A Multi-Threshold Neural Network for Frequency Estimation

L. S. Irlicht and Ian C. Bruce *and G. M. Clark

The Bionic Ear Institute, Australia

lsi@mail.medoto.unimelb.edu.au

ABSTRACT

Human perception of sound arises from the transmission of action-potentials (APs) through a neural network consisting of the auditory nerve and elements of the brain. Analysis of the response properties of individual neurons provides information regarding how features of sounds are coded in their firing patterns, and hints as to how higher brain centres may decode these neural response patterns to produce a perception of sound. Auditory neurons differ in the frequency of sound to which they respond most actively (their characteristic frequency), in their spontaneous (zero input) response, and also in their onset and saturation thresholds. Experiments have shown that neurons with low spontaneous rates show enhanced responses to the envelopes of complex sounds, while fibres with higher spontaneous rates respond to the temporal fine structure. In this paper, we determine an expression for the Cramer-Rao bound for frequency estimation of the envelope and fine structure of complex sounds by groups of neurons with parameterised response properties. The estimation variances are calculated for some typical estimation tasks, and demonstrate how, in the examples studied, a combination of low and high threshold fibres does not improve the estimation performance of a fictitious 'efficient' observer, but may improve the estimation performance of neural systems, such as biological neural networks, which are based on the detection of dominant interspike times.

1. Introduction

The auditory system forms a remarkably efficient neural network for the processing of sound. An understanding of how this system can perform tasks such as the separation of simultaneous sounds, the efficient processing of speech, and the identification of speakers, will lead to advances in the design of artificial neural networks for similar tasks, and also aid in the design of hearing prostheses such as cochlear implants. Identification of the mechanisms by which the auditory system codes properties of sounds is a first step to such an understanding.

The auditory system can functionally be broken up into two major sections. The first section transduces sound waves into neural firing patterns, and comprises the outer, middle and inner parts of the ear. The sound pressure waves cause vibrations of the eardrum which are then transmitted via the ossicles of the middle ear to fluid within the cochlea (inner ear). This results in travelling waves which propagate along the surface of the basilar membrane (BM). Analogous to a continuous filter bank, the mechanical structure of the BM causes each travelling wave to reach its maximum amplitude at a position determined by its frequency. Each auditory nerve fibre is excited by the vibration of a narrow region of the basilar membrane, and is consequently tuned to a specific frequency, termed its 'characteristic frequency'.

The second section is a multi-layer neural network which undertakes the bulk of the processing. It consists of the auditory nerve itself, and the auditory areas of the brain stem and cortex. In this paper, we will mainly be interested in the response properties of the input layer of the network - the auditory nerve (AN).

There are approximately 30,000 fibres in the auditory nerve. When sufficiently stimulated, an increase in a fibre's membrane's permeability to sodium ions is initiated, and the corresponding influx of sodium causes a sudden jump in its transmembrane potential known as a spike or action potential (AP). Since these spikes are largely identical, it is generally accepted that sound properties are coded by the place (the characteristic frequency of the neuron on which a spike occurs) and the timing of the spikes. Thus, any theory of neural sound coding must explain how the temporal (time-period) and/or spatial firing patterns are decoded to produce auditory percepts. Most do this by proposing that perceptual information is coded in one or another aspect of the neural firing pattern, such as the spike rate or the distribution of the interspike times, measured across either a single neuron or a population of neurons.

However, auditory nerve neurons differ in more

---

*Dept. of Otolaryngology, University of Melbourne

than just their characteristic frequencies. They also differ in their spontaneous firing rates, and in their response thresholds. These response differences suggest that auditory sound coding could be based on more than just the CF of the neurons. In fact, physiological experiments demonstrate that when stimulated by a complex sound, fibres with low spontaneous rates predominantly respond to the envelope, and those with high spontaneous rates to the fine temporal structure of the sound [1, 2].

