Temporal variability of spectro-temporal receptive fields in the anesthetized auditory cortex

Arne F. Meyer[1]*, Jan-Philipp Diepenbrock[2], Frank W. Ohl[2,3] and Jörn Anemüller[1]

[1] Medizinische Physik and Cluster of Excellence Hearing4all, Department of Medical Physics and Acoustics, Carl von Ossietzky University, Oldenburg, Germany
[2] Department of Systems Physiology of Learning, Leibniz Institute for Neurobiology, Magdeburg, Germany
[3] Department of Neuroprosthetics, Institute of Biology, Otto-von-Guericke University, Magdeburg, Germany

Temporal variability of neuronal response characteristics during sensory stimulation is a ubiquitous phenomenon that may reflect processes such as stimulus-driven adaptation, top-down modulation or spontaneous fluctuations. It poses a challenge to functional characterization methods such as the receptive field, since these often assume stationarity. We propose a novel method for estimation of sensory neurons' receptive fields that extends the classic static linear receptive field model to the time-varying case. Here, the long-term estimate of the static receptive field serves as the mean of a probabilistic prior distribution from which the short-term temporally localized receptive field may deviate stochastically with time-varying standard deviation. The derived corresponding generalized linear model permits robust characterization of temporal variability in receptive field structure also for highly non-Gaussian stimulus ensembles. We computed and analyzed short-term auditory spectro-temporal receptive field (STRF) estimates with characteristic temporal resolution 5–30 s based on model simulations and responses from in total 60 single-unit recordings in anesthetized Mongolian gerbil auditory midbrain and cortex. Stimulation was performed with short (100 ms) overlapping frequency-modulated tones. Results demonstrate identification of time-varying STRFs, with obtained predictive model likelihoods exceeding those from baseline static STRF estimation. Quantitative characterization of STRF variability reveals a higher degree thereof in auditory cortex compared to midbrain. Cluster analysis indicates that significant deviations from the long-term static STRF are brief, but reliably estimated. We hypothesize that the observed variability more likely reflects spontaneous or state-dependent internal fluctuations that interact with stimulus-induced processing, rather than experimental or stimulus design.

Keywords: sensory coding, receptive field, inferior colliculus, auditory cortex, time-varying, zero-mean prior, generalized linear model

1. INTRODUCTION

Neurons in the auditory system must encode information about sensory stimuli exhibiting a wide range of statistical properties. Recent evidence suggests complex time-varying encoding mechanisms that facilitate this task, including adaptation to stimulus statistics in terms of regularity (Ulanovsky et al., 2003, 2004; Malmierca et al., 2009; Netser et al., 2011), (spectro-temporal) contrast (Escabi et al., 2003; Rabinowitz et al., 2011, 2012), sound level (Dean et al., 2005; Lesica and Grothe, 2008a), task-related plasticity (Fritz et al., 2003), learning-induced plasticity (Weinberger, 1993; Ohl and Scheich, 2005), or attentive modulation of the response (Ding and Simon, 2012; Mesgarani and Chang, 2012). These processes have been observed on time scales ranging from milliseconds to hours. An understanding of the underlying mechanisms is essential and requires techniques that allow robust characterization of neural coding, in particular on short time intervals.

The receptive field (RF) constitutes the classic, functional model relating sensory stimulus and evoked response of a neuron, for reviews see (Schwartz et al., 2006; Wu et al., 2006; Sharpee, 2013). However, the RF is primarily a function of bottom-up processing assuming the response to be unmodulated in the considered time interval, corresponding to estimation of a linear time-invariant system. Temporal changes in RF structure have been analyzed by producing RF estimates for different parts of an experiment (Fritz et al., 2003; Sharpee et al., 2006), or by recurrent linear filtering (Brown et al., 2001; Stanley, 2002). Techniques that exploit sparsity of RFs in a linear model have been developed to achieve distinct RF estimates with a comparably small amounts of data (Sahani and Linden, 2002; Machens et al., 2004; Park and Pillow, 2011). The resulting estimators allow the investigation of processing on smaller time scales than standard estimators. However, the representation of non-Gaussian stimuli in cortical areas mediated by non-linear transformations may often preclude application of these methods (David et al., 2004; Sharpee et al., 2006; Christianson et al., 2008; Meyer et al., 2014).

Here, we propose a novel approach to studying time-varying encoding mechanisms of sensory neurons. The focus is on

neurons in the auditory system, which are commonly described by the spectro-temporal receptive field (STRF), the most probable spectro-temporal stimulus pattern that generated a measured spike response (Aertsen and Johannesma, 1981). Changes in STRF structure, as compared to it's static component, have been found to be small in studies of sound level encoding (Lesica and Grothe, 2008a), spectro-temporal contrast (Escabi et al., 2003; Rabinowitz et al., 2011), naturalistic stimulus processing in background noise (Lesica and Grothe, 2008b) and task-related plasticity in auditory cortex (Fritz et al., 2003). The time-varying "local" STRF, i.e., the STRF pattern best characterizing neuronal encoding within a short time-interval of experimental data, may therefore be represented as a superposition of it's long-term average and a time-dependent deviation. The former, denoted here as "static" STRF, is computed from the whole ensemble of available data, using methods that allow robust RF estimation for stimulus ensembles with second- and even higher-order correlations across stimulus dimensions (Paninski, 2004; Sharpee et al., 2004; Truccolo et al., 2005; David et al., 2007; Calabrese et al., 2011; Fitzgerald et al., 2011; Meyer et al., 2014). The superposition Ansatz, thus, reduces the problem of identifying the "local" STRF to the problem of identifying temporally-localized deviations from the long-term "static" STRF and the formulation of a suitable statistical model.

A common technique in RF estimation is to bias RF estimates toward solutions that are more probable a priori (Sahani and Linden, 2002; David et al., 2007; Park and Pillow, 2011). Sparsity is a special form of prior that biases RF parameters toward zero, e.g., by assuming a zero-mean Gaussian distribution of RF parameters (Machens et al., 2004). We adopt this technique and formulate a prior that biases local STRF estimates toward an a priori more probable STRF, specifically the static STRF estimate obtained from a long stimulus sequence as described above. For a given neuron, the static STRF prior is likely more informative than, e.g., a sparseness prior since RFs show comparably small variation between different stimulus conditions (e.g., Fritz et al., 2003; Sharpee et al., 2006). This prior is denoted "adaptive prior" because it allows the static RF estimate to adapt to temporally localized structures in the data. The formulation in terms of a prior allows us to incorporate the approach into linear as well as non-linear models, e.g., the generalized linear model [GLM, Nelder and Wedderburn (1972)], which allows RF estimation under more general conditions than linear models (Paninski, 2004; Truccolo et al., 2005; Pillow et al., 2008).

We derive closed-form solutions for the linear case and show that when the stimulus ensemble contains strong second-order correlations, a combination of zero-mean and adaptive prior, denote "mixed prior," may improve local RF estimates significantly. Further, we combine adaptive and mixed priors with a GLM. The resulting estimator allows conservative but robust characterization of time-varying neural processing, even for highly non-Gaussian stimuli like human speech that may lead to highly biased STRF estimates for linear estimators. We also apply the proposed approach to recordings from auditory midbrain and auditory cortex in anesthetized Mongolian gerbils. We address the following questions. First, do neurons at different levels in the auditory system perform time-invariant integration of

stimulus features when probed with a complex, dynamic stimulus ensemble? Second, if so, does a time-varying model of sensory feature integration provide a better description of neural processes than a static model? Finally, if response characteristics of neurons are time-dependent, do changes occur deterministically or largely at random? Finding answers to these questions may provide more insight into time-varying encoding of biologically relevant information.

2. MATERIALS AND METHODS

2.1. NEURAL CODING MODEL

Characterizing a neuron's response to sensory stimuli involves presenting stimuli from a fluctuating stimulus ensemble and recording the evoked response. To simplify matters, we assume that the response has already been discretized in time bins of duration Δ, and $r_i \in \{0, 1\}$ indicates whether or not a spike has been observed in the ith time bin. The history of stimulus features preceding the response in the ith time bin, e.g., intensity values of an image patch or the spectro-temporal density of a sound, is summarized in the vector $\mathbf{s}_i$.

In a simplified model, we may assume that a neuron integrates stimulus features in terms of a linear filter $\mathbf{k}$, the receptive field (RF), and the filtered stimulus, $z_i \equiv \mathbf{s}_i^{\text{T}} \mathbf{k}$, is transformed into a neural response using a static, memoryless non-linearity f. Such a cascade is known as linear-nonlinear (NL) model (Hunter and Korenberg, 1986; Chichilnisky, 2001), and the probability of observing a spike in time bin i is given by

$$p(\text{spike}|\mathbf{s}_i) \propto f(z_i). \tag{1}$$

A binary spiking response may be obtained by an inhomogeneous Poisson process that is modulated by $p(\text{spike}|\mathbf{s})$ (Chichilnisky, 2001). **Figure 1A** illustrates the cascade model for the STRF.

Neural processing, in particular in cortical areas, has been found to be adaptive, and a static RF may not be sufficient to account for neural responses (Brown et al., 2001; Stanley, 2002; Escabi et al., 2003; Sharpee et al., 2006; Rabinowitz et al., 2011; Mease et al., 2013). A common approach is to estimate the parameters of the model Equation (1) for different parts of the data, assuming a short-term time-invariance approximation of the system (Fritz et al., 2003; Sharpee et al., 2006). Low temporal resolution is associated with the approach since a sufficient amount of data is needed to estimate the RF, thus adaptation that occurs on a faster timescale may not be identified reliably. Here, we propose an alternative approach that allows the linear RF in Equation (1) to adapt to temporally localized structures in the data according to the additive model

$$\mathbf{k}_t = \mathbf{k} + \Delta\mathbf{k}_t, \tag{2}$$

where a temporally localized deviation $\Delta\mathbf{k}_t$ is superimposed on the long-term static RF estimate $\mathbf{k}$. We note that t indicates a time interval (in contrast to a single observation) and $\mathbf{k}_t$ is the "local receptive field" (local RF) that is assumed static during this (brief) time interval. Thus, each local RF $\mathbf{k}_t$ characterizes the response to a contiguous subset of the stimulus-response ensemble denoted by $\{r_i, \mathbf{s}_i : i \in N_t\}$.

