Imperial College London

Final Report for MEng Project

Selectivity properties of central auditory neurons

Author:
Raymond Cadoc Miles

Supervisor:
Dr. Andrei Kozlov

Submitted in partial fulfilment of the requirements for the award of BEng/MEng in Biomedical Engineering from Imperial College London

May 2021 Word Count: ....

Feedback box for project markers:

What I liked about the report:
What could/should be improved:

Abstract

In this report we will investigate the relation between the auditory stimuli and the neural response through the extraction of Spectrotemporal receptive fields (STRFs) from neurons in the mouse auditory cortex. This will be done by comparing maximal entropy models (MNE) obtained from either the resized spectrogram, sparse filtered features or reduced dimension features obtained through Uniform Manifold Approximation and Projection (UMAP). The MNE model was also used to produce spiking rates and will be compared to the real spike data from the neurons.

The three methods will attempt to identify what features of the audio input the neurons are tuned to as well as observe if relevant neuronal receptive fields can be extracted with such algorithms.

Acknowledgements

Heartfelt thanks to Dr Andrei Kozlov and Sihao Lu for supervision and feedback throughout the project

Contents

1 Introduction

The characterisation of the encoding of stimuli in the nervous system is a fascinating area of research. Since meaningful stimuli are a smaller subset of all possible stimuli reducing the dimensions of the recorded stimulus using different representations may help to identify relevant aspects of the stimulus. Mice have been shown to emit and react to ultrasonic vocalisations[9] the study of which may have implications for other mammalian models.

This fact linked with USVs being modulated and having behavioural effects in different social contexts makes this a fascinating problem to decode and extract significant lower dimensional representations from tuned neurons in the auditory cortex thus shining a light on this process.

The project will leverage the Maximum Noise Entropy method, a statistical inference method that works well with natural stimuli to produce models of the conditional response distribution [16].

Three MNE based models will be trained by using different representations of the stimulus, either using the spectrogram of the sound signal, the sparse filtered spectrogram or a reduced dimensional representation of the signal using Uniform Manifold Approximation and Projection. The model will be evaluated by producing a spiking frequency output in response to a test set that will be correlated with the real neuronal spiking for that stimulus.

2 Dataset and Ethical Concerns

As was stated in the planning report [20] the data used in this report originates from the Kozlov lab at Imperial College London and was extracted through in vivo experiments on mice handled by Dr. Mark Steadman who has the required licenses under the Home Office. The audio stimuli were obtained from Chabout et al and include the recordings of male mice that were presented with various stimuli [8] . The neural response data used in this project were obtained from extracellular recordings in the mice auditory cortex. Recording electrodes were inserted in the cortical surface between A1 (primary auditory cortex) and A2 (secondary auditory cortex). The mice were anesthetised with fentanyl, midazolam and medetomidine. The data obtained was recorded and pre-processed by Dr Mark Steadman.The project will only deal with the aforementioned dataset and there is no additional data used or further ethical concerns regarding the dataset.

3 Background

3.1 USVs

Murine USVs are commonly used as a metric of social interest, social behaviour and motivation in mice models of human diseases and disorders such as autism, and communication disorders amongst others [7, 11]. Additionally they have been used as markers of condition in a variety of diseases including Parkinson's and ataxia [24]. Mice produce discrete sound units in their ultrasonic vocalisations (USVs) termed syllables by Holy and Guo [14] in the range of 30-110kHz with repeated discrete frequency transitions present.

USVs have been seen to be classifiable by the temporal features of the vocalisation, change from infancy to adulthood. USVs have been observed to be context dependent, for example, with differences in syllable duration and complexity recorded in the male's song in the presence of females or in the presence of female urine alone with female mice

preferring one call over another[6]. Some such calls were included in the USV dataset used [8]. Another example of context dependent USV is observed during pup displacement from the litter, with the pup displaying a stress induced alarm call that results in the mother searching for the pup's location and retrieving it to the nest [9].

These are evidence to a certain level of informational content to the USVs, with some evidence even pointing to limited vocal modification abilities [3].

The hearing range of mice for the purposes of removing inaudible frequencies from the recordings was assumed to be from 2KHz to 88KHz for the 60dB recording from the paper Koay et al[15].

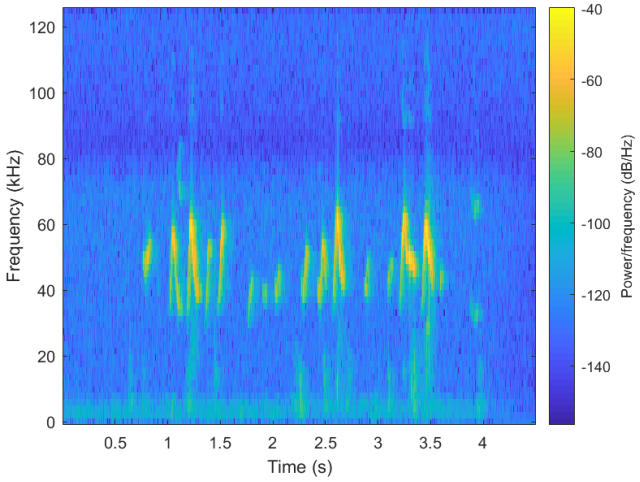


Figure 1: *An image of an USV spectrogram extracted from the database*

3.2 Receptive Fields

Receptive fields are portions of the sensory space that can elicit neuronal responses when stimulated. There have receptive fields identified in the visual, somatosensory auditory and even olfactory sensory domains.

Auditory spectrotemporal receptive fields (STRFs) are receptive fields that are sensitive to audio signals over time that can evoke neuronal responses, and while responses may be quantified by sub-threshold activity of the neuron this report will only consider action potentials as a response. STRFs may be complex and can contain many excitatory and inhibitory features as has been shown in starlings composite receptive fields. [16]

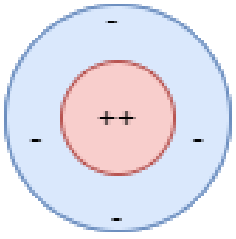


Figure 2: *An example of a on-centre off-surround receptive field similar to those found in Retinal Ganglion cells in the visual system*

The receptive field of a neuron is the set of stimuli that modulate its responses [25], and while mapping may be done from all possible stimuli to the neuronal responses, the field that deals with the approximation of this relationship is neural coding.