The neural response thresholds are highly correlated to the spontaneous rates [3, 4], but reasonably independent of characteristic frequency. Thus the brain stem receives information from fibres which may approximately be parameterised in a two-dimensional response space - where one parameter represents characteristic frequency, and the other represents threshold. Much work has been done to understand how the responses of fibres with different characteristic frequencies are synthesised for the task of frequency estimation [5, 6, 7], but very little analysis has been applied to understanding the role that fibres with different thresholds play in the same task.

In this paper we investigate the importance of having a multi-threshold system. This is achieved by generating a model of neural response to a complex sound, and investigating via Cramer-Rao bounds [8], and via the distribution of interspike times, the accuracy to which information about the frequencies within the sound may be estimated, based either on the output of high-threshold and/or low-threshold fibres.

2. Signal and Network Models

2.1. Signal Model

Consider a common estimation task performed by the auditory system: the analysis of the frequency components of a speech signal. Such signals are composed of complex sounds which exhibit a number of resonances (formants), all modulated by a voicing pitch. Perceptual experiments show that if the actual fundamental of the voicing pitch is missing from the spectrum, then the estimated voice pitch corresponds to the smallest difference between the harmonics present. This is a well noted auditory phenomenon known as the "missing fundamental" [1, 9].

Here we explore the estimation of the voice pitch from the temporal characteristics of neural response for a signal where two harmonics of the voice pitch are present, but not the fundamental.

The input to the neural network is taken to be the sound pressure wave passed through a linear filter, the cochlea. The filter characteristics of the cochlea to a 700 Hz tone is shown in Figure 1.

Thus, the filtered signal $s(t)$ is expressed as:

$$s(t) = 1 + \sum_{i=1}^{2} A_i sin(2\pi f_i t + \phi_i) \qquad (1)$$

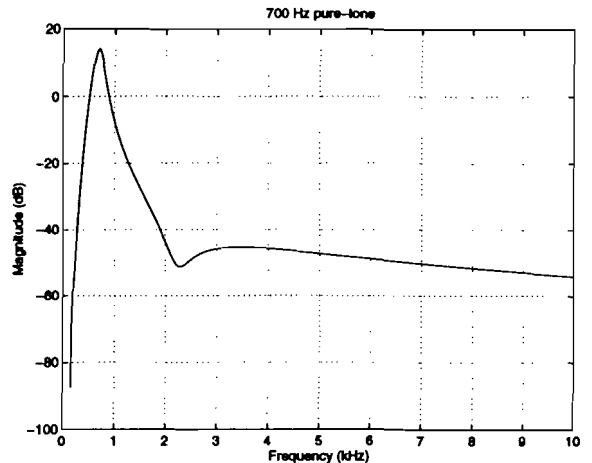where f_1 and f_2 are the harmonic components of the voice pitch present in the filtered signal.

Fig. 1: Filter characteristics of the cochlea to a 700 Hz tone.

The task is to estimate the voice pitch, $f_2 - f_1$, and its harmonics, f_2 and f_1. This could be done either by estimating f_1 and f_2 simultaneously and calculating the difference, or by introducing the $f_2 - f_1$ component to the signal via a nonlinearity and estimating the voice pitch directly from the modified signal.

2.2. Neural Network Model

The response of neurons of the auditory nerve may be modelled by an inhomogeneous Poisson process [10], where the intensity (response rate) is described by means of a compressively nonlinear (sigmoidal) function, which is brought about by a number of nonlinearities involved in AP thresholding and generation. From the form of input-output curves derived from physiological data [11] we take tanh(.) to be a suitable sigmoidal function. Although spontaneous rate is routinely used to classify fibre responses, a threshold shift can better explain the differing responses [12].