FIGURE 1 | Estimation of time-varying spectro-temporal receptive field (STRF) parameters in the linear-nonlinear model. (A) In the linear-nonlinear model, it is assumed that stimulus examples, represented here by spectro-temporal patches sampled from a speech spectrogram, are filtered by a linear filter $\mathbf{k}$, the STRF. The output of the linear stage is transformed by a static non-linearity into a spike rate. In the generative model, a binary spike response is obtained by a subsequent inhomogeneous Poisson process. **(B)** A typical approach to infer the parameters of the model is the maximum a posteriori (MAP) estimate. The MAP allows to incorporate prior information, e.g., a prior that enforce sparseness by biasing STRF parameters toward zero by assuming a zero-mean Gaussian distribution of STRF parameters $\mathbf{k}$. In the scenario of time-dependent neural processing, the MAP estimate obtained using all data, which constitutes the most probable static solution, may in turn be used as informative prior for a "local" (or time-dependent) MAP estimate $\mathbf{k}_t$. Local evidence will result in deviations from the static STRF. **(C)** Illustration of the relation between static STRF $\mathbf{k}$ (black arrow) and local STRF $\mathbf{k}_t$ (gray arrow) in two stimulus dimensions s_i and s_{i+1}. In case the likelihood (reddish contour lines) systematically deviates from the static prior (filled gray contours and gray marginal distributions), the maximum of the MAP estimate will be shifted by $\Delta\mathbf{k}_t$ resulting in a time-varying local STRF. The relatively flat marginal distributions of the sparse (zero-mean) Gaussian prior (dashed lines) indicate that the sparse solution would be dominated by the likelihood, which may be difficult to estimate for small sample sizes. For visualization purposes, the contour lines show likelihood and static prior on a logarithmic scale.

2.2. PRIOR-BASED LEARNING OF TIME-VARYING RECEPTIVE FIELD PARAMETERS

Let us assume that we have estimated a static RF $\mathbf{k}$ for a neuron. Even in case neural processing is time-varying, $\mathbf{k}$ constitutes the most probable stimulus pattern to which the neuron is sensitive across the whole stimulus-response sequence. Thus, $\mathbf{k}$ may be considered the most likely a priori description of the neuron for every part of the data. However, local evidence in the data may result in systematic deviations from the static RF. If we assume that small fluctuations are more probable than large fluctuations, the relation between static RF and local RFs may be formulated as prior over local RF parameters. This is outlined in **Figure 1B**. Having no prior reason to favor either positive or negative deviations, we use an isotropic multivariate Gaussian prior centered around the static RF,

$$p(\mathbf{k}_t|\mathbf{k}, \sigma_t) = \mathcal{N}\left(\mathbf{k}_t|\mathbf{k}, \sigma_t^2 \mathbb{I}\right), \tag{3}$$

with standard deviation σ_t and identity matrix $\mathbb{I}$. The time index indicates that σ_t may vary from part to part.

Figure 1C illustrates the relation between local deviation $\Delta\mathbf{k}_t$ from the static RF and the prior on local RF parameters. If the static RF $\mathbf{k}$ is in agreement with the local data given by the likelihood, σ_t will be small and $\mathbf{k}_t$ will be very close to $\mathbf{k}$. Otherwise, the prior distribution will be rather wide, allowing stronger deviations from $\mathbf{k}$. The estimate may become less reliable due to the higher dispersion of the prior distribution. We will demonstrate that allowing local RF parameters to be either adaptive or zero-mean may increase robustness in such situations.

The probabilistic formulation in terms of a prior allows to apply the principle to a wide range of models that can be formulated in the maximum a posteriori (MAP) framework. This includes linear as well as non-linear models such as the GLM. For a general form of the likelihood function, $p(\mathbf{r}_t|\mathbf{S}_t, \mathbf{k}_t)$, where the matrix $\mathbf{S}_t$ contains all stimulus examples in part t and $\mathbf{r}_t$ the corresponding response values, the MAP estimate can be written as

$$\hat{\mathbf{k}}_t|\mathbf{k}, \sigma_t = \arg\max_{\mathbf{k}_t} p(\mathbf{r}_t|\mathbf{S}_t, \mathbf{k}_t)p(\mathbf{k}_t|\mathbf{k}, \sigma_t). \tag{4}$$

where $p(\mathbf{k}_t|\mathbf{k}, \sigma_t)$ is the prior on $\mathbf{k}_t$ and the MAP estimate is the mode of the posterior. Thus, we use a highly informative prior on the local RF to obtain a robust estimate of the maximum of the posterior. As data size increases, the likelihood overwhelms the prior and converges, similarly to the static estimator, toward the maximum likelihood estimate.

2.3. THE LINEAR-GAUSSIAN CASE

If the response of a neuron may be described by a linear function, the stimulus-response relation can be written as

$$r_i = \mathbf{s}_i^{\mathrm{T}}\mathbf{k} + \epsilon_i, \quad \epsilon_i \propto \mathcal{N}(0, \sigma^2), \tag{5}$$

where ϵ_i is a zero-mean Gaussian white noise (GWN) sample with standard deviation σ. Here, r is assumed to be a continuous variable, e.g., the average number of spikes for several stimulus repetitions. For a complete measurement with N stimulus-response pairs, the likelihood that the response $\mathbf{r} = [r_1, r_2, r_3, \ldots, r_N]$ is generated by $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \ldots, \mathbf{s}_N]^T$ and the RF $\mathbf{k}$ is given by

$$p(\mathbf{r}|\mathbf{S}, \mathbf{k}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{r} - \mathbf{Sk})^T (\mathbf{r} - \mathbf{Sk})\right) \quad (6)$$

with maximum likelihood (ML) estimate

$$\hat{\mathbf{k}} = \left(\mathbf{S}^T\mathbf{S}\right)^{-1} \left(\mathbf{S}^T\mathbf{r}\right). \quad (7)$$

The above estimator allows RF estimation from Gaussian stimulus ensembles. Correlations across different stimulus dimensions are "rotated out" using the inverse of the covariance matrix (Paninski, 2003).

2.3.1. The "zero-mean" Gaussian prior

Sparseness is a distinct property of RFs and exploiting sparseness may improve performance of an estimator significantly (Sahani and Linden, 2002; David et al., 2007; Calabrese et al., 2011; Park and Pillow, 2011). A simple form of sparseness assumes an isotropic multivariate Gaussian distribution centered at zero with standard deviation σ_a, $p(\mathbf{k}|\sigma_a) = \mathcal{N}\left(\mathbf{k}|\mathbf{0}, \sigma_a^2\mathbb{I}\right)$. The MAP estimate is given by

$$\hat{\mathbf{k}}_{\text{zero-mean}}|\alpha = \left(\mathbf{S}^T\mathbf{S} + \alpha\mathbb{I}\right)^{-1} \left(\mathbf{S}^T\mathbf{r}\right) \quad (8)$$

with $\alpha \equiv \frac{\sigma^2}{\sigma_\alpha^2}$ (Hoerl, 1962). The problem in Equation (8) is also known as ridge regression, regularized linear regression, and penalized least squares (Hastie et al., 2001). Regularization is controlled by the hyper parameter α. For $\alpha \to \infty$, Equation (8) effectively computes the cross-correlation between stimulus and response, the spike-triggered average [STA, deBoer and Kuyper (1968)]. The "naive" ML estimator in Equation (7) arises for $\alpha = 0$. Often, α is found by cross-validation (Machens et al., 2004).

2.3.2. The "adaptive" Gaussian prior

The time-varying RF model in Equation (2) assumes that the time-dependent RF $\mathbf{k}_t$ can be described by time-dependent deviations from the static RF $\mathbf{k}$. In a probabilistic model, these deviations from the static RF can be expressed in the form of a prior that uses the static RF as most probable solution. Thus, instead of a Gaussian prior distribution with zero-mean, we may use a Gaussian centered around the static RF $\mathbf{k}$ and use this as prior distribution in the MAP estimate in Equation (4).

In the linear-Gaussian model, the MAP estimate under the adaptive prior is given by

$$\hat{\mathbf{k}}_t|\mathbf{k}, \beta_t = \left(\mathbf{S}_t^T\mathbf{S}_t + \beta_t\mathbb{I}\right)^{-1} \left(\mathbf{S}_t^T\mathbf{r}_t + \beta_t\mathbf{k}\right) \quad (9)$$

with $\beta_t \equiv \frac{\sigma^2}{\sigma_{\beta_t}^2}$, and $\sigma_{\beta_t}^2$ is the variance of the Gaussian prior in Equation (3). The estimator includes the sum of the STA and a term that enforces adaptation to the static solution. The balance between both terms, which determines the amount of adaptation to $\mathbf{k}$, is controlled by the hyperparameter β_t. Again, β_t may be found by cross-validation on the training data.

2.3.3. Mixing both priors: the "mixed" Gaussian prior

In many situations, it is necessary to control the amount of regularization and adaptation independently. In particular in case the employed stimulus ensembles contains strong second-order correlations across stimulus dimensions. However, when the adaptive prior distribution has a high dispersion, i.e., a large standard deviation, the estimator in Equation (9) will approach the non-regularized "naive" ML estimator (Equation 7) instead of the regularized ridge solution (Equation 8). In many situations, the regularized solution provides a more robust characterization of response properties than the ML estimator (Sahani and Linden, 2002; Sharpee et al., 2006; Park and Pillow, 2011).

