

A linear systems view to the concept of STRFs

Mounya Elhilali⁽¹⁾, Shihab Shamma^(2,3), Jonathan Z Simon^(2,3,4),

Jonathan B. Fritz⁽³⁾

⁽¹⁾Department of Electrical & Computer Engineering, Johns Hopkins University

⁽²⁾Department of Electrical & Computer Engineering, University of Maryland, College Park

⁽³⁾Institute for Systems Research, University of Maryland, College Park

⁽⁴⁾Department of Biology, University of Maryland, College Park

1. Introduction

A key requirement in the study of sensory nervous systems is the ability to characterize effectively neuronal response selectivity. In the visual system, striving for this objective yielded remarkable progress in understanding the functional organization of the visual cortex and its implications to perception. For instance, the discovery of ordered feature representations of retinal space, color, ocular dominance, orientation, and motion direction in various cortical fields has catalyzed extensive studies into the anatomical bases of this organization, its developmental course, and the changes it undergoes during learning or following injury. In the primary auditory cortex (A1), response selectivity of the neurons (also called their *receptive fields*) is ordered topographically according to the frequency they are most tuned to, an organization inherited from the cochlea. Beyond their tuning, however, A1 responses and receptive fields exhibit a bewildering variety of dynamics, frequency bandwidths, response thresholds, and patterns of excitatory and inhibitory regions. All this has been learned over decades of testing with a large variety of acoustic stimuli (tones, clicks, noise, and natural sounds) and response measures (tuning curves, rate-level functions, and binaural maps).

A response measure that has proven particularly useful beyond the feature-specific measures enumerated earlier is the notion of a generalized *spatiotemporal*, or equivalently for the auditory system, the *spectro-temporal receptive field* (STRF). It is distinguished from other measures by its broader descriptive power

(encompassing both dynamics and spectral selectivity) and its relatively noncommittal nature (not requiring too much prior knowledge such as frequency tuning or threshold). In the last two decades, the STRF has become widely employed in all sensory systems (Aertsen & Johannesma, 1981; De Valois & De Valois, 1988; DeAngelis, Ohzawa, & Freeman, 1995; Gosselin & Schyns, 2002; Hochstein & Shapley, 1976; Sripathi, Yoshioka, Denchev, Hsiao, & Johnson, 2006), and especially in the auditory system where it had found its original promise, development, and examination of its value and liability (Aertsen & Johannesma, 1981; De Boer, 1967; de Boer & de Jongh, 1978; Sripathi, Yoshioka, Denchev, Hsiao, & Johnson, 2006). The STRF has primarily been viewed as a linear characterization of the complex stimulus-response transformations seen in sensory neurons. However, there are well-understood limitations of the STRF ability to capture all the essential details of a sensory neuron's response. These stem from the existence of a host of known nonlinearities such as spiking threshold, rate saturation, synaptic-depression) that can complicate the interpretation of the STRF and render its representation stimulus dependent or meaningless. Nevertheless, the utility of the STRF has inspired deeper and broader examination of its constraints and pitfalls, of ways to circumvent them (Atencio, Sharpee, & Schreiner, 2008; Christianson, Sahani, & Linden, 2008; David, Hayden, & Gallant, 2006; Nagel & Doupe, 2008) or even exploit them to learn more about the nonlinearities in the system (Ahrens, Linden, & Sahani, 2008; Klein, Simon, Depireux, & Shamma, 2006; Theunissen, Sen, & Doupe, 2000). In the following sections, we explore the mathematical bases of the STRF, its richness and limitations, and ways to estimate it reliably. We then briefly review its utility in a wide range of investigations, including the cortical encoding of sound and the exploration of learning and rapid plasticity during behavior.

2. Systems approach to characterizing cortical neurons

Approaches to understanding sound processing in the auditory system, particularly at the level of auditory cortex, have often treated neurons as “black-box” devices; i.e. systems whose internal workings are opaque, but that can be

mathematically described solely in terms of their input-output relationship. Each neuron is then understood as a transducer that is driven via sensory stimulation by external sounds to evoke a neural response. In absence of any stimulation, the neuron is in its spontaneous (or quiet) state. When presented with a sensory input, the neuron is driven to spike either actively above its spontaneous state; or in a suppressive manner below its spontaneous state. The pattern of neural firing depends on the characteristics of the sensory stimulation and its match to the neuron's own tuning properties.

By understanding how a stimulus can drive a neuron to fire in an excitatory or inhibitory fashion, one gains better insight into the functional role of this neuron in sensory information processing, its tuning characteristics, and predictive knowledge of the neuron's behavior under any stimulation conditions.

Clearly, the ability to fully describe the workings and behavior of neurons (even if treated as black boxes) is often impossible. The transducer view of neurons is only possible when the neuron's output is controlled *solely* by its sensory input. Most sub-cortical stages can be reduced –somewhat– to this transducer view. Such is not the case for sensory cortex, where many neurons appear to play a role not only in sensory encoding, but also in memory, decision and behavior; therefore integrating information across a neural network spanning one or many neural loci. As the circuitry becomes more intricate, it is often impossible to understand the role of a single neuron in abstraction of the complex dynamics of the network in which it exists.

With this caveat in mind, one can still employ a number of tools to gain elementary insight into the sensory information encoding of cortical neurons, the topic of interest in this issue. Systems theory provides the mathematical language to characterize the behavior of neurons (or systems in general), and provide a better understanding of their function (Bender, 2000). For many complex systems, such as the nervous system, this characterization often focuses on investigating the dependencies of the output function (neural response) on the input function (input stimulation). This approach adopts the viewpoint of *functionals*, in which the signal as a whole is considered, rather than its values at given points in time. In other

words, the neuron is viewed as a transducer of an input signal $s(t)$ to a neural response $r(t)$ via a mapping F :

$$r(t) = F[s(t)] \quad (1)$$

$F[.]$ is a functional (function of a function) of the input signal $s(t)$. By treating the neuron as a black-box and limiting its characterization to its input-output function, we are deriving a good understanding of its behavior without incorporating any knowledge of its internal mechanisms (e.g. ionic channels), its biophysical principles or anatomical structure. Such simplification is neither uncommon (Isermann & Munchhof, 2011; Rugh, 1981; Sastry, 1999), nor unwarranted when dealing with a complex structure such as cortical neurons.

The simplest form of systems is linear and time-invariant (LTI). A linear system obeys two key rules: homogeneity and superposition. The former means that the response to stimulus s that is scaled by a factor α is also scaled by the same amount. The latter means that a cell driven by the sum of two stimuli $s_1 + s_2$ responds to the sum of the responses to the individual stimuli. A time-invariant system is one whose behavior depends on its state, independent of time; so that the system's response to a given stimulation does not depend on the time it is applied.

LTI systems that are single-input, single-output can be represented by the familiar convolution integral:

$$r(t) = \int_{-\infty}^{\infty} h(\tau)s(t - \tau)d\tau \quad (2)$$

where $s(t)$ is the input and $r(t)$ the output of the system. The kernel $h(t)$ is called the impulse response of the system, which is the limit of the sequence of responses when the system is excited with input that tends towards an impulse $\delta(t)$.

Intuitively, the convolution represents the correlation between the system's impulse response $h(\tau)$ and the time-reversed input around a given instant $s(t - \tau)$ to produce the output at time t . This operation is repeated for all time t , as the input signal continues to evolve. Thus, the system produces the strongest response when the input matches the system function $h(t)$. In this regard, $h(t)$ is sometimes thought of as a filter that selectively operates on the input signal in order to enhance or suppress certain components. Alternatively, one can think of $h(t)$ as a set of

weights which defines the output at time t as the weighted sum over the values of the input. If the system is not causal, this dependency would extend to the past, present and future values of $s(t)$; while for a causal system, this dependency is limited to only past and present values of $s(t)$. The ease of working with linear, time-invariant systems stems from the fact that $h(t)$ gives a *full* description of the black box model. It provides the mapping between *any* input signal and its corresponding output signal. If we know $h(t)$, we know everything there is to know about the system and can fully predict its behavior.

An equivalent representation of the convolution equation (Eq. 2) operates in Frequency domain. Denote the Fourier transform of $s(t)$ as $S(f)$, defined as

$$S(f) = \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft} dt \quad (3)$$

where $j = \sqrt{-1}$. Taking the Fourier transform of both sides of Eq. 2 yields

$$R(f) = H(f)S(f) \quad (4)$$

$H(f)$ is called the transfer function of the system. It is related in a one-to-one mapping to the system impulse response $h(t)$ (via the Fourier transform); and just like $h(t)$, knowing $H(f)$ means knowing everything there is to know about the behavior of the LTI system under study. The Fourier-domain equation reinforces the notion that linear systems only output frequencies that are uniquely determined by the same frequencies in the input; in other words, the output $R(f_0)$ depends solely on $S(f_0)$ and no other components of $S(f)$. The output $R(f_0)$ is simply an amplification of $S(f_0)$ by an amount dictated by $H(f_0)$. Therefore, the output of a linear system to a pure tone is also a pure tone of the same frequency; possibly amplified or dampened, and likely shifted in phase.

Such linearity cannot be expected or realistic in a system so plagued with nonlinearities such as the auditory system. A pure tone typically induces responses at the harmonics of the input frequency. The output to two pure tones is generally composed of the sum and differences of the harmonics of the input frequencies. This and a plethora of other nonlinear phenomena have been demonstrated throughout the auditory pathway, including two-tone suppression, dynamic intensity compression, lateral inhibition, masking and masking release phenomena as well as

unpredictable responses to natural stimuli (Goldstein, 1967; Nelken, 2004; Pickles, 2008).