Thus the Poisson rate of the nth neuron, $r_n(t)$, may be described by:

$$r_n(t) = r_0 + \tanh\left(\alpha_n \left[s(t) - \beta_n\right]\right) \qquad (2)$$

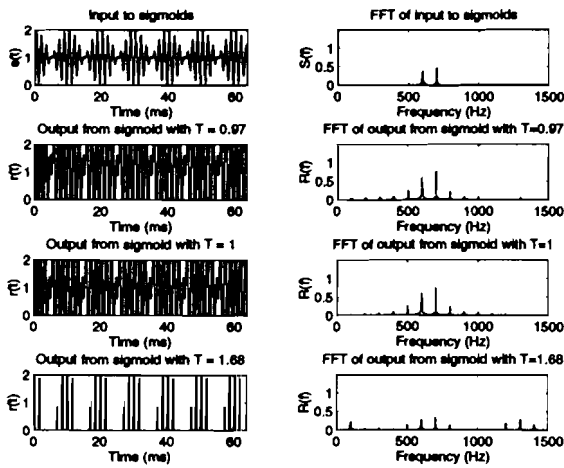where $s(t)$ is the cochlear filtered signal defined in the previous subsection.

Fig. 4: Ex. 2: Highly modulated $s(t)$. Output values of the sigmoids with optimal thresholds for 600, 700 and 100 Hz.

largely unknown, however statistical methods can yield information about the ability of various neural structures to estimate properties of the sound. In turn, these abilities help shed light on likely mechanisms for the information processing capabilities of the auditory system.

One method of analysing the ability of various mechanisms to code parameters is via the application of the Cramer-Rao Bound [8]. This permits a lower-bound to be given for the variance of the optimal unbiased estimator for the parameter in question. Of course, such an optimal estimator may not exist, or even be compatible with the structures of the auditory system. Such an analysis is still useful, however, because it can rule out mechanisms which do not convey the required information.

The following Lemma is based on calculations performed in [13], for the estimation of a pure tone.

Lemma 2.1 *Consider an observation for duration T of an inhomogeneous Poisson process with rate $r(t, f, A)$. Then the Cramer-Rao inequality can be expressed as:*

$$\frac{1}{\hat{\sigma}_f^2} \leq \int_0^T \frac{1}{r(t, f, A)} \left[\frac{\partial r(t, f, A)}{\partial f} \right]^2 dt$$

This result can be extended to define the Fisher Information Matrix $\mathbf{I}(\boldsymbol{\theta})$, for the estimation of the unknown vector parameter $\boldsymbol{\theta} = [A_1, \omega_1, \phi_1, A_2, \omega_2, \phi_2]'$, where A_i, ω_i, ϕ_i are the parameters of $s(t)$ as described in Equation 1.

Lemma 2.2 *Consider observations of a number of inhomogeneous Poisson processes, with rates $r_n(t, \boldsymbol{\theta})$. In this case, the Cramer-Rao inequality can be expressed as:*

$$[\mathbf{I}_n(\boldsymbol{\theta})]_{ij} = \int_0^T \frac{1}{r_n(t, \boldsymbol{\theta})} \frac{\partial r_n(t, \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial r_n(t, \boldsymbol{\theta})}{\partial \theta_j} dt$$

$$\sigma_n^2(\theta_i) \geq [\mathbf{I}_n^{-1}]_{ii} \tag{3}$$

Remark 1: A standard result of Cramer-Rao theory, shows that the information matrix of the combined results of independent experiments equals the sum of the information matrices of each individual experiment. Thus, under the assumption of conditional independence of auditory nerve responses, a calculation of the Fisher Information Matrix of the output of two or more neurons can be achieved by summing the individual matrices. This facilitates easy comparison of the output of various groups of fibres, and the ability to take the output of one fibre, and select the fibre which minimises the estimator variances based on the combined information of both fibres.

Thus, the evaluation of Lemma 2.2 where the response rates are taken from the neural and signal models of Section 2.1 and 2.2 enables calculation of bounds on estimator performance based on the outputs of a number of neurons.

In the case of a sigmoidal response function (2), and sinusoidal signal model (1), the integral of (3) does not appear to be analytically tractable, and consequently it is not solvable for generalised conditions. However, it is numerically solvable for any given parameters, and in a later section we numerically investigate estimator variance for some representative situations.