One framework for the approximation of the receptive field and estimation of the response function is the Linear-nonlinear (LN) model. The LN model is comprised of two stages: a linear stage in which the stimulus is projected into a set of vectors/filters and a second stage where a nonlinear transformation yields a function of the linear projections of the first states that results in a firing rate that may be further elaborated by feeding into a spiking generator for a Poisson firing.

To obtain the linear projections there is a variety of methods amongst which there are two large classes: the Spike triggered characterisation methodologies and information-theoretic methods.

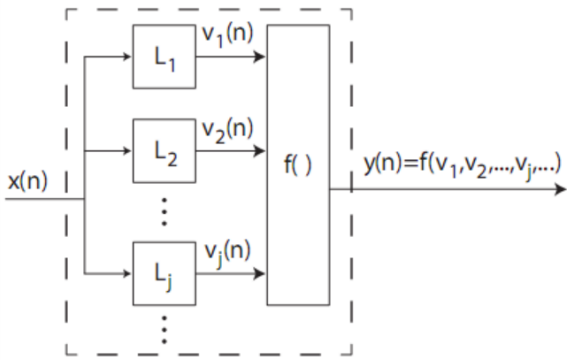


Figure 3: *An example of a LN model schematic. The stimulus x(n) is processed by Lj linear filters and their output is sent to a nonlinear gain function f() producing a response y(n) that may be fed in to a spike generator. The figure was adapted from Marmarelis 1989 [18]*

3.3 Spike Triggered Characterisation

A neuron that is responsive to a stimulus will fire upon receiving features from the said stimulus and therefore such features must be correlated to some measure with the neuronal response.

Spike-weighted moments of the stimulus such as spike-triggered average (STA) and spike-triggered covariance (STC) can be used to compute an estimate of the receptive field of a neuron.

STA takes the average of the stimulus weighted by its response (equation 1)

$$\phi_{sta} = \frac{1}{n_t} \sum_t n(t)s(t) \tag{1}$$

In equation 1 n(t) is the number of spikes at time t, n_T is the total number of spikes, s(t) is the 0-mean stimulus preceding the spike. STA can only extract a single component of the receptive field to account for multiple components [1] while additional second order moments of the feature distribution may be considered using techniques such as the STC.

STC is obtained by taking the difference between the spike-triggered covariance matrix equation 2 and the stimulus covariance matrix in equation 3.

$$C_p = \frac{1}{M-1} \sum_t s(t)(s(t))^T \tag{2}$$

$$C_s = \frac{1}{n_T - 1} \sum_t n(t)(s(t) - \phi_{sta})(s(t) - \phi_{sta})^T \tag{3}$$

The difference matrix is computed and eigen-decomposed obtaining $C_{diff} = \Omega\Lambda\Omega^T$ with eigenvectors Ω and Λ eigenvalues that indicate the variances of the spike-triggered distribution. Positive eigenvalues correspond to a larger variance of the spike-triggered distribution than the underlying stimulus distribution and negative eigenvalues correspond a smaller variance compared to the same.

Such spike-triggered methods depend on the stimuli being either spherically symmetric stimuli distributions (STA) or an uncorrelated Gaussian distribution (STC). Despite the prolificacy of Spike-triggered methods (STICA etc) natural stimuli that are often correlated and non-Gaussian, and although methodologies to decorrelate the inputs have been attempted [2], information theoretic methods were developed that can overcome such shortcomings. For example, the Maximally Informative dimensions method that can be applied to non-Gaussian stimulus distributions, however, is computationally demanding as the number of features increase. Another information theoretic methodology is the Quadratic Mutual information method [27], that provides a cost function for optimization that is more computationally viable than the Mutual Information, that is maximised when the mutual information is also maximum.

3.4 Spike Triggered Characterisation

4 Methods

4.1 Spectrogram

The spectrogram of the Ultrasonic Vocalisations and Starling song were calculated through the short-time Fourier transform with a Hanning window of 128 points and 64 overlapped samples and converted to decibels as can be seen in figure 15 in the appendix.

The inaudible frequencies below 2kHz and above 88kHz for the mice were removed [15] from the spectrogram (which also removed frequencies above the nyquist range of 125kHz). After this the spectrogram was downsampled with bicubic interpolation in the frequency axis and time axis before further processing, since ultrasound vocalisations syllables are around 30 to 200ms [14], This was taken into account for downsampling. An equivalent data preparation stage was undertaken to downsample spectrograms for the Sparse Filtering algorithm. The datapatches were assembled into the column-list form required by the Sparse Filtering algorithm and each column was normalised to have 0 mean and standard deviation.

4.2 Sparse Filtering

The sparse filtering algorithm [21] was used to extract features from the USVs. Sparse filtering is an unsupervised learning algorithm that optimises the sparsity of L2-normalised features while avoiding explicit modelling of the data distribution. It takes in a data matrix that has rows representing observations (inputs) and columns representing features (weights). The algorithms minimizes the objective function 4 using the a limited memory BFGS function optimizer [17] .

$$\sum_{i=1}^{M} \left\| \hat{\mathbf{f}}^{(\mathbf{i})} \right\|_1 = \sum_{i=1}^{M} \left\| \frac{\tilde{\mathbf{f}}^{(\mathbf{i})}}{\left\| \tilde{\mathbf{f}}^{(\mathbf{i})} \right\|_2} \right\|_1 \tag{4}$$

The desired feature matrix will have few active elements per column (population sparsity), few active features for each example (lifetime sparsity) and the distributions of each

row should be similar (high dispersal) to avoid having features that contribute significantly differently than others and thus may provide less information.

The algorithm first normalises each feature to have similar activity by dividing each feature by its l-2 norm across the columns (making each feature equally active) and then we normalise the features for each example so they lie on the unit l2 ball. Finally, the L1 penalty is used to optimize for sparseness.

