

THE EFFECT OF DIFFERENT SPECTRO-TEMPORAL REPRESENTATIONS OF AN INPUT AUDITORY STIMULUS ON THE FITTING OF A POINT PROCESS MODEL OF AUDITORY NEURONS

Eric Plourde¹ and Emery N. Brown^{2}*

¹Université de Sherbrooke
Department of Electrical and Computer Engineering
Sherbrooke, QC J1K 2R1, Canada - email: eric.plourde@usherbrooke.ca

²Neuroscience Statistics Research Laboratory
Massachusetts General Hospital, Harvard Medical School, M.I.T.
Boston, MA 02114, USA - email: enb@neurostat.mit.edu

ABSTRACT

We compare the effect of the use of three different spectro-temporal representations of an input auditory stimulus on the fitting of a point process model of auditory neuron firing. The three spectro-temporal representations considered are the spectrogram, a gammatone filterbank and the Hilbert spectrum. We firstly investigate how the model fits the recorded neuronal data when using either one of the three representations and secondly how well do the estimated parameters of the model correspond to their experimentally measured counterparts. It is observed that all three representations yield a model that fits well the recorded data. However, the characteristic frequencies obtained with the spectro-temporal parameters of the model using the gammatone filterbank corresponds better to the experimentally measured characteristic frequency than the characteristic frequency obtained with the models using the other two spectro-temporal representations. Therefore, it is concluded that the quality of the fitted parameters can be affected by the choice of the spectro-temporal representation and that, as could have been expected, the gammatone filterbank seems to more accurately extract the relevant spectro-temporal characteristics of the input auditory stimulus.

1. INTRODUCTION

One approach to better understand the factors that are important for inducing neurons to spike is to derive and fit statistical models containing the most salient factors contributing to the spiking. Such point process models have been proposed in [1, 2] where the neurons intrinsic dynamics and the spectro-temporal properties of the input auditory stimulus are considered.

Several different approaches can be used to obtain a spectro-temporal representation of the input stimulus. The

choice of the representation could have some impact on both the goodness of fit of the model to experimentally recorded neural spikes and the quality of the fitted parameters of the model in terms of their correspondence to experimentally measured values. Here we analyze the effect on the model fitting of three different spectro-temporal representations, namely the spectrogram as used in [2], a bank of gammatone filters as used in [1] and the Hilbert spectrum. The first one is a classical spectro-temporal representation, the second one is thought to mimic the spectral decomposition performed by the ear while the third one is designed to obtain a spectro-temporal representation of a non-linear signal such as speech. The objective of the study is thus to investigate how important is the chosen type of spectro-temporal representations in the fitting of the model and the analysis of its fitted parameters.

It is found that the models using either of the three different representations fit the recorded data similarly well. However, the characteristic frequency, i.e. the frequency of the input signal to which the neuron responds best, obtained with the spectro-temporal parameters of the model using the gammatone filterbank more closely match the experimentally measured characteristic frequency than the ones obtained when using the other two spectro-temporal representations. This could be due to the fact that the gammatone filterbank extracts more relevant spectro-temporal characteristics of the input auditory stimulus than the other two representations. The choice of spectro-temporal representation therefore affects the quality of the fitted parameters.

The paper is organized as follows. In Section 2, we briefly review the previously proposed statistical model. In Section 3, the fitting of the model using the three different spectro-temporal representations is analyzed and Section 4 concludes the work.

*This work was supported by the Fonds québécois de la recherche sur la nature et les technologies and the National Institutes of Health under grant DP1-OD003646.

2. PREVIOUSLY PROPOSED STATISTICAL MODEL

In this section, we briefly review the point process model proposed in [1]. The model incorporates both a spectro-temporal representation of the input speech signal as well as the intrinsic dynamics of the neuron through its past spiking activity.

A point process model can be completely defined by its conditional intensity function (CIF) $\lambda[n|\Psi]$ where n is the discrete time index and Ψ includes relevant covariates of the model. The CIF is for a point process a history-dependent generalization of the rate function of a Poisson process.