When dealing with nonlinear systems, the convolution equation can be generalized to a more universal class of systems using Volterra series¹ (Volterra, 1959):

$$r(t) = \vartheta_0 + \int_{-\infty}^{\infty} \vartheta_1(\tau)s(t-\tau)d\tau + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \vartheta_2(\tau_1, \tau_2)s(t-\tau_1)s(t-\tau_2)d\tau_1d\tau_2 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \vartheta_3(\tau_1, \tau_2, \tau_3)s(t-\tau_1)s(t-\tau_2)s(t-\tau_3)d\tau_1d\tau_2d\tau_3 + \dots \quad (5)$$

$\vartheta_n(\tau_1, \tau_2, \dots, \tau_n)$ is called the n^{th} order Volterra kernel with independent variables $\tau_1, \tau_2, \dots, \tau_n$. It describes the n^{th} order nonlinearity of the system (Mathews & Sicuranza, 2000; Rugh, 1981). n is called the order of the system, and can in theory be infinite. Note that Volterra series still assume that the system is time-invariant. It is clear that the convolution integral (Eq. 2) is a special case of the Volterra series equation (Eq. 5) by setting $\vartheta_n(\tau_1, \tau_2, \dots, \tau_n) = 0$ for $n > 1$.

Now, if we take the Fourier transform of both sides of Eq. 5, we get:

$$Y(f) = \mathcal{V}_0\delta(f) + \mathcal{V}_1(f)X(f) + \int_{-\infty}^{\infty} \mathcal{V}_2(f_1, f-f_1)X(f_1)X(f-f_1)df_1 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{V}_3(f_1, f_2, f-f_1-f_2)X(f_1)X(f_2)X(f-f_1-f_2)df_1df_2 + \dots \quad (6)$$

where $\mathcal{V}_n(\tau_1, \tau_2, \dots, \tau_n)$ is the n -dimensional Fourier transform of $\vartheta_n(\tau_1, \tau_2, \dots, \tau_n)$. Just like LTI systems, knowledge of the Volterra kernels (in either the time-domain or frequency-domain) allows one to quantify the strength of the system nonlinearity and to quantify the nonlinear energy transfer.

Generally, Eq. 5 or Eq. 6 can be difficult to handle and practically challenging to determine especially for a system with unknown order n . The number of terms in

¹ Volterra series descriptions are limited to nonlinear systems that obey certain smoothness constraints; particularly systems with fading memory (Sandberg, 1983). A general description of series analysis of systems can be found in work by contemporaries of volterra, including Brilliant and McFee (Brilliant, 1958; George, 1959; McFee, 1961; Wiener, 1958).

the operators does grow both with degree of nonlinearity in the system as well as input dimensionality. This greatly limits their applicability to low-dimensional systems with mild nonlinearities; and even then, the computation tends to be quite tedious and involved.

A simpler formulation of the same functional has been proposed by Wiener, which reformulated the Volterra series representation into sums of multidimensional convolutions that are *mutually uncorrelated*; i.e. orthogonal with respect to zero-mean Gaussian white noise –GWN– (Wiener, 1958). By construction, the Wiener functional of order n can be obtained from the original Volterra kernels by a procedure similar to Gram-Schmidt orthogonalization; and is nothing but linear combinations of Volterra kernels up to order n (Ogunfunmi, 2007; Rugh, 1981). Their orthogonal nature also allows the Wiener series to be truncated after n functionals, giving the best n^{th} order polynomial approximation to the system (given GWN input). Higher order Wiener functionals are independent and can be added without affecting the estimate of the lower order ones already computed. On the other hand, the relationship between the Wiener kernels and the system properties is more blurred. For instance, the first-order Wiener functional is not the only term containing the linear convolution; all odd order functionals contain a linear term; so truncating the Wiener series would misrepresent the contribution of the linear and other nonlinear components to the system's behavior.

The use of GWN stimulation and Wiener-series has proven quite successful in modeling neural processing in a number of sensory systems; particularly at the peripheral level (Eggermont, 1993; Johnson, 1980; Marmarelis & Marmarelis, 1978; Wickesberg, Dickson, Gibson, & Geisler, 1984). While somewhat successful, such white-noise technique has proven quite inadequate and inapplicable to central auditory loci, particularly auditory cortex for a number of reasons. The most obvious one being that white noise stimulation is quite stationary and fails to elicit responses from cortical neurons.

The spectro-temporal receptive field (STRF):

Cortical neurons are more responsive to rich sounds with non-static spectro-temporal structures. As sounds journey up the auditory system to sensory cortex,

they come readily represented into their dynamic spectra. The cortical input is mapped onto a spectro-temporal representation; as tonotopically-organized temporal neural activity. It is therefore important to tie the spectro-temporal space with the functional definition presented above.

To take the structure of the input to cortical neurons into account and also begin to gauge the role of the spectro-temporal structure of sound, we can revisit the formulation of the Volterra series model of a system; particularly its bilinear or quadratic term. From Eq. 5, the second order term of the Volterra series can be written as:

$$\begin{aligned}
V_2[s(t)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \vartheta_2(\tau_1, \tau_2) s(t - \tau_1) s(t - \tau_2) d\tau_1 d\tau_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{\vartheta}_2(\tau, \theta) s\left(t - \tau - \frac{\theta}{2}\right) s\left(t - \tau + \frac{\theta}{2}\right) d\tau d\theta \quad , \tau = \frac{\tau_1 + \tau_2}{2}, \theta = \tau_2 - \tau_1 \quad (7) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} STRF_{V_2}(\tau, f) W(t - \tau, f) d\tau d\theta
\end{aligned}$$

where we recognize the quadratic term as the Wigner-Ville distribution² of the input signal $s(t)$; $W_s(t, f) = \int s\left(t - \frac{\theta}{2}\right) s\left(t + \frac{\theta}{2}\right) e^{-j2\pi f\theta} d\theta$; i.e. the Fourier transform of the signal's autocorrelation function. Here, $STRF_{V_2}$ is defined as:

$$STRF_{V_2}(\tau, f) \equiv \int_{-\infty}^{\infty} \hat{\vartheta}_2(\tau, \theta) e^{-j2\pi f\theta} d\theta \quad (8)$$

and can be thought of as Fourier transform of the second-order Volterra kernel. Eq. 7 gives a derivation of an STRF based on the Wigner-Ville distribution; and can be shown to be easily related to STRFs derived from other time-frequency representations of the signal via a filtering operation (Theunissen, Sen, & Doupe, 2000). The class of bilinear time-frequency representations of signals is interconvertible with respect to one another, as smoothed or convolved versions of each other. Whether the Wigner-Ville, Choi-Williams distribution, the Gabor transform or

² The presence of complex conjugation of $s(t)$ in the equations is ignored since all stimuli are real-valued.

the spectrogram representation, they can all be written in general form as (Cohen, 1995; Grochenig, 2001; Sang-Won Nam & Powers, 2003):

$$C_s(t, f) = \iiint_{-\infty}^{\infty} \Psi(\xi, \theta) e^{-j2\pi\xi(t-\tau)} \underbrace{s\left(\tau + \frac{\theta}{2}\right) s^*\left(\tau - \frac{\theta}{2}\right) e^{-j2\pi f\theta}}_{\sim \text{Wigner-Ville distribution}} d\theta d\tau d\xi \quad (9)$$

where $\Psi(\xi, \theta)$ is a kernel that determines the influence of cross-terms on the time-frequency function. Note that $\Psi(\sigma, \tau) = 1$ for the Wigner-Ville distribution. The choice of time-frequency representation to use depends on ease of interpretation and applicability to the sound class of interest (Eggermont & Smith, 1990; Kim & Young, 1994; Theunissen, Sen, & Doupe, 2000). It is also important to consider representations with proper support to enable mapping back and forth between $STRF_{V_2}$ (i.e. the Fourier transform of the second-order Volterra kernel) and any other equivalent STRF representation of the neuron (see (Klein, Depireux, Simon, & Shamma, 2000) for a discussion).

Regardless of the choice of spectro-temporal representation of the stimulus $s(t)$, Eq. 7 shows that the second-order Volterra functional can be thought of as a linear system that processes the Wigner-Ville distribution of the stimulus (or other equivalent representations).

3. Construction of Spectro-temporal Receptive Fields

Reverse correlation:

The most common way of constructing STRFs uses the classic *reverse correlation* method (De Boer & Kuyper, 1968). This technique amounts to computing the cross-correlation between an input stimulus $S(t, f)$ and neural response $r(t)$; i.e. $C(t, f) = \langle S(\tau - t, f)r(\tau) \rangle$, where $\langle \cdot \rangle$ corresponds to averaging over time. If the input signal is a spectro-temporal GWN, its autocorrelation function is a pure two-dimensional impulse, i.e, $\Phi(\tau, x) = \delta(\tau, x)$, since the only nonzero correlations are those between each time-frequency channel and itself; hence corresponding to an impulse function. One can in fact show that when driven with white noise, this cross-correlation $C(t, f)$ corresponds to the *impulse response* of the neuron

(Bracewell, 1999); therefore fully determining the system function (of course, assuming the system obeys principles of linearity and time-invariance).

Computing the cross-correlation $C(t, f)$ can be technically done in a number of ways. The term “reverse-correlation” emphasizes the notion that one centers this computation around the response $r(t)$, and looks back at the stimulus that precedes it (Fig. 1). As the neuron produces a spike, reverse-correlation postulates that the neuron must have “heard” a sensory input that it “likes”, therefore driving it to fire. So one looks back at the immediate past of the input just preceding the spike and saves it. As more spikes are generated, the inputs $S(t, f)$ preceding spikes are averaged out. Features that are consistently driving the neuron to fire will add up while features that are irrelevant will “wash-out”, leaving a crisp representation of the neuron’s preference of time-frequency features. Anchoring the computation around the system output makes reverse-correlation very adequate for spiking systems, such as neurons, whose outputs translate to a binary code of 0’s and 1’s; hence limiting the computation to only few instances when the neuron fires a spike. All instances during which the neuron is operating sub-threshold are irrelevant for this computation; since they contribute only zero to the summation. The time averaging operation reduces to a computation that grows linearly with the number of spikes generated by the neuron. One can in fact follow the evolution of this computation in real-time. As sound progresses over time, a sequence of stimulus patterns can be averaged in an online fashion each time the neuron spikes, until a pattern emerges; or some measure of estimation convergence is satisfied. This method has been applied widely throughout the neuroscience literature in studying behavior of sensory neurons (De Boer & Kuyper, 1968; Johnson, 1980; Jones & Palmer, 1987; Ringach, Sapiro, & Shapley, 1997).

