Review Article

Incorporating behavioral and sensory context into spectro-temporal models of auditory encoding

Stephen V. David

Oregon Hearing Research Center, Oregon Health & Science University, 3181 SW Sam Jackson Park Rd, MC L335A, Portland, OR 97239, United States

ARTICLE INFO

ABSTRACT

For several decades, auditory neuroscientists have used spectro-temporal encoding models to understand how neurons in the auditory system represent sound. Derived from early applications of systems identification tools to the auditory periphery, the spectro-temporal receptive field (STRF) and more sophisticated variants have emerged as an efficient means of characterizing representation throughout the auditory system. Most of these encoding models describe neurons as static sensory filters. However, auditory neural coding is not static. Sensory context, reflecting the acoustic environment, and behavioral context, reflecting the internal state of the listener, can both influence sound-evoked activity, particularly in central auditory areas. This review explores recent efforts to integrate context into spectro-temporal encoding models. It begins with a brief tutorial on the basics of estimating and interpreting STRFs. Then it describes three recent studies that have characterized contextual effects on STRFs, emerging over a range of timescales, from many minutes to tens of milliseconds. An important theme of this work is not simply that context influences auditory coding, but also that contextual effects span a large continuum of internal states. The added complexity of these context-dependent models introduces new experimental and theoretical challenges that must be addressed in order to be used effectively. Several new methodological advances promise to address these limitations and allow the development of more comprehensive context-dependent models in the future.

Contents

1. Introduction

The spectro-temporal receptive field (STRF) has proven to be valuable tool for understanding how information about sound is represented and transformed as it passes through the network of auditory areas from brainstem to cortex (Aertsen and Johannesma, 1981; De Boer and Kuyper, 1968; deCharms et al., 1998; Kowalski et al., 1996). The STRF describes neural function as a filter, in that the response to any arbitrary stimulus at a point in time can be predicted as a weighted sum of the stimulus spectrogram in the immediately preceding time window. Stimuli matched to the STRF will evoke large responses, and less-well matched stimuli will produce weaker or no response. Each neuron is characterized by a different STRF, and the population of neurons constituting a brain area provides a bank of filters, each reporting the occurrence of a distinct sound feature. The model of auditory cortex as a spectro-temporal filterbank remains a dominant paradigm for central auditory representation (Chi et al., 2005; Singh and Theunissen, 2003; Yang et al., 1992). This filterbank model has inspired and continues to inspire algorithms for sound processing and signal processing more generally (Hermansky, 1998; Mesgarani and Shamma, 2005).

While sensory coding models have provided valuable insight into how the auditory system extracts useful information from sound, most models do not account for changes in internal behavioral state. Instead, they describe auditory responses exclusively as a function of the incoming stimulus. It has long been known that extensive anatomical projections from central cortical and neuromodulatory centers are situated to provide top-down control of processing in ascending auditory areas. Moreover, numerous studies have shown that changes in behavioral state (task engagement, selective attention, arousal, e.g., Fritz et al., 2003; Kuchibhotla et al., 2016; McGinley et al., 2015; Rodgers and DeWeese, 2014) and, more broadly, the behavioral context (including relatively slow changes in the acoustic environment, e.g., Dean et al., 2005; Rabinowitz et al., 2011; Ulanovsky et al., 2003) can influence sound-evoked activity. A new challenge facing the field of auditory research is to develop encoding models that integrate the influence of sensory and behavioral context. If ignored, these changes in response properties will simply appear to be noise in the auditory response. Conversely, a model that can explain these context-related effects will provide new insight into the computational strategy and neural circuitry by which top-down feedback controls auditory processing.

For the current review, the term "context" spans a wide range of timescales, falling roughly into two categories. *Sensory context* effects reflect relatively rapid adaptation to statistics of the acoustic environment, including regularities (Ulanovsky et al., 2004) and the dynamic range of noise (Dean et al., 2005; Mesgarani et al., 2014; Rabinowitz et al., 2012). *Behavioral context* effects reflect slower changes following engagement in a behavioral task (Fritz et al., 2003; Mesgarani and Chang, 2012), learning of new representations (Ohl et al., 2001; Polley et al., 2006), and, even on the developmental timescale, following peripheral hearing loss (Buran et al., 2014; Chambers et al., 2016; Noreña et al., 2003). While

sensory and behavioral context clearly reflect different neurophysiological processes, ranging from automatic adaptation to the complex goal-directed behavior, they both have the net effect of changing the way neurons encode sound. Thus, for the purposes of this review, these effects can be viewed as similar modulatory processes that occur over a large continuum of timescales.

The idea of integrating contextual variables into encoding models, while appealing, introduces substantial combinatorial complexity to the problem. Measuring the response to many stimuli across many contexts drastically increases the amount of data and experimental control required to accurately estimate a complete set of model parameters. Thus, while context is important, there are practical experimental controls and model architecture designs that make studying this problem tractable.

This review begins with an overview of context effects known to influence activity in the auditory system. It then provides a tutorial on basic methods for computing the linear STRF and a brief survey of nonlinear models that build on the linear STRF. Next, it presents several studies that illustrate the full range of contextual factors that can be incorporated into encoding models. Finally, it discusses the very real technical and conceptual challenges posed by context-dependent models and new analytical and experimental approaches that promise to address these problems in the future.

A Python software library for fitting and comparing performance of context-dependent encoding models is available online: https://bitbucket.org/lbhb/nems/.

2. Sources of contextual effects in auditory processing

Exploration of context-dependent auditory encoding models has begun only relatively recently, but numerous processes are known to modulate sound-evoked activity in auditory brain areas, particularly in auditory cortex. Classically, behavioral studies emphasize discrete changes in context that reflect switching between task conditions. In contrast, studies of sensory context often emphasize graded changes in state that reflect continuous, smooth contextual variables. These distinct analytical approaches have implicated different circuit mechanisms for contextual effects. However, a comprehensive model of auditory processing should encompass both sensory and behavioral context. This section reviews findings from both lines of research, with the goal of establishing a more general framework for contextual effects on auditory encoding.

2.1. Sensory context

In studies of sensory context, a dominant idea has been that the auditory system adapts to ongoing, and presumably irrelevant, regularities in the acoustic environment in order to enhance responses to novel and potentially important sounds. This phenomenon is illustrated most simply with oddball tone stimuli. When a standard tone of fixed frequency is presented repeatedly, it is typically perceived as less salient over time. Then, when an oddball tone with a different frequency occurs at a random time in the sequence, it pops out perceptually, and neural responses are

correspondingly large. In human electrophysiological field recordings, an enhanced oddball response is observed in the mismatch negativity (MMN, Näätänen et al., 2007). In single-unit physiology recordings, a similar phenomenon is described as stimulus-specific adaptation (SSA, Pérez-González and Malmierca, 2014; Ulanovsky et al., 2003), although a direct correspondence between MMN and SSA is still debated. In addition to frequency, many other sound features (bandwidth, temporal modulations, phoneme identity) can be used to generate oddball responses, and the salience of the pop-out indicates what sound features the brain considers to be expected or unexpected in the context established by the standard stimulus. These oddball effects are believed to arise from a combination of feedforward adaptation and local cortical inhibition (Ayala and Malmierca, 2012; Natan et al., 2015) and may be related to mechanisms for gain control (Rabinowitz et al., 2011).

The idea of adaptation to regular, predictable inputs has motivated research on natural sound encoding. Several studies have demonstrated that encoding models estimated using natural sounds show distinct properties from models fit using traditional synthetic noise and tonal stimuli (David et al., 2009; Nagel and Doupe, 2008; Theunissen et al., 2000; Woolley et al., 2005). The idea that the auditory system adapts to statistical regularities in natural sounds generalizes to real-world problems, such as noise-invariant encoding of speech and other natural sounds. A series of studies recently measured single unit activity in the auditory midbrain and cortex during presentation of speech and other vocalizations in a noisy background (Mesgarani et al., 2014; Moore et al., 2013; Rabinowitz et al., 2013). A consistent observation was that neural responses were partially invariant to the noise, i.e., that they encoded more information about the foreground signal relative to the background noise. The properties of an encoding model that can produce this noise-invariant response are still debated, but developing such a model could provide a valuable tool for development of automated speech processing systems that are robust to noise.

A different line of research has focused on a sensory context from a different angle. Rather than focusing on the segmentation of foreground versus background, this work asks how the brain learns the statistical regularities in the environment in order to encode all sounds more accurately. When sounds are presented within a limited dynamic range, neurons across the auditory system adapt their response properties, in a way that is consistent with optimal encoding of sounds within that range. This phenomenon was first demonstrated with adaptation to sound level in the inferior colliculus (Dean et al., 2005), but a similar pattern of adaptation adaption has been reported for spatial tuning, suggesting that it may be a general coding strategy used by the auditory system (Dahmen et al., 2010). Dynamic adaptation to sound level has also been demonstrated in cortex (Watkins and Barbour, 2008) and in the auditory nerve (Zilany and Carney, 2010); thus it occurs at multiple levels of processing and is not an entirely central mechanism. At face value, the idea of improved coding within the current contextual space seems useful. Conceptually, however, it opposes the idea of a degraded representation of the background stimulus observed in studies of SSA and speech in noise. Encoding models that account for these phenomena in a more general framework may help link these apparently contradictory theories of sensory context.

In addition to contextual signals from the acoustic environment, signals from other sensory modalities can also modulate auditory-evoked activity. These signals may serve to prime the auditory system to detect signals with spatio-temporally correlated cross-modal features (e.g, a flash that predicts occurrence of a tone; Brosch et al., 2005; Schroeder and Foxe, 2005; Wallace et al., 1992). As the system adapts to different sensory contexts, these cross-modal signals may also support optimal integration of cues, as information from one modality becomes more or less reliable than information from another (Fetsch et al., 2013).

2.2. Behavioral context

Top-down signals reflecting behavioral state also influence auditory representations. Sound-evoked activity changes following a transition between passive listening and behavior (Fritz et al., 2003; Niwa et al., 2012; Otazu et al., 2009; Ryan and Miller, 1977), as well as a switch between tasks (Fritz et al., 2005; Rodgers and DeWeese, 2014). These contextual effects can be specific to different sound features—spectral (Fritz et al., 2003), temporal (Jaramillo and Zador, 2011), or spatial (Lee and Middlebrooks, 2011). They can reflect distinct aspects of the task, including effort (Atiani et al., 2009), selective attention (Downer et al., 2017; Hocherman et al., 1976; Schwartz and David, 2018), or reward contingencies (David et al., 2012). More generally, changes in state that are not directly related to a task can also influence auditory activity, e.g., arousal (McGinley et al., 2015), sleep (Edeline et al., 2000; Issa and Wang, 2011), and anesthesia (Gaese and Ostwald, 2001; Massaux and Edeline, 2003). Finally, a large literature has demonstrated long-term changes in auditory representations over the course of training (e.g., Ohl et al., 2001; Polley et al., 2006) and following hearing loss (e.g., Aizawa and Eggermont, 2007; Chambers et al., 2016).

2.3. Mechanisms of contextual effects

The diversity of sensory and behavioral context effects suggest that multiple modulatory signals influence auditory coding. Many possible sources of contextual signals have been proposed. Some adaptation effects may arise from local inhibitory circuits (Guo et al., 2017; Moore and Wehr, 2013; Natan et al., 2015). Lateral connections between cortical areas may mediate multisensory and motor signals (Schneider et al., 2014; Wallace et al., 1992). Neuromodulatory systems have been implicated in relatively rapid changes in auditory activity, most prominently the cholinergic and noradrenergic systems (Bakin and Weinberger, 1996; Edeline, 2003; Kilgard and Merzenich, 1998). They are also believed to mediate long-term changes in coding (e.g., dopamine, Happel et al., 2014; oxytocin, Marlin et al., 2015). The source of top-down signals that guide effects of behavioral context is uncertain, but the prefrontal cortex has been implicated, either through direct feedback to auditory cortex (Winkowski et al., 2013) or acting through neuromodulatory systems (Fritz et al., 2007a).

3. The nuts and bolts of spectro-temporal encoding models

The earliest encoding models applied to the auditory system used spike-triggered averaging, also known as reverse correlation, to determine the average sound pressure waveform that evoked a neuronal spike (De Boer, 1968; Marmarelis and Marmarelis, 1978). While useful in peripheral areas, this model fails to account for sound-evoked activity that is not locked to the phase of the stimulus waveform, which is true in most central auditory areas. A critical advance came with the development of the spectro-temporal receptive field (STRF), in which reverse correlation is applied to the sound spectrogram, rather than the raw sound waveform (Aertsen and Johannesma, 1981; Eggermont, 1993). The STRF has been used to study coding across the auditory system, from brainstem to cortex (deCharms et al., 1998; Klein et al., 2000; Miller et al., 2001; Reiss et al., 2007). This general approach for studying neural encoding, developed in the auditory system, has since been applied to other sensory modalities (DiCarlo and

Johnson, 2000; Jones and Palmer, 1987; Nagel and Wilson, 2011; Ramirez et al., 2014).

The vast majority of work on encoding models has focused exclusively on the sensory-evoked component of neural activity. Neural data are often averaged across repeated stimulus presentations, effectively removing variability due to changes in internal behavioral state. Only relatively recently have these static, context-independent models been extended to account for the internal behavioral factors that produce variability independent of the stimulus. The following subsections describe the basic formulation of context-independent models and then how they can be extended to account for contextual influences.

3.1. Context-independent spectro-temporal encoding models

An encoding model is a mathematical function, H, that describes how a time-varying stimulus, $s(x,t)$, produces an increase or decrease in neural activity, $r(t)$ (Fig. 1a),

$$r(t) = H[s(x,t)] + \varepsilon(t) \tag{1}$$

For single-unit recordings, $r(t)$ is typically the peri-stimulus time histogram (PSTH) response averaged across repeated presentations of the stimulus. In addition to the predicted response, there remains a residual, $\varepsilon(t)$, which reflects activity that cannot be explained by the model. The residual is a combination of stimulus-evoked activity that the model fails to predict and variability in the neural response independent of the stimulus. The basic problem of encoding models is to find the optimal H that minimizes the fraction of neural activity that cannot be predicted and thus is left in the residual.

