Abstract—We study the impact of different encoding models and spectro-temporal representations on the accuracy of Bayesian decoding of neural activity recorded from the central auditory system. Two encoding models, a generalized linear model (GLM) and a generalized bilinear model (GBM), are compared along with three different spectro-temporal representations of the input stimulus: a spectrogram and two bio-inspired representations, i.e. a gammatone filter bank (GFB) and a spikegram. Signal to noise ratios between the reconstructed and original representations are used to evaluate the decoding, or reconstruction accuracy. We experimentally show that the reconstruction accuracy is best with the spikegram representation and worst with the spectrogram representation and, furthermore, that using a GBM instead of a GLM significantly increases the reconstruction accuracy. In fact, our results show that the spikegram reconstruction accuracy with a GBM fitting yields an SNR that is 3.3 dB better than when using the standard decoding approach of reconstructing a spectrogram with GLM fitting.

I. INTRODUCTION

Understanding the neural representation of an auditory stimulus as well as the decoding mechanisms of the mammalian brain is essential to improve the performance of current auditory implants. One way to improve our knowledge of these aspects is to design and analyze neural decoding algorithms. In neural decoding, the goal is to reconstruct the sound stimulus from the neural signals. Because of the stochastic nature of the neural representations, the Bayesian approach has been widely used for neural decoding [1], [2]. For example, in [3], Ramirez et al. used a Bayesian approach to reconstruct the spectrogram of songbird stimuli using the neural activity recorded from auditory midbrain neurons. In this approach, the instantaneous firing rate of a neuron is obtained through a generalized linear model (GLM) that considers the neuron’s past firings as well as a spectro-temporal representation of the input stimulus.

Ahrens et al. [5] proposed a modification to the GLM where the spectro-temporal representation of the input stimulus is transformed through an additional non linearity function to provide a more efficient matching between the input and the model. This modified GLM was termed the generalized bilinear model (GBM). The GBM is thought to model the nonlinear synaptic or dendritic responses in a neuron. In [5], the GBM was applied to neurons from the somatosensory cortex of rodents and it was shown that the fitted parameters were generally more accurate when using a GBM than a GLM. Considering the differences in the nature of the inputs in the auditory and somatosensory systems, the validation of such an approach in the auditory system remains to be confirmed. More importantly, to the knowledge of the authors, the use of GBMs have not been evaluated in the context of neural decoding.

Furthermore, the spectro-temporal representation involved in the GLM, and its modified GBM version, is usually a spectrogram (e.g. [3], [6], [7]). However, the spectrogram may not be the most efficient representation to study the auditory system since it does not match well the processing performed by the peripheral auditory system. In fact, different spectro-temporal representations including a spectrogram, a gammatone filter bank (GFB) and a Hilbert spectrum were recently studied in a neural encoding setting, i.e. where one aims at modeling the neural activity from the input stimulus. It was shown that the best fitting of the GLM was obtained with the GFB and not the spectrogram [8].

In this paper, we perform Bayesian neural decoding on the neural activity recorded from the inferior colliculus, a structure from the auditory midbrain. To improve the decoding performance, we firstly investigate the use of a GBM instead of a GLM and secondly, compare the decoding obtained when using different spectro-temporal representations, including the spectrogram, the GFB and a novel bio-inspired representation, termed the spikegram [9]. We show that the use of a GBM instead of a GLM improves quite substantially the decoding accuracy of neural activity measured in the auditory system. We also show that the decoding accuracy obtained with the spikegram is higher than when using the GFB. Both of these yield better decoding accuracies than when using the spectrogram.