Since the features are normalised, if one were to increase, the other components would decrease accordingly due to the L1 norm being minimised. Therefore, some features will be large while most will be close to 0 thus enforcing population sparsity. Normalising examples constrains the expected square value thus enforcing high dispersal. Finally, optimising for population sparsity and dispersal was shown to lead to lifetime sparsity [21].

$$\hat{\mathbf{f}}^{(\mathbf{i})} = \tilde{\mathbf{f}}^{(\mathbf{i})} / \left\| \tilde{\mathbf{f}}^{(\mathbf{i})} \right\|_2 \tag{5}$$

$$\tilde{\mathbf{f}}_{\mathbf{j}} = \mathbf{f}_{\mathbf{j}} / \left\| \mathbf{f}_{\mathbf{j}} \right\|_2 \tag{6}$$

In summary the objective function to minimize is:

$$\sum_{i=1}^{M} \left\| \hat{\mathbf{f}}^{(\mathbf{i})} \right\|_1 = \sum_{i=1}^{M} \left\| \frac{\tilde{\mathbf{f}}^{(\mathbf{i})}}{\left\| \tilde{\mathbf{f}}^{(\mathbf{i})} \right\|_2} \right\|_1 \tag{7}$$

Similar encoding principles can be found in neural circuits in the form of divisive normalisation and sparseness. Sparseness is a principle where each neuron responds to a relatively small number of stimuli and few neurons in the population are activated by said stimuli. Sparseness has been found in cortical responses and has been observed to enhance the ability of downstream neurons to perform classification tasks [4]. Divisive normalisation, where neural activity is suppressed in proportion to the activity of its surrounding group of neurons, is believed to be a neural computation principle that is found in many brain regions and applied to different sensory inputs. [5] Using sparse filtering with a neural network has been shown to produce receptive fields that have similar characteristics to receptive fields found in starling auditory neurons [16] and thus was chosen as a method to reduce the dimensionality of the spectrogram data and to train an MNE model.

4.3 UMAP

UMAP is a dimensionality reduction technique that fits the data to a Riemannian Space in which the data would be uniformly distributed. In this space the connections between datapoints is probabilistically weighted by their distance on the manifold. The embeddings are calculated whilst trying to preserve as much of the topological structure of the original graph as possible. This technique has been shown to scale more efficiently to significantly larger datasets than the previous state of the art dimensionality reduction technique t-SNE [19] finding applications in machine learning materials science and biological research [19].

Therefore, it is well suited to reduce the dimensionality of long spectrogram recordings. UMAP has also been shown to be able to provide lower dimensional clustering that may provide information on the characteristics of the stimuli. It has been used on various animal vocalisations by Sainburg et al [23] and was found to successfully cluster them into low dimensional latent spaces while maintaining information on the species, population and even the individual responsible for the vocalisation. Additionally, UMAP may be

used to extract representations on the stimuli before feeding the data into the models to further characterise input and output relationships as was done here.

The UMAP algorithm as expressed in the original Python implementation by McInnes et al [19] as well as in the utilised Matlab implementation [10] takes in 4 main hyperparameters:

- N the number of neighbours chosen to approximate the local metrics
- D the target embedding dimensionality
- Min-dist the separation between points in the embedding space
- N-epochs the number of training epochs to use when optimizing the reduced-dimension representation

N is the local scale in which the manifold is approximated to be Euclidean. If N is chosen to be too high, manifold features that occur at a smaller scale than within the N nearest-neighbours of points will be lost, while large scale manifold features might not be captured as accurately when they can't be accurately observed by patching together charts of the scale of N. This parameter determines a trade-off between large-scale and small-scale features of the data.

The target embedding dimensions hyperparameter was used to produce the dimensional input required for the MNE which was chosen to be the same as the dimensionality of the other methodologies. The number of epochs was tuned experimentally at 300.

Min-dist governs the appearance of the embedding by determining the distance between points in the lower dimensional representation. Low values produce higher packing density which can be more accurate but harder to visualise.

4.4 MNE

Another method for receptive field analysis is the Maximal Noise Entropy method that doesn't require uncorrelated Gaussian stimuli and can extract any number of features a neuron is tuned to without suffering from the curse of dimensionality the Maximally Informative Dimensions method is subject to.

MNE minimises the mutual information between the response y and the stimulus s (given by equation 8) thus limiting bias from arbitrary stimulus distributions.

$$I(y;\boldsymbol{s}) = \sum_{y}\sum_{\boldsymbol{s}} P(\boldsymbol{s})P(y \mid \boldsymbol{s}) \log\left(\frac{P(y \mid \boldsymbol{s})}{P(y)}\right) \tag{8}$$

The mutual information can be written in terms of the response entropy and the noise entropy as $I(y;s) = H_{resp}–H_{noise}$. Since P(y) is determined by averaging the response and Hresp is defined in equation 9 to only depend on P(y), the entropy of the response is fixed. To minimize the mutual information the noise entropy (equation 10) must thus be maximised (with respect to the conditional probability $P(y|s)$). To be able to determine $P(y|s)$ the statistics of the model are set so as to match the mean, the STA and the STC of the data.

$$H_{resp} = -\int \mathrm{d}y P(y)\log(P(y)) \tag{9}$$

$$H_{\text{noise}} = -\int P(\boldsymbol{s})d\boldsymbol{s}\int P(y \mid \boldsymbol{s})\log(P(y \mid \boldsymbol{s}))dy \tag{10}$$

Assuming a binary response as well as a second order MNE model, the logistic function (equation 11) gives us the probability of spiking.

$$P(y = 1|s) = \frac{1}{1 + e^{(a+\boldsymbol{h}\cdot\boldsymbol{s}+\boldsymbol{s}^T\ \mathrm{J}\boldsymbol{s})}} \tag{11}$$

The unknown weights may be found by maximising the log likelihood of the data which has been shown to be convex and therefore guaranteed to converge.[12] The data is divided into training and testing sets, with overfitting avoided through early stopping.