2.5. Interspike-Time Analysis

Although the mechanisms by which the auditory system codes frequency are still largely unknown [14], it has been hypothesised that one method may be via the detection of dominant time-intervals between neural responses - effectively picking the period of the response waveform. This could be achieved via a series of delay lines and coincidence detectors [15, 16]. How would the thresholding neurons effect this kind of system?

Although the Cramer-Rao bounds of the previous sections can limit the variance of estimators based on neural responses, they can not indicate the degree to which the auditory system's variances follow the optimal bounds, and consequently are not necessarily an accurate measure of how useful the output of a selected neuron is to the auditory system. To investigate this question, we utilise the distribution of inter-spike times.

Lemma 2.3 *Consider an Inhomogeneous Poisson Process with rate $s(t)$, over the time interval $[0, T]$.*

2.3. Filtering Properties of the Neural Model

Changing the steepness and position of the sigmoidal tanh(.) allows the simulation of a range of neural responses with various onset and saturation thresholds, and these nonlinear responses will attenuate or magnify various components of the sound spectrum. Fourier analysis can be used to find the threshold value that minimises a cost function which measures the relative magnitude of a specific frequency component at the output of the sigmoid. Such a cost function can include a penalty function which prevents the absolute magnitudes of the major components from being overly attenuated.

For the estimation task described in Section 2.1, we are interested in the components at frequencies f_1 and f_2, and the missing fundamental of the voice pitch, $f_2 - f_1$, and consequently perform the Fourier analysis described above to determine thresholds which accentuate each of these components.

The relative magnitudes of the two voice-pitch harmonics present in $s(t)$ will depend on their magnitudes in the sound pressure wave and on the filter characteristics of the cochlea at the place of the fibre's input. It is therefore possible to have a range of modulation depths in the signal. Here we investigate two signals with magnitudes chosen arbitrarily to produce a slightly modulated $s(t)$ (Example 1) and a highly modulated $s(t)$ (Example 2).

Example 1: Slightly modulated $s(t)$

Consider the signal and neural response:

$$s(t) = 1 + \frac{1}{6}\sin(2\pi 600t) + \frac{5}{6}\sin(2\pi 700t)$$
$$r(t) = 1 + \tanh\left(10(s(t) - T)\right)$$

where T is a threshold shift.

Example 2: Highly modulated $s(t)$

Consider a neural response the same as for Example 1, but with the signal:

$$s(t) = 1 + \frac{1}{2}\sin(2\pi 600t) + \frac{1}{2}\sin(2\pi 700t)$$

For both examples, Fourier analysis of the output of each sigmoid was used to maximise the relative size of its components at the frequencies 600, 700 and 100 Hz from among the parameterised sigmoid function:

$$r(t) = \tanh\left(10(s(t) - T)\right)$$

The optimal threshold values, T, are shown in Table 1.

The sigmoids are shown in Figure 2, and the input and output of the sigmoids and their Fourier

Frequency (Hz)	600	700	100
Ex. 1: Optimal T	1.73	1.00	1.75
Ex. 2: Optimal T	0.97	1.00	1.68

Table: 1: Optimal Thresholds for Examples 1 and 2

transforms are shown in Figures 3 and 4. The implications of the filtering properties of the sigmoidal nonlinearity will be investigated in the next section.
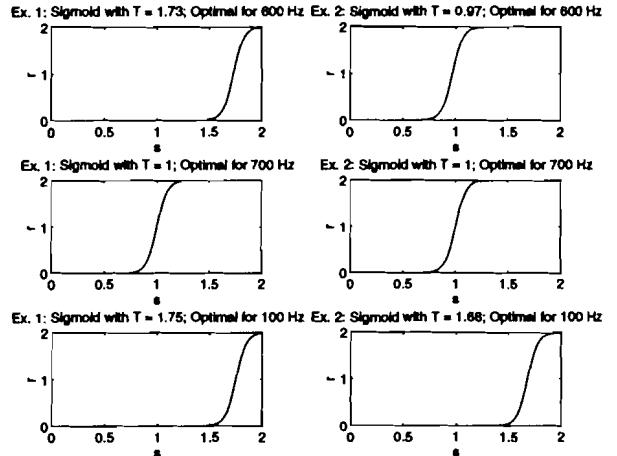