[Figure 1 about here]

It is important to note here that the reverse-correlation method yields the second order Wiener kernel; which is in principle not equivalent to the second order Volterra kernel or the STRF as derived in Eq. 7. However, the distinction between

the two is often ignored as neurophysiologists often assume that the system's order does not extend to higher orders n ; hence making the two metrics equivalent.

Aside from the mathematical foundation of the reverse correlation technique, it is important to appreciate the intuition that if the time-frequency space is sampled randomly and uniformly, and assuming the system is LTI, knowing how each time-frequency bin drives a neuron gives us a complete description of the neuron's behavior (De Boer & Kuyper, 1968). Aside from the absurdity of trying to reduce complex neurons to simple LTI systems (as discussed throughout this book), the spike-triggered method is further limited by its reliance on a uniform sampling of time-frequency space. Nevertheless, the limitations of GWN have been circumvented by using more interesting stimuli that share a lot of the characteristics of GWN such as moving ripples, TORCs or other modulated broadband sounds (Escabi & Read, 2003; Klein, Depireux, Simon, & Shamma, 2000; Kowalski, Depireux, & Shamma, 1996; Miller, Escabi, Read, & Schreiner, 2002). The spike-triggered averaging methodology has also been extended to more general sound classes such as animal vocalizations, natural sounds and human speech. Though it yields a biased representation of the STRF, compensation can be applied by normalizing the representation using second-order statistics of the input (Aertsen & Johannesma, 1981; David, Mesgarani, Fritz, & Shamma, 2009; Eggermont, Aertsen, & Johannesma, 1983; Ringach, Sapiro, & Shapley, 1997; Theunissen, Sen, & Doupe, 2000).

Estimation errors

It is clear that the receptive field is not an invariant characterization of the spectro-temporal properties of auditory neurons. It is an approach that is riddled with problems. It varies with stimulus type including bandwidth, temporal dynamics, intensity, richness of spectro-temporal properties or any other properties that determine the effectiveness of a stimulus in driving a neuron; an issue considered in chapter 3. It also changes with different levels of analysis (see chapter 4), and poses great challenges for validation as a true transfer function of neurons (see chapter 7). Nevertheless, it remains one of the great tools available to neurophysiologists to provide a window (albeit narrow) into complex processing

taking place throughout the sensory system, particularly at the level of auditory cortex.

A number of issues should however be carefully considered when working with the concept of STRF: **(i)** Sensory stimulation is a crucial component to the procedure of deriving receptive fields. It is a critical factor in determining whether a neuron is driven to its full dynamic range or under-stimulated. In addition, as has been shown in a number of studies, different stimulus ensembles with different spectro-temporal characteristics appear to thrust neurons into different operating states; yielding STRFs that are *different* linearizations of the true behavior of a neuron around different points (Machens, Wehr, & Zador, 2004). **(ii)** What form of time-frequency representation is appropriate to use when dealing with cortical neurons? This issue has to be carefully considered; especially at the level of auditory cortex where numerous investigation have demonstrated a lack of phase locking to rapidly changing stimuli (>30Hz). Stimuli such as ripples and broadband noises are usually designed with that knowledge in mind. Even natural sounds, animal vocalizations and speech exhibit similar slow dynamics. Such phenomenon has often justified the use of smooth envelope profiles of the stimulus spectrogram (or other forms of time-frequency representations); rather than the spectrogram itself. In doing so, the emergent STRF captures *only* the slow spectro-temporal characteristics of the sound that usually modulates a richer structure composed of the signal carrier and its fine-structure interaction terms as the sound travels through the auditory pathway. Carefully considering the appropriate space for representing the signal can unravel more about processing by cortical neurons; particularly their seeming multiplexing ability at eliciting responses to the sound envelope, which itself gates phase-locked responses to the sound fine structure (Elhilali, Fritz, Klein, Simon, & Shamma, 2004). **(iii)** In most studies employing STRFs as neural measure, the linearity assumptions impose inclusion of only steady-state portions of the neural response. In doing so, one bypasses the use of the phasic behavior of neurons at onset; which is believed to drive the neurons around a different operating point, hence heightening the nonlinear behavior (Shechter, Dobbins, Marvit, & Depireux, 2009). What constitute *steady-space* response is not well understood; and this issue is further compounded

by processes of neural adaptation that changes the neuron's behavior as the sensory input is extended over time. **(iv)** All derivations of receptive fields (including the general Volterra series approach) have an underlying assumption that the system is time-invariant. Such assumption is challenged by compelling experimental evidence for receptive field plasticity, both during development (de Villers-Sidani, Chang, Bao, & Merzenich, 2007; Hubel & Wiesel, 1963) and the adult brain (for review, see (J. Fritz, Elhilali, & Shamma, 2005; Weinberger, 2007)) or induced by arousal, attention, and stimulus and behavioral context. **(v)** The STRF is ultimately an average measure; how much data is needed to derive a 'good' estimate of a neuron's STRF; depends on a number of ill-understood issues including the neuron's variability, degree of nonlinearity (or alternatively the plausibility of the linearity assumption for that particular neuron) as well as other sources of noise or variability (e.g. inherent system variability, experimental noise). **(vi)** It is important to keep in mind is that STRF derivations presented earlier, based on Volterra series, Wiener series or reverse correlation are approximating a functional that maps a deterministic input $s(t)$ onto a stochastic output: a sequence of spike trains. When applying the methods above to the nervous system, it is customary to average the output process to measure the underlying spiking rate of the neuron. Such approximation does not come without its perils. It causes great variations in the estimates of the STRF function. One has to carefully consider the long-term variations of the spiking rate and whether time-averaging is valid for the specific experimental settings investigated.

Interpretation of the spectro-temporal receptive field

One of the chief impacts of the STRF technique was generalizing the concept of a classic response field (RF) to the spectro-temporal domain (Depireux, Simon, Klein, & Shamma, 2001; Escabi & Schreiner, 2002; Klein, Depireux, Simon, & Shamma, 2000; Miller, Escabi, Read, & Schreiner, 2002; Nelken & Young, 1997; Theunissen, Sen, & Doupe, 2000; Yeshurun, Wollberg, Dyn, & Allon, 1985). It is greatly related to frequency tuning curves – a classic descriptor used to probe the spectral properties of a neuron; or indirectly to measure the matrix of synaptic weights to the neuron

from tonotopically-organized inputs. The correspondence between a frequency tuning curve and the spectral axis of a neuron's STRF is not exactly one-to-one. Each is derived using different sets of stimuli with different dynamics and is likely probing the system under different states. Nevertheless, there is great agreement between the two. The STRF undoubtedly overcomes a number of limitations that beset the classic tuning curve.

The STRF procedure operates under the assumption that the system under investigation integrates information linearly; but makes no assumptions about the relationship or independence of the spectral and temporal attributes of the response. As such, it is indeed more informative than the classic tuning curve. The x-axis represents time or delay between sensory stimulation and neural response. The y-axis represents the frequency tuning; ideally spanning the entire range of hearing of the animal. It is important to note that the STRF reflects the effect of the dynamic spectrum of sound on the sensory neuron. It is the conjunction of its time and frequency characteristics, rather than just the individual contributions of each of them. Because of its two-dimensional construction, it can be interpreted in two complementary ways: rows vs. columns. The former views the STRF as an array of frequency-dependent dynamical filters; while the latter interprets the STRF as time-dependent spectral weighting function.

The STRF reveals numerous characteristics about the underlying sensory neuron investigated, all in one function:

- Spectral characteristics (features shared to some extent with the classic tuning curve): its spectral tuning, its frequency-response bandwidth, presence of excitatory peaks or inhibitory sidebands,
- Temporal characteristics: the response latency, refractoriness, modulation rate tuning;
- Spectro-temporal characteristics: frequency-modulation direction selectivity (for FM sweeps moving upwards or downwards).

4. The Rosetta stone of STRFs

In this section, we consider the interpretation of the STRF form, and in particular

how its temporal and spectral dimensions might be interdependent, how one can infer from the STRF form its sensitivity to spectral and temporal features such as directional and spectral bandwidth sensitivity.

Rank and Separability

The *rank* of a two dimensional function, such as an STRF, captures one aspect of how simple the function is. When a two dimensional function is the simple product of two one- dimensional functions, i.e. $h(t, x) = f(t)g(x)$, this captures an important notion of simplicity. When this occurs, the function is of rank 1. When the sum of two products is required, e.g. $h(t, x) = f_A(t)g_A(x) + f_B(t)g_B(x)$, the function is of rank 2. (In cases of rank 2 and higher, we demand that each temporal function $f_i(t)$ be linearly independent of every other temporal function $f(t)$, and the same for the spectral functions g ; otherwise we could have used a smaller number of terms.) A rank 2 function is clearly not as simple as a rank 1 function, but nevertheless can be expressed rather concisely. In general, the *rank* of any two dimensional function is the minimum number of simple products needed to describe the function (When the functions are approximated as discrete, the definition of rank is identical to the definition of the algebraic rank of a matrix).

An STRF of rank 1, also called *fully separable*, can be written:

$$h^{FS}(t, x) = f(t)g(x) \tag{10}$$

which has this simple interpretation: the temporal processing of the STRF is performed independently of the spectral processing (and of course vice versa). A simple model of a neuron with this property is that its spectral processing is due purely to inputs from presynaptic neurons with a range of center frequencies, while the temporal processing is due to integration in the soma of all inputs arriving from all dendrites. For many peripheral neurons, this model is a good one. An STRF of rank 1 is also called fully separable because its processing separates cleanly into independent spectral and temporal processing stages. The example STRF in Figure 2a is fully separable, and this can be verified by noting that all spectral (vertical)

cross- sections have the same shape (the shape of the spectral function $g(x)$), differing only in amplitude (and possible sign). Similarly, all temporal (horizontal) cross-sections have the same shape (the shape of the temporal function $f(t)$), differing only in amplitude (and possibly sign).