Assuming that both priors are independent, the "mixed" prior, given by the product of zero-mean and adaptive prior,

$$p(\mathbf{k}_t|\mathbf{k}, \sigma_{\alpha_t}, \sigma_{\beta_t}) \propto \mathcal{N}\left(\mathbf{k}_t|\mathbf{0}, \sigma_{\alpha_t}^2\mathbb{I}\right) \times \mathcal{N}\left(\mathbf{k}_t|\mathbf{k}, \sigma_{\beta_t}^2\mathbb{I}\right), \quad (10)$$

may improve local RF estimates in such situations. Noting that the product of two Gaussian distributions is also Gaussian, the resulting estimate for $\mathbf{k}_t$ can be written as (see Appendix for details)

$$\hat{\mathbf{k}}_t|\mathbf{k}, \lambda_{\alpha_t}, \lambda_{\beta_t} = \left(\mathbf{S}_t^T\mathbf{S}_t + (\lambda_{\alpha_t} + \lambda_{\beta_t})\mathbb{I}\right)^{-1} \left(\mathbf{S}_t^T\mathbf{r}_t + \lambda_{\beta_t}\mathbf{k}\right), \quad (11)$$

where $\lambda_{\alpha_t} \propto \frac{\sigma^2}{\sigma_{\alpha_t}^2}$ and $\lambda_{\beta_t} \propto \frac{\sigma^2}{\sigma_{\beta_t}^2}$. Thus, the MAP estimate depends on two hyper parameters, and the estimator may vary between zero-mean ($\lambda_{\beta_t} = 0$) and adaptive solution ($\lambda_{\alpha_t} = 0$).

2.3.4. Simulated example

To demonstrate the performance of the MAP estimators with the different priors, we simulated data in the linear-Gaussian model (cf. Equation 5). The linear filter of the model cell represents a temporal onset RF with symmetric positive and negative amplitudes. We probed the cell with a short Gaussian noise sequence ($N = 1000$ samples). The correlation length of the Gaussian noise what about the length of the temporal filter. Spikes were created from the RF-filtered stimulus by an inhomogeneous Poisson process. Suppose, we have also probed the neuron with a longer sequence and the hypothetical static RF estimate systematically deviates from the RF used to generate the data for the short sequence. In such a case, the "adaptive" and "mixed" priors may use the static RF estimate for the long sequence as the prior during inference of RF parameters for the short sequence.

Figure 2 shows RF estimates based on the different priors. The hyperparameters in the adaptive and mixed priors were found using cross-validation. The estimate obtained using the zero-mean Gaussian prior approximately recovers the true RF but is very fuzzy due to the small number of samples in the short

FIGURE 2 | RF estimation from simulated responses using the different priors in a linear-Gaussian model. A strongly correlated Gaussian noise sequence ($N = 1000$ samples) was filtered with a temporal linear RF (red solid line). Spikes were generated by from the filtered stimulus by an inhomogeneous Poisson process resulting in 124 spikes. **(A)** The RF estimate obtained using a sparse (zero-mean) Gaussian prior on RF parameters. The estimate is very noisy due to the small number of observations. **(B)** Suppose we have estimated a static RF for the same cell, and the RF shows a different scaling and a slightly different latency (gray line). The adaptive prior biases RF coefficients toward the static RF. The resulting RF estimate is very smooth and partially adapts to the rescaled amplitudes but does not account for temporal shift and asymmetric amplitude scaling. **(C)** The mixed prior uses an additional hyper parameter that allows to control the trade-off between zero-mean and adaptive priors. The resulting estimator allows to account for the temporal shift at the expense of a slight increase in noise. Numbers in the upper left corner indicate correlations between estimated and true RF.

sequence. The RF produced by the adaptive prior is very smooth and reveals strong adaptation to the static RF, which exhibits asymmetric amplitude scaling and a slight shift in latency. The mixed prior combines the benefits of both estimates, resulting in an estimate that closely resembles the structure of the true RF. Thus, incorporation of prior knowledge, even in case the true local RF systematically deviates from the static RF, may significantly increase the quality of the RF estimate obtained from small data samples.

2.4. THE NON-GAUSSIAN CASE: EXTENDING THE GENERALIZED LINEAR MODEL (GLM)

The GLM framework allows RF estimation under more general conditions than linear estimators (Nelder and Wedderburn, 1972). In particular, the GLM allows for different non-linearities, incorporation of spike-interactions in the form of a post-spike filter (Paninski, 2004; Truccolo et al., 2005; Calabrese et al., 2011), and correlations between different neurons (Pillow et al., 2008). Under a GLM, the likelihood for the response r is chosen from an exponential family of distributions, and the expected output is given by

$$\mathbb{E}[r|\mathbf{s}] = f(z), \qquad (12)$$

where f is an invertible non-linearity and $z = \mathbf{s}^T\mathbf{k} + b$ relates the input stimulus $\mathbf{s}$ to the input of the non-linearity through the RF $\mathbf{k}$ and an optional offset term b.

Here, we adopt the concept of the adaptive prior to a special case of GLM, the linear-nonlinear Bernoulli model (Truccolo et al., 2005; Gerwinn et al., 2010). In the Bernoulli GLM, it is assumed that the output is distributed according to a Bernoulli distribution with likelihood

$$\mathcal{L}(\{p_i, r_i\} : i \in N) = \prod_{i \in N} p_i^{r_i} (1 - p_i)^{1 - r_i} \qquad (13)$$

with $p_i \equiv p(x_i)$, and $r_i \in \{0, 1\}$ denotes the presence ($r_i = 1$) or absence ($r_i = 0$) of a spike, respectively. The canonical non-linearity corresponding to the Bernoulli assumption is given by the logistic function

$$f(z) = \frac{1}{1 + e^{-z}}. \qquad (14)$$

The log-likelihood for a local RF estimate at time t is given by

$$\log \mathcal{L}_t(\{r_i\} | \mathbf{k}_t, b_t; \{\mathbf{s}_i\} : i \in N_t) = \sum_{i \in N_t} \left[r_i z_i - \log \left(1 + e^{z_i} \right) \right] (15)$$

with $z_i \equiv \mathbf{s}_i^T\mathbf{k}_t + b_t$. In an analogy to the linear case, we may extend the GLM by a zero-mean, adaptive or mixed prior. For the mixed prior, the local MAP estimate is given by

$$\begin{aligned}
\mathbf{k}_t^* | \mathbf{k}, \alpha, \beta = \arg\max_{\mathbf{k}_t, b_t} \sum_{i \in N_t} \left[r_i z_i - \log \left(1 + e^{z_i} \right) \right] \\
- \frac{\alpha}{2} \mathbf{k}_t^T \mathbf{k}_t - \frac{\beta}{2} \left(\mathbf{k}_t - \mathbf{k} \right)^T \left(\mathbf{k}_t - \mathbf{k} \right) \qquad (16)
\end{aligned}$$

with regularization hyperparameters α and β. The Gaussian prior is not the conjugate prior to the Bernoulli distribution. Consequently, there is no closed-form solution to Equation (16). However, the penalized log-likelihood function is convex, and the parameters can be found using gradient descent. We used a trust region Newton conjugate gradient algorithm (Lin et al., 2008) to find a solution for a set of hyper parameters α and β. The values for α and β that maximize the Bernoulli log-likelihood in a 5-fold cross-validation (CV) scheme were used to estimate the final model parameters.

2.5. ELECTRO-PHYSIOLOGICAL RECORDINGS
2.5.1. Ethics statement
All experiments were conducted in accordance with the international National Institutes of Health Guidelines for Animals in

Research and with ethical standards for the care and use of animals in research defined by the German Law for the protection of experimental animals. Experiments were approved by the local ethics committee in the state Saxony-Anhalt, Germany.

2.5.2. Physiology experiment

Single-unit recordings were made in inferior colliculus (IC) and in the primary auditory cortex (A1) in ketamine-anesthetized Mongolian gerbils. Recordings were performed in an acoustically and electrically shielded recording chamber. Sounds were delivered by an amplifier to a calibrated Canton Plus XS.2 speaker. Single-unit recordings were made using a single tungsten electrode (3–4 MΩ) and digitized using a multichannel recording system (Multichannel Acquisition Processor; Plexon). Single spiking units were verified off-line (SpikeSorter; Plexon). Only units with stable waveform throughout the recording and spike count rate of at least two spikes per second were included in the analysis. This resulted in 30 units for IC and A1, respectively. For a detailed description of experimental procedures see Happel et al. (2010) and Meyer et al. (2014).

Stimuli consisted of frequency-modulated (FM) tone complexes (Meyer et al., 2013, 2014). The FM tones were arranged in consecutive 100 ms blocks, and each block contained four FM tones with randomly chosen starting and ending frequencies. At the beginning and the ending of each block a half cosine ramp of 5 ms was applied to prevent onset and offset artifacts. The frequencies were drawn from the interval 0.5 kHz to 16 kHz such that their distribution is flat on a logarithmic scale. Thus, the average spectrum is approximately constant for the auditory filter bank used for analysis (see below). The sampling frequency of the acoustic waveforms was 44.1 kHz. Each stimulus had a duration of 10 s and subsequent stimuli were interleaved with 100–1000 ms of silence. The total length of the stimulus sequence was between 300 s and 500 s for the IC recordings and between 180 s and 1000 s for the A1 recordings.