Let $s_{k,j}$ be the value of a spectro-temporal representation of the sound stimulus with frequency band j at time $k\Delta$ for $j = 1, \dots, J$. Define the relevant history of the sound stimulus for predicting the current spiking propensity as $H_{k,j} = \{s_{k,j}, \dots, s_{k-L,j}\}$, assuming a dependence that goes back L time periods. Similarly, define the relevant spiking history for predicting the current spiking propensity as $H_{k,J+1} = \{n_{k-1}, \dots, n_{k-P}\}$, assuming a dependence that goes back P time periods. Let $H_k = \{H_{k,1}, \dots, H_{k,J+1}\}$. If we assume that there is a functional F which describes the relation between H_k and the CIF $\lambda(k\Delta|H_k)$ then we can obtain the following CIF:

$$\lambda(k\Delta|H_k, \beta) = \exp \left\{ \beta_0 + \sum_{j=1}^J \sum_{l=0}^{L-1} \beta_{l,j} s_{k-l,j} + \sum_{p=1}^P \beta_{p,J+1} n_{k-p} \right\} \quad (1)$$

where $\beta = \{\beta_0, \beta_{0,1}, \dots, \beta_{L-1,J}, \beta_{1,J+1}, \dots, \beta_{P,J+1}\}$ is the $(JL + P + 1) \times 1$ vector of Volterra kernels. We interpret the Volterra series expansion as the sum of the outputs of $J + 1$ linear filters having Volterra kernels as the impulse responses. The kernels $\beta_{l,j}$ are the analogs of the Spectro-Temporal Receptive Fields (STRF) used to characterize auditory neurons. The kernel $\beta_{p,J+1}$ models the effect of the spiking history and β_0 governs the mean spiking rate. The model is then regularized and fitted to the spiking data using the TR-IRLS algorithm [1, 3] to obtain the different parameter values.

In this paper, we fit the statistical model to neural spiking activity recorded in the auditory nerves of anesthetized cats following the presentation of an input sentence spoken by a male voice and sampled at 10 kHz [4]. We use a generalized linear model (GLM) in a ridge regression framework to address properly the ill-posed inverse nature of this estimation problem and avoid overfitting [1]. The dataset is composed of the spike train responses of 55 distinct neurons each recorded across $R = 20$ trials. As in [1], we use values of $P = 40$, $L = 104$, and $J = 25$.

3. COMPARISON OF THE THREE DIFFERENT SPECTRO-TEMPORAL REPRESENTATIONS

Any spectro-temporal representation of the input auditory stimulus could theoretically be used in the proposed model as the $s_{k,j}$'s. In this section we investigate the use of three different spectro-temporal representations in the model: the spectrogram as used in [2], a gammatone filterbank as used in [1] and the Hilbert spectrum. In particular, we present the effect of this choice on the goodness of fit of the model and the quality of the estimated parameters where the quality is defined as the degree of correspondence between the estimated parameters and their experimentally measured counterpart.

3.1. Description of the spectro-temporal representations

Spectrogram

The discrete-time short-time Fourier transform (STFT) of the input stimulus is obtained by successively applying overlapping finite windows on the input signal and performing a Fourier transform on each windowed signal. Taking the magnitude squared of the result yields the spectrogram of the input stimulus. A size $N = 128$ discrete-time Fourier transform is used here.

Gammatone filter bank

This representation is obtained by applying a filter bank to the input speech signal. The filter bank consists of adjacent bandpass gammatone filters [5], the later having been designed to mimic the filtering performed by the cochlea. The bandwidth of the filters are chosen according to [6] in order to represent adequately the processing performed in the cat's cochlea. Table 1 presents the center frequency and bandwidth of each filter in the filter bank.

Hilbert spectrum

The Hilbert spectrum [7] is obtained by first performing an empirical mode decomposition (EMD) on the stimulus, applying a Hilbert transform on each mode of the decomposition and then computing the instantaneous frequency for every time step. The combination of the instantaneous frequencies present in all modes at a certain time thus gives a representation of the frequency content of the stimulus at that time. The advantage of this spectro-temporal representation is that it does not assume that the data is piecewise stationary as in the STFT case. It attempts to yield an instantaneous account of the frequencies in the signal.

Some further processing of these spectro-temporal representations are necessary to use them in the model as $s_{k,j}$'s (see Fig. 1). Firstly, in order to limit the number of parameters to estimate, it is necessary to limit the number of frequency bands present in the different representations. This can be done either by choosing the spectro-

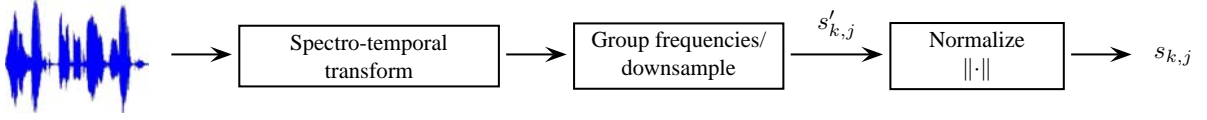


Fig. 1. Normalized spectro-temporal representation.

