Problem summary

The auditory system has to represent sound across 6 orders of magnitude of sound pressure (120dB). At present we only partly understand how the system is able to function so well across this range. One mechanism to deal with this is compression in the cochlea. The passive mechanical filtering properties of the basilar membrane are modified by an active process which amplifies very quiet sounds. Gain decreases very rapidly with sound pressure level such that a doubling in level might only lead to a 10% increase in the amplitude of vibration on the basilar membrane. Thus at the level of the cochlear there exists a mechanism which increases the dynamic range of the cochlea by at least 60dB at some frequencies. A common source of hearing deficit is damage to this active mechanism, which produces a raise in thresholds but also a reduction in the dynamic range of the system such that whilst quiet sounds are much quieter, loud sounds are just as loud. In the brain, problems remain in the representation of broadband spectra. The majority (70%) of auditory nerve (AN) fibres are highly sensitive and have a dynamic range which only changes in firing rate at very low levels. The dynamic range of the firing rate of these neurons can be as little as 15dB. Thus at normal signal levels (> 60dB SPL) almost all auditory nerve fibres are firing at their saturated rate. Recordings of the auditory nerve fibre responses to steady state vowels confirm that the representation of spectrum in auditory nerve fibres varies with level and is highly degraded at high sound levels. This being the case, how is it that we can recognise sounds and understand speech across the majority of our dynamic range? A population of cells in the ventral cochlear nucleus represent spectrum in a much more level robust manner than the auditory nerve, but the mechanism for this is not known. A computational model of the auditory periphery coupled to a model of the ventral cochlear nucleus (and in particular so-called chopper cells) would be ideally suited to address this problem.
Report summary

In this report we present a record of activity during the course of the study group. It is in no way a complete piece of work and is perhaps best regarded as a minuted version of discussions that took place throughout the week.

The main achievement of this work is a mathematical exploration of the selective listening hypothesis via an analytical and numerical study of a model dendrite driven by realistic spike train data. Peripheral discussions ranged over mechanical models of the basilar membrane, single neuron coincidence detection, preservation of rate-place codes and network studies of interacting neurons.
1 Introduction

The auditory system has to represent sound across 6 orders of magnitude of sound pressure (120dB). We only partly understand how the system is able to function so well across this range. One mechanism to deal with this is compression in the cochlea. The passive mechanical filtering properties of the basilar membrane are modified by an active process which amplifies very quiet sounds. Gain decreases very rapidly with sound pressure level such that a doubling in level might only lead to a 10% increase in the amplitude of vibration on the basilar membrane. Thus at the level of the cochlea there exists a mechanism which increases the dynamic range of the cochlea by at least 60dB at some frequencies. A common source of hearing deficit is the damage to this active mechanism, which produces a raise in thresholds but also a reduction in the dynamic range of the system such that whilst quiet sounds are much quieter, loud sounds are just as loud.

In the brain, problems remain in the representation of broadband spectra. The majority (70%) of auditory nerve (AN) fibres are highly sensitive have a dynamic range which only change in firing rate at such low levels. The dynamic range of the firing rate of these neurons can be as little as 15dB. Thus at normal signal levels (> 60dB SPL) most auditory nerve fibres are all firing and mostly at their saturated rate. Recordings of the auditory nerve fibre responses to steady state vowels confirm that the representation of spectrum in auditory nerve fibres varies with level and is highly degraded at high sound levels [1].

This being the case, how is it that we can recognise sounds and understand speech across the majority
of our dynamic range? The internal representation should be changing radically, making the task of any recogniser upstream difficult.

Not all auditory nerve fibres have such a poor dynamic range. A minority of auditory nerve fibres, those with higher thresholds and low spontaneous rates, have larger dynamic ranges because they cover the dynamic range occupied by cochlear compression. Thus the AN can be divided into two (or sometimes three) populations, which together cover the full dynamic range of the cochlea: very sensitive high spontaneous rate (HSR) fibres with narrow dynamic ranges covering low sound levels, and less sensitive low spontaneous rate fibres (LSR) covering high levels. LSR fibres preserve the representation of vowel spectra better at high levels. Adapted from [2].

Some cells (chopper cells) in the ventral cochlear nucleus show a greatly improved rate representation of spectra across signal levels compared with AN fibres. Adapted from [3, 4]. These cells project directly to the inferior colliculus and may underlie level robust representations of spectrum in the auditory system.