2.5.3. Data preprocessing

As a first step of the data analysis, all stimuli were transformed into the time-frequency domain using a gammatone filter bank with octave-like frequency resolution (Hohmann, 2002). The magnitude of the complex-valued filter bank output was resampled to $f_s = 400$ Hz for IC data and $f_s = 250$ Hz for A1 data, yielding a spectrogram representation of the acoustic stimulus. The resulting bin sizes used to quantize spike times were $\Delta t = 2.5$ ms for IC data and $\Delta t = 4$ ms for A1 data. To account for dynamic properties of the cochlea, the resampled filter bank outputs were compressed using a static logarithmic function (Gill et al., 2006).

Stimulus examples were created from the spectrogram by recasting spectro-temporal patches preceding the response in a specific time window as vectors $\mathbf{s}$. Thus, the vector $\mathbf{s}_i$ contained the part of the spectrogram that preceded the time step $i\Delta t$, $i = 1, 2, 3, \ldots$, up to T_{win}. We used $T_{\text{win}} = 50$ ms and $T_{\text{win}} = 100$ ms for IC and A1 data, respectively. The spike trains were aligned to the time bins of the filtered stimuli, i.e., a spike at time t_s was assigned to the ith time bin $[(i-1)\Delta t, i\Delta t)$ if $(i-1)\Delta t \leq t_s < i\Delta t$. We never observed more than one spike per time bin for

both IC and A1 recordings. Thus, the response can be considered essentially binary, and the Bernoulli GLM is an adequate candidate for modeling the data.

2.5.4. Correlation between STRFs

We quantified similarity of STRFs in terms of the correlation between the STRF estimates (Sharpee et al., 2004; Fitzgerald et al., 2011). For two D-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$, the correlation is given by

$$\text{cc} = \frac{\mathbf{u}^{\text{T}}\mathbf{v}}{\|\mathbf{u}\|_2 \, \|\mathbf{v}\|_2} \tag{17}$$

where $\|\mathbf{u}\|_2 = \left(\sum_i^D u_i^2\right)^{1/2}$ denotes the L_2-norm. The quantity cc describes the correlation between $\mathbf{u}$ and $\mathbf{v}$, with cc $= 1$ denoting perfect correlation and cc $= 0$ denoting uncorrelated (orthogonal) vectors. For simulated data, $\mathbf{u}$ and $\mathbf{v}$ may represent true and estimated RF, respectively. On neural recordings, we used this quantity to describe similarities between static and local STRF estimates.

2.5.5. Evaluation of prediction performance

We quantified prediction performance on neural data by keeping back a random set of test data and performing predictions on these data. The remaining data were used as training set. Predictions were done using static STRFs or time-varying STRFs (based on local STRF estimates). We used a cross-validation scheme similar to repeated random subsampling. For every data set, a random sample of 10% of the data was kept back as validation set and a static STRF was estimated using the 90% training data. We strove to have the same fraction of spikes examples in both the training and the validation set. Local STRFs were estimated by subdividing the whole data set into non-overlapping parts (20 s), removing the validation examples, and learning STRF parameters using the estimated static STRF in the mixed prior. Thus, the validation set has not been used for learning both static and local STRFs. This procedure has been repeated for different part offsets t_0, i.e., $t_0 = 0, 5, 10, 15$ s (cf. **Figure 4**).

Predictions on the validation sets were quantified using a modified version of the Bernoulli log-likelihood in Equation (15). In the setting of time-varying computations based on the STRF, the filtered stimulus may not depend on a fixed filter $\mathbf{k}$ but on some filter $\mathbf{k}_t$ that may change over time (cf. Equation 2). Thus, the time-dependent Bernoulli log-likelihood, $\log \mathcal{L}_{\mathcal{T}}$, is given by

$$\log \mathcal{L}_{\mathcal{T}}\left(\{\mathbf{r}_t\} \mid \{\mathbf{k}_t, b_t, \mathbf{S}_t\} : t \in \mathcal{T}\right) = \sum_{t \in \mathcal{T}} \sum_{i \in \tilde{N}_t} r_i z_i$$
$$- \log\left(1 + e^{z_i}\right), \tag{18}$$

where $z_i \equiv \mathbf{s}_i^{\text{T}}\mathbf{k}_t + b_t$, $\tilde{N}_t$ is the test set for each part, and $\mathcal{T}$ includes all parts. $\mathbf{k}_t$ and b_t are local STRF parameters valid for the respective stimulus-response examples. In case the local model $\mathbf{k}_t$ provides a better description of time-varying processing than the static model $\mathbf{k}$, predictions based on $\mathbf{k}_t$ should yield higher predictive power than predictions based on $\mathbf{k}$.

2.5.6. Clustering of time-varying STRFs

Local STRFs were clustered using a Gaussian mixture model (GMM) with diagonal covariance matrices. For each data set, all local STRF estimates (the bootstrap estimates derived using a random 90% subset in each part) were pooled and a multivariate mixture of Gaussians, where each Gaussian had the same dimensionality as the STRFs, was fit to the data. The parameters were found using the expectation-maximization (EM) algorithm (Dempster et al., 1977).

We varied the number of multivariate Gaussians from 1 to 20 and used the number that minimized the Bayesian information criterion (BIC). The BIC is defined as

$$\text{BIC} = -2\log(\mathcal{L}) + d\log(N),\qquad(19)$$

where $\mathcal{L}$ is the likelihood of the data given the model, d the number of parameters to be fitted, and N the total number of local STRFs (Murphy, 2012). Thus, BIC effectively penalizes overly complex models. The EM algorithm was started 10 times with different initial conditions, and the best solution was used for the analysis. We also tested non-diagonal covariance matrices which did not improve the results. Further, we found that using the Akaike information criterion (AIC) resulted in an artificially high number of clusters with highly correlated cluster centers.

3. RESULTS

3.1. SIMULATED DATA

The performance of different priors in the Bernoulli GLM is demonstrated with model simulations. We used a linear-nonlinear model cell with Gabor function-shaped linear STRF and introduced temporal variability by increasing the strength of inhibitory filter components over time (cf. **Figure 3A**), akin to, e.g., successive stimulus selectivity sharpening (Sadagopan and Wang, 2010). As an example for natural stimuli, we used 300 s of human speech from the TIMIT corpus (Garofolo et al., 1993). Speech utterances were transformed into the time-frequency domain using a gammatone filter bank with logarithmic frequency resolution between 500 Hz and 8000 Hz. The magnitude of the filter bank outputs was logarithmically compressed and subsampled to 250 Hz. After filtering speech spectrograms with the chosen time-varying linear STRF, a non-linear response was obtained by applying a static sigmoid function to the filtered stimuli. Spikes were generated by an inhomogeneous Poisson process on the transformed output.

Figure 3B shows the ground-truth time-varying STRF averaged across time and **Figure 3C** the static STRF estimate obtained from the entire stimulus-response sequence using a GLM with zero-mean Gaussian prior. The estimated static STRF closely resembles the mean true STRF yielding a correlation of 0.98. Local STRF estimates were obtained by subdividing the whole sequence into 10 non-overlapping parts, each of length 30 s, and estimating STRFs using zero-mean, adaptive, and mixed priors, respectively. The static estimate (cf. **Figure 3D**) has been used as a prior during inference of GLM parameter with the adaptive and mixed prior approaches.

Figure 3C shows the temporal evolution of the linear filter estimate as obtained with a Bernoulli GLM with zero-mean, adaptive,

and mixed priors. STRF estimates produced by the GLM with zero-mean prior tend to be very noisy due to the small number of samples and the highly non-Gaussian structure of speech. In contrast, both adaptive and mixed prior allowed a robust tracking of the true time-varying STRF for the different parts. Correlation between estimated STRF and mean true STRF for each part (cf. Equation 17) verify that the GLM with mixed prior performs slightly better than the GLM with adaptive prior. Note that STRF estimates based on adaptive and mixed prior are biased toward the static STRF. Thus, the proposed approach produces conservative estimates and may underestimate changes in the STRF in some situations.

Reliability of time-varying STRF estimates and, thus, our ability to track them, is determined jointly by variability of the underlying neuronal system and the part length of estimation windows used. To assess the impact of the latter on tracking the simulated system, we repeated the above experiment with part lengths between 5 s and 30 s and evaluated mean correlation with ground truth STRF and its standard deviation; **Figure 3E** summarizes results obtained with lengths 5 s, 15 s, 30 s. As Expected, standard deviation of STRF estimates tends to increase toward smaller part lengths, with a large increase observed near 10 s length. The results suggest that the proposed method allows time-varying characterization of the investigated system with time-constants as small as about 10 s.

The time-varying STRF estimates revealed by the GLM with adaptive prior (middle row in **Figure 3C**) and with mixed prior (bottom row in **Figure 3C**) yield similar correlation values with the true underlying time-varying STRF. However, the GLM with mixed prior reveals a slightly better performance for parts in which the true STRF tends to be more sparse. A summary comparing performance of adaptive and mixed prior GLMs for different sets of human speech stimuli and different non-linearities is shown in **Figure 3F**. Across all conditions ($N = 100$), the GLM with mixed prior reveals systematically higher correlation values with the true time-varying STRF ($p < 3 \cdot 10^{-13}$; paired Wilcoxon test). Thus, the additional hyperparameter in the mixed prior allows more flexible estimation of time-varying STRFs. The benefit of the mixed prior, however, is rather small and on huge data sets the adaptive prior, which has only one hyperparameter, may provide a good estimate of time-varying processing.