Filter number	Center frequency [Hz]	Bandwidth [Hz]
1	20	55
2	52	63
3	88	72
4	130	82
5	176	94
6	229	108
7	288	124
8	355	142
9	431	163
10	516	185
11	612	211
12	721	240
13	844	272
14	984	307
15	1141	346
16	1318	390
17	1518	438
18	1745	491
19	2000	550
20	2289	615
21	2615	687
22	2983	766
23	3400	854
24	3870	952
25	4400	1060

Table 1. Gammatone filter bank center frequencies and bandwidth.

temporal transform parameters appropriately or by grouping different adjacent frequency bands together. Secondly, the neural responses are captured at a resolution of 1 ms, if not accounted for during the spectro-temporal transform, the representation must thus be downsampled to such a time resolution. Finally, we would like to compare the parameters corresponding to the different frequency bins against each other in order to assess how much each frequency contributes to the spiking. Therefore, their corresponding representation needs to be normalized in some sense. For each frequency bin, the spectro-temporal representation is thus normalized by its Euclidean norm, yielding the desired normal basis (albeit not necessarily orthogonal). The different processing steps leading to the $s_{k,j}$'s are illustrated in Figure 1.

3.2. Goodness of fit

To evaluate the model goodness-of-fit, we used the time rescaling theorem with rescaled times computed from the estimated CIF [3]. If the latter is a good approximation to the true CIF of the point process, then the rescaled times will be independent and uniformly distributed on the interval $[0, 1)$.

We used the autocorrelation function (ACF) of the transformed rescaled times to assess their independence. ACF results (not presented here) have shown that the rescaled times computed from the CIFs estimated using each of the three different spectro-temporal representations were indeed all independent.

To assess the uniformity of the rescaled times, we used a cumulative function based on the normalized Kolmogorov-Smirnov (KS) statistic. The normalized KS statistic is given by:

$$\hat{D} = \sup \left| \frac{\hat{F}(x) - F(x)}{\hat{B}} \right| \quad (2)$$

where $F(x)$ is a cumulative distribution function of rescaled spiking times obtained from the fitted CIF, $\hat{F}(x)$ is an empirical uniform distribution function and \hat{B} is the 95 % confidence bound. A cumulative function $C(\hat{D})$, based on the normalized KS statistic, can be defined as follows:

$$C(\hat{D}) = \sum_{i=1}^{N_o} I_{\hat{D}_i < \hat{D}} \quad (3)$$

where N_o is the total number of neurons in the dataset, \hat{D}_i is the normalized KS statistics of the i^{th} neuron and

$$I_{\hat{D}_i < \hat{D}} = \begin{cases} 1 & \text{if } \hat{D}_i < \hat{D} \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

The best model will thus be the one with the steepest convergence of $C(\hat{D})$ to N_o and more so if $\hat{D} < 1$.

Fig. 2 shows the cumulative function $C(\hat{D})$ for models fitted with the three different spectro-temporal representations: spectrogram, Hilbert spectrum and gammatone filter bank. As can be observed, the fitting obtained with the three representations are quite good and very similar. The results obtained here thus suggest that as long as information regarding the spectrum is present in one form or another, the proposed model can achieve a good fitting.

3.3. Baseline and history parameters ($\beta_o, \beta_{p,J+1}$)

The baseline (β_o) and history ($\beta_{p,J+1}$) parameters of the models fitted with the three proposed spectro-temporal rep-