It is not known how these cells produce these improved representations. It has been hypothesised that they may be able to selectively listen to nerve fibres of different dynamic ranges at different sound levels. Further, simulations have shown that in principle a shunting inhibition is able to select between the different types fibres [5]. This works as follows: LSR (insensitive high-level) fibres innervate the soma, broadly tuned shunting inhibitory inputs innervate proximal dendrites and HSR (sensitive low-level) fibres innervate the dendritic tree more distally. At low sound levels the shunt is weak or ineffective and the sensitive HSR fibre inputs cause the neuron to fire, thus coding spectrum well at low signal levels. At higher levels, the HSR fibres are saturated and the rate representation of spectra at these inputs is highly degraded, but these inputs are now shunted. The LSR fibres are now active and are not shunted,
and thus the neurons output represents these inputs. Thus the neuron selectively listens to different populations of fibres at different sound levels.

In fact no model has ever fully tested this hypothesis by feeding vowel sounds through a model of the periphery and then into a model of a VCN neuron with this architecture. Neither have any empirical studies shown that this mechanism is indeed at work. We do not know the relative roles or importance or shunting vs. additive inhibition, or the consequences of different fibre types being mounted at different points along the dendrites on vowel representation.

There is another source of information about spectra which is robust to sound level. Although the average firing rate of HSR fibres saturates at low levels, temporal information remains. The timing of individual spikes codes the filtered waveform (the carrier frequencies below a few kHz and envelope information at higher frequencies). Furthermore, the shapes of the filters and non-linearities in the AN responses mean that there is a competitive interaction between nearby frequencies for dominance in the temporal code. This acts to effectively code frequencies associated with peaks in the spectra across populations of nearby fibres.

Thus there exists a robust description of spectrum in the temporal code of HSR AN fibres and there is the possibility to extract these temporal features.

One possible mechanism to take advantage of the temporal coding exploits the abrupt phase shifts that occur close to BF. As a wave travels along the basilar membrane it grows in amplitude, reaching a peak at a point dependent on the stimulus frequency. Once it has reached this peak the speed of the travelling wave and the amplitude decreases rapidly. Even in AN fibres near to BF which are saturated this still results in rapid shifts in the phase of responses across different fibres.

Shamma [7] proposed that if you simply differentiate the instantaneous firing rate across a population of fibres, you get peaks corresponding to the rapid phase shifts at BF. This representation of spectral peaks remains stable across a wide range of stimulus levels. He consolidated this in a theory in which

Figure 3: Averaged normalised firing rates for a population of chop-T cells in VCN in response to the vowel /e/ across a range of stimulus levels.
Figure 4: The Average Localised Synchronised Rate across a large population of AN fibres in response to the vowel /e/ at different stimulus levels. Shows the strength of the phase locking to the stimulus component corresponding to the best-frequency (in rate terms) of each fibre. Adapted from [2].

Figure 5: Dominant component analysis of the auditory nerve responses to the vowel /i/ at 80dB SPL. Adapted from [6]. X-axis: Neuron BF. Y-axis: the strongest component in the FFT of the PSTHs. Shows that across a wide range of BFs, the formant frequencies (and fundamental) are dominant in the timing of spikes.
lateral inhibitory networks acted upon the fine temporal structure in auditory nerve fibres. Sumner [8] showed further that the particular shape of the convolution function affected the details of the representation. Since a simple differential operates only locally, at high levels there is a tendency for individual harmonics to be represented (if they are spectrally resolved) regardless of their position relative to formants. A suitably wide inhibitory receptive field takes advantage of the fact that away from BF, the phase shift is small (because propagation is rapid) and firing is almost synchronous across a range of BFs. This synchrony can drive a coincidence detection mechanism, such that low spatial frequencies produce strong inhibition. The stronger is a particular component in a spectra, the larger is the range of BFs over which activity is phase locked to it, and away from BF the phase shift is quite gradual, producing more synchronous activity. Thus the more it drives inhibition to suppress neighbouring components. This produces a competitive interaction in the mutual inhibition of neighbouring components such that the stronger a component is relative to its neighbours the more it suppresses them. Thus it exploits synchrony suppression on the auditory nerve to maintain not only the location of spectral peaks but also their relative amplitude.

It is in principle possible that VCN neurons are processing information in just this way. Broadly tuned interneurons, which convey temporal information very well, inhibit the chopper cells. However, the models employed simply convolved firing rate probability (not spikes) on the auditory nerve with a function incorporating positive and negative regions, followed by half wave rectification and summing across time. It is not clear that the speed of processing is available in the biological substrate. Further, Bruce et al.¹ (unpublished) have performed some simulations which suggest that stochastic auditory nerve fibres are not be reliable enough to support the required temporal resolution. Finally, all these simulations were performed with linear filters in the periphery. Real auditory filters broaden at higher levels, with a proportional shallowing of the phase response.

