Towards a non-linear model for neurons in the rodent auditory cortex

James Muir

February 11, 2007

Supervised by Jennifer Linden and Maneesh Sahani.

Word count = 3350, plus appendix.

Contents

1 Introduction

The auditory system is a focus of considerable attention for physiologists and physicists alike. Over the last 200 years the passage of sound through the outer ear, mapping of its' frequency on the basilar membrane (in the cochlear) and signal transduction by hair cells has been well established. The electrical signals produced are passed - in broad terms - to the brain, and the result is our perception of sound, or *hearing*. However, at higher-level stages of this neural process our understanding decreases rapidly. Herein lies the *biological problem*; our current knowledge of how the brain processes auditory signals is just the tip of an iceberg.

The region of the brain responsible for receiving and processing auditory signals is called *auditory cortex*. Recent experimental work [5] has been performed on rodents to examine auditory cortical neurons by characterizing their Spectrotemporal Receptive Fields (STRFs). These are matrices whose elements quantify how important a sound at particular frequency and time is to the firing of the neuron. Their construction implies a *linear* neuron function. Computational techniques are integral to this work; the STRF is estimated from the data through linear regression and Bayesian techniques can then employed to improve these estimates [9]. While STRFs calculated in this manner are in themselves interesting, the most significant result of this work was found through the use of a cunning statistical method to measure the linearity of these neurons, or equivalently, how well the STRF accounts for the "signal power". It is found that this is no higher than $\sim 40\%$ [10]. A non-linear model for neurons in the rodent auditory cortex is therefore required.

In this essay I will first review the work summarised above, discussing how computational techniques can help with experimental design, data analysis and, ultimately, suggest the direction for future work. I will then outline a possible extension to the model - namely, the use of the sigmoid function as an output non-linearity (see right) - and describe how this would be implemented and tested within the framework of *Generalized linear models*.

the sigmoid
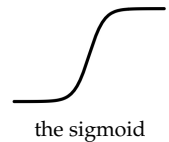
2 A review of the work

A complete description of the human brain is not within our grasp - it contains around 10^{12} neurons and each of these form many connections with other neurons. However, the brain is now known to be highly compartmentalised, and this has allowed research to be focussed on particular brain functions. While a full understanding of these is thought to be a long way off, progress is being made and advancements to medical science can be made *en route*. The auditory system is no different; we would certainly like to know more about it - particularly the rôle of the auditory cortex - and in studying it may shed more light on deafness, tinnitus, etc.

This section is a largely review of recent interdisciplinary work [5, 9, 10] on the primary auditory field (AI) and anterior auditory field (AAF) of rodent auditory cortex. These areas are known to be tonotopic, but individual neuron function has not been determined.

2.1 Receptive fields & experimental design

Receptive fields are a very useful concept in sensory neuroscience. They aim to provide a description of what a neuron is responsive to. In hearing, spectrotemporal receptive fields (STRFs) are arrays of weights that - when presented with an auditory stimulus - give a prediction for the neuronal output. Writing the STRF as D, the *predicted* neuron output as $\widehat{y}$ and the input as X,

$$\widehat{y}(t) = \sum_{\tau, f} D_{\tau f} X_{t-\tau, f} \tag{1}$$

where τ measures the time before t, the neuron firing time, and f is the stimulus frequency. It is important to note that this estimate is entirely *linear*; the STRF describes a linear neuron.

Receptive field calculation requires experimental work. Armed with sufficient stimulus-input and neuron-output data, mathematical regression methods can be employed to establish the STRF that links them. Obviously, this has to be done on an organism with a functioning auditory system[1]. The mouse is a good candidate, and has been extensively studied; powerful genetic tools exist for manipulating its neural circuitry and five distinct areas of its auditory cortex have been identified. Rats were also

---

[1]which is preferably both 1. good and 2. not too dissimilar to the human system

used in these studies. Advances in experimental technique have allowed the activity of single (or small groups of) neurons to be monitored using electrodes. Their cortical location can be readily identified. Generating sounds is relatively easy; high-grade speakers can present the rodents with acoustic stimuli. All that is left to do is decide *what stimulus to play*.