3.2. EXPERIMENTAL DATA

To investigate time-varying processing in auditory brain areas, we probed single units in the inferior colliculus (IC) and the primary auditory cortex (A1) in anesthetized Mongolian gerbils with a dynamic, broadband stimulus sequence, whose spectral extent encompassed the frequency response of each unit. The sequence was composed of consecutive 100 ms blocks of frequency-modulated (FM) tone complexes with randomly chosen starting and ending frequencies (for details see Materials and Methods). The spectrogram for a 1 s example segment is shown in **Figure 4A**. Every 10 s the sequence was interleaved with a period of silence (length between 100 ms and 1000 ms). Average spectra for 10 s stimuli (**Figure 4B**) demonstrate the fairly stationary long-term characteristic of the stimulus ensemble.

FIGURE 3 | Time-adaptive STRF estimation from simulated responses to human speech. (A) Model cell with time-varying STRF whose inhibitory components gradually increase over time. Shown are averaged linear filters for non-overlapping parts of length 30 s. Responses were generated by filtering 300 s of speech spectrograms with the time-varying STRF and applying a static non-linearity to the filtered stimuli. **(B)** The temporal average of the ground-truth time-varying STRF. **(C)** STRF estimates for the different parts of data obtained using a GLM with zero-mean (top), adaptive (middle), and mixed (bottom) prior. For the adaptive and the mixed prior, the static STRF in **(D)** has been used during inference of STRF parameters. The GLM with zero-mean prior produces highly unreliable STRF estimates for the different parts. Both adaptive and mixed prior GLMs allow robust tracking of the time-varying linear filter. Numbers in the lower left corner of each filter estimate indicate correlation between estimated local STRF and mean STRF for each part. The GLM with mixed prior performs marginally better in some cases with strong deviation of the time-varying from the static STRF. **(D)** The static STRF estimated from the whole stimulus-response sequence. **(E)** Dependence of mixed prior performance on part length. Shown are mean and standard deviation of correlations between estimated local and average true STRF across 10 trials. As part length decreases, local STRF estimates become more unreliable as indicated by the noticeable increase in standard deviation. Part lengths of 15 s and longer allowed robust tracking of time-varying processing. **(F)** Comparison of performance of a GLM with adaptive and mixed prior for different human speech stimulus ensembles and different non-linearities. Across all conditions ($N = 100$), the GLM with mixed prior reveals a better reconstruction of the true time-varying STRF (Wilcoxon rank test).

For each unit, a static STRF estimate was obtained using the whole stimulus sequence and a GLM with zero-mean prior (**Figure 4C**). The stimulus sequence was divided into overlapping parts of length $T_{part} = 20$ s and temporal shift $T_{shift} = 5$ s. Based on the partitioned data, local STRFs were estimated using the Bernoulli GLM with mixed prior (see Equation 16), with the static STRF estimate used as mean in the mixed GLM prior distribution. The principle is illustrated in **Figure 4D**.

To quantify fluctuations of local STRF estimates, we used correlation in terms of normalized subspace projection between static and local STRFs (cf. Equation 17). To obtain approximate confidence intervals we repeatedly estimated each local STRF using a random subset of 90% of the data. The results reported here represent mean and standard deviation across 10 repetitions. **Figure 5A** shows temporal evolutions of correlation with the static STRF and number of spikes in each part for three example IC units. The first two units reveal virtually no deviation from the static STRF, although the second unit exhibits adaptation as indicated by a decrease in spike count. The third unit is an example of an IC unit with rather

FIGURE 4 | Experimental design. Single units in the inferior colliculus (IC) and the primary auditory cortex (A1) in anesthetized Mongolian gerbils were probed with a dynamically fluctuating stimulus ensemble. Acoustic stimuli were composed of random frequency-modulated (FM) tone complexes arranged in 100 ms blocks. The stimulus sequence was interleaved every 10 s by a period of silence (100–1000 ms). **(A)** The spectrogram for a 1 s example. Amplitude values are given in decibel. **(B)** Average spectra for 10 s stimuli indicating stationary characteristics on medium and long time scales. Note the smaller range of amplitude values compared to **(A)**. **(C)** For every unit, an STRF was estimated using the whole stimulus sequence and a GLM with zero-mean prior. Silence periods were not included in the analysis. **(D)** A time-varying STRF was constructed by subdividing the whole stimulus sequence in overlapping parts of length T_{part} and temporal shift T_{shift}. For every part, a local STRF was estimated using a GLM with mixed prior (using the static STRF), yielding a sequence of time-varying STRFs.

strong fluctuations. Even in this case, the correlation does not fall below 0.8. The local STRF examples for the three units reveal hardly any changes in STRF structure (right column of **Figure 5A**).

Most A1 units generally showed local STRFs with a higher degree of variability over time than IC units, ranging from time-invariant (first example in **Figure 5B**) to highly fluctuating local STRFs (third example in **Figure 5B**). We found parts that show a high degree of variability (indicated by a high standard deviation) and parts that reveal only small variations across the bootstrap estimates. The example STRFs essentially revealed changes in the strength of excitatory STRF components whereas inhibitory regions remained approximately constant. Thus, balance between excitation and inhibition may play an important role in time-varying integration of stimulus features in the cortex. However, we did not find a consistent pattern across all units, and a detailed analysis would require devised experiments beyond the scope of this study.

3.2.1. Population analysis
Figure 6A summarizes mean correlation of time-varying with static STRF estimates and corresponding standard deviation for 30 IC units. For the majority of units, mean correlation was not noticeably different from one never below 0.95. In A1, mean correlations and standard deviations revealed much stronger fluctuations across the neural subpopulation (**Figure 6C**). A noticeable number of units yielded a mean correlation lower than 0.9 and exhibited a considerably larger standard deviation than in IC. We found that only 2% (4%) of the local STRFs pooled across all IC parts ($N = 2835$) have a correlation smaller than 0.9 (0.95) with the static STRF. In A1, about 20% (34%) of the local STRFs for all parts ($N = 2768$) have a correlation of less than 0.9 (0.95). For each part, we used the average correlation across the 10 bootstrapped local STRF estimates.

The variability of A1 responses in terms of the STRF gives rise to the question whether the observed fluctuations may also be explained by the number of spikes. Thus, we calculated the correlation coefficient (denoted by r in **Figure 5**) between

FIGURE 5 | Examples for time-dependent STRF analysis in auditory midbrain and cortex. For each part, we repeatedly estimated a local STRF using 90% of the data. To quantify the similarity between local and static STRFs, we used the correlation in terms of normalized subspace projection. Correlation values shown here are mean correlations across all 10 local bootstrap estimates. **(A)** Each row shows time-dependent correlations (denoted by cc) between local STRFs and static STRF (black line) and the corresponding number of spikes in the parts (blue line) for thee IC units. The gray shaded area indicates one standard deviation across the local bootstrap estimates. The arrows indicate the position of the local STRF example shown to the right, together with the static STRF. Numbers in the right lower corners indicate correlations between the cc trace and the number of spikes. **(B)** The same analysis for three A1 units. There is a considerable increase in fluctuations compared to IC examples. The local STRF examples primarily reveal changes in excitatory components at best frequency.

time-evolving correlation of local and static STRF and the number of spikes in each part. **Figures 6B,D** show the distribution of correlation coefficients for IC and A1 units, respectively. In both regions, variability in the STRF is largely uncorrelated with the number of spikes as indicated by the Gaussian-like distributions concentrated around zero (IC) or close to zero (A1). Consequently, a change in spike count does not imply a change in feature integration in terms of the STRF (and vice versa).

3.2.2. The time-varying STRF estimate yields higher predictive power than the static STRF

To verify that the time-varying STRF model provides a meaningful characterization of time-dependent neural processing, we compared prediction performance of local and static STRFs. If changes in STRF structure were exclusively induced by sensory stimuli, we could fit the model on a specific stimulus sequence and validate its predictive power on a repeated sequence.

FIGURE 6 | Population summary for 30 IC and 30 A1 units. (A) Mean and standard deviation of correlations between local STRFs with the static STRF for each IC data set. Data sets have been sorted by mean correlation. Shaded gray area indicates one standard deviation. **(B)** Distribution of correlations of

STRF fluctuations and the number of spikes in each part for 30 IC units. The distribution is approximately Gaussian centered at zero. **(C,D)** The same analysis as in **(A,B)** but for 30 A1 units. A1 units reveal a higher degree of variability compared to IC.

However, as we assume significant modulation of the response by top-down and random network processes (and potentially also other influences, e.g., the employed ketamine anesthesia), we need to adopt a different approach to quantify model performance.

Here we used a cross-validation scheme similar to repeated random subsampling (see Materials and Methods). Briefly, we split the whole stimulus-response sequence into a training (90%) and a validation (10%) set and compared predictions based on the static STRF (estimated using all training data) with predictions based on local STRF estimates (estimated for non-overlapping parts). The static STRF estimate has been used as a prior for local STRF estimates. Thus, the validation set has not been used for estimation of STRF parameters. We repeated the procedure for different randomly selected training/validation sets and different part shifts to cover all samples and all local STRFs.

The Bernoulli log-likelihood (BLL, Equation 15) is a natural measure to evaluate predictions on binary data. Thus, we adapted the BLL to the time-varying scenario (cf. Equation 18). To normalize the log-likelihood for each unit, we subtracted the log-likelihood calculated for predictions on the validation set based on the full STRF estimated using all data (including the validation set). Thus, a value greater than zero indicates that cross-validated local STRF estimates have higher predictive power than the full model that allows the best possible fit on the data by a time-invariant Bernoulli GLM.

Figure 7A shows cross-validated normalized BLL for IC recordings. For the majority of units (26 out of 30), local STRF estimates revealed higher predictive power than the static STRF. In particular, there is a number of units exhibiting relatively high STRF variability (a mean correlation of local STRFs with the static STRF of about 0.95) that showed a high increase in prediction performance. Note that correlation values refer to correlations

between local and static STRFs on the training data. Across all units the increase in predictive power is significant ($p < 1.5 \times 10^{-3}$; paired Wilcoxon test). Similar results were obtained for the 30 A1 units shown in **Figure 7B**. 28 out of 30 units yielded higher cross-validated normalized BLL for local STRFs than for the static STRF method. The increase in predictive power is highly significant ($p < 2 \times 10^{-5}$; paired Wilcoxon test). Thus, our results indicate that time-varying STRF estimates based on the GLM with mixed prior provide a robust characterization of time-varying neural integration mechanisms, in particular for cortical data.