The paper is organized as follows: in section II, the methodology of the paper is presented including a brief overview of the Bayesian neural decoding approach, a description of the GBM as well as of the studied spectro-temporal representations. Results are presented in section III and section IV concludes the paper.
II. METHODOLOGY

A. Bayesian Neural Decoding

We briefly review here the Bayesian neural decoding approach as proposed in [3]. The maximum a posteriori (MAP) estimate of the spectro-temporal representation of the input stimulus, \( n \), is obtained by maximizing the posterior distribution, \( p(n|s, \theta) \), where \( n \) is the neural activity for all neurons at all times and \( \theta \) are the covariates. This posterior distribution can be evaluated using Bayes’ rule:

\[
p(s|n, \theta) = \frac{p(n|s, \theta)p(s)}{p(n | \theta)}.
\]

Therefore, to evaluate the MAP estimator, one has to find the likelihood of the spike train given the spectro-temporal representation of the stimulus, \( p(n|s, \theta) \), as well as the prior distribution of \( s, p(s) \).

The prior distribution, \( p(s) \), is considered as a spectro-temporally correlated Gaussian prior such that

\[
p(s) = \frac{1}{(2\pi)^{T/2}|C|^{1/2}} \exp\left(-\frac{(\hat{s} - \hat{\mu})^T C^{-1}(\hat{s} - \hat{\mu})}{2}\right)
\]

where \( F \) and \( T \) denote the number of frequency and time bins in the spectro-temporal representation respectively, \( \hat{s} \) denotes the column vector of power density across all time and frequency bins, \( \hat{\mu} \) is the average power density across frequency bins at any time sample and \( C \) is the covariance matrix between all time and frequency bins and is of dimension \( FT \times FT \). All parameters of the prior are estimated from a test set of sounds.

To estimate the likelihood, it is assumed that neurons are conditionally independent given the stimulus such that

\[
p(n|s, \theta) = \prod_{i=1}^{N} \prod_{t=1}^{T} p\left(n_i(t) | s, \theta, n_{i1}(1), n_{i2}(1), \ldots, n_{i(t-1)} \right)
\]

where \( n_i(t) \) is the number of spikes fired by neuron \( i \) at time \( t \) and \( N \) is the total number of neurons. The likelihood for a given neuron \( i \) is modeled as a Poisson distribution:

\[
p\left(n_i(t) | s, \theta, n_{i1}(1), n_{i2}(1), \ldots, n_{i(t-1)} \right) = \exp(-r_i(t)dt)\left(\frac{r_i(t)dt}{n_i(t)}\right)^{n_i(t)}
\]

where \( r_i(t) \) is the instantaneous firing rate at time \( t \) of the \( i \)th neuron. In (4), the variable \( r_i(t) \) is modeled using a GLM with an exponential nonlinearity

\[
r_i(t) = \exp\left(b_i + \sum_{j=0}^{F-1} \sum_{\tau=0}^{M-1} k_i(j, \tau) s(j, t - \tau)
\right)
\]

where \( b_i \) models baseline firing and is independent from the stimulus; \( k_i \) is the linear spectro-temporal receptive field (STRF) [6] of neuron \( i \) with \( M \) time bins \( (M < T) \) and \( F \) frequency bins; \( h_i(j) \) is referred as the history filter of a neuron with length \( J \) and includes information about firing rate, saturation, refractory periods and bursting behavior. The nonlinearity function, i.e. the exponential, ensures that the firing rate will not take negative values.

The posterior distribution in (1) can then be evaluated by using (2) and (3). The different GLM parameters, including baseline firing, STRF and history filter, are fitted to neural spike trains using maximum likelihood (ML). In both ML fitting and MAP estimation, the conjugate-gradient algorithm is applied [4].