The experimental method outlined above is both expensive and time consuming. The physiological conditions required can only be maintained for around half an hour; after this, the results become meaningless and this site cannot be used again. There is need to gather as much data useful to STRF calculation in the limited time available. A good stimulus should, therefore, be information-rich, i.e. sample the entire rodent frequency range throughout its presentation. However, the parameter space for the regression method also needs consideration; time and frequency need to be divided into discrete *bins* since there is a trade-off between STRF "resolution" and the computational cost of STRF calculation.

These considerations have inspired the creation and use of the *Dynamic Random Chord* (DRC), a kind of "musical white-noise" in which tone pulses are played at random times, frequencies and intensities. Each of these quantities are discretized; frequencies are in semitone (1/12 octave) bins[2] and time bins are chosen as 20ms[3]. Each pulse lasts for 20ms and is ramped up and down by a 5-ms cosine gate. With some basic programming, a computer can produce a non-repeating DRC stimulus to be presented continuously to the rodent's auditory system. An algorithm for generating a simple DRC is given in the appendix.

2.2   Data analysis & STRF estimation

To perform any regression or inference, there is a critical requirement for sufficient data. Also, the data needs to representative of the system, and effort should be made to reduce the effect of **noise** on the estimate, which is ubiquitous in this type of experimental measurement. In light of this, the DRC described above is played for 60 seconds (3000 time bins) and then repeated[4], with no gap between repeats. The neuronal response can thus

---

[2]two frequency ranges were explored separately due to practical limitations; 2-32kHz (48bins) and 25-100kHz (24 bins)

[3]a shorter time bin would increase computational demand but may also approach the timescale for neural reset after firing

[4]N=20 for the low-f DRC presentation and N=10 for high-f

be summed over the repeats, the average being indicative of the *firing rate* for each time bin. This dampens the effects that experimental noise has on the estimate. Crucially, repeating also facilitates segregation of signal and noise components of the response, which will be discussed in section 2.4.

The STRF to be determined is chosen span a lag-time of 300ms (15 time bins) - acoustic history beyond this limit is deemed irrelevant to the neuron. It is discretized into either 48 (low-f) or 24 (high-f) frequency bins; each frequency range has their own STRF. Thus, we have a "worst-case" of $15 \times 48 = \textbf{720}$ weights to be determined, and $\simeq \textbf{3000}$ input-&-output values. There is sufficient data for STRF determination via regression.

The computational problem to be solved is multiple linear regression, and is usually performed by a process known as *Iteratively re-weighted least squares* (IRLS). The large parameter space required to produce a useful STRF makes this process computationally expensive; a *stochastic hill-climbing* algorithm may be required to improve speed and ensure convergence in the search for the global minimum of this large-dimensional space. The basic idea, however, is to minimize the sum of least-squares (S) between the actual neuronal response and that predicted by the STRF (see 1):

$$S = \sum_t \|y(t) - \widehat{y}(t)\|^2 \tag{2}$$

where t runs to 60s in 20ms steps. The weights that make up the STRF are varied until the algorithm converges on its best estimate. Improvements to this are discussed in the next section; on its own, this method is fairly crude. However, insights into the basic form of the STRF can still be made (these are confirmed with the more advanced methods):

- The areas studied (AI and AAF) have a tonotopic arrangement. This was reflected in the results; generally, individual neurons respond best to a bandwidth around a characteristic centre, f_c.

- Temporal structure was also revealed; the STRF contains an inhibitory peak which switches to an excitatory peak at around τ_p=50*ms* (AAF) and τ_p=80*ms* (AI), suggesting that the area AAF is specialised for faster temporal processing.