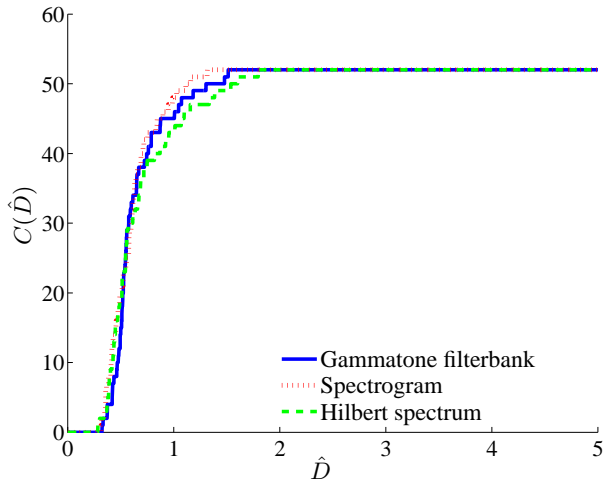


Fig. 2. Cumulative function $C(\hat{D})$ for the following spectro-temporal transforms: spectrogram, Hilbert transform and gammatone filter bank.

representations were compared. The value of these parameters (not presented here) were found to be extremely similar when using either one of the different spectro-temporal representations.

3.4. Spectro-temporal parameters ($\beta_{l,j}$)

We now look at the fitted values of the spectro-temporal parameters, i.e. $\beta_{l,j}$, for the three spectro-temporal representations of the input stimulus studied here.

Figures 3-5 show the scatter plots of the frequency corresponding to the parameter $\beta_{l,j}$ with the highest value, indicated here as the fitted characteristic frequency (CF), vs. the experimentally measured CF of the corresponding neuron. The characteristic frequency of an auditory neuron is the stimulus frequency to which it responds best. As can be observed, there is a better correlation of the fitted CF with the experimentally measured CF when using a gammatone filter bank (Fig. 4) as the spectro-temporal representation than when using either a spectrogram (Fig. 3) or a Hilbert spectrum (Fig. 5). This is quantitatively confirmed by the computed root mean square error (rmse) between the different points and the diagonal, indicated at the bottom right of each figure. This could be due to the fact that the gammatone filterbank extracts more relevant spectro-temporal characteristics of the input auditory stimulus than the other two representations. The quality of the $\beta_{l,j}$ parameter thus depend on the chosen spectro-temporal representation of the input sound stimulus. Instead of simply considering the parameter $\beta_{l,j}$ with the highest value to assess the fit and compute the rmse, one could have chosen other means such as the center of mass of the $\beta_{l,j}$'s. However, as can be observed in [1], the center of mass of the fitted $\beta_{l,j}$'s is generally quite close to the $\beta_{l,j}$ with the highest value and, therefore, the rmse results should be similar. It is relevant to note that we only report here the correspondance of the fitted characteristic frequency to the experimentally measured one. In fact, we did not comment on the other dimension of the fitted $\beta_{l,j}$

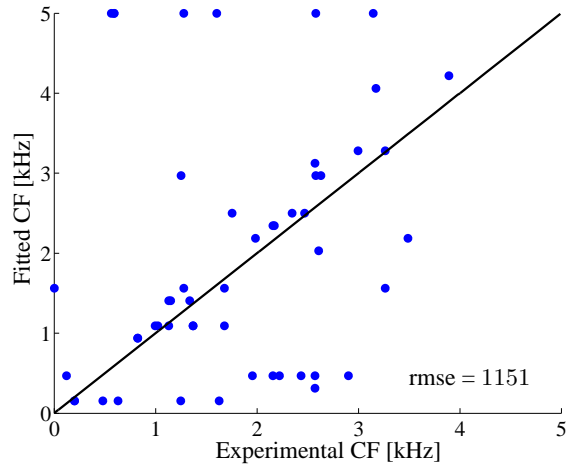


Fig. 3. Scatter plot of the frequencies corresponding to the parameter $\beta_{l,j}$ with the highest value (i.e. the fitted CF) vs. the experimentally measured CF of the corresponding neuron when using the spectrogram as the spectro-temporal representation of the input stimulus. The diagonal line indicates a perfect correspondence.

parameters, corresponding to time, for which we did not have an experimental equivalent.

4. CONCLUSION

In this paper, we analyzed the effect of using three different spectro-temporal representations on the fitting of a point process model of auditory nerve data. It is observed that the three different spectro-temporal representations studied achieve equally well fitted models indicating therefore that they represent equally well the recorded spikes. However, it is found that there is a better correspondence between the experimentally measured characteristic frequency and the one obtained from the model when using the gammatone filterbank than when using the other two representations. This could be due to the fact that the gammatone filterbank extracts more relevant spectro-temporal characteristics of the input auditory stimulus. Therefore, it is concluded that the quality of the estimated parameters can be affected by the choice of the spectro-temporal representation but not so much the goodness of fit of the model.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their helpful comments. They would also like to thank Bertrand Delgutte for providing the data to perform this study as well as Rob Haslinger and Demba Ba for kindly providing respectively their KS plot and TR-IRLS algorithms.