[Figure 2 about here]

An STRF of rank 2 is somewhat less simple and is somewhat less straightforward to interpret.

$$h^{R2}(t, x) = f(t)g(x) + f(t)g(x) \quad (11)$$

One interpretation comes from noting that this $h^{R2}(t, x)$ can be written as the sum of two fully separable STRFs, $h^{FS}(t, x) = f(t)g(x)$ and $h^{FS}(t, x) = f(t)g(x)$. This implies a possible, but less than satisfying, interpretation: the neuron has exactly two neural inputs, each of which has a fully separable STRF, and then simply adds them. Below we will present more realistic interpretations, consistent with known physiology. An STRF described by a generic two-dimensional function would not necessarily have the same physiologically motivated interpretations or models.

An STRF of general rank N can be written as the sum of N terms:

$$h^{RN}(t, x) = f_A(t)g_A(x) + f_B(t)g_B(x) + \dots f_z(t)g_z(x) \quad (12)$$

As the rank of an STRF increases, more and more complexity is permitted. Figure 2b demonstrates a simulated STRF of high rank (though still well localized in time and spectrum). STRFs of this complexity are not seen in A1 (Klein, Simon, Depireux, & Shamma, 2006). In general, *higher rank* implies more STRF complexity. On the other hand, *lower rank* suggests there is a specific property (constraint) that causes this simplicity.

Singular Value Decomposition Analysis of the STRF

Singular Value Decomposition (SVD) is a method that can be applied to any finite dimensional matrix (e.g. a discretized version of the STRF) to establish both its rank

and a unique re-expression of the matrix as the sum of terms whose number is the rank of the matrix (Hansen, 1997; Press, Flannery, Teukolsky, & Vetterling, 1992). The SVD decomposition of a matrix M takes the form of the sum of N terms:

$$M_{ij} = \Lambda_A u_{Ai} v_{Aj}^T + \Lambda_B u_{Bi} v_{Bj}^T + \dots + \Lambda_Z u_{Zi} v_{Zj}^T \quad (13)$$

where N is the rank of the matrix, u and v are vectors normalized to have unit power, and each Λ is the term's RMS power. If we discretize the STRF into a finite number of frequencies and time steps, $x = \{x_i\} = \{x_1, x_2, \dots, x_M\}$ and $t = t_j = (t_1, \dots, t_N)$, so that $h(t_j, t_j) = \{h_{ij}\} = h(t_1, \dots, t_N; x_1, x_2, \dots, x_M)$, we see the N terms,

$$h_{ij} = \Lambda_A u_A(x_i) v_A(t_j) + \Lambda_B u_B(x_i) v_B(t_j) + \dots + \Lambda_Z u_Z(x_i) v_Z(t_j) \quad (14)$$

where N is the rank of the STRF, the u and v vectors are normalized to have unit power, and each Λ is the RMS power of its term. This is the same as Eq. 12, except that time and frequency have been discretized, and thus the STRF has been discretized also.

What makes SVD unique among decompositions is that 1) it automatically orders the terms by decreasing power: $\Lambda_A > \Lambda_B > \dots > \Lambda_Z$; 2) each column u_A, u_B, \dots, u_Z is orthogonal to all the others; 3) each row $v_A^T, v_B^T, \dots, v_Z^T$ is orthogonal to all the others. The mathematical specifics are described well in textbooks (see, e.g., (Press, Flannery, Teukolsky, & Vetterling, 1992) and will not be covered here. Mathematically, SVD is intimately related to Principle Component Analysis (PCA), and both are used for a variety of analytic purposes, including noise reduction (Hansen, 1997). Physiologically, however, (as with PCA) there is in general no universal physiological interpretation to specific components, since distinct physiological contributions to the singular value expansion are, in general, not constrained to be orthogonal to each other.

Since measured STRFs are made with noisy measurements (the noise arising from both neural variability and instrument noise), the true rank of the STRF must be estimated. There are a variety of methods to do this (Stewart, 1993) but they all use the same conceptual framework: once the power of the noise is estimated, then

all SVD components with power greater than the noise can be considered signal, and the number of components satisfying this criterion is the estimate of the rank. This estimate of rank is biased (more noise results in a lower rank estimate), but it has been shown that for range of Signal-to-Noise ratios and the STRFs used in this study, noise is not an impediment to measuring high rank (Klein, Simon, Depireux, & Shamma, 2006).

SVD also motivates us to recast Eq. 12 into its N terms in continuous form

$$h^{RN}(t,x) = \Lambda_A v_A(t) u_A(x) + \Lambda_B v_B(t) u_B(x) + \dots + \Lambda_Z v_Z(t) u_Z(x), \quad (15)$$

where the u and v functions have unit power and each Λ is the RMS power of its term. Compared to Eq. 12, it more complex, but it is less arbitrary: decompositions of the form of Eqs. 10, 11, and 12 are not unique since amplitude can be arbitrarily shifted between the temporal and spectral components. In Eq. 14, all amplitude information is explicitly shared within each term by each Λ_i coefficient. When the technique of SVD, which is designed for discrete matrices, is applied to continuous two-dimensional functions, as in the case of Eq. 14, it is called the Singular Value Expansion (Hansen, 1997). We will go back and forth between the continuous and discretized versions of the STRF without loss of generality (so long as N is finite), depending on which formalism is more beneficial.

There are a few important exceptions, however, to the statement above that there are typically no physiological interpretation to the specific components in the Singular Value Expansion. This occurs when one of the temporal (or spectral) component functions is the Hilbert Transform of another. In this case, as will be shown below, the two functions have identical spectral characteristics, and are orthogonal only because they differ in phase.

Hilbert Transform & Partial Hilbert Transforms/Rotations

We now discuss the Hilbert transform, a standard tool in signal processing, and necessary for the phenomenon of temporal symmetry. The Hilbert transform of a function produces the same function but with all its phase components shifted by

90°. This can be seen in the Fourier domain. For a function $f(t)$ with Fourier transform $F(\omega)$, i.e.

$$\begin{aligned} F(\omega) &= \mathcal{F}_\omega[f(t)] = \int dt f(t) e^{-j\omega t} \\ f(t) &= \mathcal{F}_t^{-1}[F(\omega)] = (2\pi)^{-1} \int d\omega F(\omega) e^{j\omega t}, \end{aligned} \quad (16)$$

where the Hilbert transform, designated by \mathcal{H} or $\hat{\cdot}$, is defined by

$$\hat{f}(t) = \mathcal{H}[f(t)] = \mathcal{F}_t^{-1}[\text{sgn}(\omega) e^{j\pi/2} F(\omega)], \quad (17)$$

where $e^{j\pi/2}$ is a rotation by 90° in the complex plane (the role of $\text{sgn}(\omega)$ guarantees that the Hilbert transform of a real function is itself a real function). This rotation of phase by 90° means that the Hilbert transform of any sine wave is a cosine wave, and the Hilbert transform of any cosine wave is the negative sine wave, but unlike differentiation, the amplitude is unchanged by the operation.

An important property of the Hilbert transform is that it is orthogonal to the original function, and yet it still has the same frequency content (aside from the DC component, i.e. its mean, which is zeroed out). $\hat{f}(t)$ is said to be “in quadrature” with $f(t)$; a demonstration is illustrated in Figure 3. For the remainder of this section, we assume that any function $f(t)$ which will be Hilbert transformed has mean zero (or has had its mean subtracted manually).

[Figure 3 about here.]

The double application of a Hilbert transform, since applying two successive 90° rotations is equivalent to one 180° rotation, is just a sign inversion.

$$\mathcal{H}[\mathcal{H}[f(t)]] = \hat{\hat{f}}(t) = -f(t) \quad (18)$$

It is also useful to define a partial Hilbert transform. As pointed out above, a Hilbert transform of a function can be viewed as a 90° rotation in a mixing angle plane, so one can define a partial version of the transform

$$f_{\theta}(t) = \sin\theta f^{\wedge}(t) + \cos\theta f(t). \quad (19)$$

In this convention, note that

$$\hat{f}(t) = f^{\pi/2}(t), \quad f(t) = f^0(t), \quad \text{and} \quad \hat{\hat{f}}(t) = f^{\pi}(t) = -f(t).$$

Thus a partial Hilbert transform still has the same frequency content as the original function, but its phase “rotation” is not restricted to 90° and can be any angle on the complex plane.

Physiological examples of the Hilbert transform have been demonstrated in the visual system, and have been named “lagged” cells (De Valois, Cottaris, Elfar, Mahon, & Wilson, 2000; Humphrey & Weller, 1988; Mastronarde, 1987a; Mastronarde, 1987b). These lagged cells are located in the Lateral Geniculate Nucleus (LGN), one of the visual thalamic nuclei. We will continue this nomenclature and call any neuron whose impulse response is the Hilbert transform of another the lagged version of the latter. We will further generalize, and call any neuron whose impulse response is the partial Hilbert transform (Hilbert rotation) of another, the “partially lagged” version of the latter. Note that the lag is a phase lag, not a time lag.

The full and partial Hilbert transform/rotation is not restricted to the time domain and is equally applicable to the spectral domain (Simon, Depireux, Klein, Fritz, & Shamma, 2007).

Temporal Symmetry

An important class of STRFs consists of those for which all temporal cross-sections (i.e. each cross section at a constant spectral index x_c), of the given STRF are related to each other by a simple scaling, g , and rotation, θ , of the same time function:

$$h(t, x_c) = g_{x_c} f^{\theta_{x_c}}(t), \quad (20)$$

where each scaling and rotation can depend on x . Since this is then true for all spectral indices x , we call the system Temporally Symmetric and write it in the functional form:

$$h^{TS}(t, x) = g(x)f^{\theta(x)}(t) \quad (21)$$

The meaning is still the same: all temporal cross-sections are related to each other by a simple scaling and rotation of the same time function. There is only one function of t , i.e. $f(t)$, and Hilbert rotations of it (demonstrated in Figure 4).

[Figure 4 about here.]