B. Generalized Bilinear Model

In the GLM, one nonlinearity function (the exponential in (5)) has been used. In the generalized bilinear model proposed in [5], an additional nonlinearity called the input nonlinearity \( f(\cdot) \) is also applied to the input stimulus \( s(f, t) \), such that the estimated firing rate \( r_i(t) \) becomes

\[
r_i(t) = \exp\left(b_i + \sum_{j=0}^{F-1} \sum_{\tau=0}^{M-1} k_i(j, \tau) f(s(j, t - \tau))
\right)
\]

where the input nonlinearity is parametrized over some piecewise linear functions \( f_i(\cdot) \) as below:

\[
f(\cdot) = \sum_{f} b_i f_i(\cdot)
\]

and the \( b_i \)'s are the coefficients of the input filter. The input nonlinearity is therefore not fixed and the coefficients \( b_i \)'s need to be fitted along with the baseline firing, STRF and history filter. The \( f_i(\cdot) \)'s were chosen according to [5]. Compared to GLM, in GBM, the fitted input nonlinearity models the nonlinear synaptic or dendritic responses in a neuron. Moreover, the addition of nonlinearity to the input provides an additional mean to better match the input to the model. It is shown in [5] that the GBM is a more accurate neural encoding model than the GLM.

C. Spectro-Temporal Representations

The spectrogram is usually the chosen spectro-temporal representation of the input signal in the GLM (\( s(\cdot) \) in (5)), yet it may not be the most suitable one. In this paper, we investigate the use of three different spectro-temporal representations: 1- spectrogram, 2- gammatone filter bank, 3- spikegram, i.e. a bio-inspired representation that models auditory characteristics such as masking and is a sparse representation of the input stimuli. In the following, we briefly describe these three representations.

1) Spectrogram: The squared magnitude of the sliding short-time Fourier transform (STFT), usually in decibel (dB), is called a spectrogram. In the STFT, the signal is windowed with overlapping shifted windows and a fast Fourier transform is applied to each windowed signal. Here a Hamming window with a length of 1.3 ms, a 0.3 ms overlap and 1 ms shift is used. The spectrogram includes 15 equally spaced frequency bins which cover the frequency range between 200 Hz - 24200 Hz.
2) **Gammatone filter bank:** The GFB mimics the filtering of the basilar membrane of the human auditory system [10]. The GFB is obtained through filtering the input signal with a bank of gammatone filters [10]. Here, we use the approach proposed in [10] to generate a gammatone filter bank with 15 channels covering the frequency range from 200 Hz to 24200 Hz.

3) **Spikegram:** Spikegrams have been recently proposed to represent audio signals [12]. It has been shown that the brain has a sparse activity due to energy efficiency constraints [11]. A spikegram is a loose spectro-temporal representation of the neural activity at higher auditory centers that takes into account this sparseness and that considers also the masking of sounds occurring in the peripheral auditory system. The bases used in this representation are obtained through the perceptual matching pursuit (PMP) algorithm that aims at representing the signal with the smallest reconstruction error and the greatest sparsity [9]. The 2D plot of all obtained sparse coefficients in the time channel plane is called a spikegram.

### III. RESULTS AND DISCUSSION

The dataset used in this study consists of multi-unit neural recordings from the inferior colliculus of ketamine-anesthetized guinea pigs; the detailed experimental setup can be found in [13]. Recordings were performed simultaneously at 32 different sites following the successive presentation of 6 different guinea pig vocalizations, having both voiced and unvoiced characteristics, for 20 trials each (Fig. 1 (a) shows the spectrogram of one of these stimuli). The electrode array was composed of two parallel shanks with 16 electrodes each, we labeled the electrodes as no. 1 to 16 for those on the first shank and no. 17 to 32 for those on the second shank.