5 Results

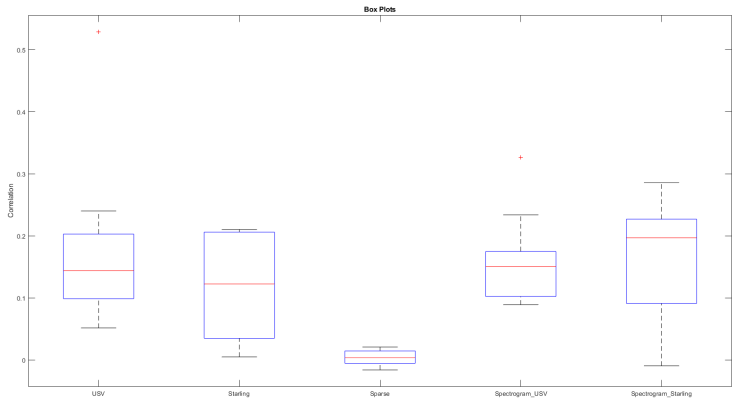


Figure 4: *Boxplots of the correlation coefficients for all the methods.*

The three MNE models were used to create spiking data obtained from equation 11, that was then related to the neuronal data by calculating the correlation between the two rates of fire. The models will be referenced in terms of handling of their input data as Sparse MNE, Spectrogram MNE and UMAP MNE. The MNE obtained from the down-sampled spectrogram gave us the correlations in figure 5 from which the 15 best performing Neurons, defined by having the highest correlation coefficients when exposed to the USVs, were extracted and used for the following analysis.

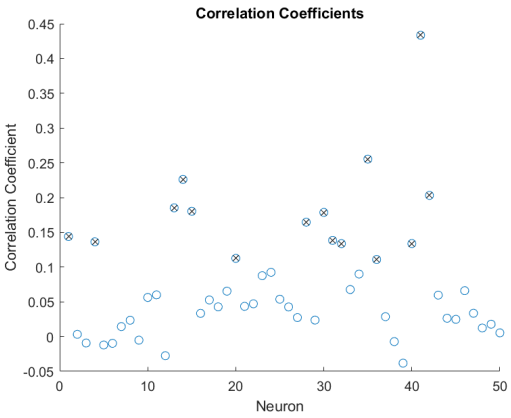
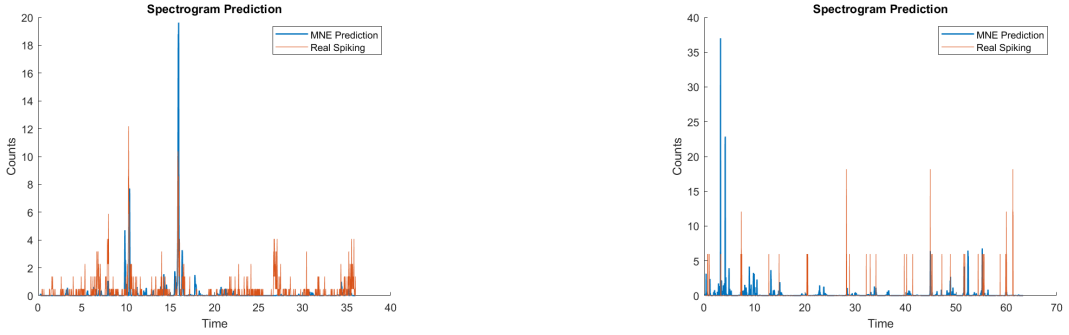


Figure 5: *The correlations coefficients for all 50 units with the 15 best performing neurons are displayed with neurons selected for further analysis identified by the crosses. The correlations were obtained by the spectrograms as inputs*

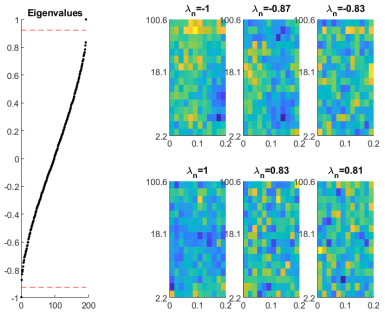


(a) Best Spectrogram Prediction.

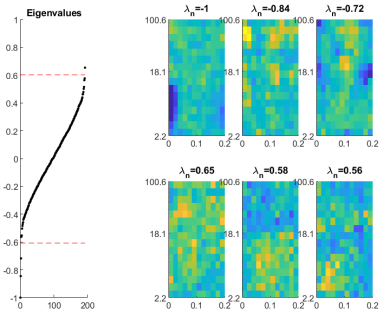
(b) Worst Spectrogram Prediction.

Figure 6: The best and the worst predictions from the Spectrogram MNE Model, with the predicted firing rate plotted alongside the observed firing rate for a mouse USV

The spectrogram MNE model was hypothesized to be the best performing due to the limited amount of manipulation on the data conserving the most structure therefore it was used to select the neurons to further characterise the dataset with the global mean of USV the dataset being found to be 0.0726 with standard deviation 0.0687 as opposed to the Starling Song that had mean 0.05 with standard deviation of 0.089.



(a) Highest Spectrogram Correlation.



(b) Lowest Spectrogram Correlation.

Figure 7: The 6 most significant excitatory features (top row) and inhibitory features(bottom row) for the best Spectrogram prediction (a) and the worst Spectrogram prediction (b). The EigenSpectrum is plotted on the right side of the images.

The dotted lines of the Eigenspectrum refer to the largest magnitude values obtained from 500 random Gaussian matrices with the same means and variances as the matrix J (that gives an indication of feature strength) and therefore values outside that range are likely to have appeared outside of chance.

5.1 Spectrogram Model

Mean correlation for USV throughout the reduced dataset was 0.1507 with a standard deviation of 0.0649, while the mean of the correlation for starling song was 0.1628 with standard deviation 0.1014 for the 6 neurons that were part of the dataset and responded to starling song.