Interestingly, this structure is similar to that of receptive fields commonly found in vision[2], which are often approximated by the *Gabor function*:

$$G(x,y) \sim \cos(x - \phi) \exp^{-x^2 + y^2}. \tag{3}$$

This is plotted in figure 1 and shows the basic features of the STRF that were found.
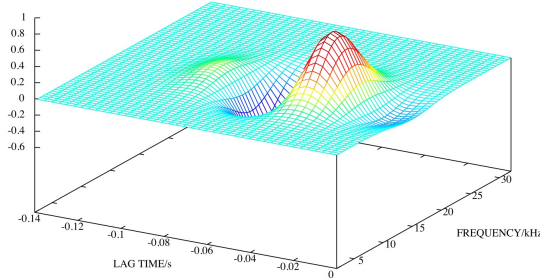

Figure 1: The Gabor function, $G(\tau_p, f_c) \sim \cos(\tau_p - \tau) \exp^{-(\tau - \tau_p)^2 + (f - f_c)^2}$

2.3   Improving the estimates

Bayesian techniques are becoming increasingly popular in application to problems where something is learned, predicted or inferred from data, particularly in the presence of **noise** which corrupts the data and makes the model susceptible to *overfitting*. The simplest of these is maximization of the *likelihood* of the data given the parameters and assumptions made. If we assume Gaussian output noise of constant variance σ^2 and nothing about the parameters, then maximizing the logarithm of the likelihood is equivalent to the minimizing the sum of least-squares. This is shown in the appendix.

However, this framework can be far more powerful than standard regression methods if we can place so-called "priors" (probability distributions) on the weights. Recently, techniques have been developed to impose *sparsity* and *smoothness* constraints on arrays of weights that are inferred from **noisy** data. Both these conditions are desirable in STRF calculation; we want our model, while explaining the data, to be as simple as possible (Occam's Razor) and expect that weights have similar values to their neighbours. These methods, known as Automatic Relevance and Smoothness

Determination respectively, are summarised below.

Automatic Relevance Determination (ARD). This was originally developed by David Mackay in 1992[5] [7] and sets out to identify (and zero) irrelevant weights. ARD is achieved by initially modelling each weight as a zero-mean Gaussian, with *hyperparameters* (to be determined) specifying how far the weight is allowed to vary from zero. The "evidence" is optimised through a tricky integration[6] over the weights, and the hyperparameters are found as a result. Any hyperparameters beneath a specified tolerance level have their weight set to zero and deemed irrelevant to the model – changing this weight from zero has no affect on the output. In this way, a set of *pseudo*-priors can be specified for the weight matrix, with irrelevant weights zeroed. While this is computationally costly[7], it frequently improves the result of the subsequent maximization.

Automatic Smoothness Determination (ASD). This is a novel technique [9] that has been developed - in light of the success of ARD - to hyperparameterise the "smoothness" of the STRF across the (unrelated) dimensions of time and frequency. As in ARD, the evidence is optimized to determine the hyperparameter values. Likelihood maximization then takes place under a framework that effectively constrains weights to vary smoothly on scales specified by the hyperparameters.

As might be expected, these disparate approaches can be combined into a potentially powerful technique. Loosely speaking, ASD is called first to determine smoothness scales, and ARD is then set to work on this pre-defined basis. This is referred to as ASD/RD.

2.4   How linear are the neurons?

To quantify the success of the methods described above, some statistical effort is called for to segregate the noise component present in the STRF estimates, which is symptomatic of overfitting. Then, the degree to which the STRF accounts for the data - or, equivalently, the *linear*ity of the neuron - can be determined. A novel method has been developed [10] to achieve just this, which works under the assumption of trial-independent, zero-mean

---

[5]but not named until 1996 by R. M. Neal, who extended the technique

[6]often made tractable using Laplace's method or through a Monte-Carlo approach [8]

[7]the cost associated with each iteration in the optimization scales as d^3 where d is the number of weights [8]

Gaussian noise. This allows the so-called "signal power" to be written in terms of the response only. The signal power captured by the STRF estimate can then be quantified using 9-from-10 cross validation on the data. The performance for each method of STRF estimation is tabulated below.