We evaluate the reconstruction using both a visual comparison of the reconstructed and original representations as well as the Signal to Noise Ratio (SNR) [3] between the original spectro-temporal representation of the sound stimulus $s(t)$ and the reconstructed one $\hat{s}(t)$. The SNR is computed as:

$$\text{SNR} = 10\log_{10} \left( \frac{\sigma_e^2}{\sigma_s^2} \right)$$  \hspace{1cm} (8)

where the reconstruction error $e(t)$ is the difference between the true spectrogram, $s(t)$, and estimated spectrogram, $\hat{s}(t)$, and $\sigma^2$ denotes the variance term. We evaluated the decoding, i.e. the reconstruction of the spectro-temporal representations, for all stimuli. For each reconstructed stimulus, we extracted the prior information and also fitted the GLM and GBM parameters using all stimuli in the dataset excluding the one being reconstructed. In all figures of this section, the information of all 32 recording sites are used for stimulus reconstruction, the time bins are 1 ms. One can notice that there are weak correlations between the time bins in the spectro-temporal representations of Fig. 1 (a), Fig. 2 (a) and Fig. 3 (a). This is to match the poor correlation between the time bins of the recorded neural activity.

![Fig. 1: Spectrogram reconstruction for one of the stimuli where (a) is the input spectrogram and (b) and (c) are respectively the decoded stimulus representation with a GLM and a GBM when using the neural activity from all 32 recording sites. Time and frequency bins equal respectively 1 ms and 1.6 kHz.](image)

**TABLE I:** The reconstruction SNRs (in dB) for all stimuli when using the GLM and the GBM. The results are given as the mean ± one half a standard deviation.

<table>
<thead>
<tr>
<th>Stimulus Number</th>
<th>SNR-GLM</th>
<th>SNR-GBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.62 ± 0.21</td>
<td>5.06 ± 0.27</td>
</tr>
<tr>
<td>2</td>
<td>4.96 ± 0.19</td>
<td>5.10 ± 0.51</td>
</tr>
<tr>
<td>3</td>
<td>6.13 ± 0.30</td>
<td>7.14 ± 0.41</td>
</tr>
<tr>
<td>4</td>
<td>3.09 ± 0.08</td>
<td>4.33 ± 0.17</td>
</tr>
<tr>
<td>5</td>
<td>5.08 ± 0.29</td>
<td>5.79 ± 0.34</td>
</tr>
<tr>
<td>6</td>
<td>3.72 ± 0.22</td>
<td>4.24 ± 0.28</td>
</tr>
</tbody>
</table>

### A. GBM vs. GLM

First, we compare the decoding results when using a GBM vs. a GLM for a spectrogram representation. In Fig. 1, the decoding result for the spectrogram reconstruction of one of the stimulus is shown where (a) shows the spectrogram representation of the input stimulus and (b), (c) show the decoded stimulus using a GLM and a GBM respectively. As can be observed in Fig. 1, the decoded spectrogram using a GBM (Fig. 1 (c)) is much more similar to the original spectrogram (Fig. 1 (a)) than the decoded spectrogram using GLM (Fig. 1 (b)). In order to evaluate quantitatively the spectrogram reconstruction when using a GBM vs. a GLM, we computed the SNRs as given by (8). Table I presents the SNRs for each of the 6 stimuli when using the GLM and the GBM. As can be observed, a better reconstruction is obtained for all cases when using the GBM vs. the GLM; the average SNR increasing from 4.60 dB for the GLM to 5.27 dB for the GBM.
B. Reconstruction Results for Different Spectro-Temporal Representations

In this subsection, we compare the reconstruction results for different spectro-temporal representations including the spectrogram, the GFB and the spikegram representations. In Fig. 2, the reconstruction when using a GFB is shown for the same stimulus as in Fig. 1. As is seen, the reconstructed GFB, especially when using a GBM, is more similar to the GFB of the original signal than when comparing the reconstructed spectrogram with the original spectrogram. One reason could be that the bases are gammatone which mimics the auditory filter bank of the mammalian ear, hence these bases produce a representation that better corresponds to the processing performed by the peripheral auditory system. Moreover, the reconstruction obtained using the GBM is again better than the one obtained when using the GLM. In fact, the mean SNR using the GLM is 5.30 dB and the one using the GBM is 6.26 dB.