To test how prediction performance depends on the part length used to derive local STRF estimates, we repeated the above analysis for different part lengths $T_{\text{part}} = 5 \, \text{s}, 10 \, \text{s}, 20 \, \text{s}, 30 \, \text{s}$ (cf. **Figure 4**). The results are summarized in **Figures 7C,D** for IC and A1 recordings, respectively. In IC, we found a considerable increase in prediction performance of the time-varying model compared to the static model for part lengths greater or equal to $20 \, \text{s}$. In the cortex, however, the increase in prediction performance is also evident for smaller part lengths. These findings are consistent with the results obtained using simulated data (cf. **Figure 3E**). This suggests that the proposed approach is able to capture time-varying processes in the cortex with a characteristic temporal resolution of $5–30 \, \text{s}$.

3.2.3. Characterization of fluctuations in time-varying STRFs

The increase in predictive power for local STRFs suggests that changes in STRF structure are correlated with changes in response properties. However, this does not quantify whether or not these variations are systematic, e.g., the local STRF may fluctuate randomly or between distinct states. To quantify changes in local STRFs, we clustered local bootstrap estimates for each unit using a Gaussian mixture model (GMM; see Materials and Methods). If local STRFs were fluctuating between distinct states, we should

FIGURE 7 | Predictive power of time-invariant and time-varying STRF model. We evaluated the predicted Bernoulli log-likelihood (BLL) of local STRFs and static STRF for each unit using repeated random subsampling (see text for details). For each unit, we normalized the log-likelihood by the log-likelihood of the full model learned using all data (including the test data; denoted by "full"). A value greater than zero indicates that even the best static Bernoulli GLM model may not account for potentially time-varying neural processing. **(A)** Normalized BLL for 30 IC units. For the majority of units, predictive power of local STRF estimates is higher than for the static STRF. Across all units, the observed increase in predictive power is significant (Wilcoxon rank test). The black solid line indicates the linear fit. Color encodes mean correlation between local STRFs and static STRF for each unit. Note that due to random sampling,

the correlation has been calculated for slightly different subsets than in **Figure 6**. **(B)** Normalized BLL for 30 A1 units. Compared to IC, there is a considerable increase in predictive power of local STRFs (Wilcoxon rank test). For some units, predicted normalized BLL is greater than zero, implying that a static model is not sufficient to explain the data. **(C)** Average increase in cross-validated log-likelihood of the time-varying over the static model as a function of the interval used to derive the local STRFs for IC data. Only units that contained at least five spikes in each interval were used in the analysis (here: $N = 26$). **(D)** The same as in **(C)** but for the 22 A1 units that contained at least five spikes in each time interval. The time-varying model provided a better description of cortical recordings than the static model, even on small time scales of 5–10 s. * indicates statistical significance (Wilcoxon rank test; $\alpha = 0.05$).

be able to identify a small number of GMM clusters, and these clusters should repeatedly occur over time. However, in case fluctuations were not governed by such processes, the number of clusters should be rather large, and clusters should not occur more than once.

Figure 8A shows the temporal evolution of local STRF clusters for the two latter cortical example units in **Figure 5B**. In general, we were able to identify distinct clusters that almost never occur multiple times. The temporal extent of many clusters is about the length of the time window of local data. This is likely a result of correlations induced by overlapping parts of the data used to derive local STRFs. However, in some units the temporal extent was clearly larger, and we also find clusters that occurred multiple times (**Figure 8B**). In IC, we were able to identify only a very small number of clusters, and the extent of the clusters was much smaller than in A1 (**Figure 8C**). Thus, in both regions we did not find evidence for gradual changes in the STRF or distinct repetitive neural states.

4. DISCUSSION

4.1. SUMMARY OF FINDINGS

The purpose of this study was to analyze time-varying neural processing in terms of the STRF. Therefore, we developed a probabilistic approach that characterizes time-varying computations by time-dependent deviations from the standard static STRF. The formulation of time-varying STRF inference in terms of a prior allowed us to apply the proposed approach to a range of

models, including linear models and the GLM. Using simulated responses to natural stimuli, we demonstrated that a Bernoulli GLM with mixed prior, a prior that can either be adaptive or zero-mean, provides a means to analyze time-dependent computations, even in case the stimulus ensemble has strong second- and even higher-order correlations across stimulus dimensions. Hence, the proposed approach allows the investigation of time-varying cortical responses to behaviorally relevant stimuli, e.g., con-specific vocalizations or human speech, on medium time scales (10–30 s).

We applied the concept of the time-varying STRF to recordings from IC and A1 in anesthetized Mongolian gerbils. A1 neurons showed a noticeably higher degree of fluctuations in the STRF compared to neurons in IC. In both regions, fluctuations could not be explained by the number of spikes, suggesting that spiking probability is only partially determined by linear feature integration. To test this, we also analyzed how other parameters in the Bernoulli GLM, namely sigmoid non-linearity and spiking threshold, change in situations in which the linear time-varying STRF was highly correlated with the static STRF (data not shown). We found that a change in spike count was mostly correlated with a shift in threshold. In a small subset of units, however, spiking was also affected by a scaling of the non-linearity. Note that the STRF describes effective neuronal processing and its parameters may not directly correspond to a specific neural implementation. Nevertheless, by investigating how the parameters of the linear-nonlinear model change over time, as we have

FIGURE 8 | Clustering of time-varying STRFs. (A) Temporal evolution of clusters of bootstrapped local STRFs for the latter two A1 example units in **Figure 5B**. The majority of clusters appeared only once during the whole stimulus sequence. The temporal extent of the clusters indicates that consecutive local STRFs describe very similar states. On a global scale, however, fluctuations appear at random and are not systematic. Clusters have been sorted by total size in descending order. Color encodes the number of bootstrap STRF estimates in the different clusters per time step (white: 0, black: 10). **(B)** Summary of cluster statistics across all clusters for 30 A1 units. Each histogram relates the number of occurrence of clusters to the total temporal extent of the clusters. For each unit, the total time the time-varying STRF was in a specific state (cluster) was estimated by counting all time steps in which more than half of the bootstrap estimates were assigned to the corresponding cluster. Note that we did not include clusters with the static STRF into the analysis (cluster 1 in both examples in **A**). **(C)** Same analysis as in b for 30 IC units. In IC, only a few clusters were identifiable, and virtually all clusters did not occur more than once per stimulus sequence. Thus, in both regions a time-varying STRF analysis does not reveal repeated patterns or long-term changes in STRF structure.

done for the linear part, we may obtain insight into the effective mechanisms underlying these changes.

The time-varying STRF model provided significantly better prediction of IC and A1 neuron responses compared to the time-invariant STRF model. The proposed approach provides a simple and powerful way of extending existing response models for auditory neurons, thus capturing time-dependent coding of these neurons. For a subset of units in both IC and A1, we found that predictions on the validation set based on the time-varying model were better than predictions by a time-invariant model that has also been trained on the validation set. Consequently, for these units a static model seems to be insufficient to account for the neural response. This finding underlines both validity of the proposed approach and the importance of time-varying models to studying neural processing.

A cluster analysis of time-variant STRFs revealed that fluctuations are consistent, i.e., local STRFs estimated in consecutive, overlapping time frames were assigned to the same cluster. Note that the employed GMM did not use temporal information during clustering. As a result, any temporally localized clusters were exclusively induced by structures in the data. These clusters, however, did not show a repeating pattern. In most cases, clusters did not occur multiple times. Consequently, fluctuations were not completely random but there is no evidence for systematic or gradual changes on longer time scales. We would like to emphasize that the employed stimulus ensemble is stationary on longer

time scales (cf. **Figure 4B**) and that experimental design did not induce behavioral relevance to any stimulus part. In case the stimulus ensemble includes relevant context information and strongly varies over time, we would expect both gradual or systematic changes and further evidence of time-dependence in the STRF.

4.2. POSSIBLE ORIGINS OF FLUCTUATIONS IN THE STRF
The origins of the observed fluctuations may be manifold, and thus we can only speculate. In both IC and A1, neurons were probed using the same stimulus ensemble, and experimental conditions and protocol were identical. Consequently, the observed increase in fluctuations in cortical compared to midbrain neurons is likely a result of characteristic differences between both regions.

Information-theoretic STRF analyses suggest that single cortical neurons may encode multiple features corresponding to multiple STRFs (Atencio et al., 2008, 2012). A possible question in relation to the above findings is whether neurons encode these features simultaneously or whether they "switch" between features at different time instants. We identified distinct local STRF clusters, but these clusters did not frequently occur over time. In most cases, clusters occurred only once. This suggests that the observed fluctuations represent a single fluctuating feature rather than an additional acoustic feature.

Recent studies provide evidence that A1 neurons may respond to features that are weakly correlated with the spectro-temporal amplitude pattern (Chechik et al., 2006; Chechik and Nelken,

2012). For the employed fluctuating stimulus ensemble, such correlations would presumably vary across time as observed for cortical neurons. Another potential source is an increase in context sensitivity through adaptive bottom-up processing. The amplitude of the FM tones was identical throughout the whole stimulus sequence, but the local spectro-temporal contrast varied on a time scale up to seconds (cf. **Figure 4A**). Such changes may already be sufficient to evoke a change in response properties (Rabinowitz et al., 2011, 2012). Although changes in spectro-temporal contrast have been shown to be mainly compensated for by the neural non-linearity, there may be an interaction with other time-varying processes. Further, the fixed block structure of the stimulus ensemble may evoke adaptation to stimulus statistics in terms of regularity (Ulanovsky et al., 2003, 2004), and this type of adaptation has been found to be more pronounced in the cortex than in IC (Malmierca et al., 2009). The static STRF, a simplified model that primarily assumes time-invariant bottom-up processing, may not be an adequate description of neural processing in such situations.