Typically for auditory encoding models, the input stimulus is the spectrogram, $s(x,t)$, which describes the time-varying energy at each frequency, x, over time. In the classic formulation of a spectrogram, which is used in a wide range of signal processing applications outside of neuroscience, frequency channels are linearly spaced (Gill et al., 2006). However, frequency tuning in the mammalian cochlea and auditory nerve is approximately log-spaced, and a "cochleogram," with logarithmically-spaced frequency channels is often used for a more accurate model of the sensory periphery (see (Aertsen and Johannesma, 1981; Katsiamis et al., 2007) and Fig. 2).

Most studies of auditory encoding models have focused on single-unit spiking data. However, the same approach can be applied to a variety of neural signals. Encoding models have been used to characterize selectivity of single-unit membrane potential (Machens et al., 2004), multiunit activity (Eggermont, 1998), local field potentials (Eggermont et al., 2011; Hullett et al., 2016), magnetoencephalographic (MEG) signals (Ding and Simon, 2012), fMRI BOLD signals (Boumans et al., 2008; Moerel et al., 2012), and even psychophysical behavior (Varnet et al., 2013). While the basic encoding model framework can apply to any neural signal, details of the best implementation can depend on the signal under scrutiny. For example, signals reflecting large neural populations, such as MEG and BOLD, may require a different formulation of the static nonlinearity or noise model than single-unit spike recordings (see below).

The current review focuses on a particular encoding model, the linear-nonlinear spectro-temporal receptive field (LN STRF), which is widely used in studies of peripheral and central processing (Fig. 1b and Calabrese et al., 2011; David et al., 2009; Holdgraf et al., 2016; Rabinowitz et al., 2012). This model consists of two stages. First a linear filter, the STRF, computes a weighted sum of the stimulus spectrogram over frequency and the preceding time to produce a linear prediction of the time-varying spike rate, $r_{\mathrm{lin}}(t)$,

$$r_{\mathrm{lin}}(t) = \sum_{x=1}^{X} \sum_{u=0}^{U} h(x,u)s(x, t-u) \tag{2}$$

The weights, $h(x,u)$, indicate the gain applied to the stimulus frequency channel x at time lag u. Positive values indicate components of the stimulus correlated with increased neural response, and negative values indicate components correlated with decreased response.

The linear filter provides a basic picture of the spectral tuning and latency of evoked responses, and many studies focus only on the properties of this filter (David et al., 2012; Fritz et al., 2003; Miller et al., 2002; Reiss et al., 2007). However, a purely linear model does not account for some basic nonlinear properties of neurons, namely that they have a threshold activation for spiking and their responses saturate for very strong stimuli. Thus in a second stage of the LN model, the output of the linear filter passes through a static nonlinearity to predict the spike rate response:

$$r_{LN}(t) = f[r_{\mathrm{lin}}(t)] \tag{3}$$

The nonlinear function f typically has a sigmoid shape, which can be specified with a number of different forms (Thorson et al., 2015). A nonlinearity used commonly in the generalized linear model framework is a logistic sigmoid,

$$f[r_{\mathrm{lin}}(t)] = a + \frac{b}{1 + \exp[-(r_{\mathrm{lin}}(t) - c)/d]} \tag{4}$$

for which the terms a, b, c, and d are free parameters that define the threshold, amplitude and slope of the nonlinearity (Rabinowitz et al., 2012).

In addition to providing a description of the stimulus patterns that evoke neural activity, the LN STRF and other encoding models have the ability to predict the neural response to an arbitrary novel stimulus, even if it was not used to fit the model itself. For the LN STRF, prediction amounts to applying the stimulus spectrogram as input to Equations (1) and (2). The accuracy of the model can then be assessed quantitatively by comparing the predicted response and actual response to that stimulus. In theory, a perfect model should predict the time-varying response exactly, leaving zero residual (i.e., $\varepsilon(t) = 0$ in Eq. (1)).

More practically, measures of prediction accuracy can be used to compare the relative performance of two alternative encoding models. For example, the benefit of a specific formulation of the static nonlinearity or a constraint on the weights in the linear filter can be assessed by whether incorporating them into a model produces an improvement in prediction accuracy (Thorson et al., 2015). Prediction can also be used to assess a model's generalizability. For example, a model can be fit using a synthetic noise stimulus and its accuracy tested with a natural sound (Theunissen et al., 2000). Finally, relevant to the current review, prediction accuracy can also be used to determine how well a model generalizes across different behavioral contexts (Holdgraf et al., 2016; Rabinowitz et al., 2012; Schwartz and David, 2018).

Several different metrics have been used to measure prediction accuracy, including the correlation coefficient (Pearson's R), mean-squared error, Poisson log-likelihood, and mutual information (Atencio et al., 2008; Calabrese et al., 2011; Eggermont et al., 1983; Sahani and Linden, 2003). Implicit in each of these metrics are different assumptions about noise in neural activity that cannot be predicted from the stimulus (i.e., the residual, $\varepsilon(t)$ in Eq. (1)). This is a nuanced statistical issue, but the choice of metric can bias comparisons of model performance. Both the correlation coefficient and mean-squared error assume that the residual is Gaussian noise, a
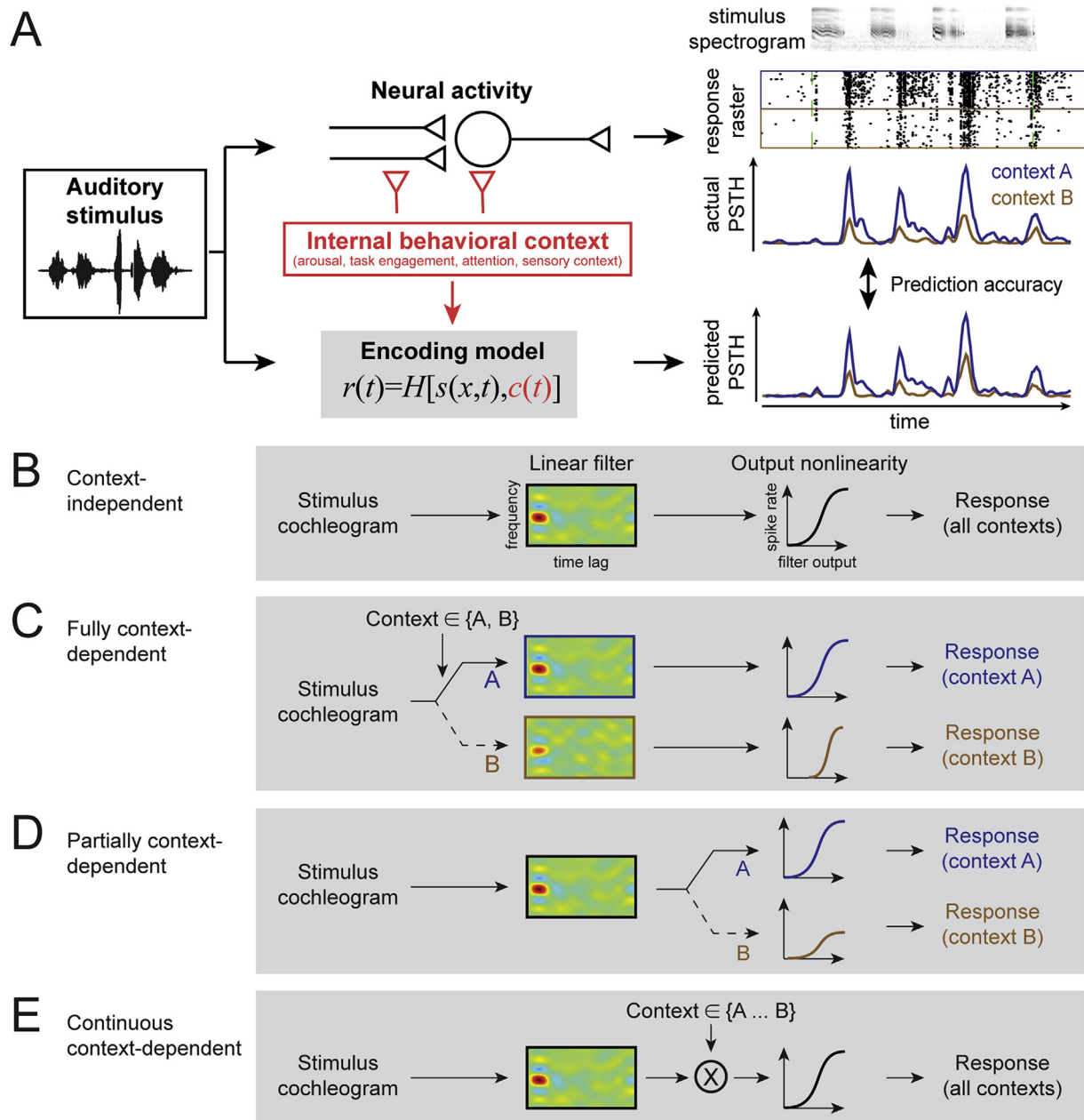
Fig. 1. Framework for context-dependent auditory encoding models. **a.** Information from an input sound stimulus (acoustic waveform at left, spectrogram at top right) is represented by the output time-varying spike rate of a neuron (right). However, spiking activity also reflects the effects of internal brain state, or context (red), which can modulate sound-evoked activity. Thus evoked activity can differ between context A (blue) and B (brown). Averaging neural activity across repeated presentations of the same sound produces a peristimulus-time histogram (PSTH) response, which in this example shows a clear difference in amplitude between behavioral contexts. The goal of the encoding model is to predict the PSTH in each context (bottom right). **b.** Traditional spectro-temporal models are context-independent and predict the same sound-evoked activity regardless of context. The LN STRF models the neural response as a linear weighted sum of the stimulus spectrogram followed by an output nonlinearity to account for spike threshold and saturation. **c.** A fully context-dependent model assumes that the encoding can change arbitrarily between contexts, and thus a separate set of model parameters are fit for each context. **d.** A partially context-dependent model assumes that just a subset of parameters is context-dependent while the remaining parameters are fixed across contexts. In this example, linear filter weights are fixed while the parameters of the output nonlinearity are context-dependent. **e.** A continuous context-dependent model allows context to change smoothly among many states. This contrasts with the more typical approaches in c and d, where context represents a discrete change in brain state between a small number of conditions. In this LN STRF example, a continuous context variable modulates the amplitude of the linear filter output.

very general assumption that lends itself to straightforward analysis. However, because variability in neuronal spiking can be described as a Poisson process, Poisson log-likelihood has been proposed as a better model of noise in spiking neurons (Calabrese et al., 2011; Paninski et al., 2004). At a further extreme, information theoretic measures such as mutual information make even fewer assumptions about neural noise and represent, in theory,

unbiased assessment of the portion of the neural activity that can be predicted by the model. However, accurate measures of information are themselves prone to bias and require careful implementation (Treves and Panzeri, 1995). While different prediction metrics vary in their details, they all provide a means of determining the best model fit and comparing model performance. Moreover, they tend to all produce roughly similar results,
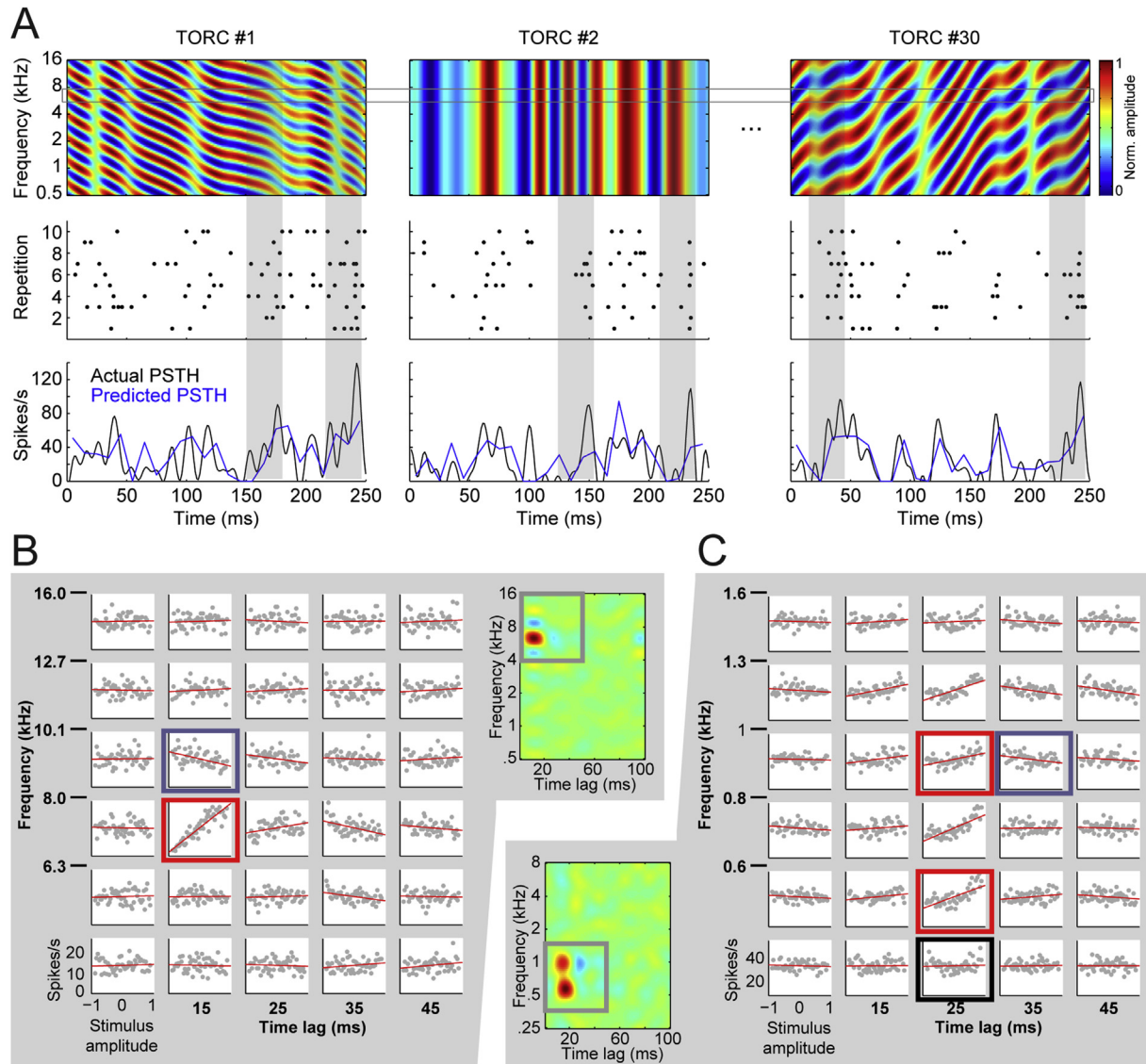
Fig. 2. STRF estimation by linear regression. **a.** Top row shows the log spectrogram of three temporally orthogonal ripple combination (TORC) stimuli, noise-like stimuli that efficiently sample spectro-temporal stimulus space for reverse correlation. Middle row shows the raster plot of spike events recorded from a neuron in A1 during repeated presentation of each TORC. Bottom row shows the PSTH response (black), computed as the average spike rate at each time. The STRF identifies which stimulus spectro-temporal features correlate with increases or decreases in the response. Gray shading delineates time periods preceding relatively large evoked spike rates. Blue curves overlaid in the bottom panel show the PSTH predicted by the STRF. **b.** Each scatter plot compares the response against the stimulus spectrogram at a different frequency and time lag preceding the response (stimulus amplitude normalized between −1 and 1). Individual points indicate average response after binning by stimulus amplitude. The STRF is estimated by computing the correlation between the stimulus and response, averaged over time (Eq. (9)). The slope of a line fit to each scatter plot indicates the corresponding weight in the STRF. Positive slopes indicate stimulus components correlated with increased spike rate (red box), and negative slopes indicate components correlated with a decrease in spike rate (blue box). The weights are plotted in a heat map (right, interpolated to facilitate visualization). Gray box indicates the subset of weight calculations illustrated in the scatter plots. **c.** Example STRF calculation for a second A1 neuron, plotted as in B. This neuron shows excitatory tuning over a wider range of frequencies and inhibition at a later time lag than excitation. Several correlation plots also indicate nonlinear stimulus-response relationships (e.g., black box) that are not fit well by a line. These deviations reflect a nonlinear stimulus-response relationship that cannot be explained by a linear STRF. (Data reanalyzed from (David et al., 2012).).