Fig. 2: Left: Ex. 1 - Slightly modulated $s(t)$. Sigmoids with optimal thresholds for 600, 700 and 100 Hz (top to bottom). Right: Ex. 2 - Highly modulated $s(t)$. Sigmoids with optimal thresholds for 600, 700 and 100 Hz (top to bottom).
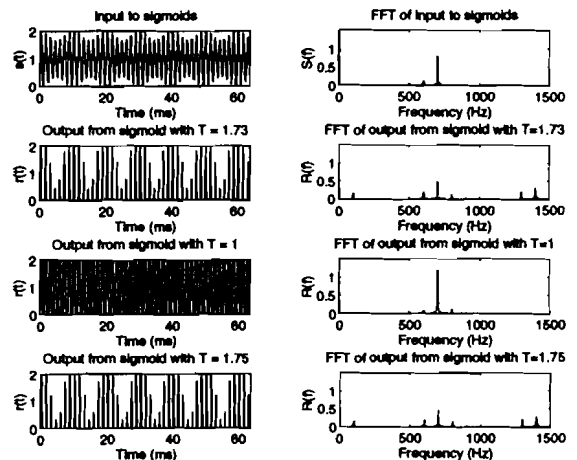

Fig. 3: Ex. 1: Slightly modulated $s(t)$. Output values of the sigmoids with optimal thresholds for 600, 700 and 100 Hz.

2.4. Cramer-Rao Bounds for Neural Frequency Estimation

The auditory system takes the responses of some 30,000 auditory nerve neurons, and can produce estimates of the amplitudes, A_i, and frequencies ω_i of the sound $s(t)$. Exactly how this is achieved is

Then the distribution of spikes occurring with a gap of τ is:

$$D(\tau) = \frac{\int_0^{T-\tau} s(t)s(t+\tau)dt}{\left[\int_0^T s(t)dt\right]^2 / 2}$$

Remark 2:

This distribution measures the frequency of spikes occurring with a time difference of τ, regardless of the existence of spikes within the interval. It is consistent with the type of estimator proposed earlier in this section. An alternative expression for the distribution of inter-spike times can also be generated.

The results of Lemma 2.3 are used in a later section to calculate the effect of the sigmoidal nonlinearity on the interspike-time distribution. Similar to the Cramer-Rao bounds, the integral appears analytically intractable, but can easily be calculated numerically for specific examples.

3. Results

In this section, the effect of various neural thresholds on the frequency estimation task of Section 2.1 is determined for the case of an efficient estimator, and also for an interspike-time based estimator.

3.1. Estimation by an Efficient Estimator

Analytical descriptions of the inner terms of the integral given in (3) are derived for a signal of the form expressed in (1). The integrals, however, appear not to be analytically tractable, and therefore numerical integration was implemented using an adaptive recursive Newton Cotes 8 panel rule. Cramer-Rao Bounds were calculated for Examples 1 and 2 of Section 2.3, with all three possible combinations of low-threshold ($T = 1.00$) and high-threshold ($T = 1.80$) sigmoid pairs (L+L, H+H and L+H). The bounds were evaluated over 20 ms (Table 2) and 100 ms (Table 3).

