

Efficient Coding Approach Towards Non-Linear Spectro-Temporal Receptive Fields

Pranav Sankhe^{1*} and Prasanna Chaporkar²

¹Center for Information and Neural Networks, Osaka, Japan.

²Electrical Engineering Department, IIT Bombay, Mumbai, India.

*Corresponding author(s). E-mail(s):
pranavsankhe40@gmail.com;

Abstract

Linear Non-Linear(LN) models are widely used to characterize the receptive fields of early-stage auditory processing. We apply the principle of efficient coding to the LN model of Spectro-Temporal Receptive Fields(STRFs) of the neurons in primary auditory cortex. The Efficient Coding Principle has been previously used to understand early visual receptive fields and linear STRFs in auditory processing. Efficient coding is realized by jointly optimizing the mutual information between stimuli and neural responses subjected to the metabolic cost of firing spikes. We compare the predictions of the efficient coding principle with the physiological observations, which match qualitatively under realistic conditions of noise in stimuli and the spike generation process.

Keywords: STRF, Auditory Processing, Efficient Coding, LN Model

Information theoretical concepts like entropy and mutual information are useful measures to describe the coding efficiency of sensory systems. Applications of the efficient coding principle to early visual and linear auditory systems have helped explain the characteristics of neuronal receptive fields (Srinivasan et al., 1982; Atick et al., 1990; Lewicki, 2002; Smith & Lewicki, 2006). In particular, neural responses have been found to be optimized to transmit as much information as possible about the naturally occurring stimuli. Furthermore,

this optimization is subjected to the cost of generating the neural response (Laughlin et al., 1998; Lennie, 2003). For instance, early visual representations in the retina, LGN, and primary visual cortex have been shown to be consistent with the principles of information maximization in the presence of noise. In the auditory system, the spectro-temporal receptive fields (STRFs) have also been shown to maximize the information received from the stimulus along with minimizing the cost of generating spikes. (Lewicki, 2002; Smith Lewicki, 2006)

The efficient coding principle has been used to explain the dependency of STRFs on the stimulus properties (Lesica & Grothe, 2006; Zhao, 2011). In addition to that, multiple experimental studies have shown throughout auditory code that includes the cochlea, the auditory nerve, and the auditory cortex follow optimal coding (Attias et al., 1997; Rieke et al. 1995; Lewicki, 2002;). Interestingly enough, evidence for the efficient coding hypothesis has also been demonstrated by behavioural experiments on human auditory perception (Ming et al., 2009).

Traditional STRFs are linear kernels that characterize the spectral and temporal selectivity of a neuron to the incoming stimulus. More specifically, STRF explains the sensitivity of a neuron at a particular latency t to acoustic frequency f by relating the spike rate of the neuron with the time-frequency representation (spectrogram) of the stimulus. However, along the auditory pathway, there are several nonlinearities in the neural responses which the linear STRFs fail to capture. For example, the neurons in auditory systems of primates, bats and birds (Rauschecker et al., 1995; Ohlemiller et al., 1996; Scheich et al., 1979) are tuned explicitly to only complete animal vocalizations. Additionally, auditory neurons also have been found to be selective to the amplitude of sounds in the same frequency range (Nelken et al., 2008) and to sounds of varying levels of spectral contrast (Barbour et al., 2003). These nonlinearities in auditory coding have inspired several nonlinear models of STRFs, such as the Linear-Nonlinear (LN) model. In this paper, we focus on applying the efficient coding principle to a non-linear model of STRF of neurons in the primary auditory cortex.