especially in the domain of relatively simple models, such as the LN STRF (Thorson et al., 2015).

3.2. Incorporating behavioral context into the STRF

Equation (1) describes a model in which neural activity is determined entirely by the stimulus. However, central sensory activity is not static, and depends on the state of the brain prior to sensory input. Thus, a more general encoding model describes neural activity as a function of both the stimulus and time-varying context, $c(t)$,

$$r(t) = H[s(x,t), c(t)] \tag{5}$$

This context variable can reflect a wide range of factors independent of the immediate stimulus. Individual studies typically focus on a single aspect of context, ranging from recent acoustic experience (e.g., David and Shamma, 2013; Holdgraf et al., 2016; Rabinowitz et al., 2012), changes in task engagement or attention (e.g., Da Costa et al., 2013; David et al., 2012; Ding and Simon, 2012; Fritz et al., 2003; Nourski et al., 2016), and long-term effects of hearing loss (e.g., Noreña et al., 2003).

Practical model architectures for studying context-dependence

of sound encoding are illustrated for a simple LN STRF in Fig. 1b–e. A static, *context-independent model* provides a baseline assessment of sound encoding, assuming that encoding model parameters do not change with behavioral context (Fig. 1b). In the LN STRF framework, this means that neural activity in any behavior condition is predicted by the model defined in Eqs. (2) and (3). In most behavioral studies, animals are trained to switch discretely between two behavioral states (e.g., active versus passive, Fritz et al., 2003; or attend location versus frequency, Rodgers and DeWeese, 2014). A *discrete fully context-dependent model* fits the entire model separately for each attention condition (Fig. 1c),

$$r(t) = H_c[s(x, t)], \quad c \in \{A, B\} \tag{6}$$

If the effects of context can be isolated to a subset of changes in sound-evoked activity, the model can be constrained so that only some parameters are context-dependent. A *discrete partially context-dependent model*, is an encoding model in which only a subset of the parameters is fit separately for each behavior state. For the example in Fig. 1d, the linear filter, and thus its output, are the same for both behavior conditions, but the parameters of the static nonlinearity are fit separately for each behavior condition,

$$r(t) = f_c[r_{\text{lin}}(t)], \quad c \in \{A, B\} \tag{7}$$

This example would account for a context-dependent change in the overall excitability, or gain, of a neuron without a change in its spectro-temporal selectivity.

While manipulation of context through behavioral control provides a convenient and tractable discretization of context, it does not fully capture the variability of internal states. Several aspects of behavioral context, such as arousal (e.g., McGinley et al., 2015), motor control (Schneider et al., 2014), attention (Rabinowitz et al., 2015), and motor planning (Runyan et al., 2017), can change smoothly among many different values. If a continuous contextual variable can be quantified, then it can be incorporated directly into a *continuous context-dependent model* (Fig. 1e). For the example in the figure, the context variable scales the output of the linear filter prior to the output nonlinearity,

$$r(t) = f[c(t)r_{\text{lin}}(t)] \tag{8}$$

The continuous context variable could, in theory, be incorporated into the linear STRF itself, but this more complex model would require a large dataset for accurate estimation.

These alternative model architectures vary in their complexity and thus the required amount of data for accurate fitting. The optimal fit algorithm may also depend on model architecture. Thus a robust method for comparing model performance is necessary to determine if the additional contextual terms benefit model performance (Wu et al., 2006). Cross-validation tests with held-out data provide a general framework for evaluating model performance, and significant improvements in performance can be measured against the context-independent model. However careful selection of validation data that samples adequately across different contexts, as well as sensory conditions, is required to assess performance robustly. A software library for fitting and comparing performance of different context-dependent encoding models is available online: https://bitbucket.org/lbhb/nems/.

3.3. STRF estimation by reverse correlation

Reverse correlation is the canonical method for STRF estimation, providing the best mean-squared error estimate of the STRF in the case when the stimulus is white noise (i.e., has no statistical regularities that produce first-order correlations in frequency or time, Aertsen and Johannesma, 1981; Eggermont, 1993; Klein et al., 2000). Natural stimuli, as well as other stimuli with spectral and temporal correlations, require additional steps to obtain an accurate STRF (see below and Theunissen et al., 2001). A key analytical requirement for reverse correlation is that neural activity is recorded during presentation of a wide variety of spectro-temporal patterns. Gaussian white noise is a useful stimulus for studies in the periphery (Aertsen and Johannesma, 1981). However, it tends not to evoke strong responses in primary auditory cortex (A1) and other central areas. Several alternative types of stimuli have been developed, including random chords (Pienkowski et al., 2009; Sahani and Linden, 2003) and ripples (Klein et al., 2000; Miller et al., 2001). These stimuli contain harmonic and temporally modulated sound features that tend to evoke much stronger responses in central areas. They also are constructed to maintain noise-like statistical properties, permitting straightforward reverse correlation analysis.

Fig. 2a shows the spectrogram from three segments of temporally orthogonal ripple combinations (TORCs), noise stimuli designed to efficiently sample the diversity of stimuli required for reverse correlation (Klein et al., 2000). Neural activity recorded during repeated sound presentation (middle, Fig. 2a) is averaged to produce a PSTH response (bottom, Fig. 2a), which fluctuates over time. Large values in the PSTH indicate segments of the TORC stimuli that evoke strong neural responses (vertical gray bars).

The STRF in Eq. (2) requires a coefficient, or gain, for each spectral channel, x, and time lag, u. Reverse correlation can be thought of as multiple regression, where a line is fit to a scatter plot of the stimulus at a particular time lag, $s(x, t\text{-}u)$, and the corresponding response, $r(t)$ (Fig. 2b). The slope of each line indicates the weight of the respective STRF coefficient. In the case of white noise stimuli, the slope can be estimated directly from the correlation between stimulus and response,

$$h(x, u) = \frac{1}{T} \sum_{t=1}^{T} r(t)s(x, t - u) \tag{9}$$

For visualizing the STRF, the slope of the line fit to each scatter plot is represented in a heat map (Fig. 2b–c). Here, positive slopes are colored red, indicating frequencies and time lags associated with increased neural activity, and negative slopes are colored blue, associated with decreased activity.

The first example (Fig. 2b) shows data from an A1 neuron whose response to TORCs can be described fairly well by a linear STRF. Data from the same neuron are plotted in Fig. 2a. The scatter plots are well-fit by regression lines. The slopes of the lines are near zero in most cases, indicating that tuning is confined to a small range of frequencies and time lags. A best frequency is apparent between 6 and 8 kHz, where the regression line has a steep positive slope, and an inhibitory sideband is apparent in the adjacent frequency band. The ability of this STRF to predict sound-evoked activity is clear in the relatively close match between the actual and predicted PSTHs in Fig. 2a.

A second example (Fig. 2c) shows data for another A1 neuron with strong evoked responses to TORCs, but which exhibits clear nonlinear properties that are not well described by the STRF. As in the previous example, excitatory and inhibitory tuning are clearly observable in the scatter plots. However, the scatter plots show curvature that is not captured by the regression line. U-shaped responses are visible for some frequencies and time lags (black box), indicating an excitatory response to both loud and quiet sounds that cannot be captured by a linear model.

For non-white noise stimuli, such as natural sounds, additional computations are required to remove bias introduced by correlated

structure. Several different methods can be used in this case, including normalized reverse correlation (Theunissen et al., 2001), ridge regression (Machens et al., 2004), and gradient-based techniques (Atencio et al., 2008; Calabrese et al., 2011; David et al., 2007; Meyer et al., 2014). Each requires specification of a small number of "hyperparameters" for tuning the STRF estimate. Methods for choosing these hyperparameters are relatively straightforward but are more nuanced than reverse correlation. Gradient-based methods are the most powerful and flexible of these approaches. They are also the most computationally intensive, a limitation which has become less severe with ongoing improvements in computer power. These estimation methods are not detailed here, but are described elsewhere (Machens et al., 2004; Paninski et al., 2004; Theunissen et al., 2001).

4. Context-dependent encoding models for top-down changes in behavioral state

Although it has long been established that internal behavioral state can influence sensory coding (Hocherman et al., 1976; Ryan and Miller, 1977), it is only relatively recently that behavioral context has been integrated into neural sensory encoding models. Early studies of behavior-dependent coding were completed in the visual system, following the demonstration of selective attention effects in visual cortex (Moran and Desimone, 1985). Researchers mapped spatial receptive fields in visual cortical area V4 and found that they shifted to represent the retinotopic area at the locus of attention (Connor et al., 1997). Studies of spatial attention also considered the possibility of effects on feature selectivity. McAdams and Maunsell measured orientation tuning curves in V4 during changes in spatial attention (McAdams and Maunsell, 1999). They observed that the baseline firing rate and gain of evoked responses could change, but orientation tuning remained largely stable. Studies in area MT, which is associated with coding of visual motion, extended this idea into the domain of feature selectivity. Animals were trained to attend to the direction of motion in a dot pattern and direction selectivity could be measured during task performance (Treue and Martinez-Trujillo, 1999). Subsequent studies of feature attention in V4 have argued that shifts can occur in visual selectivity following changes in feature-based attention (David et al., 2008).

Early studies of changes in feature selectivity in the auditory system focused on long-term changes following learning and pharmacological manipulation. Studies pairing cholinergic modulation with pure tone presentation during recording from auditory cortex identified changes in neuronal frequency tuning that reflected the frequency of the paired tone (Bakin and Weinberger, 1996). Ohl and colleagues measured changes in spectro-temporal selectivity of LFP activity in gerbils as they learned acoustic discriminations (Ohl et al., 2001; Ohl and Scheich, 1997). Generally, the long-term changes in auditory tuning following these manipulations reflected behaviorally important sound features, supporting the idea that the auditory system is able to engage a matched filter mechanism to enhance discriminability between task-relevant features.

Subsequent work by Fritz et al. (2003) explored changes on a shorter timescale, "rapid plasticity" that occurred during engagement in a task and that reversed to a baseline tuning in a subsequent passive condition. These effects occurred in the same temporal regime as the effects of selective attention in the visual system. Their task required animals to detect a target tone with fixed frequency at a random position in a sequence of TORC stimuli (Fig. 3a, Klein et al., 2000). Thus, STRFs could be estimated by reverse correlation from activity during behavior and compared to STRFs in a passive condition. The use of identical stimuli in both behavioral contexts controlled for any possible bias from the use of different stimuli to measure STRFs in the different conditions. When aversive conditioning was used for behavior (animals were required to cease licking a water spout to avoid a mild tail shock) the STRFs showed a selective enhancement of responses to the target frequency. Subsequent experiments measured the effects of more complex targets consisting of tone chords and of including tones of different frequency in the reference period (Fritz et al., 2007b, 2005). Across all of these experiments a consistent pattern was observed: during behavior, STRFs underwent plasticity that produced enhanced responses to target sound features and decreased responses to features in the reference stimuli. As in the case of long-term effects of learning (Ohl et al., 2001; Ohl and Scheich, 1997) and effects of attention in the visual system (David et al., 2008; Treue and Martinez-Trujillo, 1999), these changes are consistent with a matched filter for perceptual enhancement. Changes in the STRF reflect enhanced discriminability between neural responses to the different task categories. These effects represent an optimal strategy for enhancing discriminability, under the constraint of maintaining an overall constant level of sound-evoked activity across the entire system (Mesgarani et al., 2010).