Using the definition of the Hilbert rotation Eq. 19, we can re-express Eq. 21 to explicitly show that a temporally symmetric STRF is rank 2 (i.e. is the sum of two linearly independent product terms):

$$\begin{aligned} h^{TS}(t, x) &= g(x)f^{\theta(x)}(t) \\ &= g(x)\cos\theta(x)f(t) + g(x)\sin\theta(x)\hat{f}(t) \\ &= f(t)g_A(x) + \hat{f}(t)g_B(x) \end{aligned} \quad (22)$$

where

$$\begin{aligned} g_A(x) &= g(x)\cos\theta(x) \\ g_B(x) &= g(x)\sin\theta(x) \\ \tan\theta(x) &= \frac{g_B(x)}{g_A(x)} \\ g^2(x) &= g_A^2(x) + g_B^2(x). \end{aligned}$$

We will often use the form of Eq. 22, which is completely equivalent to Eq. 21. In Eq. 22 it is explicit that a temporally symmetric STRF has rank 2 and cannot have higher rank.

Temporally symmetric STRFs, because of the symmetry relating the two temporal functions, fall into a very special subcategory of rank 2 STRFs. For these

STRFs, it turns out there *are* physiological interpretations for the individual components. The neural benefit of temporally symmetric STRFs is to allow modulation spectral selectivity without also being hampered by modulation phase selectivity (Simon, Depireux, Klein, Fritz, & Shamma, 2007). Not also that not all rank 2 STRFs are temporally symmetric: there are a wide variety of simple counter-examples (Simon, Depireux, Klein, Fritz, & Shamma, 2007).

For systems, which are not exactly temporally symmetric but are of rank 2, or for systems, which have been truncated by SVD to rank 2, we can define an index of temporal symmetry, η_t . This index ranges from 0 to 1, where $\eta_t = 1$ for the temporally symmetric case and $\eta_t = 0$ when the two time functions are temporally unrelated. First we put Eq. 22, which is explicitly rank 2, into the form of Eq. 14:

$$h^{R2}(t,x) = \lambda_A v_A(t) u_A(x) + \lambda_B v_B(t) u_B(x) \quad (23)$$

Since the u and v functions have unit power, we define the index of temporal symmetry to be the magnitude of the normalized complex inner product between the two temporal analytic signals (Cohen, 1995)

$$\eta_t = \left| \int \frac{1}{2} (v_A(t) + j\hat{v}_A(t))^* (v_B(t) + j\hat{v}_B(t)) dt \right|, \quad (24)$$

where $*$ is the complex conjugate operator. The rank 1 case, since it is automatically temporally symmetric, is also given the value $\eta_t = 1$. Temporal symmetry's cousin, spectral symmetry, is described in (Klein, Simon, Depireux, & Shamma, 2006).

Temporal Symmetry and Quadrant Separability

When the STRF is separable, the Spectro-Temporal Modulation Transfer Function (MTF_{ST}), the two-dimensional Fourier Transform of the STRF, is also separable, because the Fourier transform of the STRF is given by the simple products of the

Fourier transforms of $f(t)$ and $g(x)$. It was noticed that even when the STRF is not separable, the quadrants of the MTF_{ST} are still individually separable (Depireux, Simon, Klein, & Shamma, 2001; Klein, Simon, Depireux, & Shamma, 2006; Kowalski, Depireux, & Shamma, 1996), but neither the significance nor the origin of this property was well understood. It was demonstrated in (Simon, Depireux, Klein, Fritz, & Shamma, 2007) that when the STRF is of rank 2, quadrant separability and temporal symmetry are intimately related mathematical properties, manifesting differently as temporal symmetry in the STRF and quadrant separability in the MTF_{ST} . Mathematically, quadrant separability in the MTF_{ST} can manifest in two other ways in a rank 2 STRF: spectral symmetry (the spectral analog of temporal symmetry) or total directional selectivity. To our knowledge, spectral symmetry has not been observed physiologically. Total directional selectivity, in contrast, has been much discussed in the vision literature when STRF is interpreted as the visual Spatio-Temporal Response Field: both temporal and spectral functions are added in quadrature (Adelson & Bergen, 1985; Barlow & Levick, 1965; Borst & Egelhaaf, 1989; Chance, Nelson, & Abbott, 1998; De Valois, Cottaris, Elfar, Mahon, & Wilson, 2000; Emerson & Gerstein, 1977; Heeger, 1993; Maex & Orban, 1996; McLean & Palmer, 1994; Smith, Snowden, & Milne, 1994; Suarez, Koch, & Douglas, 1995; Watson & Ahumada, 1985). Mathematically, quadrant separable STRF of rank 4 are also possible, but we are not aware of their presence in any physiological system.

5. STRFs and cortical processing

The STRF concept has proven valuable in the understanding of auditory cortical processing and its applications in audio systems ranging from speech and speaker recognition, to enhancements of speech signals and cochlear implants. Their first valuable contribution has been to provide an effective summary of the representation of sound features in the primary auditory cortex (Atencio, Sharpee, & Schreiner, 2008; Theunissen, Sen, & Doupe, 2000; Versnel, Kowalski, & Shamma, 1995). A key finding has been the realization that the huge variety of auditory cortical responses in frequency tuning, dynamics, and other features can be

integrated in models of auditory processing that view the cortex as a ‘multiscale’ spectro-temporal analyzer that decomposes the signal into a rich representation that can be effectively harnessed for a variety of applications (Chi, Ru, & Shamma, 2005). For example, complex signals such as speech and music can be analyzed in this cortical model to understand how phonetic information is represented (Mesgarani, David, Fritz, & Shamma, 2008), or how timbres of different musical instruments are encoded (Patil, Pressnitzer, Shamma, & Elhilali, 2012). The uniqueness of the representation of the different types of signals allows us further to separate speech from interfering noise, to detect the presence of speech compared to non-speech (Mesgarani, Slaney, & Shamma, 2006), and to separate multiple speakers (Elhilali & Shamma, 2008).

STRFs have proven valuable in plasticity and development studies as a summary of a neuron’s response properties *even if incomplete*. The reason is that in many cases, it is not the completeness of the representation that is of interest, but rather *how* the (admittedly incomplete) representation changes during behavior or learning. In this role, the STRF has demonstrated a robustness and versatility that was not initially appreciated. For instance, auditory cortical STRFs often provide an estimate of the *relative* sensitivity of a neuron to different frequencies and temporal dynamics. Therefore, when examining the effects of attention and learning, changes in neuronal responsiveness are often unpredictable in their dynamics or spectral configurations, or are very small and augment already very noisy responses. The STRF counters both of these challenges. Thus, by virtue of its generalized nature (not being specific to a particular feature or response dimension), changes in the neuronal sensitivity in one of multiple dimensions can readily be spotted and characterized (J. Fritz, Elhilali, Klein, & Shamma, 2003). Furthermore, these response changes are often very difficult to discern or interpret (as spectral or temporal changes) in the firing rates themselves. STRFS by their nature are simultaneous *relative* measures of a neuron’s responsiveness to many frequencies and dynamics. Consequently, the effect of attention and learning (and behavior in general) is often seen as changes in the STRF shape that can be readily interpreted as changes in the relative sensitivity of the neuron to different frequencies (J. B. Fritz,

Elhilali, & Shamma, 2005), or in the temporal parameters of the STRF such as its latency (J. Fritz, Elhilali, & Shamma, 2005).

Another interesting use of the STRFs is in characterizing global response properties of a large population of cells or cortical areas for the purpose of summarizing its representational properties, and how that may change due to learning and plasticity. Specifically, in many situations, cortical neurons encode considerable details about the underlying sensory stimuli, and the encoded information is likely to change with stimulus context and behavioral conditions in a manner that is difficult to discern across large sets of single neuron data because of the complexity of cortical receptive fields. STRFs help overcome this problem because they facilitate methods of stimulus *reconstruction* to study how complex sounds are encoded in A1. In such a method, measured STRFs are used to *invert* the responses and hence map population responses to an estimate of the stimulus spectrogram (Mesgarani, David, Fritz, & Shamma, 2009). This enables one to perform a direct comparison between an original stimulus spectrogram and its reconstruction following learning or from responses during attention to particular cues in a behavioral paradigm. In fact, by estimating the fidelity of such reconstructions using generalized spectrally and temporally modulated noise stimuli, one can determine the range over which A1 neurons can faithfully encode such spectro-temporal features. Finally, contrasting stimulus reconstructions under different behavioral states can reveal a novel view of the rapid changes in spectro-temporal response properties induced by different attentional and motivational states.

There is much ongoing research to develop new STRF measurement procedures and structures, and to combine them with new methodologies that go beyond the simple methods employed earlier. For instance, generalizing the STRFs to include nonlinear phenomena (Atencio, Sharpee, & Schreiner, 2008; Christianson, Sahani, & Linden, 2008; Nagel & Doupe, 2008), inhibitory interactions (Schinkel-Bielefeld, David, Shamma, & Butts, 2012), and more flexible structures are among the goals being pursued. Also, STRFs are now being regularly estimated with a variety of specialized complex sounds such as bird vocalizations and speech (David, Mesgarani,

& Shamma, 2007; Theunissen, Sen, & Doupe, 2000), measurements that yield STRFs that are tailored to describe more faithfully cortical responses to these sounds. Furthermore, STRFs have now transcended their initial use in single unit recordings and are currently being widely used as a tool to summarize responses in human subjects recorded with fMRI (Zatorre & Belin, 2001), MEG (Ding & Simon, 2012), EEG (Power, Reilly, & Lalor, 2011), and ECoG (Mesgarani & Chang, 2012).

6. Conclusion

As the divergent impressions of the seven blind men touching the elephant suggests, there are many ways to characterize the receptive field properties of auditory neurons, ranging from classical frequency response areas with responses to tone bursts covering a wide range of frequencies and amplitudes, to click rate tuning for temporal MTF, to pure amplitude tuning at CF, to STRFs derived from artificial stimuli (white noise, dynamic ripples, TORCs) or natural sounds.

As this chapter shows, the STRF is a useful and powerful measure to characterize steady state response properties of auditory neurons, combining a broad description of both temporal dynamics and spectral selectivity. However, although the STRF provides an excellent linear characterization of complex stimulus-response transformations observed in neurons throughout the auditory pathway, there is considerable range in the degree to which neural responses can be described or approximated by a linear model.