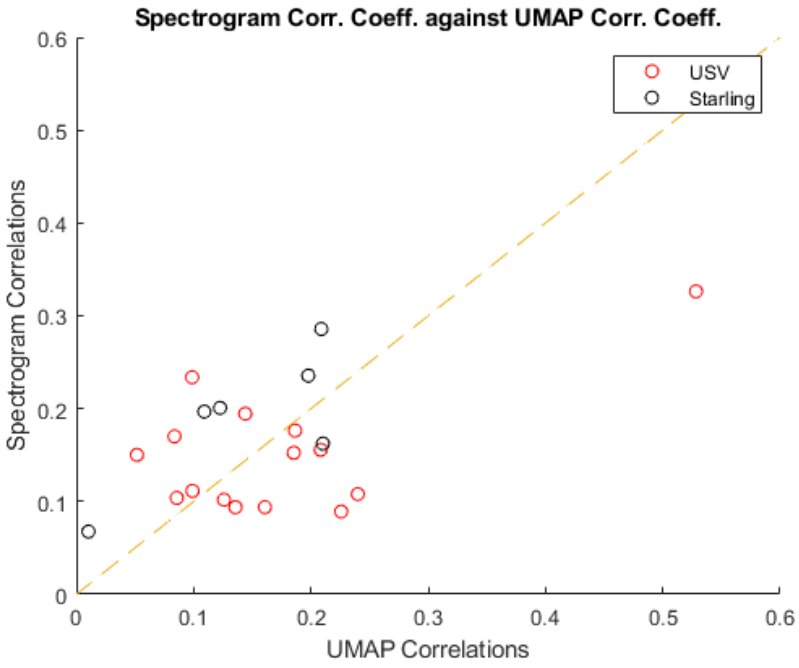


Figure 8: *Spectrogram Correlation Coefficients against UMAP Correlation Coefficients.*

5.2 Sparse Filtering Model

The sparse filtering model performed below expectations due to the training being insufficient, using larger sound files or multiple sound files would possibly improve the performance by extracting more meaningful representations of the sounds.

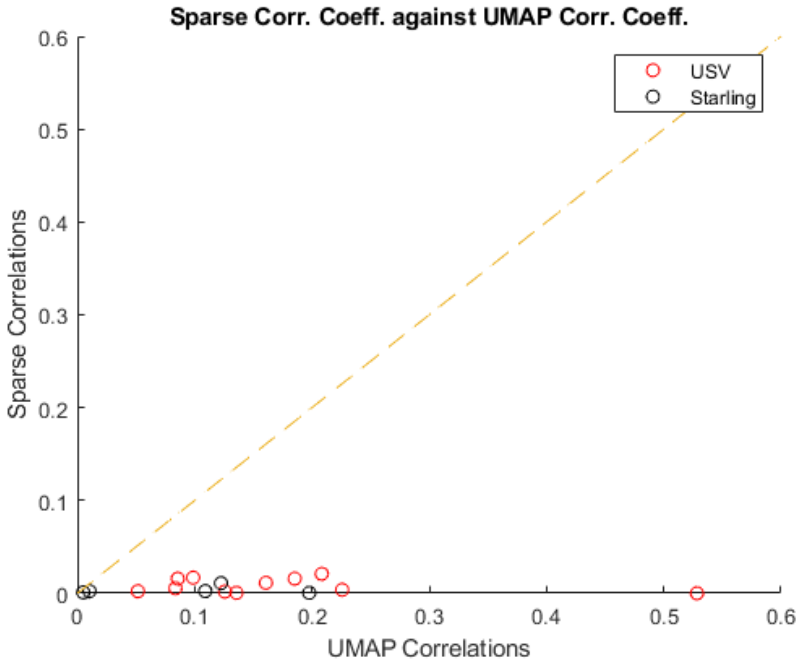


Figure 9: *Correlation Coefficients obtained from Sparse MNE against Correlation Coefficients obtained from UMAP MNE*

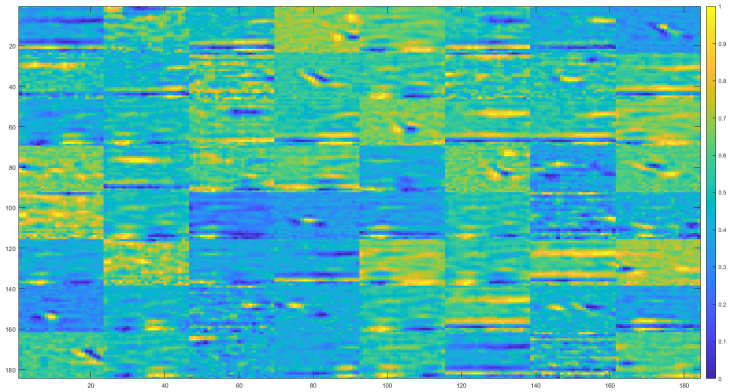
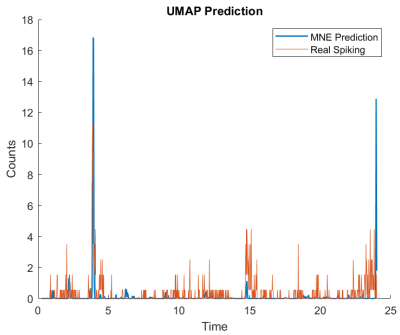


Figure 10: *An example of a set of features extracted from the spectrogram from Sparse Filtering, the shape of certain parts of the patches presents angles that are reminiscent of the slopes in the USVs*

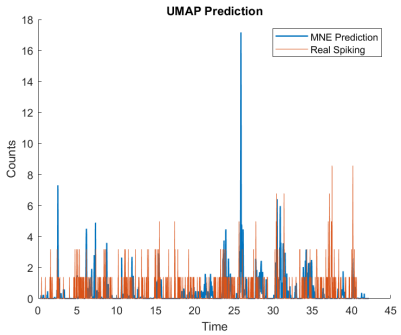
5.3 UMAP Model

The UMAP model performed above expectations predicting the real data with higher correlations that even the model built on the spectrogram itself. An example of a high

correlation prediction and a low correlation prediction can be seen in figure 11a and 11b respectively.



(a) Best UMAP Prediction.



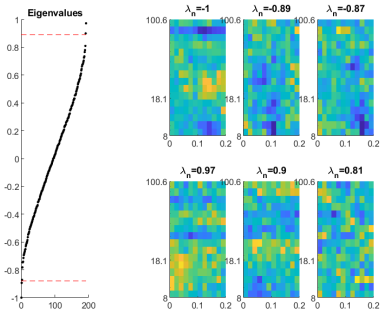
(b) Worst UMAP Prediction.