Ex.	T	600 Hz	700 Hz	100 Hz	$\sum$
	L+L	7.24e-3	1.73e-4	7.80e-3	1.52e-2
1	H+H	4.32e-3	3.77e-4	5.05e-3	9.75e-3
	L+H	5.38e-3	2.31e-4	5.97e-3	1.16e-2
	L+L	4.01e-4	4.03e-4	9.00e-4	1.70e-3
2	H+H	1.23e-3	1.16e-3	2.65e-3	5.04e-3
	L+H	5.89e-4	5.82e-4	1.34e-3	2.51e-3

Table: 2: Cramer-Rao Bounds for Examples 1 and 2: 20 ms

In some cases the low-threshold sigmoids produced the smallest bound, in others the high-threshold, but in none of the examples studied did the combination of thresholds (L+H) produce the

Ex.	T	600 Hz	700 Hz	100 Hz	$\sum$
	L+L	5.59e-5	1.16e-6	5.72e-5	1.14e-4
1	H+H	2.85e-5	3.04e-6	3.16e-5	6.31e-5
	L+H	3.77e-5	1.67e-6	3.95e-5	7.89e-5
	L+L	2.85e-6	2.88e-6	5.46e-6	1.12e-5
2	H+H	8.32e-6	8.27e-6	1.69e-5	3.35e-5
	L+H	4.23e-6	4.27e-6	8.24e-6	1.67e-5

Table: 3: Cramer-Rao Bounds for Examples 1 and 2: 100 ms

smallest error. This indicates that for the situations investigated, an 'efficient' estimator will operate best when the thresholds are identical, although different situations will have different optimal thresholds, possibly validating the need for auditory nerve fibres with different onset and saturation thresholds.

3.2. Estimation from the Interspike–Time Distribution

One possible mechanism for frequency estimation would be to utilise the interspike-time distribution as examined in Section 2.5 to measure the dominant interspike time observed between two fibres [17]. Figure 5 shows the relative frequency of occurrence of interspike times for Examples 1 and 2, with a low-threshold and a high-threshold sigmoid.
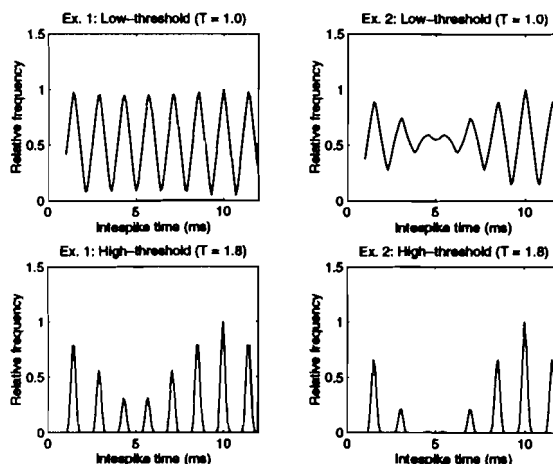

Fig. 5: Left: Ex. 1 - Slightly modulated $s(t)$. Interspike-time distribution for low-threshold (T = 1.0) and high-threshold (T = 1.8) sigmoids (top and bottom). Right: Ex. 2 - Highly modulated $s(t)$. Interspike-time distribution for low-threshold (T = 1.0) and high-threshold (T = 1.8) sigmoids (top and bottom).

For both Examples 1 and 2, the interspike time corresponding to the weighted average of the periods resulting from the 600 and 700 Hz components of the signal is emphasised by the low-threshold sigmoids, and to the period of the 100 Hz "missing fundamental" by the high-threshold sigmoid.

This suggests that for a system measuring inter-spike times, a combination of low-threshold and high-threshold fibres is useful to estimate all three frequency components when using this estimation technique, particularly if further filtering is used to extract only one frequency per fibre (eg. [7]).

4. Conclusion

The neurons of the input layer of the auditory system (the auditory nerve) may be parameterised in terms of their response properties including the frequency of sound to which they best respond, and their response thresholds. For the task of frequency estimation, we have measured the importance of combining the output of auditory nerve neurons with differing thresholds . The resulting Cramer-Rao bounds permit the calculation of frequency estimation variance for any given neural parameters and show, for the calculated examples, that an 'efficient' observer of the output of two neurons may not necessarily benefit from a mixing of differently thresholded neurons. However, it has been hypothesised that auditory frequency estimation is largely based on the detection of action potentials with suitably defined delays, and we demonstrate how combining the output of differently thresholded neurons may improve the frequency estimation capabilities of such a system. Due to the numerical nature of the calculations, these results are not calculated parametrically, but are demonstrated for a number of specific examples. An open question is the extension of these results to a general case - thereby specifying conditions under which the combined threshold responses are more useful than single threshold responses, and conditions under which they are not.

