A mathematical analysis of sound coding in the auditory nerve: Spatiotemporal summing mechanisms

I. C. Bruce a L. S. Irlight b G. M. Clark a

a The Human Communication Research Centre, Department of Otolaryngology, The University of Melbourne, 384-388 Albert Street, East Melbourne VIC 3002, AUSTRALIA
b The Australian Bionic Ear and Hearing Research Institute, 384-388 Albert Street, East Melbourne VIC 3002, AUSTRALIA

1 Introduction
An understanding of mechanisms by which the human auditory system codes acoustical information has application to automated speech recognition and cochlear implants. In this paper, we formulate and analyze a mathematical model of the auditory system. The results are used to evaluate certain key theories of intensity and frequency coding.

The peripheral auditory system codes sound properties in the firing patterns of the auditory nerve (AN). The 30,000 AN fibers are each tuned to a particular frequency of sound, and the response of each fiber is a sudden jump in its electrical potential known as a spike or action potential (AP). Since these spikes are largely identical, sound properties must be encoded by the place and timing of the spikes.

Many theories of intensity and frequency coding reflect the convergence of groups of AN fibers on individual neurons of the brain stem by assuming spatial summation of nerve fiber activity [7][8]. This motivates the analysis of population, or summed responses. However, due to the large number of fibers in the AN, physiological investigations of such responses are quite difficult.

Mathematical models can easily deal with summed stochastic systems, and can also be used to determine the information content of various aspects of neural firing patterns [4]. In order to best achieve this, the model developed here is general enough to encompass a wide range of coding theories, analytical to facilitate investigation of information content and parametric to permit such an investigation under various scenarios.

Previous models such as that of [1] show the validity of summing mechanisms, but do not combine all the desired features.

In Section 2, the neural response models are developed. Section 3 details the properties of a summing mechanism. Section 4 applies the model to explore the feasibility of various theories of hearing. Conclusions are drawn in Section 5.

2 Model of AN response
The AN’s response can be closely approximated by a series of stochastically distributed identical spikes, where the spike probability is conditioned on the time since the last spike. It can therefore be modelled via a Self-Exciting Point Process (SEPP) model [5]. Here the firing probability is related to the properties of the individual neuron, the intensity of the stimulus, and the time since the neuron last fired.

Thus, following [3], we make the following assumption:

Assumption 2.1 The stochastic process describing the timing of the action potentials is a SEPP [5], with instantaneous rate (intensity) equal to \( s(t) r(\gamma) \), where \( s(t) \geq 0 \) is a stimulus related function, and \( r(t-t_{N}) \geq 0 \) is a recovery function which lowers the rate of action potential generation as a function of \( t - t_{N} \), the time since last action potential.

It is possible to define an “expected” rate of neural response, \( \lambda(t) \overset{\Delta}{=} E[s(t) r(\gamma)]/t \), which could be observed as the average of the rates of a large number of identical neurons subjected to identical stimuli, or as the Per-Stimulus Time Histogram (PSTH) which is the averaged response of one neuron subject to repeated identical stimuli.

Approximation 2.1 \( r(\gamma) = U(t-a) \), where \( U(.) \) is the unit step function, and \( a \) is the non-paralyzable dead time (absolute refractory period) of the fiber, i.e. the minimum possible time between neural firings.
Approximation 2.1 does not significantly affect the statistics of neural response [2]. It permits the analysis of neural response in terms of the stimulus and refractory functions, and is applied here to facilitate calculation of the variance of the neural response. To specify these functions, it is necessary to experimentally examine the PSTH ($\lambda(t)$) of neurons subjected to a variety of acoustical stimuli. The results indicate an approximation to the discharge rate of the $n^{th}$ AN fiber in response to a pure tone:

$$\lambda_n(t) = A_n + B_n \sin(\omega t + \phi_n)$$

**Remark**

(i) Together with the preceding SEPP theory, Equation (1) forms the basis of the neural models of this paper.

(ii) The Synchronization Index (SI) measures the amount with which the neural response is synchronized to the phase of the stimulus. It can be expressed as $S_f = \frac{F_n}{F_0}$, and is obtained experimentally from the period histogram.

Based on further experimental and computer results not reported here, and to facilitate the development of a general summation model, the following approximation is made.