The STRF description is a very appealing approach to taming the cortical zoo of diverse neuronal receptive fields. The richness and variety of STRFs makes it possible to encode arbitrary sounds and responses can be used for stimulus reconstruction (Mesgarani & Chang, 2012; Mesgarani, David, Fritz, & Shamma, 2009). Even within A1 there is considerable diversity between modules of narrow and broadly tuned neurons, and also laminar diversity - neurons in granular layers are “more linear” i.e. they are better describe by a model with one relevant stimulus feature (Sharpee, Atencio, & Schreiner, 2011). Input neurons in layer IV have stimulus feature preferences that are more separable in frequency/time space than those in supra or infra-granular levels. In

contrast, output layers in A1 show STRFs with more complex shapes, such as elongated subfields or repeating bands of excitation or suppression (Atencio & Schreiner, 2010).

But, not all neurons in A1 have sufficient linearity in their spiking response to inputs that they can be well described by STRFs – nevertheless, these highly non-linear or “Jackson Pollock” cells may play a very important role in auditory processing. The trade-off of the STRF approach is that you lose the ability to characterize a fairly large percentage of auditory responsive cells in A1. What percentage of cells in A1 have STRFs? Estimates of the percentage varies from ~30% to 70%, and this percentage goes down considerably as you move to higher cortical areas. However, there are still beautiful STRFs even several synapses higher up – as shown in the example neuron recorded from macaque monkey rostral STG (Figure 5) (see (Hackett, 2011) for neuroanatomical location of rSTG).

[Figure 5 about here]

Concern about the lack of general applicability to all cortical cells combines with wariness of the STRF that arises from its limited predictive power (Laudanski, Edeline, & Huetz, 2012), and also because of the presence of nonlinearities (spiking threshold, rate saturation, rectification, synaptic depression, modulation of responses by stimulus context, attentional effects and other state-dependent effects) that cast doubt on the linear assumptions underlying the STRF. Various approaches have been taken to go beyond the STRF by new analysis techniques that add input or output nonlinearities, simplified second order interaction terms, feedback kernels, spectro-temporal contrast kernels or multiple feature dimensions (Ahrens, Linden, & Sahani, 2008; Atencio, Sharpee, & Schreiner, 2008; Calabrese, Schumacher, Schneider, Paninski, & Wooley, 2011; Rabinowitz, Willmore, Schnupp, & King, 2012). These STRF-nonlinearity models represent exciting new approaches to describing neural function in A1, and generally also have greater predictive success.

The shape of STRFs depends on the choice of stimulus and properties of the stimulus ensemble (Aertsen & Johannesma, 1981; Blake & Merzenich, 2002; David, Mesgarani, Fritz, & Shamma, 2009; Laudanski, Edeline, & Huetz, 2012; Machens, Wehr, & Zador, 2004; Theunissen, Sen, & Doupe, 2000; Valentine & Eggermont, 2004; Woolley, Gill, & Theunissen, 2006), and STRFs generated from the prevailing ripple stimuli generally lack predictive power for communication calls and other natural sounds. This realization has led to the use of different stimulus classes (conspecific vocalizations, such as birdcalls or speech) to generate STRFs that have enhanced predictive power for the sounds in these classes (David, Mesgarani, & Shamma, 2007; Laudanski, Edeline, & Huetz, 2012; Nagel & Doupe, 2008).

One valuable use of the STRF has been to explore the effects of selective attention on cortical filters (Atiani, Elhilali, David, Fritz, & Shamma, 2009; David, Fritz, & Shamma, 2012; J. Fritz, Elhilali, Klein, & Shamma, 2003; J. B. Fritz, Elhilali, & Shamma, 2005; J. B. Fritz, Elhilali, & Shamma, 2007). In addition to its effect on gain control, attention also appears to be able to independently re-shape receptive fields by rapidly changing synaptic input weightings to enhance processing of relevant stimulus features. These dynamic task-driven changes are captured and observed in ~60% of STRFs and have led to the insight that this reshaping can be described as the convolution of the STRF with a contrast filter emphasizing foreground over background stimuli (J. B. Fritz, Elhilali, & Shamma, 2007; Mesgarani, Fritz, & Shamma, 2010).

These results in the awake, behaving animal suggest that the STRF is highly adaptive and reflects top-down influences such as attention, as well as bottom-up influences such as statistics of the ongoing stimulus ensemble or soundscape. So, to make it more complicated, and interesting, the elephant in the story of the seven blind men is a dynamic elephant that changes shape in response to cognitive and stimulus context.

References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, Optics & Image Science*, 2, 284-299.
- Aertsen, A. M. H. J., & Johannesma, P. I. M. (1981). The spectro-temporal receptive field. *Biol.Cybernetics*, 42, 133-143.
- Ahrens, M. B., Linden, J. F., & Sahani, M. (2008). Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *The Journal of Neuroscience*, 28(8), 1929-1942. doi: 10.1523/JNEUROSCI.3377-07.2008
- Atencio, C. A., & Schreiner, C. E. (2010). Laminar diversity of dynamic sound processing in cat primary auditory cortex. *103*(1), 192-205.
- Atencio, C. A., Sharpee, T. O., & Schreiner, C. E. (2008). Cooperative nonlinearities in auditory cortical neurons. *Neuron*, 58(6), 956-966. doi: 10.1016/j.neuron.2008.04.026
- Atiani, S., Elhilali, M., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron*, 61(3), 467-480.
- Barlow, H. B., & Levick, W. R. (1965). The mechanism of directionally selective units in rabbit's retina. *The Journal of Physiology*, 178(3), 477-504.
- Bender, E. (2000). *An introduction to mathematical modeling*. New York: Dover.
- Blake, D. T., & Merzenich, M. M. (2002). Changes of AI receptive fields with sound density. *Journal of Neurophysiology*, 88(6), 3409-3420. doi: 10.1152/jn.00233.2002
- Borst, A., & Egelhaaf, M. (1989). Principles of visual motion detection. *Trends in Neurosciences*, 12(8), 297-306.
- Bracewell, R. (1999). *The fourier transform and its applications*, McGraw-Hill.
- Brilliant, M. (1958). *Theory of the analysis of nonlinear systems*. (Technical report No. 345). Research Laboratory of Electronics, Massachusetts Institute of Technology.
- Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L., & Wooley, S. M. N. (2011). A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS ONE*, 6(1), e16104.

- Chance, F. S., Nelson, S. B., & Abbott, L. F. (1998). Synaptic depression and the temporal response characteristics of V1 cells. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 18(12), 4785-4799.
- Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *118*(2), 887-906.
- Christianson, G. B., Sahani, M., & Linden, J. F. (2008). The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *28*(2), 446-455.
- Cohen, L. (1995). *Time-frequency signal analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- David, S. V., Fritz, J. B., & Shamma, S. A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 109(6), 2144-2149. doi: 10.1073/pnas.1117717109; 10.1073/pnas.1117717109
- David, S. V., Mesgarani, N., Fritz, J. B., & Shamma, S. A. (2009). Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. *29*(11), 3374-3386.
- David, S. V., Hayden, B. Y., & Gallant, J. L. (2006). Spectral receptive field properties explain shape selectivity in area V4. *Journal of Neurophysiology*, 96(6), 3492-3505. doi: 10.1152/jn.00575.2006
- David, S. V., Mesgarani, N., & Shamma, S. A. (2007). Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network: Computation in Neural Systems*, 18(3), 191-212. doi: 10.1080/09548980701609235
- De Boer, E. (1967). Correlation studies applied to the frequency resolution of the cochlea. *Journal of Auditory Research*, 7(2), 209-217.
- de Boer, E., & de Jongh, H. R. (1978). On cochlear encoding: Potentialities and limitations of the reverse-correlation technique. *The Journal of the Acoustical Society of America*, 63(1), 115-135.
- De Boer, E., & Kuyper, P. (1968). Triggered correlation. *Biomedical Engineering, IEEE Transactions On*, BME-15(3), 169-179.
- De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York, Oxford: University Press.
- De Valois, R. L., Cottaris, N. P., Elfar, S. D., Mahon, L. E., & Wilson, J. A. (2000). Some transformations of color information from lateral geniculate nucleus to striate

cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 97(9), 4997-5002.

- de Villers-Sidani, E., Chang, E. F., Bao, S., & Merzenich, M. M. (2007). Critical period window for spectral tuning defined in the primary auditory cortex (A1) in the rat. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 27(1), 180-189. doi: 10.1523/JNEUROSCI.3227-06.2007
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neurosciences*, 18(10), 451-458.
- Depireux, D. A., Simon, J. Z., Klein, D. J., & Shamma, S. A. (2001). Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *Journal of Neuroscience*, 21(3), 1220-1234.
- Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America*, 109(29), 11854-11859. doi: 10.1073/pnas.1205381109
- Eggermont, J. J. (1993). Wiener and volterra analyses applied to the auditory system. *Hearing Research*, 66(2), 177-201.
- Eggermont, J. J., Aertsen, A. M., & Johannesma, P. I. (1983). Quantitative characterisation procedure for auditory neurons based on the spectro-temporal receptive field. *Hearing Research*, 10(2), 167-190.
- Eggermont, J. J., & Smith, G. M. (1990). Characterizing auditory neurons using the wigner and rihacek distributions: A comparison. *The Journal of the Acoustical Society of America*, 87(1), 246-259.
- Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z., & Shamma, S. A. (2004). Dynamics of precise spike timing in primary auditory cortex. *Journal of Neuroscience*, 24(5), 1159-1172.
- Elhilali, M., & Shamma, S. A. (2008). A cocktail party with a cortical twist: How cortical mechanisms contribute to sound segregation. *Journal of the Acoustical Society of America*, 124(6), 3751-3771.
- Emerson, R. C., & Gerstein, G. L. (1977). Simple striate neurons in the cat. II. mechanisms underlying directional asymmetry and directional selectivity. *Journal of Neurophysiology*, 40(1), 136-155.
- Escabi, M. A., & Read, H. L. (2003). Representation of spectrotemporal sound information in the ascending auditory pathway. *Journal of Neuroscience*, 23(5), 350-362.