STRF estimation method	Signal power captured
ML	12%
ARD	26%
ASD	27.5%
ASD/RD	30%

The ability of the STRF to predict test data improves with the use of more sophisticated optimization techniques; sparsity and smoothness are hence valid constraints to impose. However, ASD/RD (the most successful of the techniques) only manages to capture 30% of the signal power. Even when using a similar measure which considers that the STRF could be inaccurate (the *predictive power*), this maximum is only increased to $\sim 40\%$. These conclusions suggest either 1. that a non-linear model is required to describe the rôle of auditory cortical neurons in the rodent or 2. that the DRC is somehow producing this effect. However, low percentages have also been reported using natural sounds as stimuli [6] which gives the second of these arguments less weight.

A suggested extension - to incorporate non-linearity into the neuronal model - is discussed in the next section.

3 A suggested experiment

The data analysis described above strongly suggests that a non-linear model for rodent auditory cortical neurons is required. This section motivates the use of the sigmoid as an "output non-linearity" to try, and proposes its implementation in the framework of Generalized Linear Models.

3.1 The sigmoid as an output nonlinearity

The sigmoid function is used widely in both physical and biological systems[8] to model switch-like behaviour between two levels (`off` → `on`). It is also commonly chosen as the "activation" in simple neural models [8]. Its mathematical form is

$$y = \frac{1}{1 + \exp^{-a.x}} \tag{4}$$

where a is a parameter governing the "softness" of the switching. This is shown in figure 2, below.
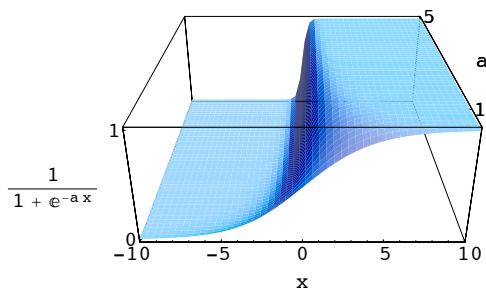

Figure 2: The sigmoid function, for values of "a" ranging from 0.5 to 5.

The sigmoid is piecewise linear except in the case $a = \infty$, where it represents a perfect switch. It has a point of inflection at $x = 0$ and its value is bounded by 0 and 1[9]. The low-x response performs soft *rectification* (which gets harder with increasing a) and the high-x behaviour is akin to *saturation*. Both of these properties are realistic for a model neuron; is it neurophysiologically reasonable for inputs below a certain threshold to excite no response and for sufficiently large inputs to be essentially equivalent.

---

[8]e.g. in 2-level systems of statistical physics, or as an approximation to the Hill function in e.g. allosteric interactions

[9]the neural output data will need normalising to a maximum of 1 for comparison

Machens *et al* [6] used the criteria of saturation and rectification to fit a static nonlinearity to their rat data, but reported only small improvements in model performance. However, this approach is model-based; it uses the STRF to generate the input (i.e. as in equation 1) to be run through the non-linearity. Their result discourages the use of a sigmoid under this method, **but**, things can be done differently. We appeal to a framework under which the "receptive-field" weight matrix (strictly, **not** the STRF) is learned from the data in a form of regression *through the sigmoid output non-linearity*. This is shown schematically below.

$$W_{\tau f} \dashrightarrow \text{min-LS algorithm}$$

$$\mathbf{X}_t \longrightarrow \varsigma \longrightarrow \widehat{y}_t$$

(sigmoid)

$$\sigma^2$$
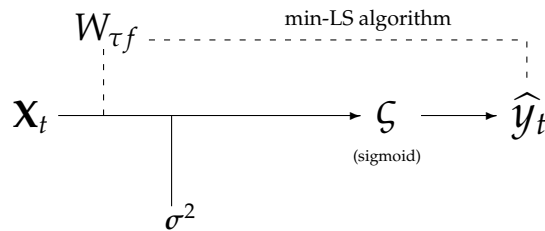
To implement this model, we turn to a branch of mathematics known as Generalized Linear Models.