**Approximation 2.2** $A_n$ and $B_n$ are constant for fibers close to the characteristic place (CP) and are denoted as $A$ and $B$. For $N$ fibers spread out evenly over a distance $d$ on the basilar membrane (BM), the phase delay of the $n^{th}$ fiber can be expressed as $\phi_n = \frac{ad}{N}$. Thus (1) can be approximated by:

$$\lambda_n(t) = A + B \sin\left(\omega t + \frac{ad}{N}\right)$$

In order to develop a spatial summing mechanism, we introduce a further approximation

**Approximation 2.3** Input fibers are given equal weighting in the summing mechanism.

Under Approximations 2.1, 2.2 and 2.3, the sum of the rates of $N$ fibers close to the CP spread out evenly over a distance of $d$ on the BM can be expressed as:

$$S(t) = \sum_{n=0}^{N-1} [A + B \sin\left(\omega t + \frac{ad}{N}\right)]$$

From (2), the rate for a small length $dx$ with a nerve fiber density of $\frac{N}{d}$ can be expressed as a continuous function of time and position $x$ on the BM.

$$\lambda(x,t) = \{A + B \sin(\omega t + \alpha x)\} \frac{N}{d} dx$$

**Approximation 2.4** $S(t)$ can be approximated by integrating the $\lambda(x,t)$ of (4) over a distance $d$.

$$I(t) = NA + \frac{2NB}{\omega d} \sin\left(\frac{\alpha d}{2}\right) \sin\left(\omega t + \frac{\alpha d}{2}\right)$$

A total spatiotemporal count, $R(t, \tau)$, also permits the analysis of coincidence detection [1], a form of spatial summation where temporal integration of the summed spatial input takes place over a short period of time:

**Lemma 1** Define $R(t, \tau)$ as the expected number of spikes from the summing mechanism in the time interval $[t - \tau, t]$. Then

$$R(t, \tau) = NA + \frac{4NB}{\omega d} \sin\left(\frac{\omega \tau}{2}\right) \sin\left(\omega t + \frac{\alpha d - \omega \tau}{2}\right)$$

3 **Model Response Properties**

The brain does not receive an “average” response as provided by the models of Section 2, but rather a single realization of a stochastic process. Thus, in order to determine the information content of the signals it receives, we must calculate the variance of the signal, thereby bounding the average error between a single realization and the process average.

The variance can be calculated exactly under the condition of a constant stimulus function, and this occurs when both the incident tone is at a constant intensity, and the synchronization index of the fiber (or the MI of a summed population of fibers) is zero. A zero SI or MI occurs naturally in two situations. The SI is at a maximum at low frequencies and drops rapidly when the frequency rises above 1 kHz. Also, as discussed in Section 4, a summed population of nerve fibers will have a low MI when the region of summation and phase delay cause the phases of the individual neural responses to cancel out.

**Lemma 2** For a fixed $s(t) = s_0$, and $r(t) = U(t - a)$, and assuming an equilibrium state, the number of spikes counted in the time interval $[t - \tau, t]$ is denoted $(N_e(t, \tau))$ and has mean and variance:
Fig. 1. Detectability vs integration time and spike rate

\[
E[N_e(t, \tau)] = \frac{s_0}{1 + as_0} \tau
\]
\[
\text{var}[N_e(t, \tau)] = 2 \sum_{n=1}^{\infty} \frac{s_0^n}{(n-1)!} U(\tau - an) \times \int_0^{\tau-an} [\tau-an-\gamma]^{n-1} e^{-s_0 \gamma} d\gamma + \frac{\tau s_0}{1 + as_0} - \left( \frac{\tau s_0}{1 + as_0} \right)^2 \]

If the single fiber or summed rate is not constant, Lemma 2 no longer applies, and other approximations such as treating \( s(t) \) as slowly varying [6], or treating the SEPP as a Poisson Process must be applied.

Expressions for the process variance permit analysis of neural information transmission capabilities. We analyze a detectibility measure \( d'_e \), which takes into account the spontaneous discharge rate. It can be used to bound the frequency of relative errors of a given size and permits a useful comparison with psychophysical results.

\[
d'_e \triangleq \frac{E[N_e(t, \tau)] - E[N_S(t, \tau)]}{0.5 \left[ \sqrt{\text{var}[N_e(t, \tau)]} + \sqrt{\text{var}[N_S(t, \tau)]} \right]}
\]

where \( N_S(t, \tau) \) is the number of spontaneous spikes observed in the period \( \tau \) before time \( t \).

Values of \( d'_e \) for a single fiber with dead time of 1ms are calculated via Lemma 2, and graphed in Figure 1. It is also useful to develop an expression for the modulation in rate experienced by the process.