Fig. 3 shows the decoding result for the same stimulus as presented in Fig. 1 and Fig. 2 but when using a spikegram representation. As can be observed, the reconstruction obtained, especially when using a GBM, is again better than when using a GLM. Moreover, for the spikegram reconstruction, the average SNR obtained when using a GLM is 6.62 dB while for the case of using a GBM it is 7.91 dB indicating a better reconstruction accuracy than when using either a spectrogram or a GFB.

C. Reconstruction Results When Using Neural Activity Recorded from Different Number of Electrodes

To further compare the decoding accuracy of GBM vs. GLM using the different spectro-temporal representations, we present in Fig. 4, the mean SNR results (over the 6 input stimuli) as a function of the number of recording sites (corresponding to the frequency resolution) used for neural decoding. Note that in Fig. 4, for each recording site (neuron) on the horizontal axis, the neural activity corresponding to 20 trials are used for neural decoding. In particular, the magenta color indicates the averaged SNR for the spikegram reconstruction, the green color indicates the averaged SNR for the GFB reconstruction and the blue one indicates the averaged results for the spectrogram representation. Moreover, the solid and star lines represent the reconstruction performed with GBM and GLM respectively. As can be observed, the averaged SNR results are always lower for the GLM than for the GBM irrespective of the number of electrodes (i.e. of the frequency resolution). Also, the reconstruction of spikegrams is always better than that of GFB and spectrograms. The fact that the decoding accuracy when using either a GFB or a spikegram is significantly better than when using a spectrogram was expected since these are two bio-inspired representations and, as such, should yield spectro-temporal representations that are closer to the ones encoded in the central auditory system [9], [11].

It is to be noted that the improvement in SNR is greater when adding the first 16 electrodes than the last 16 ones. This
is due to the fact that the inferior colliculus is organized in a laminar structure where neurons on a given lamina code for the same characteristic frequency, i.e. they respond to the same frequency. Due to the two shank construction of the electrode array (recording sites), electrode no. 1 and no. 17 were approximately on the same lamina and therefore encoded the information for similar frequencies, the same being true for electrodes 2 and 18, 3 and 19, etc. The addition of the last 16 electrodes therefore does not add new frequency information to the model yet a slight improvement in SNRs is still observed when adding those last electrodes. This improvement could be the result of a slight misplacement of the shank yielding a small difference in the encoded frequencies of the first 16 electrodes compared to the last 16 ones and therefore slightly increasing the frequency resolution. It could also be due to the addition of temporal information from the second shank. In fact, differences in temporal patterns were observed for different regions of a given isofrequency lamina.

IV. CONCLUSION

We compared the decoding, or reconstruction accuracy for a Bayesian decoding approach of the auditory system’s neural activity 1) when using a GBM vs. a GLM and 2) when using three different spectro-temporal representations (spectrogram, GFB and spikegram). It is shown both qualitatively and quantitatively (using SNRs) that the reconstruction accuracy with a GBM is better than the one obtained when using a GLM. Moreover, it is also observed that the reconstruction accuracy for the spikegram is better than for the GFB, which are both better than when using a spectrogram representation. This could be due to the fact that the spikegram extracts more relevant spectro-temporal characteristics of the input auditory stimulus than the two other representations. Therefore, it is concluded that the quality of the reconstruction is directly affected by the choice of the spectro-temporal representation and the type of model (GBM vs. GLM), the best results being obtained for the GBM with a spikegram representation of the input stimulus. In fact, when using a spikegram representation with a GBM encoding model, the SNR of the reconstruction is improved by 3.3 dB compared to the standard approach using a GLM with a spectrogram representation of the input stimulus. Our work shows that the GBM offers a great potential for auditory neural decoding and that bio-inspired spectro-temporal representations such as GFB and spikegram should be used instead of the conventional spectrogram.

ACKNOWLEDGMENT

The authors would like to thank H. Najaf-Zadeh for the PMP code as well as J. Pillow for the GLM toolbox and neural decoding code.

REFERENCES