3.2 Generalized linear models

The theory of Generalized Linear Models (GLMs) has been advanced over the last few decades to deal with regression-type problems where the model contains - despite its name - some *non-linear* function relating input and output. The language of GLMs is outlined below, as described in Dobson [4].

The model requires two ingredients: an equation linking the output and input variables - which need not be linear - and an expression for the probability distribution of the output, which must be from the **exponential family of distributions** (Normal, Poisson and Binomial distributions are all in the family). There exists a wide range of computer packages that can, from this starting point, estimate parameters (through either least-squares or maximum likelihood), check model adequacy (e.g. by calculation of residuals) and interpret the results through hypothesis testing, etc.

What all of this means may become clearer with an example. The simplest

GLM is of the form

$$E(Y_i) = \mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad ; \quad Y_i \sim N(\mu_i, \sigma^2) \tag{5}$$

i.e. the expectation value of the output, $E(Y)$ is constructed in a linear fashion from the inputs x and parameters $\boldsymbol{\beta}$, and the probability distribution is Normal about this mean due to a uniform Gaussian noise. This is known as a **Normal Linear Model**. The model used in STRF calculation is an example of this special case[10]:

$$E(Y_t) = \mu_t = \sum_{\tau,f} D_{\tau f} X_{t-\tau,f} \quad ; \quad Y_t \sim N(\mu_t, \sigma^2) \ . \tag{6}$$

To use the sigmoid as an output non-linearity, however, we need to harness the full power of GLMs. These extend to models with a non-identity **link function**, g, where

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \ . \tag{7}$$

In the model proposed, the sigmoid takes as its input the linear construction

$$\Theta_t = \sum_{\tau,f} W_{\tau f} X_{t-\tau,f}$$

where the weight matrix is given the symbol W to distinguish it from the STRF. It then returns expectation value for the output, $E(Y_t)$. The operation of the output non-linearity can be written as

$$E(Y_t) = \mu_t = \frac{1}{1 + e^{-\Theta_t}} \tag{8}$$

and this can be re-cast into GLM form by multiplying through by the inverse-sigmoid (labelled ς^{-1} for convenience):

$$\varsigma^{-1}(E(Y_t)) = \varsigma^{-1}(\mu_t) = \sum_{\tau,f} W_{\tau f} X_{t-\tau,f} \quad ; \quad Y_t \sim N(\mu_t, \sigma^2) \ . \tag{9}$$

A few things are worthy of note here. Firstly, nothing is assumed about the form of the weight matrix W; it is "learned" from the data. This makes the approach completely removed from that in [6], where the STRF is taken as given and attempts are made to fit a non-linearity to the data (which, incidentally, turns out vaguely sigmoidal). It is also important to note that there

---

[10]to get an exactly equivalent form the STRF and input each need their rows pasted together to form a single vector

is no explicit parameter which governs the sigmoid switch behaviour ('*a*' in equation 4); the sharpness is governed only by the values of the weights. This means there is no immediate danger of *non-identifiability*[11] which can be problematic when inferring input non-linearities [1].

3.3 Suggested implementation

I propose that the sigmoid-output model - cast as a GLM in equation (9) above - should be tested on rodent neural data obtained from the DRC experiments. With the weight matrix being learned through the sigmoid in this way, it is hoped that this model of the auditory cortical neuron function will perform better than the linear STRF. However, for a significant improvement to be made (i.e. beyond the $\sim 5\%$ increase in signal power accountability reported using an output-fit approach [6]), weight matrices determined under the GLM would need to capture some crucial structure that the STRFs are lacking. The questions that this experiment would aim to answer are:

- If this extra structure is present, what is its form and what are the implications for the rôle of the auditory neuron?

- Will the sigmoid-GLM procedure be able to find this structure alone, i.e. without the help of Bayesian techniques? Under GLMs, methods for assigning priors to model parameters have to rely on sampling [3]; the integrals used in ARD and ASD are feared to be intractable. Further work in this area could inspire a more complete Bayesian-GLM framework.