Definition 3.1 The MI is defined as the magnitude of a variable component with a particular frequency, divided by the variable mean. \( MI \triangleq \frac{|X(\omega)|}{X} \)

For the integral expression \( R(t, \tau) \), the MI is:

\[
MI = \left| \frac{4B \sin \left( \frac{\omega \tau}{2} \right) \sin \left( \frac{\omega d}{2} \right) \sin (\omega t + ad - \frac{\omega \tau}{2})}{A \omega \rho \alpha \omega} \right|
\]

In Figure 2, the behaviour of the analytical expression for MI is compared with the output of the computer model including a numerical spatial summation mechanism. The computer simulation is not restricted to the case where \( A_n, B_n \) and \( \alpha \) are assumed to be constant over distance \( d \), and the waveforms for \( \lambda_n(t) \) is a better estimate of the neural firing probabilities than (2).

4 Consequences of the Model

Intensity Coding
A good measure of the sound intensity of a continuous tone will not include fluctuations over time. Therefore a MI close to zero is a useful property for intensity coding. The MI = 0 when \( \omega \tau = 2\pi \) or when \( ad = -2\pi \). It is unlikely that the integration period \( \tau \) can adapt itself to any given sound wave frequency maintain a MI

\[
\begin{align*}
\frac{\text{var}[N_e(t, \tau)]}{\text{var}[N_S(t, \tau)]} & = \frac{2 \sum_{n=1}^{\infty} \frac{s_0^n}{(n-1)!} U(\tau - an) \times \int_0^{\tau-an} [\tau-an-\gamma]^{n-1} e^{-s_0 \gamma} d\gamma + \frac{\tau s_0}{1 + as_0} - \left( \frac{\tau s_0}{1 + as_0} \right)^2} \end{align*}
\]

\[
\text{var}[N_e(t, \tau)] \approx \frac{s_0}{(s_0 a + 1)^3} \tau
\]
of 0. From physiological and psychophysical investigations it is also unlikely that \( \tau \) is long enough for the MI to approach 0 through a time-averaging process. However the MI can be maintained at 0, independent of the integration time, if the spatial summation distance corresponds to a phase difference of \(-2\pi\). This could easily be “hard-wired” into the summation model.

In order to calculate the mechanism properties required to zero the MI, we assume that the phase curve at the point of the summing mechanism is identical to the phase curve seen at the BM. Figure 3 shows that the critical bandwidths are close to the CF widths for summation distances corresponding to a MI = 0. Thus our model approximately predicts the critical bandwidths. Results which indicate a lowering of variance due to small amounts of low frequency synchronization help predict the systematic error at low frequencies [6]. In addition, the summation model with dead time effects displays an increase in detectability (or discrimination) with increasing sound intensity. This explains why relative difference limens for intensity can decrease with increasing intensity, known as the “near miss” to Weber’s law in which relative difference limens remain constant over intensity [7].

**Frequency coding**

A summing mechanism with similar properties to those optimal for intensity coding could also be used to improve the spatial rate profile used in rate-place frequency coding, where the place of maximum rate, or alternatively the edge of the rate profile, codes the tone frequency.

For theories of temporal frequency coding, a high MI is needed for the coding of the tone period. This will be maintained when the summation distance is close to zero, and when the temporal integration time is very short. The small summation distance is inconsistent with optimal intensity coding parameters. Therefore, any form of temporal frequency coding needs vastly different summing mechanisms to those for intensity coding. However, calculations of the number of fibers needed for each mechanism indicate that it is quite possible for multiple mechanisms to exist in parallel.

## 5 Conclusion

The analytical models for auditory nerve response developed in this paper permit quantitative analysis of many hypothesized AN sound coding mechanisms. This closed-form model is of particular use for the optimization of coding parameters. It predicts that spatial summation over a distance corresponding to critical bandwidths provides optimal intensity discrimination. It explains how the near miss to Weber’s Law could be due to dead time effects causing increasing discrimination with increasing signal intensity. The models also provide important restrictions on possible intensity and frequency coding mechanisms.

### References


Author/s:
Bruce, I. C.; Irlicht, L. S.; Clark, Graeme M.

Title:
A mathematical analysis of sound coding in the auditory nerve: spatiotemporal summing mechanisms

Date:
1995

Citation:

Persistent Link:
http://hdl.handle.net/11343/26933