Since the initial demonstration of task-related plasticity, subsequent studies have shown that several aspects of the task other than the relevant sound features can influence rapid plasticity in auditory cortex. Changing task difficulty by varying the signal-to-noise ratio (SNR) of a tone embedded in noise changed the overall excitability of A1 neurons (Atiani et al., 2009). In a study using a task that required detection of a spatially localized target, the authors observed narrowing of spatial receptive fields upon task engagement (Lee and Middlebrooks, 2011). Finally, reversing the valance of the target in the tone detection task (using an approach paradigm in which animals were rewarded for responding to target tones rather than being punished for missing) had the unexpected effect of producing an opposite pattern of plasticity (David et al., 2012). STRFs showed a selective decrease in response at the target tone frequency (Fig. 3c). Although the sign of STRF plasticity was opposite when target valance was reversed, the relative change between target and non-target frequencies remained consistent with a matched filter model (Mesgarani et al., 2010): in both the approach and avoidance data, the plasticity enhanced the neural response to one of the task categories (target during avoidance, noise during approach). This relative enhancement permits a downstream area from A1 to decode stimulus identity better than in the baseline, passive condition (David et al., 2012). Together, these diverse results reveal that, while the task-relevant acoustic features are critical for shaping spectro-temporal tuning plasticity, other aspects of the task, including difficulty, reward value and/or motor responses, determine the sign of the modulation and overall changes in excitability.

The studies in A1 gave rise to new questions about the mechanisms that produce task-related plasticity and about the possibility of behavioral context in subcortical auditory areas. Slee & David (Slee and David, 2015) used an approach behavior identical to the one used in A1 (David et al., 2012) to measure possible effects in the ferret inferior colliculus (IC). They observed the same pattern of suppression of the STRF at target frequency as in A1. However, in many cases, changes in the STRF in IC could be described by an overall change in excitability rather than a shift in turning (Fig. 3b). A change in gain can be described by a model with only partial dependence on behavioral context (Fig. 1d). The linear STRF can be fit the same in both passive and active conditions and the change in gain can be captured by allowing just the parameters of the static nonlinearity to change.

Data from the studies in A1 and IC (David et al., 2012; Slee and David, 2015) were reanalyzed to test for systematic differences in
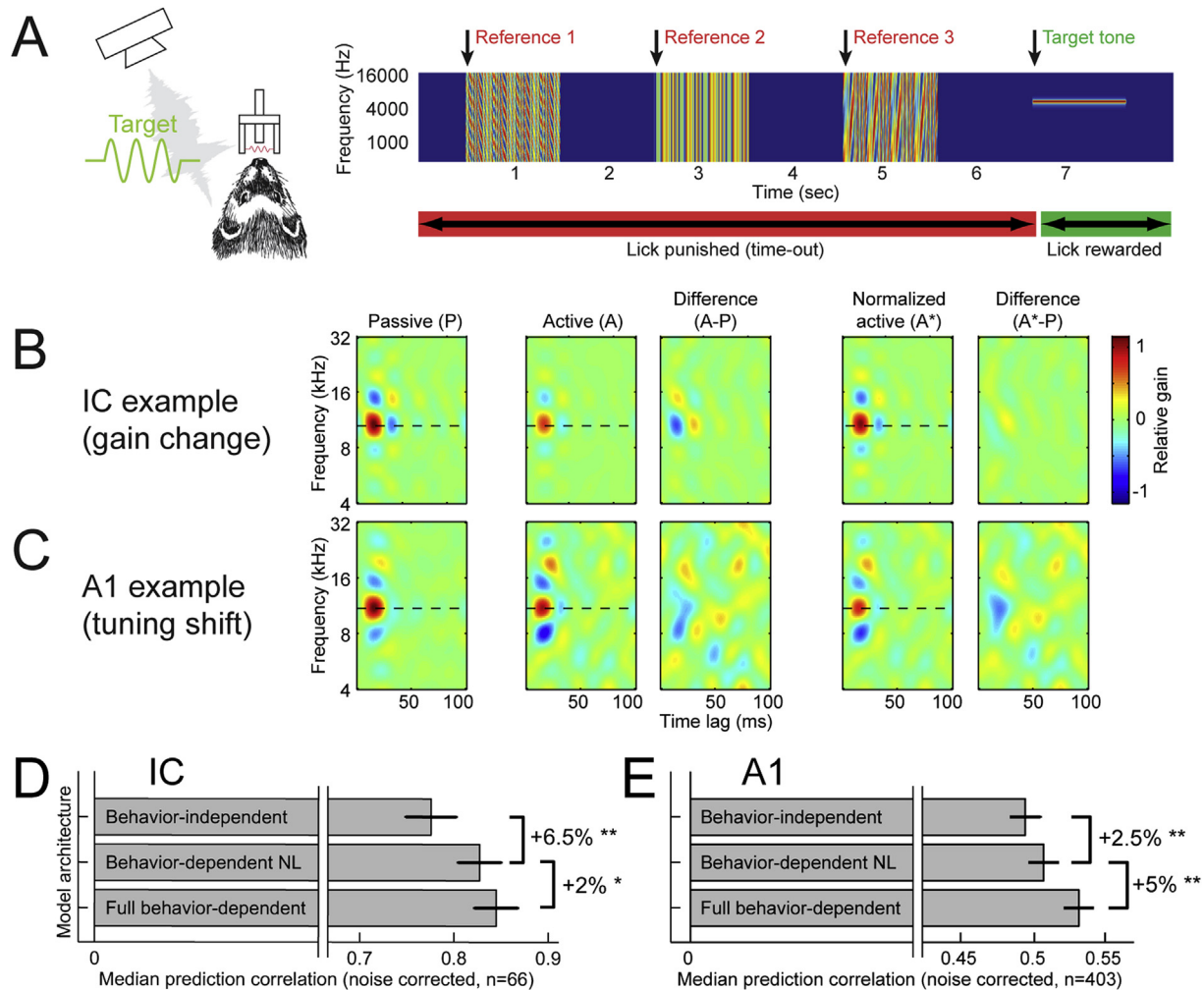
Fig. 3. Partial versus full context-dependence of LN STRFs during auditory behavior. **a.** Go/no-go tone detection task in ferret. Left, head-fixed animal responds to target sounds by licking a water spout. Right, spectrogram of example trial. A random number of TORC stimuli are presented followed by a pure tone target, with fixed frequency across a set of behavioral trials. A lick prior to target onset results in punishment with a time-out. A lick following target onset results in a liquid reward. Identical TORC stimuli were also presented during passive listening before and/or after behavior. **b.** Comparison of STRFs measured from responses of an IC neuron to the TORCs during passive listening and behavior. The passive STRF shows best frequency at 10 kHz, matched to the frequency of the target tone (dashed line, panel 1). During behavior, the STRF weights are lower (panel 2), reflecting decreased response gain (blue region in active-passive difference, panel 3). Rescaling the active STRF by the global decrease in response gain accounts for most of the behavior-dependent change (panels 4—5). Such a global gain change can be captured in a partially context-dependent model, in which the STRF is fixed across behavior conditions but the output nonlinearity is fit separately (Fig. 1d). **c.** Comparison of STRFs for an A1 neuron, plotted as in B. This neuron also shows a decrease in the STRF at the target frequency. However, normalizing for the slight increase in global gain does not account for most of the behavior-dependent change. **d.** Comparison of median LN STRF prediction correlation for a set of IC neurons across passive and active conditions, fit using three different models of context dependence. Error bars indicate one standard error. The model with a context-dependent static nonlinearity shows an increase in performance over the behavior-independent model (**p < .001, permutation test). The fully context-dependent model shows only slightly greater prediction accuracy (*p < .05), indicating that the majority of behavioral effects can be explained by the changes to the static nonlinearity. **e.** Comparison of the same models for a set of A1 neurons collected under the same behavioral conditions. Here, the fully context-dependent model is required to explain most of the behavior-dependent changes in the LN STRF. (Data reanalyzed from (David et al., 2012; Slee and David, 2015).).

local versus global effects of plasticity in the two areas. Prediction accuracy was compared between three LN STRF architectures that either ignore behavior context (context-independent model) or incorporate different forms of context-dependence. For the fully context-dependent model, separate STRF and output nonlinearity parameters were estimated for the passive and active task conditions (Eq. (6), Fig. 1c). For the partially context-dependent model, a single STRF was estimated for both conditions, and only parameters of the output nonlinearity were estimated separately per behavior condition (Eq. (7), Fig. 1d). As suggested by the example in Fig. 3b, fitting the output nonlinearity separately between passive and active conditions resulted in improved prediction accuracy over the context-independent model. The fully context-dependent model, in which the entire STRF was estimated separately, provided only a

slight additional increase in prediction accuracy. A different pattern of model performance was revealed for the A1 data. The partially context-dependent model showed a relatively small increase over the behavioral independent model, while the fully context-dependent model provided a larger increase. Thus, the partially context-dependent model explains a larger portion of behavior effects in IC than in A1.

This pattern of model performance suggests a hierarchical model of the mechanism producing selectivity changes in A1. Neurons in the midbrain IC undergo task-dependent changes in excitability that can be captured largely in the static nonlinearity. Neurons in A1 integrate inputs from many IC neurons, each of which might undergo different changes in excitability. These differential changes in input strength lead to a shift in the selectivity

in A1 that requires a model in which the STRF is also context-dependent. Thus by identifying the encoding model parameters that depend on behavioral context, it is possible to make inferences about how internal behavioral state interfaces with incoming auditory signals. Information traveling from IC to A1 passes through neurons in the medial geniculate body (MGB) of the thalamus. Studies of behavior-dependent coding in the MGB during the same behavior could reveal further insight into how auditory information is transformed as it passes through this network.

5. Context-dependent models for bottom-up effects of experience and learning

In addition to exerting top-down control on sound processing, the auditory system also exhibits the ability to adapt to changing environments in order to encode relevant sound features more effectively. A classic example of this process is hearing loss, where damage to the auditory periphery degrades sound information that reaches the brain, and central auditory areas undergo plasticity, possibly adaptive or pathological, to compensate for the reduced sensory drive (Bajo et al., 2010; Chambers et al., 2016; Noreña et al., 2003).

On a shorter timescale, normal-hearing individuals are also able to learn statistical regularities in highly distorted sounds and extract behaviorally relevant information. A classic example of this process is sine-wave speech, where signals are synthesized from dynamic chords tracking the formants of speech recordings (Dorman et al., 1997; Mcauly and Quatieri, 1986). Initially, the signals are perceived as tonal noise. After hearing the original speech signal, subjects are able to recognize the sine-wave speech, and after exposure to several examples, they are able to generalize the percept, so that they can recognize sine-wave speech without ever having heard the original signal. Similarly, noise-vocoders are used to simulate the input provided by cochlear implants (Shannon et al., 1995). Subjects presented with noise-vocoded speech initially have trouble perceiving the speech sound, but performance improves after training with lexical feedback (Hervais-Adelman et al., 2008).

Studies using recordings from single neurons in A1 to reconstruct sound spectrograms suggest that the encoding properties of cortical neurons serve to enhance natural signals over distorting noise (Mesgarani et al., 2014; Rabinowitz et al., 2013). Spectrograms reconstructed from neural responses to a noisy speech stimulus produced a clean stimulus spectrogram, in which the representation of the distorting noise diminished. Similarly, studies of selective attention during field potential recordings from humans have shown enhanced representation of the attended over non-attended speech stream in the neural signal (Ding and Simon, 2012; Mesgarani and Chang, 2012). While decoding analysis has demonstrated an enhanced representation of behaviorally relevant and attended signals, the encoding properties of neurons that support this enhancement have not been characterized.

To explore the neural encoding properties that support encoding of signals in noise, a recent study explored changes in encoding properties in human auditory cortex as subjects learned to perceive highly distorted speech (Holdgraf et al., 2016). Electrocorticographic (ECoG) recordings measured local field potential (LFP) in the superior temporal gyrus (STG) of patients implanted with subdural electrode arrays during monitoring for treatment of intractable epilepsy. The time-varying energy in the high-frequency broadband signal (HFB, 70–150 Hz), also referred to as high gamma LFP, was used to measure sound evoked neural activity at each recording site. Although the exact composition of the HFB remains unclear, it is believed to reflect multi-unit activity near the recording site (Steinschneider et al., 2008). Thus the HFB signal can be treated as a time-varying multi-unit response and used to

measure the STRF using the same methods as for single-neuron data (Hullett et al., 2016; Mesgarani and Chang, 2012; Pasley et al., 2012).

By comparing HFB activity during presentation of the distorted speech sounds before and after exposure to the clean speech, Holdgraf et al. (2016) were able to measure changes in neural activity that reflected improved representation of the speech signal. The STG is believed to be an area equivalent to parabelt auditory cortex, which represents highly processed auditory information. Thus, the STRFs measured in the STG are fairly complex and can be difficult to interpret by standard tuning measures such as best frequency and tuning bandwidth (Hullett et al., 2016; example in Fig. 4b). However, they can be used to predict the HFB response in the different behavioral contexts. To test the effects of learning from exposure to the clean speech signal, the STRF measured in the clean speech condition was used to predict responses to the noisy signals pre- and post-exposure. Evoked activity had much greater amplitude post-exposure, and the STRF predicted the stronger post-exposure response. More importantly, the STRF more accurately predicted the response in the post-exposure period, indicating that the underlying spectro-temporal filters were able to selectively amplify the speech sounds relative to the competing noise (Fig. 4b). This dynamic enhancement did not occur if subjects were presented irrelevant noise signals during the intervening period rather than the clean speech.