Response variabilities across neurons are strongly influenced by network fluctuations that operate over a range of spatial and temporal scales, extending in some cases across cortical areas (Zohary et al., 1994; Smith and Kohn, 2008). They may arise from stochastic internal sources, while complex deterministic processes have been suggested to provide a major contribution (Beck et al., 2012; Toups et al., 2012) that may be discernible from random noise (Balaguer-Ballester et al., 2014). These fluctuations are sensitive to global or local internal states (for a review, see Harris and Thiele, 2011), e.g., they are more pronounced and may also change during anesthesia (Ecker et al., 2014) or in response to behavioral salience induced by experimental design (Abolafia et al., 2013). To test whether such effects may contribute to STRF feature integration, we applied the proposed approach to recordings from a multichannel shaft electrode for FM tone complex stimuli in two preparations (data not shown; for a description of the electrode see Happel et al., 2010). If fluctuations in STRFs resulted from global states, there should be non-zero correlations between adjacent electrode channels. We found small correlations of fluctuations between the different channels in the range −0.05 to 0.34. Zohary et al. (1994) observed comparable correlations in sensory cortex for spike count data on smaller time scales (approximately 1–2 s). A theoretical analysis demonstrated that even small correlations may have a distinct impact on global brain performance. Conversely, the observed fluctuations may be a result of global phenomena, e.g., top-down modulation, the employed ketamine anesthesia, or variability in local microcircuits as demonstrated in visual areas (Hansen et al., 2012). The fact that noise correlations are virtually absent in IC (Garcia-Lazaro et al., 2013) supports this hypothesis.

4.3. LIMITATIONS, EXTENSIONS, AND POSSIBLE APPLICATIONS
In general, the proposed approach may be applied to any scenario in which sensory feature integration in neurons varies over time under the above constraints. Throughout different stimulus ensembles and different non-linearities, we found 10 s to be a robust lower limit for the proposed approach in the auditory system. Additional parameters, such as data dimensionality, strength

of correlations between different stimulus dimensions and the number of spikes in each part, may influence characteristic time scales. In particular, for continuous response signals, e.g., local field potentials within cortex (Arieli et al., 1996) or from the cortical surface (Mesgarani and Chang, 2012), the proposed approach may provide a robust characterization of time-varying processes on smaller time scales than for spiking neurons.

In the context of time-varying neural processing, state-space methods have been of special importance, in particular to study time-varying properties in the hippocampus (Brown et al., 1998, 2001). Due to their recurrent nature, state-space approaches do not require a division of the data into discrete time intervals (Stanley, 2002; Eden et al., 2004), and it is also possible to embed Markovian dynamics into the GLM (Paninski et al., 2010). However, finding solutions the resulting problem usually involves approximations to the posterior and, for small data sets, these may be numerically unstable (Paninski et al., 2010). The proposed approach does not involve any approximations and is very straightforward to implement using standard gradient descent techniques. Further, the formulation of the parameter learning problem in terms of a prior allows to use the approach in a large range of models, including the GLM, but also non-probabilistic approaches, e.g., (Meyer et al., 2014). Thus, in case time-varying processing appears on time scales of about 10 s, the proposed approach represents an alternative to state-space models, in particular for highly non-Gaussian stimulus ensembles that require careful regularization of RF parameters.

The adaptive prior assumes that changes in the STRF are rather small. This is a realistic assumption, e.g., coarse shape and best frequency of a neuron usually remain preserved, even under highly diverse conditions (Escabi et al., 2003; Fritz et al., 2003; Lesica and Grothe, 2008a,b; Rabinowitz et al., 2011; Schumacher et al., 2011). However, when deviations from the static STRF are strong, the adaptive prior will assume the unregularized ML solution. The mixed prior also includes a regularization term enforcing sparseness, thus allowing for more robust estimates in such situations. There are more sophisticated forms of priors, e.g., Laplace prior (Gerwinn et al., 2010; Calabrese et al., 2011), smoothness priors (Sahani and Linden, 2002; Machens et al., 2004), or priors enforcing highly localized RFs (Park and Pillow, 2011), that could be used to increase robustness of local STRF estimates at the expense of more elaborate and computationally expensive optimization schemes.

The hyper parameters of the Bernoulli GLM for both static and local STRF models were found using cross-validation. The parts of the data from which local STRF parameters have been estimated were rather small and even for two hyper parameters this could be done efficiently. However, a computationally efficient alternative to finding hyper parameters of a probabilistic model is Bayesian inference (see Murphy, 2012 for an overview). If neural processing can be described by a linear model, approaches based on empirical Bayes estimation may allow to find multiple hyper parameters more efficiently (Park and Pillow, 2011). In Gerwinn et al. (2010), the posterior distribution over the model parameters of a Poisson GLM has been approximated by a Gaussian using the Expectation Propagation algorithm. It has been demonstrated that the mean of the posterior distribution may have advantages

over the MAP estimate in some situations, in particular for the Laplace prior. Bayesian inference also enables the calculation of Bayesian confidence intervals that characterize the uncertainty about the optimal solution. Future research may include Bayesian inference of time-varying RF parameters.

Simultaneous recordings from several layers in A1 revealed that spectro-temporal separability, temporal precision, and feature selectivity varied with layer (Atencio et al., 2009). The proposed approach may provide complementary time-dependent information about coding strategies in different cortical layers. In particular, STRF stationarity may be a layer-dependent property as found in the visual cortex (Hansen et al., 2012). Furthermore, time-dependent analysis of multiple simultaneously recorded neurons in terms of sensory feature integration could help to understand cortical abstraction from spectro-temporal features to auditory identities (Nelken et al., 1999, 2003; Chechik and Nelken, 2012), and temporal dynamics underlying selective attention (Ding and Simon, 2012; Mesgarani and Chang, 2012).

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: Arne F. Meyer; Jan-Philipp Diepenbrock; Frank W. Ohl; Jörn Anemüller. Collected the data: Jan-Philipp Diepenbrock. Developed and implemented the methods: Arne F. Meyer. Analyzed the data: Arne F. Meyer. Wrote the paper: Arne F. Meyer. Revised the paper: Arne F. Meyer; Jan-Philipp Diepenbrock; Frank W. Ohl; Jörn Anemüller.

ACKNOWLEDGMENT

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fncom.2014.00165/abstract

REFERENCES

Abolafia, J. M., Martinez-Garcia, M., Deco, G., and Sanchez-Vives, M. V. (2013). Variability and information content in auditory cortex spike trains during an interval-discrimination task. *J. Neurophysiol.* 110, 2163–2174. doi: 10.1152/jn.00381.2013

Aertsen, A., and Johannesma, P. I. (1981). The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol. Cybern.* 42, 133–143. doi: 10.1007/BF00336731

Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* 273, 1868–1871. doi: 10.1126/science.273.5283.1868

Atencio, C. A., Sharpee, T. O., and Schreiner, C. E. (2008). Cooperative nonlinearities in auditory cortical neurons. *Neuron* 58, 956–966. doi: 10.1016/j.neuron.2008.04.026

Atencio, C. A., Sharpee, T. O., and Schreiner, C. E. (2009). Hierarchical computation in the canonical auditory cortical circuit. *Proc. Natl. Acad. Sci. U.S.A* 106, 21894–21899. doi: 10.1073/pnas.0908383106

Atencio, C. A., Sharpee, T. O., and Schreiner, C. E. (2012). Receptive field dimensionality increases from the auditory midbrain to cortex. *J. Neurophysiol.* 107, 2594–2603. doi: 10.1152/jn.01025.2011

Balaguer-Ballester, E., Tabas-Diaz, A., and Budka, M. (2014). Can we identify non-stationary dynamics of trial-to-trial variability? *PLoS ONE* 9:e95648. doi: 10.1371/journal.pone.0095648

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., and Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron* 74, 30–39. doi: 10.1016/j.neuron.2012.03.016

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., and Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neurosci.* 18, 7411–7425.

Brown, E. N., Nguyen, D. P., Frank, L. M., Wilson, M. A., and Solo, V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *Proc. Natl. Acad. Sci. U.S.A* 98, 12261–12266. doi: 10.1073/pnas.201409398

Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L., and Woolley, S. M. N. (2011). A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS ONE* 6:e16104. doi: 10.1371/journal.pone.0016104

Chechik, G., and Nelken, I. (2012). Auditory abstraction from spectro-temporal features to coding auditory entities. *Proc. Natl. Acad. Sci. U.S.A* 109, 18968–18973. doi: 10.1073/pnas.1111242109

Chechik, G., Anderson, M. J., Bar-Yosef, O., Young, E. D., Tishby, N., and Nelken, I. (2006). Reduction of information redundancy in the ascending auditory pathway. *Neuron* 51, 359–368. doi: 10.1016/j.neuron.2006.06.030

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network* 12, 199–213. doi: 10.1080/713663221

Christianson, G. B., Sahani, M., and Linden, J. F. (2008). The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *J. Neurosci.* 28, 446–455. doi: 10.1523/JNEUROSCI.1775-07.2007

David, S. V., Vinje, W. E., and Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. *J. Neurosci.* 24, 6991–7006. doi: 10.1523/JNEUROSCI.1422-04.2004

David, S. V., Mesgarani, N., and Shamma, S. A. (2007). Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network* 18, 191–212. doi: 10.1080/09548980701609235

Dean, I., Harper, N. S., and McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nat. Neurosci.* 8, 1684–1689. doi: 10.1038/nn1541

deBoer, E., and Kuyper, P. (1968). Triggered correlation. *IEEE Trans. Biomed. Eng.* 15, 169–179. doi: 10.1109/TBME.1968.4502561

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* 39, 1–38.