- Extension: is there a need for some *input* non-linearity, too? Candidates might take the form of `AND/OR` logical conditions - `AND` might demand, for example, that two distinct regions in the weight matrix be stimulus-activated for the neuron to fire.

If the model is successful, it should then be extended in application to other species. The neurological implications of this model - and its performance under distortion - could then be considered in more depth.

---

[11]e.g. if $\Theta = aW.X$ then the transformation $a \rightarrow a.\lambda, W \rightarrow W/\lambda$ results in the same output

4 Summary

Neuroscience is still young. Much is yet understood about how the brain works – the high-level neural basis of hearing is one such example. However, progress is being made. I have discussed how computational and experimental techniques can be married together to tackle the problem of "working blind", and that this synergistic approach can lead to suggestions for the direction of future research.

Conclusively, neurons in the rodent auditory cortex have a non-linear function. In light of this, I have suggested a non-linear model to be investigated.

- -

A Appendix

A.1 The Dynamic Random Chord

Algorithm 1 (*Octave* code - a simple dynamic random chord)

```
dur=0.02;                      # 20ms time-bin
t=0:(1/fs):dur;                # fs is appropriate sample rate (high)

for i=0:dur:t_max                   # time index, 20ms steps
      for j=f_min:f_step:f_max      # frequency index

        d=rand();

            if d<lim        # lim controls the sparsity of the DRC
            d=0;                    # set (most) entries to zero
            endif

        a(i,j)=d*sin(2*pi*j*t);    # amplitude set to random value
                                    # pure tone of j Hertz for 20ms
      endfor
endfor

wavwrite('DRC.wav', a, fs);     # save the dynamic random chord
```

In practise, the amplitudes are also discretized, frequency bins are spaced as $2^{n/12}$ and a more efficient algorithm would not run "element-by-element" in this way. A graphical representation of a DRC generated with the above

algorithm (`lim` = 0.85) is plotted below; black bars are pure tones of zero amplitude.

frequency

time

Figure 3: A graphical representation of an example DRC

A.2 Equivalence of Least Squares and Maximum Log-Likelihood

The STRF model (1) with constant Gaussian output noise gives the following likelihood:

$$y|X \sim \prod_t \exp^{-(y(t) - \widehat{y}(t))^2 / 2\sigma^2}$$

and the logarithm of this is simply

$$\sum_t -(y(t) - \widehat{y}(t))^2 / 2\sigma^2.$$

Maximising this expression is achieved through finding the point where its derivative (with respect to the parameters) is zero; since the noise term is constant, the equivalence of the result with that of the least squares method (2) can be made.

The logarithm of the likelihood is often preferred to the likelihood itself as it generally takes more handleable (i.e. non-infinitesimal) values.

References

[1] M. Ahrens, L. Paninski, and M. Sahani. Inferring input nonlinearities in neural encoding models. *Submitted*, 2006.

[2] P. Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

[3] D. G. T. Denison, C. C. Holmes, B. K. Mallick, and A. F. M. Smith. *Bayesian Methods for Nonlinear Classification and Regression*. Wiley Series in Probability and Statistics, 2002.

[4] A. J. Dobson. *An Introduction To Generalized Linear Models, 2nd Edition*. Chapman and Hall, 2002.

[5] J. Linden, R. Liu, M. Sahani, C. Schreiner, and M. Merzenich. Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *J Neurophysiol*, 90:2660–2675, June 2003.

[6] C. Machens, M. Wehr, and A. Zador. Linearity of cortical receptive fields measured with natural sounds. *J. Neurosci*, 24(5):1089–1100, February 2004.

[7] D. J. C. Mackay. Bayesian interpolation. *Neural Computation*, 4:415–447, 1992.

[8] D. J. C. Mackay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[9] M. Sahani and J. Linden. Evidence optimization techniques for estimating stimulus-response functions. In *Advances in Neural Information Processing Systems*, pages 301–308. MIT Press, June 2003.

[10] M. Sahani and J. Linden. How linear are auditory cortical responses? In *Advances in Neural Information Processing Systems*, pages 109–116. MIT Press, June 2003.