Figure 11: The best and the worst predictions from the UMAP MNE Model, with the predicted firing rate plotted alongside the observed firing rate for a mouse USV

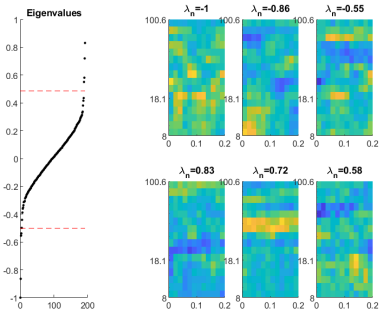
The mean correlations coefficients went from 0.0516 ± 0.098 to 0.5283 ± 0.003 for the worst and best performing neurons exposed USVs and the full population examined had a mean of 0.1706 ± 0.1137. The performance was significantly poorer for the starling song with the average correlation being 0.1211 ± 0.0902 with best and worst performances being of 0.2217 ± 0.0182 and only 0.008 ± 0.0184. A paired t-test was used in the test obtained with a value of 0.14

Model	Mean	Standard Deviation
Spectrogram	0.1507	0.0649
Sparse Filtered	0.016	0.0114
UMAP	0.1706	0.1137

Table 1: Correlation Coefficient statistics for the 15 neuron ensemble



(a) Best UMAP Prediction.



(b) Worst UMAP Prediction.

Figure 12: The 6 most significant excitatory features (top row) and inhibitory features (bottom row) for the best UMAP prediction (a) and the worst UMAP prediction (b). Eigenspectra are displayed on the left hand side of each figure

6 Discussion

The MNE based methods show that the spectrogram does contain information that may be used to inform predictive MNE models. However the representation is likely incomplete as indicated from the mean values of the Correlation Coefficients as shown in Table 1.

From the correlation coefficients obtained from the predictions of the MNE models, it can be seen that the highest correlation values are obtained when the neuron has a less uniform distribution of firing rates over time, and fires in more discrete bursts as may be observed in Figure 12a and 12b. This could be due to the fact that the neurons that do not change their firing rate during the experiment may not be responsive to the features of the stimulus but to other independent processes not present in the audio data.

One particular neuron that was interesting as an outlier had a particularly high correlation value probably because it had discrete areas of increased activity that the model was able to identify 14. Additionally, the said neuron that can be seen in figure 8 on the right of the graph had better performance to USVs trials as opposed to Starling Songs suggesting it may be selective to the USVs.

An interesting phenomenon is the correlation being higher for starling songs than USVs in certain neurons. However, due to the low number of neurons observed for the starling song trial, it might be an anomaly and needs further study. The correlation coefficients electing neurons that performed best for starling song may have excluded certain neurons that performed slightly better when exposed to starling song than to USVs. The reason these neurons would perform similarly to the USVs and not significantly lower would suggest that the receptive fields are responding to similarities in the signal that would be useful to quantify.

The further experimentation on the methodologies used will require further fine tuning, particularly in the dimensionality reduction stages. One possible reason the UMAP MNE method performed so well may have been due to the fact the data was not downsampled before being processed by the algorithm unlike the other methods. The spectrogram MNE method also fared particularly well as was expected from the fact the representation would not change particularly from the original signal and thus preserve relevant characteristics.

Sparse filtering to extract the weights to train the MNE model yielded correlations close to 0. This may be due to the Sparse filtering algorithm requiring more data to fit the weights or might be due to an incorrect training epoch number. With more data being used to obtain the sparse representation the results would likely improve and be more generalizable. The Sparse Filtered weights may have been fit the noise due to number of epochs for the small amount of data some examples of the sparse filtered weights are in figure 14 and figure 10.

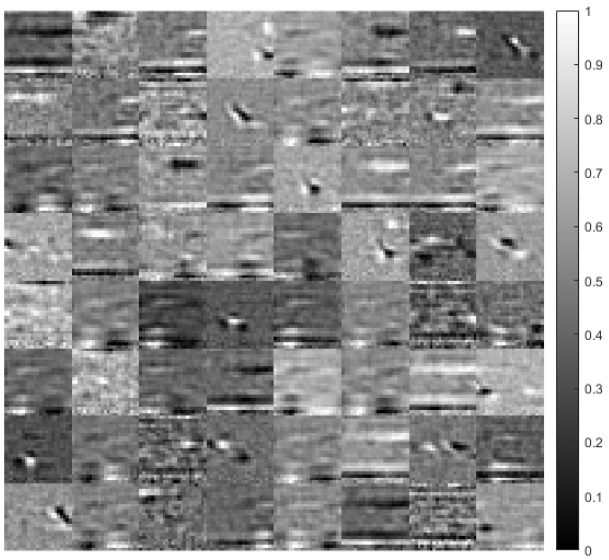


Figure 13: *An example of Sparse Filter features that contained low frequency noise*