The LN model, which is an instantiation of a generalized linear model (GLM), is used to characterize the early stages of auditory processing (Aertsen & Johannesma, 1981; Radtke-Schuller et al., 2020). LN STRFs have been used successfully to model the receptive fields of the auditory cortex, and it has been shown that compared to the linear model of STRF, the LN model estimates the neural responses significantly better (Machens et al., 2004; Simoncelli et al., 2004). LN model is conceptualized by applying a static non-linearity onto the linear approximation of the spike generating process. In the linear stage of the model, a FIR filter $h(f, \tau)$ is applied to the time-frequency representation

(spectrogram) of the stimulus ($s(f, t)$) to produce spike rate $r(t)$.

$$r(t) = \sum_{f=q}^{f=F} \sum_{\tau=0}^{\tau=T} h(f, \tau) s(f, t - \tau) \quad (1)$$

$h(f, \tau)$ is often referred to as the linear Spectro-Temporal Receptive Field (STRF) in the auditory system. Let \mathbf{r} be the output vector of N neurons representing the firing rate. In the case of populations of neurons, let \mathbf{S} be the input time-frequency representation of the audio stimulus and \mathbf{h}_i be the linear filter of neuron i . The response of neuron i is computed by taking an inner product of noise corrupted input and linear filter to obtain membrane voltage y_i . This voltage is passed through a static nonlinearity to produce the spikes, which are corrupted with output neural noise. The current study used a rectified linear function (Agarap, 2018) for the output non-linearity in the second stage of the LN model. Each neuron is characterized with an offset (b_i) and slope(g_i) for its rectified linear function. The neural response in terms of firing rate is directly calculated by applying the RELU function on the membrane voltage (y_i).

$$y_i = \mathbf{h}_i(\mathbf{S} + \mathbf{n}_s) \quad (2)$$

$$r_i = g_i(y_i + b_i) + n_r \quad (3)$$

We employ a generative linear-nonlinear model coupled with an efficient coding principle to optimize the linear filters and the static nonlinearities. The optimization problem is designed to maximize the mutual information between the natural sounds and the neural responses in the presence of input and output noise. The metabolic costs involved in generating the neural response impose constraints on neurons' firing rate, thus constraining mutual information maximization. We borrow techniques and intuitions from the work of Karklin and Simoncelli (Karklin & Simoncelli, 2011) on receptive fields of retinal ganglion cells to derive and explain the results in this paper.

We consider the problem of jointly optimizing the linear filter and nonlinearities by maximizing the mutual information ($I(S, R)$) subject to neural costs such as metabolic energy costs involved in the spike generation process. Following the approach taken in vision, the neural costs are modelled to be proportional to the average spike rate of the neuronal population (Zhaoping, 2006). λ_i parameterizes the trade-off between the cost of neural response and the information gained by generating the response. Mutual information can be expressed as the difference between average uncertainty about the identity of the stimulus given the response and total stimulus entropy, i.e., $H(S) - H(S|R)$. Since the global stimulus entropy $H(S)$ does not depend on the model, the optimization problem can be framed as maximizing the negative conditional entropy subject to the firing cost. Incorporating penalty on the firing rate, we

get an objective function:

$$-H(S|R) - \sum_i \lambda_i < r_i > \quad (4)$$

The conditional entropy $H(S|R)$ is computed by calculating the expectation of covariance of the posterior $C_{s|r}$. $C_{s|r}$ is calculated by the covariance of the prior $C(r|s)$ multiplied by the linear filters \mathbf{W} and the diagonal matrix of the nonlinear responses \mathbf{G} and added to the inverse of the global covariance matrix of the input (Karklin & Simoncelli, 2011). Using the identity that logarithm of determinant of a matrix is equal to the trace of the logarithm of the matrix, we the objective function simplifies to:

$$-E[\frac{1}{2} \log(2\pi e * \text{trace}(C_{s|r}))] - \sum_i \lambda_i < r_i > \quad (5)$$

The model is trained using across batches of data using gradient ascent on the objective function that encapsulates the coding efficiency. We use Tensorflow (Abadi et al., 2016) to perform the optimization with a learning rate of 0.001 over 80,000 iterations. We run the model for 36 neurons. The verification of convergence of the LN model is not possible by simulating data, unlike the algorithms for other generative models like PCA or ICA. Cochleagrams of natural sounds are obtained from a large collection of human speech, ferret vocalizations, and recordings of the ambient laboratory environment (Lopez Espejo et al., 2019). Cochleogram is a spectral representation of the auditory nerve firing rate (Brown & Cooke, 1994). Neural recordings from the primary auditory cortex of ferrets (Lopez Espejo et al., 2019), passively listening to the natural sounds and cochleagram of the stimulus were used to estimate the LN STRFs of auditory neurons. Spike rate data and the cochleagrams are fitted for each neuron using the gradient ascent algorithm to update the parameters. Preprocessing of the data and model estimation was done using the NEMS library in Python (David, 2018).