Ding, N., and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A* 109, 11854–11859. doi: 10.1073/pnas.1205381109

Ecker, A. S., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C. R., et al. (2014). State dependence of noise correlations in macaque primary visual cortex. *Neuron* 82, 235–248. doi: 10.1016/j.neuron.2014.02.006

Eden, U. T., Frank, L. M., Barbieri, R., Solo, V., and Brown, E. N. (2004). Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Comput.* 16, 971–998. doi: 10.1162/089976604773135069

Escabi, M. A., Miller, L. M., Read, H. L., and Schreiner, C. E. (2003). Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. *J. Neurosci.* 23, 11489–11504.

Fitzgerald, J. D., Rowekamp, R. J., Sincich, L. C., and Sharpee, T. O. (2011). Second order dimensionality reduction using minimum and maximum mutual information models. *PLoS Comput. Biol.* 7:e1002249. doi: 10.1371/journal.pcbi.1002249

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223. doi: 10.1038/nn1141

Garcia-Lazaro, J. A., Belliveau, L. A. C., and Lesica, N. A. (2013). Independent population coding of speech with sub-millisecond precision. *J. Neurosci.* 33, 19362–19372. doi: 10.1523/JNEUROSCI.3711-13.2013

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., et al. (1993). Timit acoustic-phonetic continuous speech corpus. Philadelphia, PA: Linguistic Data Consortium.

Gerwinn, S., Macke, J. H., and Bethge, M. (2010). Bayesian inference for generalized linear models for spiking neurons. *Front. Comput. Neurosci.* 4:12. doi: 10.3389/fncom.2010.00012

Gill, P., Zhang, J., Woolley, S. M. N., Fremouw, T., and Theunissen, F. E. (2006). Sound representation methods for spectro-temporal receptive field estimation. *J. Comput. Neurosci.* 21, 5–20. doi: 10.1007/s10827-006-7059-4

Hansen, B. J., Chelaru, M. I., and Dragoi, V. (2012). Correlated variability in laminar cortical circuits. *Neuron* 76, 590–602. doi: 10.1016/j.neuron.2012.08.029

Happel, M. F. K., Jeschke, M., and Ohl, F. W. (2010). Spectral integration in primary auditory cortex attributable to temporally precise convergence of thalamocortical and intracortical input. *J. Neurosci.* 30, 11114–11127. doi: 10.1523/JNEUROSCI.0689-10.2010

Harris, K. D., and Thiele, A. (2011). Cortical state and attention. *Nat. Rev. Neurosci.* 12, 509–523. doi: 10.1038/nrn3084

Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). *The Elements of Statistical Learning*. New York, NY: Springer. doi: 10.1007/978-0-387-21606-5

Hoerl, A. E. (1962). Application of ridge analysis to regression problems. *Chem. Eng. Prog.* 58, 54–59.

Hohmann, V. (2002). Frequency analysis and synthesis using a gammatone filterbank. *Acta Acust. United Ac.* 88, 433–442.

Hunter, I. W., and Korenberg, M. J. (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biol. Cybern.* 55, 135–144.

Lesica, N. A., and Grothe, B. (2008a). Dynamic spectrotemporal feature selectivity in the auditory midbrain. *J. Neurosci.* 28, 5412–5421. doi: 10.1523/JNEUROSCI.0073-08.2008

Lesica, N. A., and Grothe, B. (2008b). Efficient temporal processing of naturalistic sounds. *PLoS ONE* 3:e1655. doi: 10.1371/journal.pone.0001655

Lin, C.-J., Weng, R. C., and Keerthi, S. S. (2008). Trust region newton method for logistic regression. *J. Mach. Learn. Res.* 9, 627–650.

Machens, C. K., Wehr, M. S., and Zador, A. M. (2004). Linearity of cortical receptive fields measured with natural sounds. *J. Neurosci.* 24, 1089–1100. doi: 10.1523/JNEUROSCI.4445-03.2004

Malmierca, M. S., Cristaudo, S., Pérez-González, D., and Covey, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *J. Neurosci.* 29, 5483–5493. doi: 10.1523/JNEUROSCI.4153-08.2009

Mease, R. A., Famulare, M., Gjorgjieva, J., Moody, W. J., and Fairhall, A. L. (2013). Emergence of adaptive computation by single neurons in the developing cortex. *J. Neurosci.* 33, 12154–12170. doi: 10.1523/JNEUROSCI.3263-12.2013

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

Meyer, A. F., Diepenbrock, J.-P., Happel, M. F., Ohl, F. W., and Anemüller, J. (2014). Discriminative learning of receptive fields from responses to non-gaussian stimulus ensembles. *PLOS ONE* 9:e93062. doi: 10.1371/journal.pone.0093062

Meyer, A. F., Diepenbrock, J.-P., Ohl, F. W., and Anemüller, J. (2013). "Quantifying neural coding noise in linear threshold models," in *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on* (San Diego, CA), 1127–1130.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.

Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized linear models. *J. R. Stat. Soc. A Gen.* 135, 370–384. doi: 10.2307/2344614

Nelken, I., Rotman, Y., and Yosef, O. B. (1999). Responses of auditory-cortex Neurons to structural features of natural sounds. *Nature* 397, 154–157. doi: 10.1038/16456

Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., and Farkas, D. (2003). Primary auditory cortex of cats: feature detection or something else? *Biol. Cybern.* 89, 397–406. doi: 10.1007/s00422-003-0445-3

Netser, S., Zahar, Y., and Gutfreund, Y. (2011). Stimulus-specific adaptation: can it be a neural correlate of behavioral habituation? *J. Neurosci.* 31, 17811–17820. doi: 10.1523/JNEUROSCI.4790-11.2011

Ohl, F. W., and Scheich, H. (2005). Learning-induced plasticity in animal and human auditory cortex. *Curr. Opin. Neurobiol.* 15, 470–477. doi: 10.1016/j.conb.2005.07.002

Paninski, L., Ahmadian, Y., Ferreira, D. G., Koyama, S., Rad, K. R., Vidne, M., et al. (2010). A new look at state-space models for neural data. *J. Comput. Neurosci.* 29, 107–126. doi: 10.1007/s10827-009-0179-x

Paninski, L. (2003). Convergence properties of three spike-triggered analysis techniques. *Network* 14, 437–464. doi: 10.1088/0954-898X/14/3/304

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15, 243–262. doi: 10.1088/0954-898X/15/4/002

Park, M., and Pillow, J. W. (2011). Receptive field inference with localized priors. *PLoS Comput. Biol.* 7:e1002219. doi: 10.1371/journal.pcbi.1002219

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

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

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

Sadagopan, S., and Wang, X. (2010). Contribution of inhibition to stimulus selectivity in primary auditory cortex of awake primates. *J. Neurosci.* 30, 7314–7325. doi: 10.1523/JNEUROSCI.5072-09.2010

Sahani, M., and Linden, J. F. (2002). "Evidence optimization techniques for estimating stimulus-response functions," in *Advances in Neural Information Processing Systems (NIPS)* (Vancouver, BC), 301–308.

Schumacher, J. W., Schneider, D. M., and Woolley, S. M. N. (2011). Anesthetic state modulates excitability but not spectral tuning or neural discrimination in single auditory midbrain neurons. *J. Neurophysiol.* 106, 500–514. doi: 10.1152/jn.01072.2010

Schwartz, O., Pillow, J. W., Rust, N. C., and Simoncelli, E. P. (2006). Spike-triggered neural characterization. *J. Vis.* 6, 484–507. doi: 10.1167/6.4.13

Sharpee, T., Rust, N. C., and Bialek, W. (2004). Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput.* 16, 223–250. doi: 10.1162/089976604322742010

Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P., and Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature* 439, 936–942. doi: 10.1038/nature04519

Sharpee, T. O. (2013). Computational identification of receptive fields. *Annu. Rev. Neurosci.* 36, 103–120. doi: 10.1146/annurev-neuro-062012-170253

Smith, M. A., and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* 28, 12591–12603. doi: 10.1523/JNEUROSCI.2929-08.2008

Stanley, G. B. (2002). Adaptive spatiotemporal receptive field estimation in the visual pathway. *Neural Comput.* 14, 2925–2946. doi: 10.1162/089976602760805340

Toups, J. V., Fellous, J.-M., Thomas, P. J., Sejnowski, T. J., and Tiesinga, P. H. (2012). Multiple spike time patterns occur at bifurcation points of membrane potential dynamics. *PLoS Comput. Biol.* 8:e1002615. doi: 10.1371/journal.pcbi.1002615

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., and Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J. Neurophysiol.* 93, 1074–1089. doi: 10.1152/jn.00697.2004

Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* 6, 391–398. doi: 10.1038/nn1032

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex Neurons. *J. Neurosci.* 24, 10440–10453. doi: 10.1523/JNEUROSCI.1905-04.2004

Weinberger, N. M. (1993). Learning-induced changes of auditory receptive fields. *Curr. Opin. Neurobiol.* 3, 570–577. doi: 10.1016/0959-4388(93)90058-7

Wu, M. C., David, S. V., and Gallant, J. L. (2006). Complete functional characterization of sensory Neurons by system identification. *Annu. Rev. Neurosci.* 29, 477–505. doi: 10.1146/annurev.neuro.29.051605.113024

Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated Neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143. doi: 10.1038/370140a0