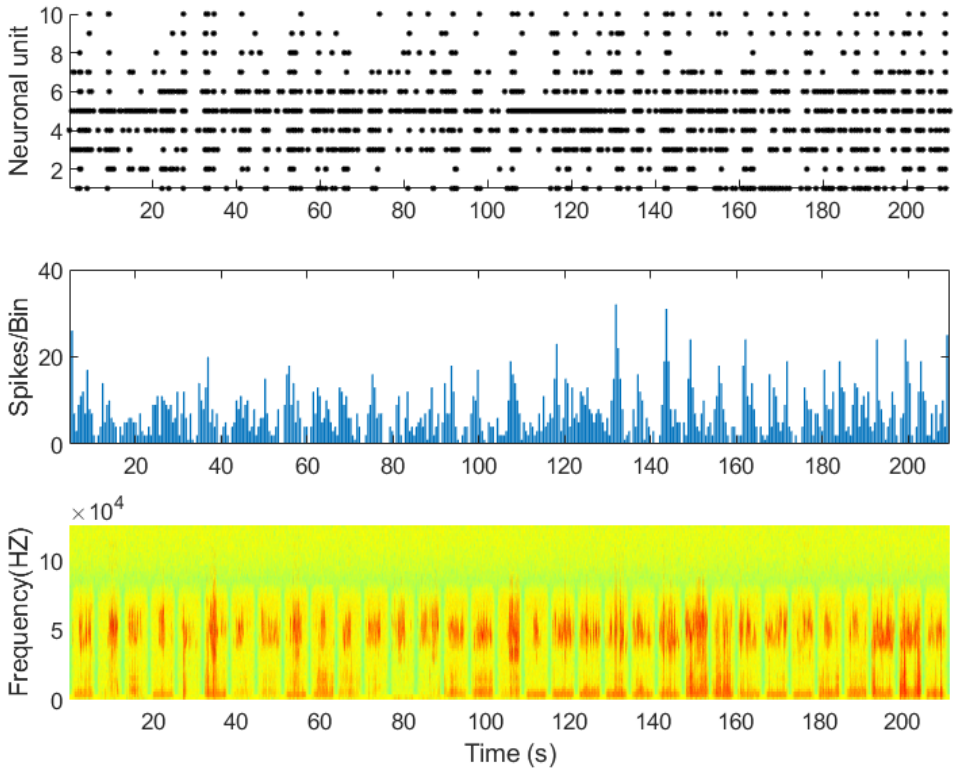


Figure 14: *Raster plot, Peristimulus time histogram and Spectrogram of the outlier neuron*

7 Future work

Due to time constrains many exciting avenues were not explored. There is a large number of parameters that would have been interesting to study further such as modifying the spectrogram calculation, changing window types, overlaps, time lengths, or tuning the number of features extracted with the sparse filtering algorithms.

The bin sizes for peristimulus time histograms that were used to train the MNE and to extract the results have a significant effect on the correlations. And the resizing of the spectrum may remove some fine detail tuned neurons may be sensitive for, so tuning the dimensionality reduction must be done with a view to preserve as much of the detail of the signal as possible.

Future work would include the aformentioned parameter analysis as well as the implementation of other predictive methodologies. Further data acquisition may be useful to explore as well as the production of modulation spectra calculated through the 2D Fourier Transform of the autocorrelation of the spectrogram [26] to study the statistical properties of the features obtained. Pooling features into additional layers with neural networks [13] may expose some new characteristics of the receptive fields although the complexity of further layers of the mice's receptive field as a non-vocal learner may not be meaningful.

Further statistical characterisation of both the predictions and of the features of the receptive fields must be undertaken using Kolmogorov-Smirnov tests for instance and further t-tests.

An avenue of interest would be using other dimensionality techniques on the data such as autoencoders or variations on the algorithms used such as the parametric UMAP [22] .

8 Conclusion

In conclusion, Maximal Noise Entropy methodologies were shown to be relevant for the characterisation of the input-output relationship of neurons and to yield spectrotemporal receptive fields that have predictive significance. While this has been shown through the average values of the correlation coefficients and for a limited set of parameters, the methods show promise and may be used and expanded to study the audio features neurons are selective for.

References

[1] Johnatan Aljadeff et al. "Analysis of Neuronal Spike Trains, Deconstructed". In: *Neuron* 91.2 (2016), pp. 221–259. ISSN: 0896-6273. DOI: https://doi.org/10.1016/j.neuron.2016.05.039. URL: https://www.sciencedirect.com/science/article/pii/S0896627316302501.

[2] Johnatan Aljadeff et al. "Spike Triggered Covariance in Strongly Correlated Gaussian Stimuli". In: *PLOS Computational Biology* 9.9 (2013), e1003206. DOI: 10.1371/journal.pcbi.1003206. URL: https://app.dimensions.ai/details/publication/pub.1027970163andhttps://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1003206&type=printable.

[3] Jarvis ED. Arriaga G Zhou EP. "Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds." In: *PLoS One.* (2012). DOI: doi:10.1371/journal.pone.0046610.

[4] Baktash Babadi and Haim Sompolinsky. "Sparseness and Expansion in Sensory Representations". In: *Neuron* 83.5 (2014), pp. 1213–1226. ISSN: 0896-6273. DOI: https://doi.org/10.1016/j.neuron.2014.07.035. URL: https://www.sciencedirect.com/science/article/pii/S0896627314006461.

[5] Matteo Carandini and David J Heeger. "Normalization as a canonical neural computation". In: *Nature Reviews Neuroscience* 13.1 (2012), pp. 51–62.

[6] McCormick D. Castellucci GA Calbick D. "The temporal organization of mouse ultrasonic vocalizations." In: *PLoS One.* 3 (Sept. 2018). DOI: 10.1371/journal.pone.0199929.

[7] Jonathan Chabout et al. "A Foxp2 Mutation Implicated in Human Speech Deficits Alters Sequencing of Ultrasonic Vocalizations in Adult Male Mice". In: *Frontiers in Behavioral Neuroscience* 10 (2016), p. 197. ISSN: 1662-5153. DOI: 10.3389/fnbeh.2016.00197. URL: https://www.frontiersin.org/article/10.3389/fnbeh.2016.00197.

[8] Jonathan Chabout et al. "Male mice song syntax depends on social contexts and influences female preferences". In: *Frontiers in Behavioral Neuroscience* 9 (2015). DOI: 10.3389/fnbeh.2015.00076.

[9] Charles Cohen-Salmon et al. "Differences in patterns of pup care in mice V—Pup ultrasonic emissions and pup care behavior". In: *Physiology and Behavior* 35.2 (1985), pp. 167–174. ISSN: 0031-9384. DOI: https://doi.org/10.1016/0031-9384(85)90331-2. URL: https://www.sciencedirect.com/science/article/pii/0031938485903312.

[10] Wayne Moore Connor Meehan Jonathan Ebrahimian and Stephen Meehan. *Uniform Manifold Approximation and Projection (UMAP).* 2021. URL: https://www.mathworks.com/matlabcentral/fileexchange/71902.

[11] Elodie Ey et al. "The Autism ProSAP1/Shank2 mouse model displays quantitative and structural abnormalities in ultrasonic vocalisations". In: *Behavioural Brain Research* 256 (2013), pp. 677–689. ISSN: 0166-4328. DOI: https://doi.org/10.1016/j.bbr.2013.08.031. URL: https://www.sciencedirect.com/science/article/pii/S0166432813005287.