- Escabi, M. A., & Schreiner, C. E. (2002). Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *22*(10), 4114-4131.
- Fritz, J., Elhilali, M., Klein, D. J., & Shamma, S. A. (2003). Dynamic adaptive plasticity of spectro-temporal receptive fields in the primary auditory cortex of the behaving ferret [Abstract]. *Society for Neuroscience Meeting, New Orleans, LA*,
- Fritz, J., Elhilali, M., & Shamma, S. (2005). Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hearing Research, 206*(1-2), 159-176.
- Fritz, J. B., Elhilali, M., & Shamma, S. A. (2005). Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *Journal of Neuroscience, 25*(33), 7623-7635.
- Fritz, J. B., Elhilali, M., & Shamma, S. A. (2007). Adaptive changes in cortical receptive fields induced by attention to complex sounds. *Journal of Neurophysiology, 98*(4), 2337-2346.
- George, D. (1959). *Continuous nonlinear systems*. (Technical report No. 355). Research Laboratory of Electronics, Massachusetts Institute of technology.
- Goldstein, J. L. (1967). Auditory nonlinearity. *The Journal of the Acoustical Society of America, 41*(3), 676-689.
- Gosselin, F., & Schyns, P. (2002). A new framework for visual categorization. *Trends Cog. Neurosci, (6)*, 70-76.
- Grochenig, K. (2001). *Foundations of time-frequency analysis*. Boston, MA: Birkhauser.
- Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hearing Research, 271*(1-2), 133-146. doi: 10.1016/j.heares.2010.01.011; 10.1016/j.heares.2010.01.011
- Hansen, P. C. (1997). *Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion*. Philadelphia: SIAM.
- Heeger, D. J. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology, 70*(5), 1885-1898.
- Hochstein, S., & Shapley, R. M. (1976). Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *The Journal of Physiology, 262*(2), 265-284.
- Hubel, D. H., & Wiesel, T. N. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology, 26*, 994-1002.

- Humphrey, A. L., & Weller, R. E. (1988). Functionally distinct groups of X-cells in the lateral geniculate nucleus of the cat. *The Journal of Comparative Neurology*, 268(3), 429-447.
- Isermann, R., & Munchhof, M. (2011). *Identification of dynamic systems: An introduction with applications* Springer.
- Johnson, D. H. (1980). Applicability of white-noise nonlinear system analysis to the peripheral auditory system. *The Journal of the Acoustical Society of America*, 68(3), 876-884.
- Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, 58(6), 1233-1258.
- Kim, P. J., & Young, E. D. (1994). Comparative analysis of spectro-temporal receptive fields, reverse correlation functions, and frequency tuning curves of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, 95(1), 410-422.
- Klein, D. J., Depireux, D. A., Simon, J. Z., & Shamma, S. A. (2000). Robust spectrotemporal reverse correlation for the auditory system: Optimizing stimulus design. *J. Neurosci.*, 20(1), 85-111.
- Klein, D. J., Simon, J. Z., Depireux, D. A., & Shamma, S. A. (2006). Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex. *Journal of Computational Neuroscience*, 20(2), 111-136.
- Kowalski, N., Depireux, D. A., & Shamma, S. A. (1996). Analysis of dynamic spectra in ferret primary auditory cortex. I. characteristics of single-unit responses to moving ripple spectra. *J. Neurosci.*, 16(15), 3503-3523.
- Laudanski, J., Edeline, J. M., & Huetz, C. (2012). Differences between spectro-temporal receptive fields derived from artificial and natural stimuli in the auditory cortex. *PloS One*, 7(11), e50539. doi: 10.1371/journal.pone.0050539; 10.1371/journal.pone.0050539
- Machens, C. K., Wehr, M. S., & Zador, A. M. (2004). Linearity of cortical receptive fields measured with natural sounds. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 24(5), 1089-1100. doi: 10.1523/JNEUROSCI.4445-03.2004
- Maex, R., & Orban, G. A. (1996). Model circuit of spiking neurons generating directional selectivity in simple cells. *Journal of Neurophysiology*, 75(4), 1515-1545.

- Marmarelis, P., & Marmarelis, V. (1978). *Analysis of physiological systems: The white noise approach*. New York: Plenum.
- Mastrorarde, D. N. (1987a). Two classes of single-input X-cells in cat lateral geniculate nucleus. I. receptive-field properties and classification of cells. *Journal of Neurophysiology*, *57*(2), 357-380.
- Mastrorarde, D. N. (1987b). Two classes of single-input X-cells in cat lateral geniculate nucleus. II. retinal inputs and the generation of receptive-field properties. *Journal of Neurophysiology*, *57*(2), 381-413.
- Mathews, V., & Sicuranza, G. (2000). *Polynomial signal processing*. New York: Wiley.
- McFee, R. (1961). Determining the response of nonlinear systems to arbitrary inputs. *American Institute of Electrical Engineers, Part II: Applications and Industry, Transactions of The*, *80*(4), 189-193.
- McLean, J., & Palmer, L. A. (1994). Organization of simple cell responses in the three-dimensional (3-D) frequency domain. *Visual Neuroscience*, *11*(2), 295-306.
- Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*(7397), 233-236. doi: 10.1038/nature11020
- Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2008). Phoneme representation and classification in primary auditory cortex. *The Journal of the Acoustical Society of America*, *123*(2), 899-909. doi: 10.1121/1.2816572
- Mesgarani, N., Fritz, J., & Shamma, S. (2010). A computational model of rapid task-related plasticity of auditory cortical receptive fields. *Journal of Computational Neuroscience*, *28*(1), 19-27. doi: 10.1007/s10827-009-0181-3; 10.1007/s10827-009-0181-3
- Mesgarani, N., Slaney, M., & Shamma, S. A. (2006). Content-based audio classification based on multiscale spectro-temporal features. *14*(3), 920-930.
- Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *J.Neurosci.*, *102*(6), 3329-3333.
- Miller, L. M., Escabi, M. A., Read, H. L., & Schreiner, C. E. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *87*(1), 516-527.
- Nagel, K. I., & Doupe, A. J. (2008). Organizing principles of spectro-temporal encoding in the avian primary auditory area field L. *Neuron*, *58*(6), 938-955. doi: 10.1016/j.neuron.2008.04.028

- Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *14*(4), 474-480.
- Nelken, I., & Young, E. D. (1997). Linear and nonlinear spectral integration in type IV neurons of the dorsal cochlear nucleus. I. regions of linear interaction. *78*(2), 790-799.
- Ogunfunmi, T. (2007). *Adaptive nonlinear system identification: The volterra and wiener model approaches (signals and communication technology)* Springer.
- Patil, K., Pressnitzer, D., Shamma, S., & Elhilali, M. (2012). Music in our ears: The biological bases of musical timbre perception. *PLoS Computational Biology*, *8*(11), e1002759. doi: 10.1371/journal.pcbi.1002759; 10.1371/journal.pcbi.1002759
- Pickles, J. O. (2008). *An introduction to the physiology of hearing* (Third ed.) Emerald Group Publishing Limited.
- Power, A. J., Reilly, R. B., & Lalor, E. C. (2011). Comparing linear and quadratic models of the human auditory system using EEG. *Conference Proceedings: ...Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2011*, 4171-4174. doi: 10.1109/IEMBS.2011.6091035
- Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1992). *Numerical recipes: The art of scientific computing* Cambridge University Press.
- Rabinowitz, N. C., Willmore, B. D., Schnupp, J. W., & King, A. J. (2012). Spectrotemporal contrast kernels for neurons in primary auditory cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(33), 11271-11284. doi: 10.1523/JNEUROSCI.1715-12.2012; 10.1523/JNEUROSCI.1715-12.2012
- Ringach, D., Sapiro, G., & Shapley, R. (1997). A subspace reverse-correlation technique for the study of visual neurons. *Vision Research*, *37*(17), 2455-2464.
- Rugh, W. (1981). *Nonlinear system theory: The volterra/wiener approach*. Baltimore, MD: Johns Hopkins Univ. Press.
- Sandberg, I. (1983). Series expansions for nonlinear systems. *Circuits, Systems and Signal Processing*, *2*(1), 77-87.
- Sang-Won Nam, & Powers, E. J. (2003). Volterra series representation of time-frequency distributions. *Signal Processing, IEEE Transactions On*, *51*(6), 1532-1537.

- Sastry, S. (1999). *Nonlinear systems* Springer.
- Schinkel-Bielefeld, N., David, S. V., Shamma, S. A., & Butts, D. A. (2012). Inferring the role of inhibition in auditory processing of complex natural stimuli. *Journal of Neurophysiology*, *107*(12), 3296-3307. doi: 10.1152/jn.01173.2011
- Sharpee, T. O., Atencio, C. A., & Schreiner, C. E. (2011). Hierarchical representations in the auditory cortex. *J. Neurosci.* *31*(5), 761-767.
- Shechter, B., Dobbins, H. D., Marvit, P., & Depireux, D. A. (2009). Dynamics of spectro-temporal tuning in primary auditory cortex of the awake ferret. *Hearing Research*, *256*(1-2), 118-130. doi: 10.1016/j.heares.2009.07.005; 10.1016/j.heares.2009.07.005
- Simon, J. Z., Depireux, D. A., Klein, D. J., Fritz, J. B., & Shamma, S. A. (2007). Temporal symmetry in primary auditory cortex: Implications for cortical connectivity. *J. Neurosci.* *27*(3), 583-638.
- Smith, A. T., Snowden, R. J., & Milne, A. B. (1994). Is global motion really based on spatial integration of local motion signals? *Vision Research*, *34*(18), 2425-2430. doi: 10.1016/0042-6989(94)90286-0
- Sripati, A. P., Yoshioka, T., Denchev, P., Hsiao, S. S., & Johnson, K. O. (2006). Spatiotemporal receptive fields of peripheral afferents and cortical area 3b and 1 neurons in the primate somatosensory system. *The Journal of Neuroscience*, *26*(7), 2101-2114. doi: 10.1523/JNEUROSCI.3720-05.2006
- Stewart, G. W. (1993). Determining rank in the presence of error. In M. S. Moonen, G. H. Golub & B. L. Moor (Eds.), *Linear algebra for large scale and real-time applications* (). Dordrecht: Kluwer Academic Publishers.
- Suarez, H., Koch, C., & Douglas, R. (1995). Modeling direction selectivity of simple cells in striate visual cortex within the framework of the canonical microcircuit. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *15*(10), 6700-6719.
- Theunissen, F. E., Sen, K., & Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.* *20*(6), 2315-2331.
- Valentine, P. A., & Eggermont, J. J. (2004). Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex. *Hearing Research*, *196*(1-2), 119-133. doi: 10.1016/j.heares.2004.05.011
- Versnel, H., Kowalski, N., & Shamma, S. A. (1995). Ripple analysis in ferret primary auditory cortex. III. topographic distribution of ripple response parameters. *J. Aud. Neurosc.*, *1*, 271-286.