Either of these mechanisms, or some other mechanism, might underlie the processing producing the

¹http://www.ece.mcmaster.ca/%7Eilbruce/
robust rate-representations in VCN chopper cells. Equally, such mechanisms may be at work in other nuclei. However, timing information at least, does not progress beyond certain nuclei in the superior olive the next stage in the auditory pathway.

The chopper neurons show the best rate representation in the VCN (in fact the best are the chop-T neurons). These neurons also show a resonant oscillatory property found in many integrate and fire type models [9]. When inputs are constant, firing is regular (they chop, hence the name). Modulations at the inputs having the same period as the preferred oscillation rate of the neuron are selectively amplified. Thus these neurons are tuned to temporal fluctuation at their inputs, up to several hundred Hertz an effective modulation gain. This might constitute the beginning of a system for extracting temporal information. Hewitt and Meddis [10] showed that if neurons having similar temporal tuning properties such as found in VCN were coincident on a neuron further up the system (such as in inferior colliculus), then that neuron produced was rate-tuned to modulation rate (whereas VCN cells were only tuned in there temporal response). They proposed that this might be a method for temporal pitch code extraction. For low-frequency neurons with similar spectral and temporal tuning properties, this mechanism might also be expected to function similar to the ALSR representation shown in figure 4. However no one has explored this idea, although Meddis and O’ Mard [11] are currently developing a physiologically based model for pitch extraction along these lines. These mechanisms do not anyway appear to provide an explanation for how VCN stellate cells show robust rate tuning to vowels (although no one has investigated the responses to vowels in simple models of VCN chopper cells).
2 Modelling the basilar membrane

The basilar membrane plays the role of amplifier and filter of incoming acoustic signals, as illustrated in Fig. 8. In brief, longitudinal pressure waves induce transverse travelling wave on the basilar membrane (BM). The location of maximum BM amplitude depends on signal frequency (characteristic frequency). Hair-cells transduce BM motion into cell depolarisation and subsequently spikes in auditory nerve fibres. Several mechanical models for this system exist (see for example [12]). In the acoustic community it is quite common to treat this early stage of sensory processing using a combination of linear and nonlinear filters that connect the velocity of the stapes to the BM velocity, as illustrated in Fig. 9 [13].

2 The stapes is the stirrup-shaped small bone that transmits sound vibrations to the inner ear.
The group discussed building a mechanistic model based on conservation of mass, momentum and energy laws, without any ad-hoc hypotheses, in order to understand the filtering properties of the BM. The basic formulation of the problem was broken into two parts (solid and fluid), as in Fig. 10. For the solid part we consider an elastic wall obeying the linear equation:

$$\rho \frac{\partial^2 U}{\partial t^2} = G \Delta U + \mu \Delta V + f(U, V, f),$$  \hspace{1cm} (1)

where $U$ is the displacement, $V$ is the velocity, $G$ is the shear modulus, $\mu$ is the viscosity, and $\rho$ is the mass density. The function $f$ represents active response of the membrane. The fluid motion is described by the compressible Navier-Stokes equations

$$\rho \left( \frac{\partial V}{\partial t} + V \cdot \nabla V \right) = -\nabla P + \nabla \cdot \sigma,$$  \hspace{1cm} (3)

where $P$ is static pressure and $\sigma$ is normal stress. The boundary condition at the interface of fluid and solid is

$$(-P I + \sigma)_s \cdot n = (-P I + \sigma)_f \cdot n, \quad V_s = V_f.$$  \hspace{1cm} (4)

### 3 Single neuron modelling

As discussed in Section 1, the HSR fibres have a small dynamic range, and the LSR a large dynamic range, see Fig. 11. Thus selective processing of different SR populations of ANFs is required to tackle the dynamic range problem. The selective listening model [5] is one means of achieving this, and is illustrated in Fig. 12. This is a single neuron dendritic model that relies on shunting inhibition to veto the input from HSR fibres with increasing sound level, allowing the LSR fibres to influence neuronal output. Although the full dynamic range is spread across different fibre types, the single neuron can select from the fibre population and achieve a large dynamic range. Here we revisit this idea again from a mathematical perspective.