Analysis of the modulation power spectra of STRFs measured in the noisy conditions before and after exposure to the clean signal revealed a possible strategy used by the brain to enhance the relevant signals. For the temporal distortions, STRFs in the post-exposure period attenuated tuning to high frequency temporal modulations which contained greater power than the speech signal itself. Conversely, in the case of spectral noise, the STRFs shifted to attenuate narrowband spectral signals. Thus the rapid, dynamic changes in STRFs revealed that the brain operates as a matched filtered, identifying stimulus bands with the greatest signal-to-noise ratio over just a few seconds of clean speech exposure and adapting filter properties to enhance those bands.

6. Modeling effects of continuously changing context

While the majority of behavioral studies seek to control changes in behavior state so that neural activity is characterized in a small number (2–4) of distinct, discrete states, the underlying neural processes that influence auditory processing may in fact be continuously varying. Studies of pupil-indexed arousal have shown that neuromodulatory effects on cortical activity can indeed vary smoothly and even non-monotonically (McGinley et al., 2015). Other studies have observed modulatory effects of motor activity on sensory responses that may fluctuate smoothly with the amount of activity in motor cortex (Rummell et al., 2016; Schneider et al., 2014). More broadly, evidence from pharmacological studies indicates that sensory cortical function can be modulated smoothly and continuously by variations in neuromodulatory tone (Bakin and Weinberger, 1996; Salgado et al., 2011). Thus, an alternative to the typical approach of modeling discrete changes in context is to introduce a continuous variable reflecting behavioral state into the encoding model (Fig. 1e).

While no study has yet integrated a continuous task-related variable into auditory encoding models, this concept has been used to study the effects of changing stimulus context. In this approach, an LN STRF is used to model encoding of complex sounds, but a second filter is used to extract contrast information from the stimulus, which is applied as a scaling term to the output of the STRF (Rabinowitz et al., 2012; Williamson et al., 2016). While one could consider this definition of sensory context simply to be a
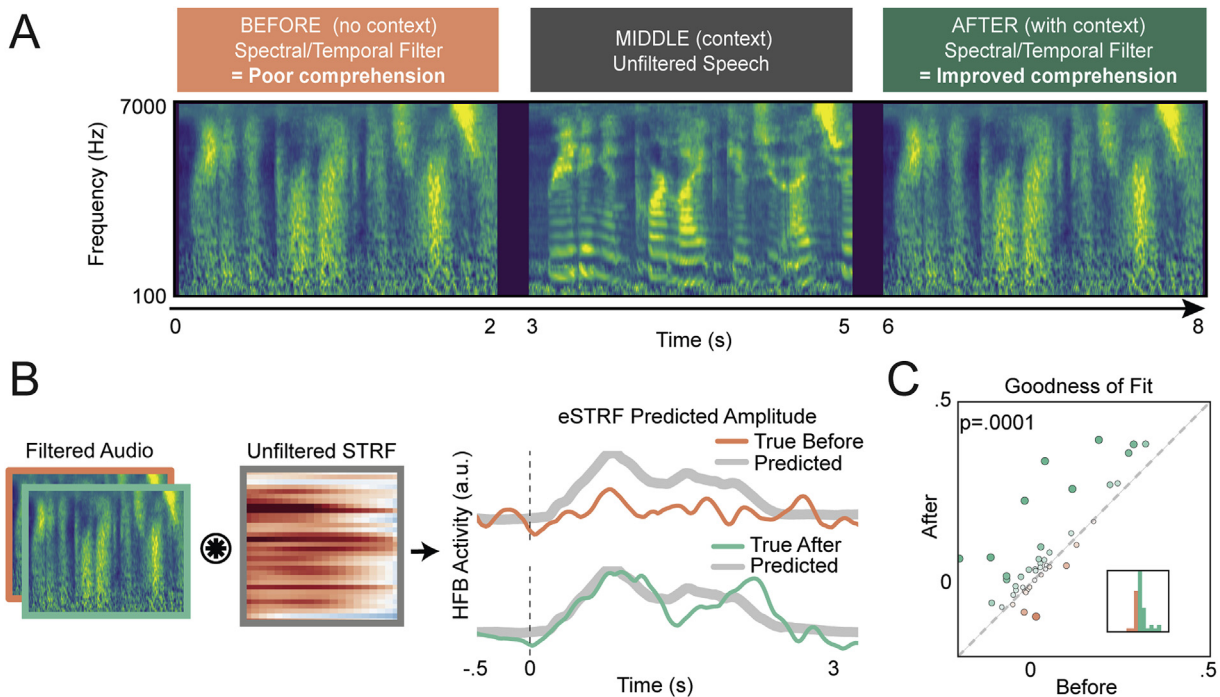
Fig. 4. Rapid changes in spectro-temporal tuning following adaptation to noisy stimulus statistics. **a.** Human subjects were asked to recognize speech that was severely degraded by temporally or spectrally modulated noise. Upon initially hearing the noisy speech, they were not able to report its content accurately. They were then presented the same sentence in quiet, followed again by the degraded speech, at which point they reported that they were able to perceive the speech in the degraded conditions. **b.** Electricorticographic (ECoG) activity was recorded from the surface of patients undergoing monitoring for epilepsy treatment during presentation of the speech stimuli. High-frequency band (HFB) activity recorded from electrodes positioned on the superior temporal gyrus (STG) showed a modest response during presentation of the first distorted speech stimulus (red curve, top right). Stronger responses were observed for the same stimulus following exposure to the speech in quiet (bottom right). HFB activity was used to estimate the ensemble STRF (eSTRF) from responses to the speech in quiet. This eSTRF was used to predict the response in the pre- and post-exposure periods (gray curves). **c.** The accuracy with which eSTRFs predicted HFB activity was greater during the post-exposure period than the pre-exposure period, consistent with more accurate encoding of the speech signal after exposure to the undistorted speech. This improved prediction accuracy was consistent with a decrease in eSTRF weights for channels with large distortions, suggesting that the auditory system dynamically forms a matched filter to extract the most reliable speech features available in the current noisy context. (Figures modified and reprinted by permission from (Holdgraf et al., 2016).).

more complex nonlinear model of sensory encoding, it represents a more general strategy for modeling the effect of internal behavioral state (reflecting stimulus history) on sound-evoked activity.

Discrete context-dependent models (Fig. 1c—d) have been used in numerous studies to characterize contextual effects of changing stimulus statistics. This includes a comparison of tuning to natural versus synthetic noise stimuli (David et al., 2009; Theunissen et al., 2000), variable spectral density (Norena et al., 2008; Schneider and Woolley, 2011), variable stimulus contrast (Rabinowitz et al., 2011) and variable sound level (Nagel and Doupe, 2008). As in the behavioral studies described above, these models implicitly assume the system shifts discretely between static "modes" of activation for different categories of stimuli. In reality, the changes between stimulus conditions reflect the output of dynamic filters that adapt to the distinct statistical properties of the stimuli. Thus, modeling the effects of stimulus context as a continuous state variable that modulates sound-evoked activity represents an alternative approach to this well-known problem. Future studies that identify behavioral context variables will build on this approach in more refined models that integrate smooth, continuous changes in behavioral state into encoding models.

A recent study developed a context-dependent encoding model to account for contrast normalization in the magnitude of sound-evoked activity in auditory cortex (Rabinowitz et al., 2012). The authors had previously reported that changing the spectral contrast of stimuli within a narrow band around a neuron's best frequency modulates the gain of sound evoked responses (Rabinowitz et al., 2011). This process is similar to adaptive gain control reported for

sound level in the auditory midbrain (Dean et al., 2008, 2005), but the current model is able to capture the dynamics of the adaptation and can generalize to stimuli that fluctuate between the different regimes, rather than simply being fit discretely within isolated stimulus regimes.

Single-unit recordings for fitting the context-dependent model were acquired from A1 of awake and anesthetized ferrets during acoustic stimulation. The stimulus was a sequence of dynamic random chords (DRC) composed of tones at 23 different frequencies ranging from 0.5 to 22.6 kHz. The entire sequence was 360 s in duration. Every 3 s the contrast in each frequency band switched randomly between high (92%) and low (33%), with a constant overall sound level. The stimulus could be used to estimate a standard linear STRF using methods similar to those described above (k_{ft}, Fig. 5a). The authors then considered several alternative models in which an additional contextual signal was defined based on the current contrast in each spectral channel. A separate linear filter was fit to the contrast signal (κ_{ft}, Fig. 5a), and its output was scaled to produce two time-varying spectral contrast signals, c_t and d_t. These time-varying signals then replaced what are typically modeled using static coefficients in the nonlinearity applied to the output of the linear filter,

$$f(x) = a + \frac{b}{1 + \exp[-(x - c_t)/d_t]} \quad (10)$$

Rather than being static, as it is in context-independent models, the output nonlinearity changed dynamically with local contrast.
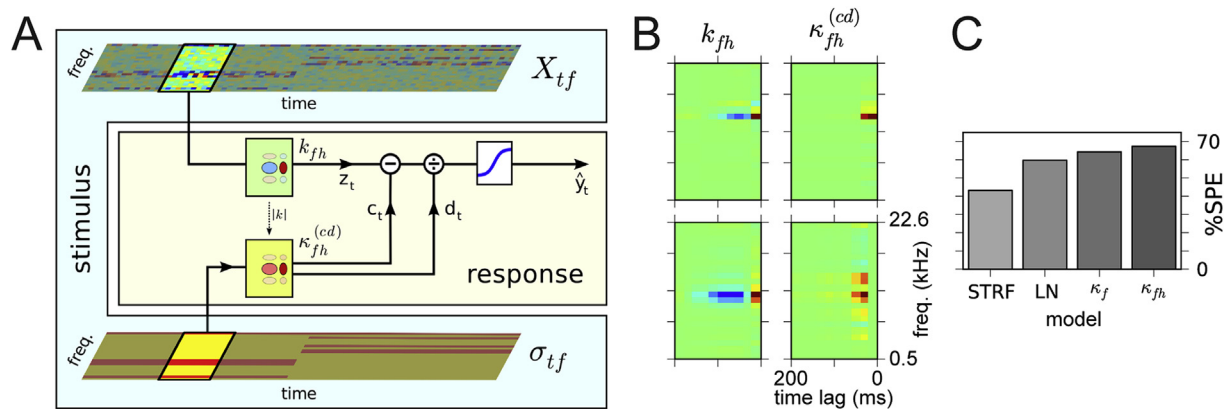
Fig. 5. Incorporating nonlinear gain control into auditory encoding models to account for local stimulus context. **a.** Activity of A1 neurons was recorded during presentation of random chord stimuli in which each frequency channel fluctuated between high and low contrast every few seconds (top). The stimulus spectrogram could be used to estimate a standard STRF (k_{fh}). Stimulus context was defined as the contrast level in each frequency band (bottom), which was used to estimate a second, contrast filter ($\kappa_{fh}^{(cd)}$), the output of which was used to scale the output of the linear STRF. **b.** Example linear filters (left) and contrast filters (right) for two neurons. The contrast filters tended to have slightly broader spectral tuning than the linear filter. **c.** Comparison of percent signal power explained (%SPE) for different models. The model incorporating a full spectro-temporal contrast filter performed more accurately than a model with a contrast filter that measured only instantaneous contrast (κ_f), but both context-dependent models performed better than the LN STRF and the STRF alone (i.e., with no static nonlinearity). %SPE provides an alternative measure of prediction accuracy, similar to prediction correlation but accounting for noise in the PSTH of the neural activity being predicted (Sahani and Linden, 2003). (Figures modified and reprinted by permission from (Rabinowitz et al., 2012).).

This dynamic gain term accounted for gain control previously observed in auditory cortex (Rabinowitz et al., 2011), but it also identified the spectral and temporal extent of its influence, which varied substantially across A1 neurons. Generally, the linear filter and contrast filter shared similar tuning, but they showed important differences, particularly in their temporal extent (Fig. 5b). Contrast filters typically had longer integration times, up to 500 ms or longer. Linear STRFs typically show integration times of less than 100 ms. Importantly, allowing an independent contrast filter improved model prediction accuracy over a model with a static nonlinearity (Fig. 5c). Thus, a model that incorporates a continuously varying context-dependent term is able to account for A1 activity better than a static LN STRF model.

7. Looking ahead: bigger data and better behavior

7.1. Overcoming data limitations

Robust and reliable control of animal behavior is challenging. Animals, of course, can be trained to perform operant behaviors that require a specific and sometimes difficult auditory discrimination (Osmanski et al., 2013). However, maintaining good performance while obtaining enough data from a stable single-neuron recording to estimate encoding model parameters requires precise experimental design, substantial patience, and a certain amount of luck on the part of the experimenter. Even when animals perform a task, it is difficult to fully control behavioral state. The behavioral strategy and degree of effort can vary over the timecourse of a single experiment (Lakatos et al., 2016; McGinley et al., 2015). Thus there is a persistent sense that more data from a single experiment and data sampling a variety of behavioral contexts is needed to form a more complete understanding of the interaction of context and auditory encoding (Fritz et al., 2007a). Given the minutes and maybe hours of recording time permitted by traditional neurophysiological recordings from awake, behaving animals, it may seem impossible to reach this goal.

Because of these limitations, studies incorporating behavioral context into encoding models have thus far been limited to the relatively simple LN STRF. More comprehensive nonlinear encoding models have been proposed, but their complexity makes it difficult to incorporate into behavioral studies, where data are often critically limited. One approach that may allow the integration of behavior into more sophisticated models is dimensionality reduction, in which encoding models are constrained to the minimum number of parameters required for accurate prediction (Thorson et al., 2015). Second-order nonlinear models can require a squaring of the number of parameters, but many of those parameters are not actually required. Subspace projections and targeted nonlinearities provide a means of reducing dimensionality while providing greater explanatory power than linear models (Atencio and Sharpee, 2017; David and Shamma, 2013). These strategies may provide the best direction toward more comprehensive context-dependent encoding models.

New experimental technologies also promise to address the data limitation problem.