To investigate the role of input and output noise, we estimated STRFs from the model in high noise and low noise conditions. The high noise conditions were set to the ones present in the physiological data. The input and output noise was set to extremely low values (-40dB) in low noise conditions. We observed that in the presence of significant output noise, the STRFs generated by the model match with the estimated STRFs from the physiological data [Fig 1]. The model was able to estimate sharp rectifying non-linearities of individual neurons in the presence of significant output noise. On the other hand, in low noise conditions, the model was unable to learn the sharp rectification. The estimated slopes of almost all the neurons evaluate to zero. Past literature has suggested that hard rectification in neurons is more useful in the presence of noise (Carandini, 2004). The results support claims that the neurons in the early auditory regions optimally encode the incoming stimulus information.

Past work has mainly focused on the linear part of the STRF (Lesica &

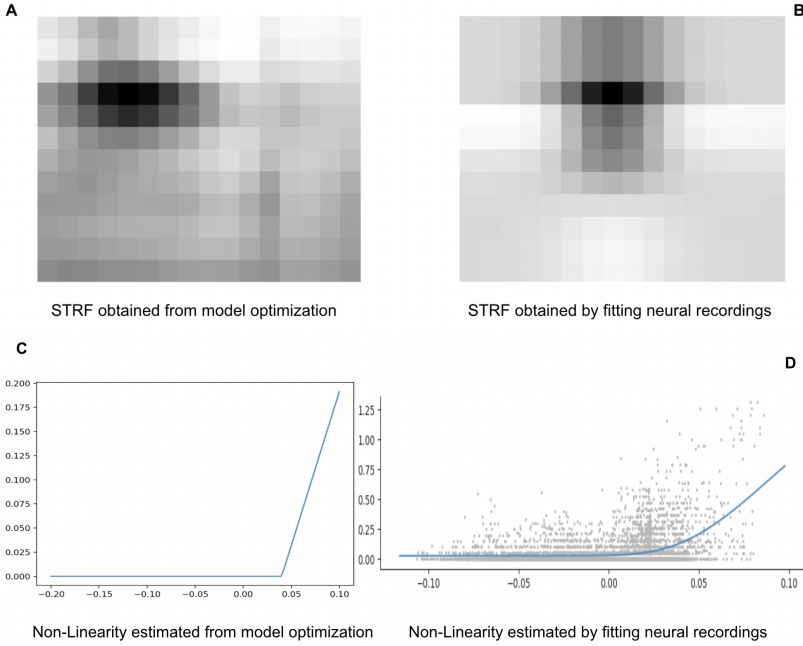


Fig. 1 Figure A is the LN STRF obtained by applying efficient coding principle whereas figure B shows the physiological LN STRF estimated from neural spike data recorded from the primary auditory cortex of one neuron of ferrets while listening to natural sounds. Figures C and D illustrate the estimated non-linearity and the measured non-linearity for a neuron respectively.

Grothe, 2006; Zhao, 2011), and our results indicate that the efficient coding principle can be applied to non-linear models of STRF. Here we show an efficient coding for non-linear STRFs under the presence of input and output noise. The linear filter of the STRF and static nonlinearities are estimated jointly for a population of neurons by maximizing the mutual information and between spike rates and auditory stimuli (cochleagrams). We demonstrate that the non-linear model STRFs derived using the efficient coding principle qualitatively matches with the physiological STRFs obtained by fitting neural data recorded from the primary auditory cortex of ferrets.