5. REFERENCES

- [1] E. Plourde, B. Delgutte, and E. N. Brown, "A point process model for auditory neurons consider-

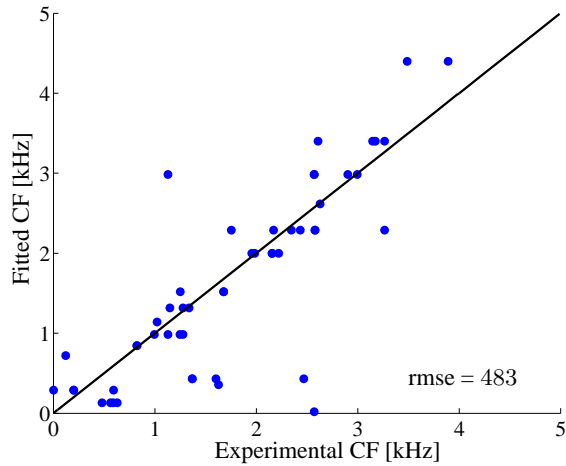


Fig. 4. Scatter plot of the frequencies corresponding to the parameter $\beta_{i,j}$ with the highest value (i.e. the fitted CF) vs. the experimentally measured CF of the corresponding neuron when using the gammatone filterbank as the spectro-temporal representation of the input stimulus. The diagonal line indicates a perfect correspondence.

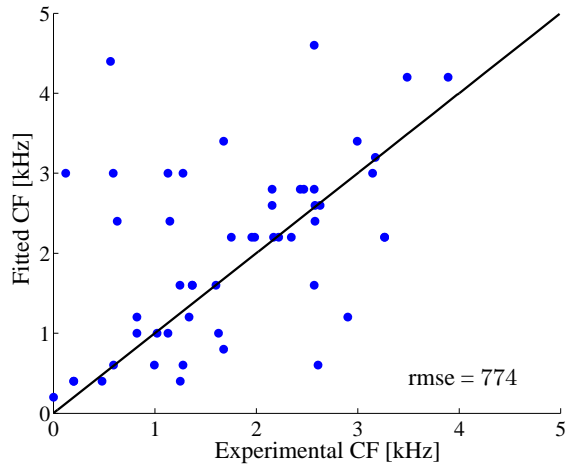


Fig. 5. Scatter plot of the frequencies corresponding to the parameter $\beta_{i,j}$ with the highest value (i.e. the fitted CF) vs. the experimentally measured CF of the corresponding neuron when using the Hilbert spectrum as the spectro-temporal representation of the input stimulus. The diagonal line indicates a perfect correspondence.

ing both their intrinsic dynamics and the spectro-temporal properties of an extrinsic signal,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1507–1510, June 2011.

- [2] A. Calabrese, J. W. Schumacher, D. M. Schneider, L. Paninski, and S. M. N. Woolley, “A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds,” *PLoS ONE*, vol. 6, no. 1, pp. 1–16, January 2011.
- [3] Paul Komarek, *Logistic Regression for Data Mining and High-Dimensional Classification*, Ph.D. thesis, Carnegie Mellon University, 2004.
- [4] B. Delgutte, B. M. Hammond, and P. A. Cariani, “Neural coding of the temporal envelope of speech: Relation to modulation transfer functions,” in *Psychophysical and physiological advances in hearing*, A. R. Palmer, A. Reese, A. Q. Summerfield, and R. Meddis, Eds. 1998, pp. 595–603, London: Whurr.
- [5] P. I. M. Johannesma, “The pre-response stimulus ensemble of neurons in the cochlear nucleus,” in *IPO symposium on hearing theory*, B. L. Cardozo, E. de Boer, and R. Plomp, Eds., Eindhoven, Netherlands, 1972.
- [6] L. H. Carney and T. C. T. Yin, “Temporal coding of resonances by low-frequency auditory nerve fibers: single fiber responses and a population model,” *J. Neurophysiol.*, vol. 60, pp. 1653–1677, 1988.
- [7] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. R. Soc. Lond. A*, vol. 454, pp. 903–995, 1998.