In traditional electrophysiological experiments, maintaining stable recordings for many minutes or hours is challenging, especially in behaving animals. Techniques for chronic electrophysiological recording from implanted electrode arrays (Cohen and Maunsell, 2009) and calcium imaging through cranial windows (Kuchibhotla et al., 2016; Runyan et al., 2017) have matured in recent years. Experiments using these methods have the potential to study the same neurons during presentation of large stimulus sets and under multiple behavioral conditions, thus supporting estimation of more comprehensive encoding models.

However, each of these new methods does suffer from its own limitations. Tissue growth can inhibit the recording quality of chronically implanted electrodes, and the resulting yield of long-term single unit recordings can be low, even for large multi-channel arrays (Polikov et al., 2005). Calcium imaging can be performed stably over many days, but temporal precision of calcium signals cannot be resolved to precise spike times (Vogelstein et al., 2010). Addressing these limitations is an active area of research, and these methods will become increasingly useful for encoding models as the associated technologies improve.

The development of high-quality subdural ECoG recordings from human patients is particularly exciting, having opened up access to subjects who can perform multiple sophisticated behaviors while recording multiunit and even occasionally single unit activity (Mesgarani and Chang, 2012; Nourski et al., 2016). Human ECoG studies promise new insight, especially into high-level auditory processing, but this approach does have limitations.

Access to single-unit data is limited, and most studies are restricted to high frequency band LFP activity, which approximates multiunit recordings from superficial cortical areas. Electrode placement is also difficult to control between subjects, and access to subjects is limited to a relatively small number of institutions. Finally, although these recordings are available outside the focus of epileptic activity, it should also be noted that these patients may show abnormal brain oscillations, even in unaffected areas (Engel and da Silva, 2012). Thus, human ECoG recordings, while valuable, may not become a widely used experimental resource.

Finally, technologies for identification and targeted manipulation of neural circuits promise an additional avenue for developing more comprehensive encoding models. Optogenetic tools can be used to tag neural subpopulations according to distinguishing molecular markers and their position in neural circuits. This approach has been used to compare tuning properties between populations (Guo et al., 2017; Moore and Wehr, 2013) and to compare behavior-dependent effects between populations (Kuchibhotla et al., 2016). Optogenetic tools can also be used to reversibly activate or inactivate specific neural populations during acoustic stimulation. This approach has been used to characterize the role of different neural populations in cortex in coding of sounds and sensory context (Natan et al., 2015; Phillips and Hasenstaub, 2016). These tools are currently limited mostly to studies in mice. However, ongoing efforts are underway to develop them for use in other species, including non-human primates (Macdougall et al., 2016).

7.2. Expanding the definition of behavioral context

On a more theoretical level, recent studies have identified non-auditory signals that influence the activity of auditory neurons, even if they are not explicitly controlled by a behavioral task. These include arousal (McGinley et al., 2015), anesthesia state (Stringer et al., 2016), motor activity (Brosch et al., 2005; Fritz et al., 2010; Schneider et al., 2014), stimulus reward value (Baruni et al., 2015; David et al., 2012), and behavioral choice (Bizley et al., 2013; Runyan et al., 2017). Simultaneous recordings from large neural populations also allow the analysis of large-scale population activity, which itself reflects changes in internal state (Okun et al., 2015; Pillow et al., 2008; Runyan et al., 2017; Stevenson et al., 2012). These signals can be incorporated into context-dependent models, along with variables reflecting task conditions controlled by the experimenter.

While these additional modulatory signals do not necessarily reflect strategies for optimal auditory signal processing or discrimination, they represent essential components of the larger neural system that supports auditory processing. Hearing exists fundamentally to serve behavior. Encoding models that incorporate these contextual variables will provide new insight into the transformation of auditory signals into decision variables and motor responses. Incorporating multiple contextual variables into a single model will also reveal how the underlying modulatory processes emerge and interact in auditory pathways.

7.3. Alternatives to the linear STRF

As theories of auditory coding have developed, the STRF has been identified as a special case of a much broader class of sensory encoding models (Eggermont, 1993; Wu et al., 2006). Encoding models represent any solution to the general problem of characterizing the functional relationship between sensory stimuli and neural responses. The STRF is fundamentally limited in its ability to describe sensory coding because it is a linear approximation of the relationship between stimulus spectrogram and evoked neural

spike response. It is well-established that linear models fail to describe a substantial portion of central neural responses to natural sounds (David et al., 2009; Machens et al., 2004; Willmore et al., 2016). Linear models have classically been appealing because their estimation is computationally tractable. However, the ongoing increase in computational power available in the lab has made model complexity less of a constraint.

The LN STRF as described by Eqs. (2) and (3) is not constrained to apply a specific formulation of the stimulus spectrogram or of the static nonlinearity applied to the linear filter output. In addition to the standard linear spectrogram (Aertsen and Johannesma, 1981; Theunissen et al., 2001), several different methods have been developed for generating *cochleograms* from raw stimulus waveforms, which simulate the filter properties of the cochlea and thus model the signals arriving from the auditory nerve more accurately (Chi et al., 2005; Gill et al., 2006; Katsiamis et al., 2007). Importantly, the best model of the periphery varies between species. In birds, the cochlea is better approximated by a spectrogram with linear-spaced frequency channels, while in mammals, log-spaced channels are more widely used. Similarly, the formulation of the output nonlinearity can vary from a simple hard threshold to a sigmoid with independent curvature near threshold and saturation (Calabrese et al., 2011; Rabinowitz et al., 2011; Thorson et al., 2015). Varying the formulations of these model elements can impact model prediction accuracy, though they rarely impact the basic pattern of selectivity observed in the linear STRF (Gill et al., 2006; Thorson et al., 2015).

A more critical limitation of the LN STRF is the linear filter itself. The example in Fig. 2c illustrates one nonlinear response property that cannot be captured by the linear STRF. Several stimulus-response scatter plots show U-shaped relationships (e.g., black box at 25 ms time lag, bottom row). A line fit to these data is not able to capture the increased response for large and small stimuli. Instead the fit has nearly zero slope and the linear STRF indicates zero gain at that frequency and time lag.

Several model architectures have been proposed to capture nonlinear encoding properties. The linear STRF can be conceived as the first-order Volterra series expansion of a nonlinear stimulus-response function (Aertsen and Johannesma, 1981; Eggermont, 1993). Some nonlinear models incorporate second-order non-linearities, in which the response is the linear weighted sum of the stimulus spectrogram plus a weighted sum of the product of values in the spectrogram (Atencio et al., 2008; Atencio and Sharpee, 2017). These approaches have been integrated with probability theory-based models of neural coding and neural network models from machine learning (Atencio et al., 2008; Calabrese et al., 2011; Pillow et al., 2008).

Other architectures incorporate elements that model specific biologically motivated computations that cannot be captured by the linear model. For example, neural responses might integrate over the output of two or more LN STRFs fit in parallel (Harper et al., 2016; Kozlov and Gentner, 2016; Schinkel-Bielefeld et al., 2012), neuronal inputs might undergo short-term synaptic depression and/or facilitation prior to filtering (David and Shamma, 2013), or inputs might undergo a more complex transformation that mimics nonlinear processing at an earlier processing stage (Willmore et al., 2016). Finally, models like the one described above (Fig. 5) that incorporate information about stimulus contrast or background noise in a gain control term are able to account for some nonlinear responses by casting them as components of the local stimulus context (Rabinowitz et al., 2012; Williamson et al., 2016).

Although it is well established that the LN STRF fails to predict a large portion of sound evoked activity, no single encoding model has yet taken the place of the STRF as the "standard model" for auditory coding. The domain of possible sounds to test is so large

and the variety of possible nonlinear model is so vast that finding a definitive replacement that consistently performs better is difficult. Generally, the more comprehensive nonlinear models require more free parameters than simpler linear models, and thus they require more data for accurate estimation. Because data are typically limited in behavioral studies, these more complex models have not yet been used to study the effects of behavioral context on coding, and the majority of this work has instead focused on the linear STRF. As technologies for chronic recording from behaving animals improve, incorporating context into these models will become more feasible. Until then, the best way to describe auditory neural encoding comprehensively remains an open question.

8. Conclusions

In their essence, all spectro-temporal encoding models reduce to the same basic concept: a function that predicts a time-varying neural signal based on the dynamic pattern of sensory stimulation. The parameters of the model fit to a neural signal provide information about the spectro-temporal features of stimuli encoded by that signal. In addition, the accuracy with which the model predicts neural activity provides an objective, quantitative measure of how well it describes the stimulus-response relationship (Wu et al., 2006).

As neuroscientists have come to appreciate the role of behavioral context on processing in the auditory and other sensory systems, the need for incorporating internal behavioral state into encoding models of sensory processing has become clear (Fritz et al., 2007a). Studies incorporating behavior have shown that attention to relevant task features can cause shifts in the selectivity of auditory neurons to emphasize the contrast between features important for performing the behavioral task (Fritz et al., 2003; Holdgraf et al., 2016; Mesgarani et al., 2010).

Contextual influences on auditory processing reflect a large number of behavioral factors and vary over a wide range of timescales. A single model that can account for all of these contextual influences remains a theoretical concept and is not practical with current experimental data, especially for mammals and other large animals. However, the idea of exhaustively tracking behavior and modeling its relationship with neural activity is being applied in smaller species (Kato et al., 2015; Robie et al., 2017). This review argues that the problem of contextual influence can be studied from a foundation of static, context-independent sensory encoding models. Encoding models can incorporate multiple discrete and continuous behavioral state variables, both experimentally controlled and passively observed. As behavioral state is monitored more closely and more powerful methods for data acquisition are developed, encoding models will grow increasingly comprehensive.

Encoding models that predict context-dependent neural activity more accurately will provide new insight into the neural circuits that support auditory behavior. When contextual signals arrive via specific neural pathways, context-dependent models can be constrained so that only parameters reflecting activity in those pathways vary with context. Thus, encoding models can be used to infer the neural circuit elements that modulate auditory processing. These inferences can subsequently be used to identify computational strategies used by the brain for sensory processing during real-world behavior.

Acknowledgments

References

Aertsen, A.M., Johannesma, P.I., 1981. The spectro-temporal receptive field. A functional characteristic of auditory neurons. Biol. Cybern. 42, 133–143.

Aizawa, N., Eggermont, J.J., 2007. Mild noise-induced hearing loss at young age affects temporal modulation transfer functions in adult cat primary auditory cortex. Hear. Res. 223, 71–82. https://doi.org/10.1016/j.heares.2006.09.016.

Atencio, C.A., Sharpee, T.O., 2017. Multidimensional receptive field processing by cat primary auditory cortical neurons. Neuroscience. https://doi.org/10.1016/j.neuroscience.2017.07.003.

Atencio, C.A., Sharpee, T.O., Schreiner, C.E., 2008. Cooperative nonlinearities in auditory cortical neurons. Neuron 58, 956–966.

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, 467–480.

Ayala, Y.A., Malmierca, M.S., 2012. Stimulus-specific adaptation and deviance detection in the inferior colliculus. Front. Neural Circ. 6, 89. https://doi.org/10.3389/fncir.2012.00089.

Bajo, V.M., Nodal, F.R., Moore, D.R., King, A.J., 2010. The descending corticocollicular pathway mediates learning-induced auditory plasticity. Nat. Neurosci. 13, 253–260. https://doi.org/10.1038/nn.2466.

Bakin, J.S., Weinberger, N.M., 1996. Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. Proc. Natl. Acad. Sci. U. S. A. 93, 11219–11224.

Baruni, J.K., Lau, B., Salzman, C.D., 2015. Reward expectation differentially modulates attentional behavior and activity in visual area V4. Nat. Neurosci. 18, 1656–1663. https://doi.org/10.1038/nn.4141.

Bizley, J.K., Walker, K.M.M., Nodal, F.R., King, A.J., Schnupp, J.W.H., 2013. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Curr. Biol. 23, 620–625. https://doi.org/10.1016/j.cub.2013.03.003.

Boumans, T., Gobes, S.M.H., Poirier, C., Theunissen, F.E., Vandersmissen, L., Pintjens, W., Verhoye, M., Bolhuis, J.J., Van der Linden, A., 2008. Functional MRI of auditory responses in the zebra finch forebrain reveals a hierarchical organisation based on signal strength but not selectivity. PLos One 3, e3184. https://doi.org/10.1371/journal.pone.0003184.

Brosch, M., Selezneva, E., Scheich, H., 2005. Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. J. Neurosci. 25, 6797–6806.

Buran, B.N., Sarro, E.C., Manno, F.A.M., Kang, R., Caras, M.L., Sanes, D.H., 2014. A sensitive period for the impact of hearing loss on auditory perception. J. Neurosci. 34, 2276–2284. https://doi.org/10.1523/JNEUROSCI.0647-13.2014.

Calabrese, A., Schumacher, J.W., Schneider, D.M., Paninski, L., Woolley, S.M.N., 2011. A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. PLos One 6, e16104. https://doi.org/10.1371/journal.pone.0016104.

Chambers, A.R., Resnik, J., Yuan, Y., Whitton, J.P., Edge, A.S., Liberman, M.C., Polley, D.B., 2016. Central gain restores auditory processing following near-complete cochlear denervation. Neuron 89, 867–879. https://doi.org/10.1016/j.neuron.2015.12.041.

Chi, T., Ru, P., Shamma, S.A., 2005. Multiresolution spectrotemporal analysis of complex sounds. J. Acoust. Soc. Am. 118, 887–906.

Cohen, M.R., Maunsell, J.H.R., 2009. Attention improves performance primarily by reducing interneuronal correlations. Nat. Neurosci. 12, 1594–1600. https://doi.org/10.1038/nn.2439.

Connor, C.E., Preddie, D.C., Gallant, J.L., Van Essen, D.C., 1997. Spatial attention effects in macaque area. J. Neurosci. 17, 3201–3214, V4. .