[12] Jeffrey D. Fitzgerald et al. "Second Order Dimensionality Reduction Using Minimum and Maximum Mutual Information Models". In: *PLoS Comput. Biol.* 7.10 (2011). DOI: 10.1371/journal.pcbi.1002249. URL: https://doi.org/10.1371/journal.pcbi.1002249.

[13] Geoffrey E Hinton and Ruslan R Salakhutdinov. "Reducing the dimensionality of data with neural networks". In: *science* 313.5786 (2006), pp. 504–507.

[14] Timothy Holy and Zhongsheng Guo. "Ultrasonic Songs of Male Mice". In: *PLoS biology* 3 (Jan. 2006), e386. DOI: 10.1371/journal.pbio.0030386.

[15] Heffner H Koay G Heffner R. "Behavioral audiograms of homozygous med(J) mutant mice with sodium channel deficiency and unaffected controls." In: *Hear Res* (2002).

[16] Andrei S. Kozlov and Timothy Q. Gentner. "Central auditory neurons have composite receptive fields". In: *Proceedings of the National Academy of Sciences* 113.5 (2016), pp. 1441–1446. ISSN: 0027-8424. DOI: 10.1073/pnas.1506903113. eprint: https://www.pnas.org/content/113/5/1441.full.pdf. URL: https://www.pnas.org/content/113/5/1441.

[17] Nocedal J. Liu D.C. "On the limited memory BFGS method for large scale optimization". In: (1989). DOI: https://doi.org/10.1007/BF01589116.

[18] V.Z. Marmarelis. "Signal transformation and coding in neural systems". In: *IEEE Transactions on Biomedical Engineering* 36.1 (1989), pp. 15–24. DOI: 10.1109/10.16445.

[19] Leland McInnes, John Healy, and James Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. 2020. arXiv: 1802.03426 [stat.ML].

[20] R. Miles. "Planning Report". In: (2021).

[21] Jiquan Ngiam et al. "Sparse Filtering". In: *Proceedings of the 24th International Conference on Neural Information Processing Systems*. NIPS'11. Red Hook, NY, USA: Curran Associates Inc., 2011, 1125–1133. ISBN: 9781618395993.

[22] Tim Sainburg, Leland McInnes, and Timothy Q. Gentner. "Parametric UMAP: learning embeddings with deep neural networks for representation and semi-supervised learning". In: *CoRR* abs/2009.12981 (2020). arXiv: 2009.12981. URL: https://arxiv.org/abs/2009.12981.

[23] Gentner TQ. Sainburg T Thielk M. *Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires.* 2020. PLoS Comput Biol. 16(10): e1008228 (doi:10.1371/journal.pcbi.1008228).

[24] Maria Luisa Scattoni, Jacqueline Crawley, and Laura Ricceri. "Ultrasonic vocalizations: A tool for behavioural phenotyping of mouse models of neurodevelopmental disorders". In: *Neuroscience and Biobehavioral Reviews* 33.4 (2009). Risk Factors for Mental Health: Translational Models from Behavioral Neuroscience, pp. 508–515. ISSN: 0149-7634. DOI: https://doi.org/10.1016/j.neubiorev.2008.08.003. URL: https://www.sciencedirect.com/science/article/pii/S0149763408001243.

[25] C. S. Sherrington. "The Integrative Action of the Nervous System". In: (1906).

[26] Nandini C. Singh and Frédéric E. Theunissen. "Modulation spectra of natural sounds and ethological theories of auditory processing". In: *The Journal of the Acoustical Society of America* 114.6 (2003), pp. 3394–3411. DOI: 10.1121/1.1624067. eprint: https://doi.org/10.1121/1.1624067. URL: https://doi.org/10.1121/1.1624067.

[27] Kari Torkkola. "Feature Extraction by Non Parametric Mutual Information Maximization". In: *J. Mach. Learn. Res.* 3.null (Mar. 2003), 1415–1438. ISSN: 1532-4435.

Appendices

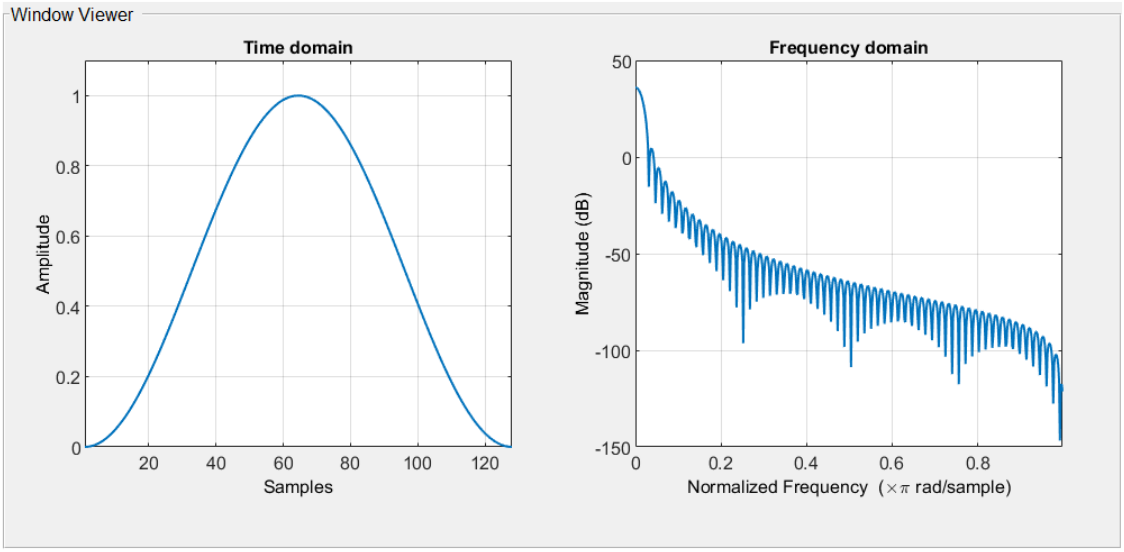


Figure 15: *Hanning Window Computed over 128 points*