References

[1] G. Langner, "Periodicity coding in the auditory system," *Hear. Res.*, vol. 60, pp. 115-142, 1992.

[2] J. W. Horst, E. Javel, and G. R. Farley, "Coding of spectral fine structure in the auditory nerve. I. fourier analysis of period and inter-spike interval histograms.," *J. Acoust. Soc. Am.*, vol. 79, pp. 398-416, February 1986.

[3] M. C. Liberman, "Auditory nerve response from cats raised in a low noise chamber," *J. Acoust. Soc. Am.*, vol. 63, pp. 442-455, 1978.

[4] I. M. Winter, D. Robertson, and G. K. Yates, "Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibres," *Hear. Res.*, vol. 45, pp. 191-202, 1990.

[5] M. I. Miller and M. B. Sachs, "Representation of voice pitch in discharge patterns of auditory-nerve fibers," *Hear. Res.*, vol. 14, pp. 257-279, 1984.

[6] D. O. Kim and K. Parham, "Auditory nerve spatial encoding of high-frequency pure tones: Population response profiles derived from d' measure associated with nearby places along the cochlea," *Hear. Res.*, vol. 52, pp. 167-80, 1991.

[7] P. Srulovicz and J. L. Goldstein, "A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum," *J. Acoust. Soc. Am.*, vol. 73, pp. 1266-1276, April 1983.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory.* Prentice Hall Signal Processing Series, New Jersey: PTR Prentice Hall, 1993.

[9] E. Javel and J. B. Mott, "Physiological and psychophysical correlates of temporal processes in hearing," *Hear. Res.*, vol. 34, pp. 275-294, 1988.

[10] M. J. Penner, "Neural or energy summation in a poisson counting model," *J. Math. Psychol.*, vol. 9, pp. 286-93, 1972.

[11] M. B. Sachs, R. L. Winslow, and B. H. A. Sokolowski, "A computational model for rate-level functions from cat auditory-nerve fibers," *Hear. Res.*, vol. 41, pp. 61-70, 1989.

[12] J. W. Horst, E. Javel, and G. R. Farley, "Coding of spectral fine structure in the auditory nerve. II: Level-dependent nonlinear responses," *J. Acoust. Soc. Am.*, pp. 2656-2681, December 1990.

[13] W. M. Siebert, "Frequency discrimination in the auditory system: Place or periodicity mechanisms?," *Proc. IEEE*, vol. 58, pp. 723-30, 1970.

[14] J. O. Pickles, *An introduction to the physiology of hearing.* London: Academic Press Inc. Ltd, 1982.

[15] J. C. Licklider, "'Periodicity' pitch and 'place' pitch," *J. Acoust. Soc. Am.*, vol. 26, p. 945, 1954.

[16] G. M. Clark, L. S. Irlicht, and T. D. Carter, "A neural model for the time-period coding of frequency for acoustic and electric stimulation," *Presented at the 16th Annual Meeting of the Australian Neuroscience Society*, January 1996.

[17] D. Au, I. Bruce, L. Irlicht, and G. M. Clark, "Cross-fiber interspike interval probability distribution in acoustic stimulation: A computer modelling study," *Ann. Oto. Rhino. Laryn.*, vol. 104, no. 9, pp. 346-349, 1995.

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

Title:
A multi-threshold neural network for frequency estimation

Date:
1996

Citation:
Irlicht, L. S., Bruce, I. C., & Clark, G. M. (1996). A multi-threshold neural network for
frequency estimation. In Proceedings of the 7th Australian Conference on Neural Networks
(Canberra, 10-12 April 1996), Canberra, A.C.T.

Persistent Link:
http://hdl.handle.net/11343/26950