Da Costa, S., van der Zwaag, W., Miller, L.M., Clarke, S., Saenz, M., 2013. Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. J. Neurosci. 33, 1858–1863. https://doi.org/10.1523/JNEUROSCI.4405-12.2013.

Dahmen, J.C., Keating, P., Nodal, F.R., Schulz, A.L., King, A.J., 2010. Adaptation to stimulus statistics in the perception and neural representation of auditory space. Neuron 66, 937–948. https://doi.org/10.1016/j.NEURON.2010.05.018.

David, S.V., Fritz, J.B., Shamma, S.A., 2012. Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 109, 2150–2155.

David, S.V., Hayden, B.Y., Mazer, J.A., Gallant, J.L., 2008. Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. Neuron 59, 509–521.

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. J. Neurosci. 29, 3374–3386.

David, S.V., Mesgarani, N., Shamma, S.A., 2007. Estimating sparse spectro-temporal receptive fields with natural stimuli. Network 18, 191–212.

David, S.V., Shamma, S.A., 2013. Integration over multiple timescales in primary auditory cortex. J. Neurosci. 33, 19154–19166. https://doi.org/10.1523/JNEUROSCI.2270-13.2013.

De Boer, E., 1968. Reverse correlation. I. A heuristic introduction to the technique of triggered correlation with the application to the analysis of compound systems. Proc K Ned Akad Wet C 71, 472–486.

De Boer, E., Kuyper, P., 1968. Triggered correlation. IEEE Trans. Biomed. Eng. 15, 159–179.

Dean, I., Harper, N.S., McAlpine, D., 2005. Neural population coding of sound level

adapts to stimulus statistics. Nat. Neurosci. 8, 1684–1689. https://doi.org/10.1038/nn1541.

Dean, I., Robinson, B.L., Harper, N.S., McAlpine, D., 2008. Rapid neural adaptation to sound level statistics. J. Neurosci. 28, 6430–6438. https://doi.org/10.1523/JNEUROSCI.0470-08.2008.

deCharms, R.C., Blake, D.T., Merzenich, M.M., 1998. Optimizing sound features for cortical neurons. Science 280, 1439–1443 (80-.).

DiCarlo, J.J., Johnson, K.O., 2000. Spatial and temporal structure of receptive fields in primate somatosensory area 3b: effects of stimulus scanning direction and orientation. J. Neurosci. 20, 495–510.

Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. U. S. A. 2012 https://doi.org/10.1073/pnas.1205381109.

Dorman, M.F., Loizou, P.C., Rainey, D., 1997. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. J. Acoust. Soc. Am. 102, 2403–2411. https://doi.org/10.1121/1.419603.

Downer, J., Rapone, B., Verhein, J., O'Connor, K.N., Sutter, M.L., 2017. Feature selective attention adaptively shifts noise correlations in primary auditory cortex. J. Neurosci. 37 (21), 5378–5392. https://doi.org/10.1523/JNEUROSCI.3169-16.2017.

Edeline, J.M., 2003. The thalamo-cortical auditory receptive fields: regulation by the states of vigilance, learning and the neuromodulatory systems. Exp. Brain Res. 153, 554–572.

Edeline, J.M., Manunta, Y., Hennevin, E., 2000. Auditory thalamus neurons during sleep: changes in frequency selectivity, threshold, and receptive field size. J. Neurophysiol. 84, 934–952.

Eggermont, J.J., 1998. Representation of spectral and temporal sound features in three cortical fields of the cat. Similarities outweigh differences. J. Neurophysiol. 80, 2743–2764.

Eggermont, J.J., 1993. Wiener and Volterra analysis applied to the auditory system. Hear. Res. 66, 177–201.

Eggermont, J.J., Aertsen, A.M., Johannesma, P.I., 1983. Prediction of the responses of auditory neurons in the midbrain of the grass frog based on the spectro-temporal receptive field. Hear. Res. 10, 191–202.

Eggermont, J.J., Munguia, R., Pienkowski, M., Shaw, G., 2011. Comparison of LFP-based and spike-based spectro-temporal receptive fields and cross-correlation in cat primary auditory cortex. PLos One 6, e20046. https://doi.org/10.1371/journal.pone.0020046.

Engel, J., da Silva, F.L., 2012. High-frequency oscillations - where we are and where we need to go. Prog. Neurobiol. 98, 316–318. https://doi.org/10.1016/j.pneurobio.2012.02.001.

Fetsch, C.R., DeAngelis, G.C., Angelaki, D.E., 2013. Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. Nat. Rev. Neurosci. 14, 429–442. https://doi.org/10.1038/nrn3503.

Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P., Shamma, S.A., Elhilali, M., 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat. Neurosci. 13, 1011–1019. https://doi.org/10.1038/nn.2598.

Fritz, J.B., Elhilali, M., David, S.V., Shamma, S.A., 2007a. Auditory attention—focusing the searchlight on sound. Curr. Opin. Neurobiol. 17, 437–455.

Fritz, J.B., Elhilali, M., Shamma, S.A., 2007b. Adaptive changes in cortical receptive fields induced by attention to complex sounds. J. Neurophysiol. 98, 2337–2346. https://doi.org/10.1152/jn.00552.2007.

Fritz, J.B., Elhilali, M., Shamma, S.A., 2005. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. J. Neurosci. 25, 7623–7635.

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

Gaese, B.H., Ostwald, J., 2001. Anesthesia changes frequency tuning of neurons in the rat primary auditory cortex. J. Neurophysiol. 86, 1062–1066.

Gill, P., Zhang, J., Woolley, S.M., Fremouw, T., Theunissen, F.E., 2006. Sound representation methods for spectro-temporal receptive field estimation. J. Comput. Neurosci. 21, 5–20.

Guo, W., Clause, A.R., Barth-Maron, A., Polley, D.B., 2017. A corticothalamic circuit for dynamic switching between feature detection and discrimination. Neuron 95, 180–194. https://doi.org/10.1016/j.neuron.2017.05.019 e5.

Happel, M.F.K., Deliano, M., Handschuh, J., Ohl, F.W., 2014. Dopamine-modulated recurrent corticoefferent feedback in primary sensory cortex promotes detection of behaviorally relevant stimuli. J. Neurosci. 34, 1234–1247. https://doi.org/10.1523/JNEUROSCI.1990-13.2014.

Harper, N.S., Schoppe, O., Willmore, B.D.B., Cui, Z., Schnupp, J.W.H., King, A.J., 2016. Network receptive field modeling reveals extensive integration and multi-feature selectivity in auditory cortical neurons. PLoS Comput. Biol. 12, e1005113 https://doi.org/10.1371/journal.pcbi.1005113.

Hermansky, H., 1998. Should recognizers have ears? Speech Commun. 25, 3–27. https://doi.org/10.1016/S0167-6393(98)00027-2.

Hervais-Adelman, A., Davis, M.H., Johnsrude, I.S., Carlyon, R.P., 2008. Perceptual learning of noise vocoded words: effects of feedback and lexicality. J. Exp. Psychol. Hum. Percept. Perform. 34, 460–474. https://doi.org/10.1037/0096-1523.34.2.460.

Hocherman, S., Benson, D.A., Goldstein Jr., M.H., Heffner, H.E., Hienz, R.D., 1976. Evoked unit activity in auditory cortex of monkeys performing a selective attention task. Brain Res. 117, 51–68.

Holdgraf, C.R., de Heer, W., Pasley, B., Rieger, J., Crone, N., Lin, J.J., Knight, R.T.,

Theunissen, F.E., 2016. Rapid tuning shifts in human auditory cortex enhance speech intelligibility. Nat. Commun. 7, 13654. https://doi.org/10.1038/ncomms13654.

Hullett, P.W., Hamilton, L.S., Mesgarani, N., Schreiner, C.E., Chang, E.F., 2016. Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. J. Neurosci. 36, 2014–2026. https://doi.org/10.1523/JNEUROSCI.1779-15.2016.

Issa, E.B., Wang, X., 2011. Altered neural responses to sounds in primate primary auditory cortex during slow-wave sleep. J. Neurosci. 31, 2965–2973. https://doi.org/10.1523/JNEUROSCI.4920-10.2011.

Jaramillo, S., Zador, A.M., 2011. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. Nat. Neurosci. 14, 246–251. https://doi.org/10.1038/nn.2688.

Jones, J., Palmer, L., 1987. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J. Neurophysiol. 58, 1233–1258.

Kato, S., Kaplan, H.S., Schrödel, T., Skora, S., Lindsay, T.H., Yemini, E., Lockery, S., Zimmer, M., 2015. Global brain dynamics embed the motor command sequence of Caenorhabditis elegans. Cell 163, 656–669. https://doi.org/10.1016/j.cell.2015.09.034.

Katsiamis, A.G., Drakakis, E.M., Lyon, R.F., 2007. Practical gammatone-like filters for auditory processing. EURASIP J. Audio, Speech, Music Process 2007, 1–15. https://doi.org/10.1155/2007/63685.

Kilgard, M.P., Merzenich, M.M., 1998. Cortical map reorganization enabled by nucleus basalis activity. Science 279, 1714–1718 (80-.).

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. Comput. Neurosci. 9, 85–111.

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. Neurophysiol. 76, 3503–3523.

Kozlov, A.S., Gentner, T., 2016. Central auditory neurons have composite receptive fields. Proc. Natl. Acad. Sci. Unit. States Am. 113, 1441–1446. https://doi.org/10.1073/pnas.1506903113.

Kuchibhotla, K.V., Gill, J.V., Lindsay, G.W., Papadoyannis, E.S., Field, R.E., Sten, T.A.H., Miller, K.D., Froemke, R.C., 2016. Parallel processing by cortical inhibition enables context-dependent behavior. Nat. Neurosci. 20, 62–71. https://doi.org/10.1038/nn.4436.

Lakatos, P., Barczak, A., Neymotin, S.A., McGinnis, T., Ross, D., Javitt, D.C., O'Connell, M.N., 2016. Global dynamics of selective attention and its lapses in primary auditory cortex. Nat. Neurosci. 19, 1707–1717. https://doi.org/10.1038/nn.4386.

Lee, C.-C.C., Middlebrooks, J.C., 2011. Auditory cortex spatial sensitivity sharpens during task performance. Nat. Neurosci. 14, 108–114. https://doi.org/10.1038/nn.2713.

Macdougall, M., Nummela, S.U., Coop, S., Disney, A., Mitchell, J.F., Miller, C.T., 2016. Optogenetic manipulation of neural circuits in awake marmosets. J Neurophysiol. Nat. Neurosci. Nat. Meth. 116, 1286–1294.

Machens, C.K., Wehr, M.S., Zador, A.M., 2004. Linearity of cortical receptive fields measured with natural sounds. J. Neurosci. 24, 1089–1100.

Marlin, B.J., Mitre, M., D'amour, J.A., Chao, M.V., Froemke, R.C., 2015. Oxytocin enables maternal behaviour by balancing cortical inhibition. Nature 520, 499–504. https://doi.org/10.1038/nature14402.

Marmarelis, P.Z., Marmarelis, V.Z., 1978. Analysis of Physiological Systems: the White Noise Approach. Plenum, New York, NY.

Massaux, A., Edeline, J.-M., 2003. Bursts in the medial geniculate body: a comparison between anesthetized and unanesthetized states in Guinea pig. Exp. Brain Res. 153, 573–578. https://doi.org/10.1007/s00221-003-1516-3.

McAdams, C.J., Maunsell, J.H.R., 1999. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area. J. Neurosci. 19, 431–441, V4. .

Mcauly, R.J., Quatieri, T.F., 1986. Speech analysis/synthesis based on a sinusoidal representation. IEEE Trans. Audio 34, 744–754. https://doi.org/10.1109/TASSP.1986.1164910.

McGinley, M.J., David, S.V., McCormick, D.A., 2015. Cortical membrane potential signature of optimal states for sensory signal detection. Neuron 87, 179–192.

Mesgarani, N., Chang, E.F., 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233–236. https://doi.org/10.1038/nature11020.

Mesgarani, N., David, S.V., Fritz, J.B., Shamma, S.A., 2014. Mechanisms of noise robust representation of speech in primary auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 111, 6792–6797. https://doi.org/10.1073/pnas.1318017111.

Mesgarani, N., Fritz, J.B., Shamma, S.A., 2010. A computational model of rapid task-related plasticity of auditory cortical receptive fields. J. Comput. Neurosci. 28, 19–27. https://doi.org/10.1007/s10827-009-0181-3.

Mesgarani, N., Shamma, S.A., 2005. Speech enhancement based on filtering the spectrotemporal modulations. In: IEEE Int. Conf. Acoust. Speech, Signal Process, pp. 1105–1108.

Meyer, A.F., Diepenbrock, J.P., Happel, M.F.K., Ohl, F.W., Anemüller, J., 2014. Discriminative learning of receptive fields from responses to non-gaussian stimulus ensembles. PLoS One 9. https://doi.org/10.1371/journal.pone.0093062.

Miller, L.M., Escabi, M.A., Read, H.L., Schreiner, C.E., 2001. Functional convergence of response properties in the auditory thalamocortical system. Neuron 32, 151–160.

Miller, L.M., Escabi, M.A., Read, H.L., Schreiner, C.E., Escabí, M.A., 2002. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex.

J. Neurophysiol. 87, 516–527.

Moerel, M., De Martino, F., Formisano, E., 2012. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J. Neurosci. 32 (41), 14205–14216.

Moore, A.K., Wehr, M., 2013. Parvalbumin-expressing inhibitory interneurons in auditory cortex are well-tuned for frequency. J. Neurosci. 33, 13713–13723. https://doi.org/10.1523/JNEUROSCI.0663-13.2013.

Moore, R.C., Lee, T., Theunissen, F.E., 2013. Noise-invariant neurons in the avian auditory cortex: hearing the song in noise. PLoS Comput. Biol. 9, e1002942 https://doi.org/10.1371/journal.pcbi.1002942.

Moran, J., Desimone, R., 1985. Selective attention gates visual processing in the extrastriate cortex. Science 229, 782–784 (80-.).