References

- [1] Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 216(1205), 427-459.
- [2] Atick, J. J., & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural computation*, 2(3), 308-320.
- [3] Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature neuroscience*, 5(4), 356-363.
- [4] Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, 439(7079), 978-982
- [5] Laughlin, S. B., van Steveninck, R. R. D. R., & Anderson, J. C. (1998). The metabolic cost of neural information. *Nature neuroscience*, 1(1), 36-41.
- [6] Lennie, P. (2003). The cost of cortical computation. *Current biology*, 13(6), 493-497.
- [7] Lesica, N. A., & Grothe, B. (2008). Efficient temporal processing of naturalistic sounds. *PLoS One*, 3(2), e1655.
- [8] Zhao, L., & Zhaoping, L. (2011). Understanding auditory spectro-temporal receptive fields and their changes with input statistics by efficient coding principles. *PLoS computational biology*, 7(8), e1002123.
- [9] Attias, H., & Schreiner, C. E. (1997). Temporal low-order statistics of natural sounds. *Advances in neural information processing systems*, 27-33.
- [10] Rieke, F., Bodnar, D. A., & Bialek, W. (1995). Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 262(1365), 259-265.
- [11] Ming, V. L., & Holt, L. L. (2009). Efficient coding in human auditory perception. *The Journal of the Acoustical Society of America*, 126(3), 1312-1320.
- [12] Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268(5207), 111-114.
- [13] Ohlemiller, K. K., Kanwal, J. S., & Suga, N. (1996). Facilitative responses to species-specific calls in cortical FM-FM neurons of the mustached bat. *Neuroreport*, 7(11), 1749-1755.

- [14] Bonke, D., Scheich, H., & Langner, G. (1979). Responsiveness of units in the auditory neostriatum of the guinea fowl (*Numida meleagris*) to species-specific calls and synthetic stimuli. *Journal of comparative physiology*, 132(3), 243-255.
- [15] Nelken, I., & Bar-Yosef, O. (2008). Neurons and objects: the case of auditory cortex. *Frontiers in neuroscience*, 2, 9.
- [16] Barbour, D. L., & Wang, X. (2003). Contrast tuning in auditory cortex. *Science*, 299(5609), 1073-1075.
- [17] Aertsen, A. M., & Johannesma, P. I. M. (1981). The spectro-temporal receptive field. *Biological cybernetics*, 42(2), 133-143.
- [18] Radtke-Schuller, S., Town, S. M., Yin, P., Elgueda, D., Schuller, G., Bizley, J. K., ... & Fritz, J. B. (2020). Dorsal prefrontal and premotor cortex of the ferret as defined by distinctive patterns of thalamo-cortical projections. *Brain Structure and Function*, 225, 1643-1667
- [19] Machens, C. K., Wehr, M. S., & Zador, A. M. (2004). Linearity of cortical receptive fields measured with natural sounds. *Journal of Neuroscience*, 24(5), 1089-1100.
- [20] Simoncelli, E. P., Paninski, L., Pillow, J., & Schwartz, O. (2004). Characterization of neural responses with stochastic stimuli. *The cognitive neurosciences*, 3(327-338), 1.
- [21] Agarap, A. F. (2018). Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*.
- [22] Karklin, Y., & Simoncelli, E. P. (2011). Efficient coding of natural images with a population of noisy linear-nonlinear neurons. *Advances in neural information processing systems*, 24, 999.
- [23] Zhaoping, L. (2006). Theoretical understanding of the early visual processes by data compression and data selection. *Network: computation in neural systems*, 17(4), 301-334
- [24] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)* (pp. 265-283).
- [25] Lopez Espejo, M., Schwartz, Z. P., & David, S. V. (2019). Spectral tuning of adaptation supports coding of sensory context in auditory cortex. *PLoS computational biology*, 15(10), e1007430.

- [26] Brown, G. J., & Cooke, M. (1994). Computational auditory scene analysis. *Computer Speech & Language*, 8(4), 297-336.
- [27] David S V. Neural Encoding Model System (NEMS) [Internet]. 2018. Available: <https://github.com/LBHB/NEMS>
- [28] Carandini, M., & Stevens, C. (2004). Amplification of trial-to-trial response variability by neurons in visual cortex. *PLoS biology*, 2(9), e264