- Volterra, V. (1959). *Theory of functionals and integral and integro-differential equations*. New York: Dover Publications.
- Watson, A. B., & Ahumada, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, Optics & Image Science*, 2, 322-341.
- Weinberger, N. M. (2007). Auditory associative memory and representational plasticity in the primary auditory cortex. *Hearing Research*, 229(1-2), 54-68. doi: 10.1016/j.heares.2007.01.004
- Wickesberg, R. E., Dickson, J. W., Gibson, M. M., & Geisler, C. D. (1984). Wiener kernel analysis of responses from anteroventral cochlear nucleus neurons. *Hearing Research*, 14(2), 155-174.
- Wiener, N. (1958). *Nonlinear problems in random theory*. New York: Technology Press and Wiley.
- Woolley, S. M., Gill, P. R., & Theunissen, F. E. (2006). Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 26(9), 2499-2512. doi: 10.1523/JNEUROSCI.3731-05.2006
- Yeshurun, Y., Wollberg, Z., Dyn, N., & Allon, N. (1985). Identification of MGB cells by volterra kernels. I. prediction of responses to species specific vocalizations. *Biological Cybernetics*, 51(6), 383-390.
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *11*(10), 946-953.

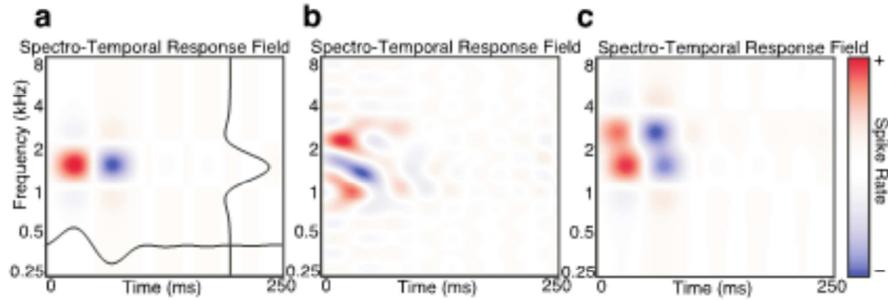


Figure 2: **a.** A simulated fully separable STRF, with spectral and temporal one-dimensional cross-sections as inserts (all horizontal slices have the same profile, as do vertical slices). **b.** A simulated high rank STRF with no particular symmetries. **c.** A simulated temporally symmetric and quadrant separable STRF of rank 2. This STRF's symmetry is not obviously visible. This STRF was created by adding to the STRF in **a** the same STRF except with the spatial cross-section shifted upward by $3/4$ octave and the temporal cross-section Hilbert-rotated (see below) by 30° .

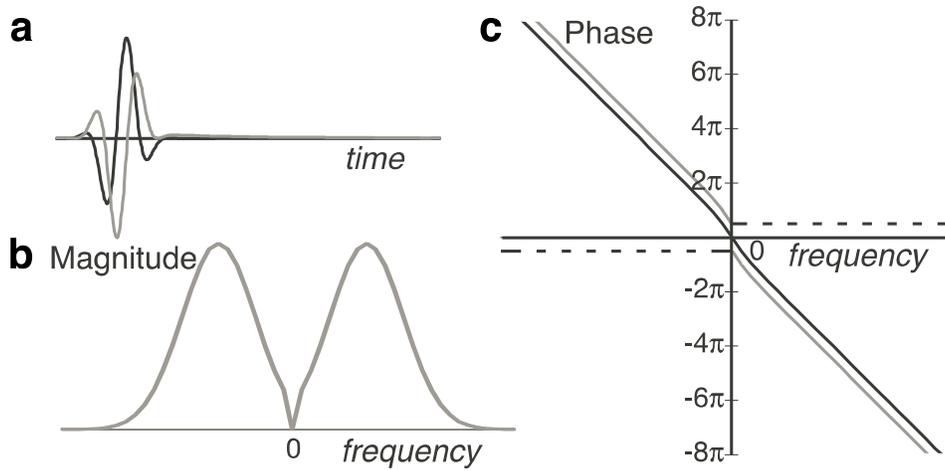


Figure 3: An example of a function and its Hilbert transform. **a.** A function (in black) overlaid with its Hilbert transform (in gray); the two are orthogonal. **b.** The magnitude of the Fourier transform of the function (in black) overlaid with the magnitude of the Fourier transform of its Hilbert transform (in gray); they overlap exactly. **c.** The phase of the Fourier transform of the function (in black) overlaid with the phase of the Fourier transform of its Hilbert transform (in gray); the difference is exactly $\pm 90^\circ$ (dashed line).

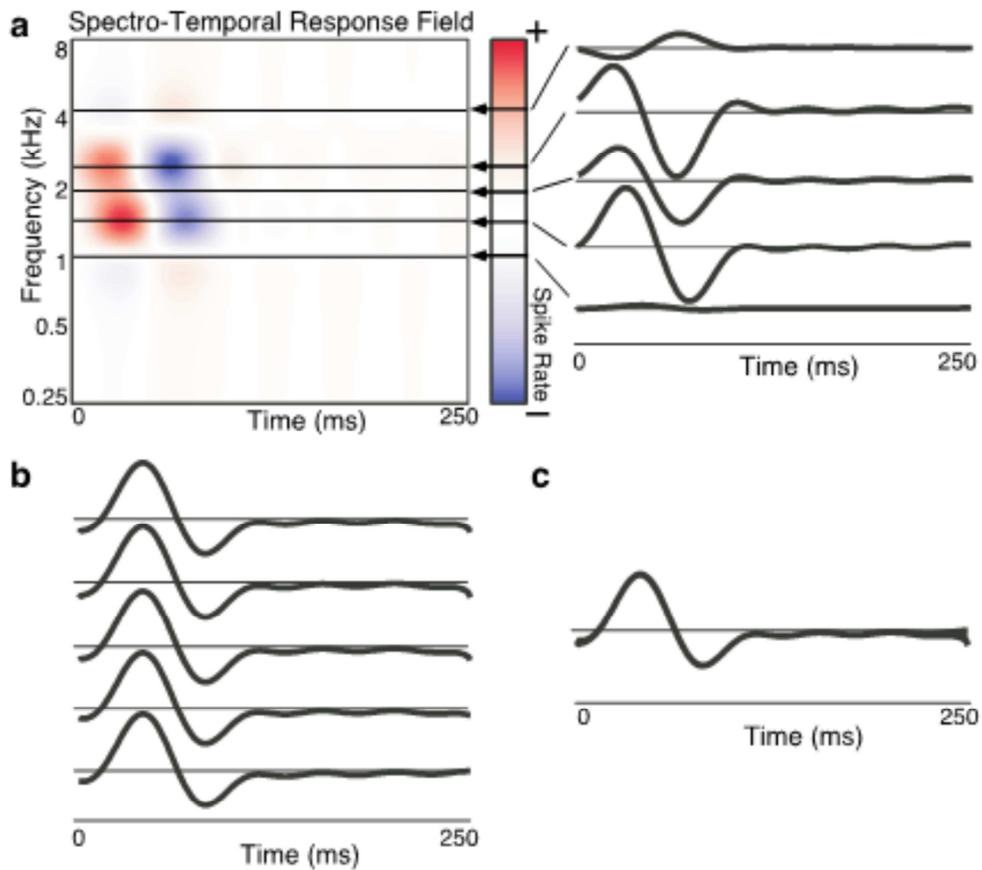
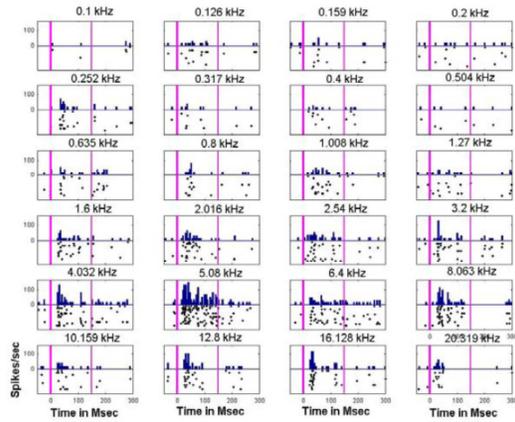


Figure 4: **a.** The simulated temporally symmetric and quadrant separable STRF from Figure 1c, and five fixed-frequency cross-sections, corresponding to five temporal impulse responses. **b.** The same five impulse responses but individually Hilbert-rotated and rescaled. **c.** The same Hilbert-rotated and rescaled impulse responses superimposed. The Hilbert-rotation phases were calculated by taking the negative phase of the complex correlation coefficient between the analytic signal of each temporal cross-section and the analytic signal of the 4th temporal cross-section.

Response to PTs



STRF from TORCs

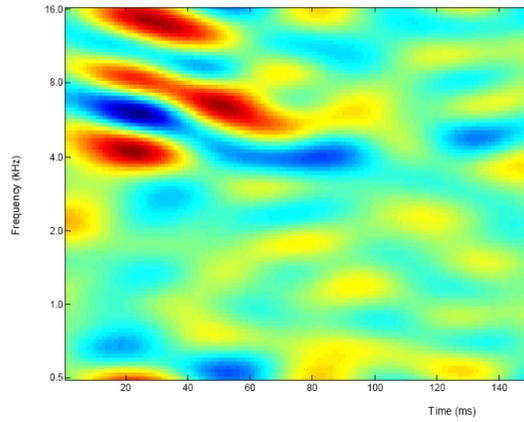


Figure 5: Example of recording in the rostral STG of the monkey, showing an STRF at least 2-3 synapses beyond A1. Most neurons in rSTG (34/53 or 64%) had receptive fields made up of several narrow frequency bands.