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. https://doi.org/10.1016/j.clinph.2007.04.026.

Nagel, K.I., Doupe, A.J., 2008. Organizing principles of spectro-temporal encoding in the avian primary auditory area field L. Neuron 58, 938–955.

Nagel, K.I., Wilson, R.I., 2011. Biophysical mechanisms underlying olfactory receptor neuron dynamics. Nat. Neurosci. 14, 208–216. https://doi.org/10.1038/nn.2725.

Natan, R.G., Briguglio, J.J., Mwilambwe-Tshilobo, L., Jones, S.I., Aizenberg, M., Goldberg, E.M., Geffen, M.N., 2015. Complementary control of sensory adaptation by two types of cortical interneurons. Elife 4. https://doi.org/10.7554/eLife.09868.

Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L., 2012. Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. J. Neurosci. 32, 9323–9334. https://doi.org/10.1523/JNEUROSCI.5832-11.2012.

Norena, A.J., Gourevitch, B., Pienkowski, M., Shaw, G., Eggermont, J.J., 2008. Increasing spectrotemporal sound density reveals an octave-based organization in cat primary auditory cortex. J. Neurosci. 28, 8885–8896.

Noreña, A.J., Tomita, M., Eggermont, J.J., 2003. Neural changes in cat auditory cortex after a transient pure-tone trauma. J. Neurophysiol. 90.

Nourski, K.V., Steinschneider, M., Rhone, A.E., 2016. Electrocorticographic activation within human auditory cortex during dialog-based language and cognitive testing. Front. Hum. Neurosci. 10, 202. https://doi.org/10.3389/fnhum.2016.00202.

Ohl, F.W., Scheich, H., 1997. Learning-induced dynamic receptive field changes in primary auditory cortex of the unanaesthetized Mongolian gerbil. J. Comp. Physiol. A. 181, 685–696.

Ohl, F.W., Scheich, H., Freeman, W.J., 2001. Change in pattern of ongoing cortical activity with auditory category learning. Nature 412, 733–736.

Okun, M., Steinmetz, N.A., Cossell, L., Iacaruso, M.F., Ko, H., Barthó, P., Moore, T., Hofer, S.B., Mrsic-Flogel, T.D., Carandini, M., Harris, K.D., 2015. Diverse coupling of neurons to populations in sensory cortex. Nature 521, 511–515. https://doi.org/10.1038/nature14273.

Osmanski, M.S., Song, X., Wang, X., 2013. The role of harmonic resolvability in pitch perception in a vocal nonhuman primate, the common marmoset (Callithrix jacchus). J. Neurosci. 33, 9161–9168. https://doi.org/10.1523/JNEUROSCI.0066-13.2013.

Otazu, G.H., Tai, L.H., Yang, Y., Zador, A.M., 2009. Engaging in an auditory task suppresses responses in auditory cortex. Nat. Neurosci. 12, 646–654.

Paninski, L., Pillow, J.W., Simoncelli, E.P., 2004. Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. Neural Comput. 16, 2533–2561. https://doi.org/10.1162/0899766042321797.

Pasley, B.N., David, S.V., Mesgarani, N., Flinker, A., Shamma, S.A., Crone, N.E., Knight, R.T., Chang, E.F., 2012. Reconstructing speech from human auditory cortex. PLoS Biol. 10, e1001251 https://doi.org/10.1371/journal.pbio.1001251.

Pérez-González, D., Malmierca, M.S., 2014. Adaptation in the auditory system: an overview. Front. Integr. Neurosci. 8, 19. https://doi.org/10.3389/fnint.2014.00019.

Phillips, E.A.K., Hasenstaub, A.R., 2016. Asymmetric effects of activating and inactivating cortical interneurons. Elife 5. https://doi.org/10.7554/eLife.18383.

Pienkowski, M., Shaw, G., Eggermont, J.J., 2009. Wiener-Volterra characterization of neurons in primary auditory cortex using poisson-distributed impulse train inputs. J. Neurophysiol. 101, 3031–3041. https://doi.org/10.1152/jn.91242.2008.

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., Simoncelli, E.P., 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454, 995–999. https://doi.org/10.1038/nature07140.

Polikov, V.S., Tresco, P.A., Reichert, W.M., 2005. Response of brain tissue to chronically implanted neural electrodes. J. Neurosci. Meth. 148, 1–18. https://doi.org/10.1016/j.jneumeth.2005.08.015.

Polley, D.B., Steinberg, E.E., Merzenich, M.M., 2006. Perceptual learning directs auditory cortical map reorganization through top-down influences. J. Neurosci. 26, 4970–4982.

Rabinowitz, N.C., Goris, R.L., Cohen, M., Simoncelli, E.P., 2015. Attention stabilizes the shared gain of V4 populations. Elife 4. https://doi.org/10.7554/eLife.08998.

Rabinowitz, N.C., Willmore, B.D.B., King, A.J., Schnupp, J.W.H., 2013. Constructing noise-invariant representations of sound in the auditory pathway. PLoS Biol. 11, e1001710 https://doi.org/10.1371/journal.pbio.1001710.

Rabinowitz, N.C., Willmore, B.D.B., Schnupp, J.W.H., King, A.J., 2012. Spectrotemporal contrast kernels for neurons in primary auditory cortex. J. Neurosci. 32, 11271–11284. https://doi.org/10.1523/JNEUROSCI.1715-12.2012.

Rabinowitz, N.C., Willmore, B.D.B., Schnupp, J.W.H., King, A.J., 2011. Contrast gain control in auditory cortex. Neuron 70, 1178–1191. https://doi.org/10.1016/j.neuron.2011.04.030.

Ramirez, A., Pnevmatikakis, E.A., Merel, J., Paninski, L., Miller, K.D., Bruno, R.M., 2014. Spatiotemporal receptive fields of barrel cortex revealed by reverse correlation of synaptic input. Nat. Neurosci. 17, 866–875. https://doi.org/10.1038/nn.3720.

Reiss, L.A.J., Bandyopadhyay, S., Young, E.D., 2007. Effects of stimulus spectral contrast on receptive fields of dorsal cochlear nucleus neurons. J. Neurophysiol. 98, 2133–2143. https://doi.org/10.1152/jn.01239.2006.

Robie, A.A., Hirokawa, J., Edwards, A.W., Umayam, L.A., Lee, A., Phillips, M.L., Card, G.M., Korff, W., Rubin, G.M., Simpson, J.H., Reiser, M.B., Branson, K., 2017. Mapping the neural substrates of behavior. Cell 170, 393–406. https://doi.org/10.1016/j.cell.2017.06.032 e28.

Rodgers, C.C., DeWeese, M.R., 2014. Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task for rodents. Neuron 82, 1157–1170. https://doi.org/10.1016/j.neuron.2014.04.031.

Rummell, B.P., Klee, J.L., Sigurdsson, X.T., Sigurdsson, T., 2016. Attenuation of responses to self-generated sounds in auditory cortical neurons. J. Neurosci. 36, 12010–12026. https://doi.org/10.1523/JNEUROSCI.1564-16.2016.

Runyan, C.A., Piasini, E., Panzeri, S., Harvey, C.D., 2017. Distinct timescales of population coding across cortex. Nature. https://doi.org/10.1038/nature23020.

Ryan, A.F., Miller, J.M., 1977. Effects of behavioral performance on single-unit firing patterns in inferior colliculus of the rhesus monkey. J. Neurophysiol. 40, 943–956.

Sahani, M., Linden, J.F., 2003. How linear are auditory cortical responses? In: Becker, S., Thrun, S., Obermayer, K. (Eds.), Advances in Neural Information Processing Systems. MIT Press, Cambridge, MA, pp. 301–308.

Salgado, H., García-Oscos, F., Dinh, L., Atzori, M., 2011. Dynamic modulation of short-term synaptic plasticity in the auditory cortex: the role of norepinephrine. Hear. Res. 271, 26–36. https://doi.org/10.1016/j.heares.2010.08.014.

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. J. Neurophysiol. 107, 3296–3307. https://doi.org/10.1152/jn.01173.2011.

Schneider, D.M., Nelson, A., Mooney, R., 2014. A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature 513, 189–194. https://doi.org/10.1038/nature13724.

Schneider, D.M., Woolley, S.M.N., 2011. Extra-classical tuning predicts stimulus-dependent receptive fields in auditory neurons. J. Neurosci. 31, 11867–11878. https://doi.org/10.1523/JNEUROSCI.5790-10.2011.

Schroeder, C.E., Foxe, J., 2005. Multisensory contributions to low-level, "unisensory" processing. Curr. Opin. Neurobiol. https://doi.org/10.1016/j.conb.2005.06.008.

Schwartz, Z.P., David, S.V., 2018. Focal suppression of distractor sounds by selective attention in auditory cortex. Cereb. Cortex 28 (1), 323–339.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.

Singh, N.C., Theunissen, F.E., 2003. Modulation spectra of natural sounds and ethological theories of auditory processing. J. Acoust. Soc. Am. 114, 3394–3411.

Slee, S.J., David, S.V., 2015. Rapid task-related plasticity of spectro-temporal receptive fields in the auditory midbrain. J. Neurosci. 35, 13090–13102. https://doi.org/10.1523/JNEUROSCI.1671-15.2015.

Steinschneider, M., Fishman, Y.I., Arezzo, J.C., 2008. Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. Cereb. Cortex 18, 610–625. https://doi.org/10.1093/cercor/bhm094.

Stevenson, I.H., London, B.M., Oby, E.R., Sachs, N.A., Reimer, J., Englitz, B., David, S.V., Shamma, S.A., Blanche, T.J., Mizuseki, K., Zandvakili, A., Hatsopoulos, N.G., Miller, L.E., Kording, K.P., 2012. Functional connectivity and tuning curves in populations of simultaneously recorded neurons. PLoS Comput. Biol. 8, e1002775 https://doi.org/10.1371/journal.pcbi.1002775.

Stringer, C., Pachitariu, M., Steinmetz, N.A., Okun, M., Bartho, P., Harris, K.D., Sahani, M., Lesica, N.A., 2016. Inhibitory Control of Correlated Intrinsic Variability in Cortical Networks. https://doi.org/10.7554/eLife.19695.001.

Theunissen, F.E., David, S.V., Singh, N.C., Hsu, A., Vinje, W.E., Gallant, J.L., 2001. Estimating spatial temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Netw. Comput. Neural Syst. 12, 289–316.

Theunissen, F.E., Sen, K., Doupe, A.J., 2000. Spectral-temporal receptive fields of non-linear auditory neurons obtained using natural sounds. J. Neurosci. 20, 2315–2331.

Thorson, I.L., Liénard, J., David, S.V., 2015. The essential complexity of auditory receptive fields. PLoS Comput. Biol. 11, e1004628 https://doi.org/10.1371/journal.pcbi.1004628.

Treue, S., Martinez-Trujillo, J.C., 1999. Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399, 575–578.

Treves, A., Panzeri, S., 1995. The upward bias in measures of information derived from limited data samples. Neural Comput. 7, 399–407.

Ulanovsky, N., Las, L., Farkas, D., Nelken, I., 2004. Multiple time scales of adaptation in auditory cortex neurons. J. Neurosci. 24, 10440–10453.

Ulanovsky, N., Las, L., Nelken, I., 2003. Processing of low-probability sounds by cortical neurons. Nat. Neurosci. 6, 391–398.

Varnet, L., Knoblauch, K., Meunier, F., Hoen, M., 2013. Using auditory classification images for the identification of fine acoustic cues used in speech perception. Front. Hum. Neurosci. 7, 865. https://doi.org/10.3389/fnhum.2013.00865.

Vogelstein, J.T., Packer, A.M., Machado, T.A., Sippy, T., Babadi, B., Yuste, R., Paninski, L., 2010. Fast nonnegative deconvolution for spike train inference from population calcium imaging. J. Neurophysiol. 104.

Wallace, M.T., Meredith, M.A., Stein, B.E., 1992. Integration of multiple sensory

modalities in cat cortex. Exp. Brain Res. 91, 484–488. https://doi.org/10.1007/BF00227844.

Watkins, P.V., Barbour, D.L., 2008. Specialized neuronal adaptation for preserving input sensitivity. Nat. Neurosci. 11, 1259–1261. https://doi.org/10.1038/nn.2201.

Williamson, R.S., Ahrens, M.B., Linden, J.F., Sahani, M., 2016. Input-specific gain modulation by local sensory context shapes cortical and thalamic responses to complex sounds. Neuron 91, 467–481. https://doi.org/10.1016/j.neuron.2016.05.041.

Willmore, B.D.B., Schoppe, O., King, A.J., Schnupp, J.W.H., Harper, N.S., 2016. Incorporating midbrain adaptation to mean sound level improves models of auditory cortical processing. J. Neurosci. 36, 280–289. https://doi.org/10.1523/JNEUROSCI.2441-15.2016.

Winkowski, D.E., Bandyopadhyay, S., Shamma, S.A., Kanold, P.O., 2013. Frontal cortex activation causes rapid plasticity of auditory cortical processing. J. Neurosci. 33, 18134–18148. https://doi.org/10.1523/JNEUROSCI.0180-13.2013.

Woolley, S.M., Fremouw, T.E., Hsu, A., Theunissen, F.E., 2005. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat. Neurosci. 8, 1371–1379.

Wu, M.C.-K., David, S.V., Gallant, J.L., 2006. Complete functional characterization of sensory neurons by system identification. Annu. Rev. Neurosci. 29, 477–505.

Yang, X., Wang, K., Shamma, S.A., 1992. Auditory representations of acoustic signals. IEEE Trans. Inf. Theor. 38, 824–839.

Zilany, M.S.A., Carney, L.H., 2010. Power-law dynamics in an auditory-nerve model can account for neural adaptation to sound-level statistics. J. Neurosci. 30, 10380–10390. https://doi.org/10.1523/JNEUROSCI.0647-10.2010.