Consider a infinite dendritic cable (with cell body at $x = 0$) with membrane potential $V(x, t)$ at position $x \in \mathbb{R}$, and $t \in \mathbb{R}^+$:

$$V_t = -\frac{V}{\tau} + DV_{xx} + I_{syn},$$  \hspace{1cm} (5)
Figure 11: The full dynamic range is spread across different fibre types: i) High-spontaneous rate: low thresholds, tiny dynamic range (80% of ANFs), ii) Medium-spontaneous rate: slightly higher thresholds, larger dynamic range, and iii) Low-spontaneous rate: high thresholds, large dynamic range.

where the synaptic input \( I_{\text{syn}} = I_{\text{syn}}(x, t) \) for a single spike arriving at point \( x_0 \) at time \( t_0 \) is given by

\[
I_{\text{syn}}(x, t) = A(x)\delta(x - x_0)\delta(t - t_0)(V_{\text{syn}} - V(x, t)).
\]

The term \( (V_{\text{syn}} - V(x, t)) \) in (6) is a shunt that pushes the membrane potential toward the reversal potential for the synapse: \( V_{\text{syn}}^E > 0 \) (excitatory) and \( V_{\text{syn}}^I < 0 \) (inhibitory). In this case the solution is readily calculated (using Fourier transforms), and because of the linearity of the problem can easily be generalised to multiple inputs at multiple positions:

\[
V(x = 0, t) = \sum_n \sum_m A(x_n)[V_{\text{syn}}^n - V(x_n, T_{mn})]G(x_n, t - T_{mn}).
\]

Here \( x_n \) are the points of synaptic contact (indexed by the label \( n \in \mathbb{Z} \)), whilst the spike train at position

Figure 12: A diagram illustrating the basic mechanism thought to underly selective listening. Key ingredients are the spatially extended dendritic tree and and shunting inhibition that can veto HSR fibres with increasing stimulus.
Figure 13: There are many standard algorithms for transforming vowel sounds into the language of spike trains. This allows us to set up a selective listening test of our dendrite model using realistic neuronal data. In the above figures the two spike train plots (on five distinct s) show typical vowel spike data. This realistic data drives a simple unbranched cable model with instantaneous synaptic conductances shunted toward a positive reversal potential (excitation on HSR and LSR fibres) and a negative reversal potential on the remaining (veto) fibres.

\[ x_n \] is written in terms of the spike arrival times \( T_m \) (and \( m \in \mathbb{Z} \)). Here \( G(x, t) \) is the Green’s function:

\[
G(x, t) = \frac{e^{-t/\tau}}{\sqrt{4\pi Dt}} e^{-x^2/(4D)} H(t),
\]

where \( H(t) \) is the Heaviside step function. The formula for \( V(x = 0, t) \) is self-referential, though it can be easily iterated numerically. Matlab code for doing this is given in the Appendix. To model the active dynamics of the cell body (soma) we use a simple integrate-and-fire model and write

\[
\frac{du}{dt} = -u + V(0, t),
\]

subject to reset: \( u(T(m)^+) = 0 \), with

\[
T(m) = \inf\{ t \mid u \geq h, T(m) > T(m - 1) + \tau_R \}.
\]

Here \( \tau_R \) is a refractory time-scale. Thus spikes at the soma (at the times \( T(m) \)) are generated whenever \( u \) reaches the threshold \( h \). Using the above we were able to explore systems like that of Fig. 12 and confirm that the selective listening model can indeed solve the dynamic range problem. An example of the passing real vowel data through the model is shown in Fig. 13. Output from the model defined by (7) is shown in Fig. 14. To check that our formulation is consistent with direct numerical simulations of the underlying PDE model we also did simulations in the NEURON simulation environment, see Fig. 15.
Network processing

First consider a single IF neuron:

\[ \frac{du}{dt} = -\frac{u}{\tau} + A \sum_{n=1}^{N} \delta(t - t_n), \]  

(11)

with solution

\[ u(t_N) = A \sum_{n=1}^{N} e^{-\frac{t_n-t_0}{\tau}}. \]  

(12)

For a periodic input spike train \( t_n = n\Delta \), then

\[ u(t_N) = \frac{A}{1 - e^{-\Delta/\tau}}. \]  

(13)

To guarantee firing (i.e. that the solution trajectory can reach threshold) we require \( u(t_N) > u_{th} \).

