, SIMON JZ, AND SHAMMA SA. Measuring the dynamics of neural responses in primary auditory cortex. *Comments Theor Biol* 5: 89–118, 1998.
- DONG D AND ATICK JJ. Temporal decorrelation: a theory of lagged and nonlagged cells in the lateral geniculate nucleus. *Network: Comput Neural Syst* 6: 159–178, 1995.
- EFRON B AND TIBSHIRANI B. *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.
- EGGERMONT JJ. Wiener and Volterra analyses applied to the auditory system. *Hear Res* 66: 177–201, 1993.
- EGGERMONT JJ. Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity. *Hear Res* 74: 51–66, 1994.
- ESCAPI MA AND SCHREINER CE. Non-linear spectro-temporal envelope processing in the cat ICC. *Assoc Res Otolaryngology Abstr* 22: 869, 1999.
- EVANS EF. Single-unit studies of mammalian cochlear nerve. *Aud Invest Sci Technol Basis* 68: 324–367, 1979.
- EVANS E, ROSS H, AND WHITFIELD I. The spatial distribution of unit characteristic frequency in the primary auditory cortex of the cat. *J Physiol (Lond)* 171: 476–493, 1965.
- GHIGLIA DC AND PRITT MD. *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*. New York: Wiley-Interscience, 1998.
- HAYKIN SS. *Adaptive Filter Theory*. Upper Saddle River, NJ: Prentice-Hall, 1996.
- HEIL P, RAJAN R, AND IRVINE D. Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. II. Organization of response properties along the “isofrequency” dimension. *Hear Res* 63: 135–156, 1992.
- HEIL P, RAJAN R, AND IRVINE D. Topographic representation of tone intensity along the isofrequency axis of cat primary auditory cortex. *Hear Res* 76: 188–202, 1994.
- IMIG T AND ADRIAN H. Binaural columns in the primary field (A1) of cat auditory cortex. *Brain Res* 138: 241–257, 1977.
- KLEIN DJ, DEPIREUX DA, SIMON JZ AND SHAMMA SA. Robust spectro-temporal reverse correlation for the auditory system: optimizing stimulus design. *J Comput Neurosci* 9: 85–111, 2000.
- KOHN A, DEPIREUX D, AND SHAMMA S. Effects of different anesthetics on the responses to broad-band sounds in the ferret primary auditory cortex. *Soc Neurosci Abstr* 22: 424.12, 1996.
- KOWALSKI N, DEPIREUX DA, AND SHAMMA SA. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single unit responses to moving ripple spectra. *J Neurophysiol* 76: 3503–3523, 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. *J Neurophysiol* 76: 3524–3534, 1996b.
- KVALE M AND SCHREINER CE. Perturbative m-sequences for auditory systems identification. *Acustica* 81: 1–2, 1995.
- MENDELSON JR, SCHREINER CE, SUTTER ML, AND GRASSE KL. Functional topography of cat primary auditory cortex: responses to frequency-modulated sweeps. *Exp Brain Res* 94: 65–87, 1993.
- MIDDLEBROOKS JC, DYKES RW, AND MERZENICH MM. Binaural response-specific bands in primary auditory cortex of the cat: topographical organization orthogonal to isofrequency contours. *Brain Res* 181: 31–48, 1980.
- NELKEN I, PRUT Y, VAADIA E, AND ABELES M. Population responses to multifrequency sounds in the cat auditory cortex: one- and two-parameter families of sounds. *Hear Res* 72: 206–222, 1994.
- PAPOULIS A. *The Fourier Integral and its Applications*. New York: McGraw-Hill, 1962.
- POLITIS DN. Computer-intensive methods in statistical analysis. *IEEE Signal Process* 15: 39–54, 1998.
- SAUL AB AND HUMPHREY AL. Spatial and temporal response properties of lagged and nonlagged cells in cat lateral geniculate nucleus. *J Neurophysiol* 64: 206–224, 1990.
- SCHREINER CE AND CALHOUN BM. Spectral envelope coding in cat primary auditory cortex: properties of ripple transfer functions. *Aud Neurosci* 1: 39–61, 1994.
- SCHREINER CE, MENDELSON J, AND SUTTER ML. Functional topography of cat primary auditory cortex: representation of tone intensity. *Exp Brain Res* 92: 105–122, 1992.
- SCHREINER CE AND SUTTER ML. Topography of excitatory bandwidth in cat primary auditory cortex: single-neuron versus multiple-neuron recordings. *J Neurophysiol* 68: 1487–1502, 1992.
- SCHREINER CE AND URBAS JV. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear Res* 32: 49–64, 1988.
- SHAMMA SA, FLESHMAN JW, WISER PR, AND VERSNEL H. Organization of response areas in ferret primary auditory cortex. *J Neurophysiol* 69: 367–383, 1993.
- SHAMMA SA, VERSNEL H, AND KOWALSKI N. Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra. *Aud Neurosci* 1: 233–254, 1995.
- SIMON JZ, KLEIN DJ, SHAMMA SA, AND DEPIREUX DA. Functional and structural implications of non-separability of spectral and temporal responses in AI. *Assoc Res Otolaryngol Abstr* 23: W2, 2000.
- SUTTER ML, LOFTUS WC, AND O’CONNOR KN. Temporal properties of two-tone inhibition in cat primary auditory cortex. *Assoc Res Otolaryngol Abstr* 19: 452, 1996.
- THEUNISSEN FE, SEN K, AND DOUPE AJ. Spectral-temporal receptive fields of non-linear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315–2331, 2000.
- VERSNEL H, KOWALSKI N, AND SHAMMA SA. Ripple analysis in ferret primary auditory cortex. III. Topographic distribution of ripple response parameters. *Aud Neurosci* 1: 271–285, 1995.
- WATSON AB AND AHUMADA AJ. Model of human visual-motion sensing. *J Opt Soc Am A* 2: 322–341, 1985.
- 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.