Hence, we see that the minimal input frequency \( F_c = 1/\Delta \) to cause firing given is given by

\[ F_c = \frac{1}{\tau \ln \left( \frac{u_{th}}{u_{th} - A} \right)} . \]  

(14)

Since we need \( N \geq 1 \) spikes to fire then

\[ F_{out} < F_{in}, \]  

(15)
Figure 15: The selective listening model of Fig. 12 can also be coded up and tested within the NEURON simulation environment, as done here. As expected, the NEURON code is consistent with our mathematical solution. The obvious advantage to using our mathematical approach as opposed to NEURON is that this does not require the numerical solution of the PDE dendrite model. The two figures show voltages at the soma and a point on the dendrite as a function of time. On the left HSR fibres and an inhibitory veto signal are included. On the right the LSR fibres are also included, showing that LSR information can influence the soma even in the presence of HSR stimulation.

where $F_{in} = \Delta^{-1}$ is the input frequency and $F_{out}$ the output frequency of the neuron. This gives rise to the so-called frequency dilemma; in a feed-forward chain of neurons, labelled by $i$, at some point $F_i < F_c$ and firing stops. One solution to this problem is to consider convergent input (and divergent output). But how do we preserve place-rate coding and avoid interference between high and low frequencies? Imagine a spike train formed from the mixing of one train with a high rate and the other with a low rate. Sometimes a spike from each train will be close in time, separated say by $\delta t$. Roughly speaking such a pair will not elicit a response from a coincidence detector neuron with $\delta t > \tau$, so that by choosing a small $\tau$ one can stop the mixing of low frequencies with high. Thus a partial mechanism that suppresses frequency mixing seems to be coincidence detection - this can suppress the mixing of low with high frequencies (but not high with low!). Consider a one-dimensional place model with characteristic frequency (CF) and choose a neuron with CF = $x$ (same place-label as ANFs). To fire upon receiving two spikes a time $T(x) = 1/CF(x)$ apart we set

$$\tau(x) = \frac{K}{x}, \quad K^{-1} = \ln \left[ \frac{A}{u_{th} - A} \right].$$

Hence the system depicted in Fig. 16 can achieve coincidence detection (and thus partially solve the frequency dilemma) if we pick an appropriate distribution of cell membrane decay constants.
Figure 16: Coincidence detection at the single neuron level can be achieved by varying the cell membrane decay constant. In a one-dimensional place network the choice $\tau(x) = K/x$ can suppress the mixing of low with high frequencies (but not high with low!). Here (AV)CN represents the antero-ventral cochlear nucleus.

5 Exploring the network consequences of local coincidence detection

The general question thus arises as to how a network of interacting coincidence detectors would behave, and whether we could quantify how good such a network would be in preserving rate-place coding. In the absence of a general theory of synaptically interacting spiking neurons computer simulations are ideally suited to exploring system level properties. Here we discuss the philosophy of such an approach to auditory neuroscience.

A computer science perspective on unknowns and changes:

Certain facts are missing

Models are built based on knowns:

$$\text{input} \rightarrow \boxed{\text{BLACK BOX}} \rightarrow \text{output}$$

Hence models

- should have room for unknowns.
- allow changes to be made and new components to be added with the minimum effort and cost.

Object-oriented approach:

1. Local individuals: Classes: entities, parameters, activities $\rightarrow$ interactions with other objects.
2. Global states: Networks: objects interact with each other, collective responses and effect.

6 Discussion

Auditory neuroscience is in need of hypotheses to test, both concerning the mechanism for level-tolerant representations within VCN, and for alternative strategies to solve the problem of level tolerance in the neural code.

In this report we have addressed a number of issues from a mathematical perspective, and in particular have answered the question “Can the selective listening hypothesis work? ”, the answer being “yes”. Two more natural questions were raised in the initial presentation of Chris Sumner:

- Can fast inhibitory processes operate to recover information?
- Are there other possible mechanisms? e.g. Within channel temporal mechanisms, network processing, distributed coding.

For these two important questions we drew no concrete conclusions.

Appendix

Matlab code for evolving equation (7):

```matlab
for k=2:length(T)
    t = T(k);
    V(k,:) = zeros(size(x));
    for kk=1:length(x0)
        for ii=1:length(t0{kk})
            I = min(find(T>=t0{kk}(ii)));
            I = I-1;
        end
    end
end
```
if(I > k)
    break;
else
    if(type(kk)==1)
        Vs=Ve;
    else
        Vs=Vi;
    end
    V(k,:) = V(k,:)+(Vs-V(I,:)).*Green(x,x0(kk),t,t0{kk}(ii));
end
end

References


