Lateral Inhibitory Networks for Auditory Processing

C.J. Sumner and D.F. Gillies
Department of Computing, Imperial College of Science Technology and Medicine,
180 Queens Gate, London SW7 2BZ, United Kingdom
Tel: +44 171 589 5111 ext 58378, E-mail: cjs2@doc.ic.ac.uk

ABSTRACT

A neural-network model is described that produces a rate-place representation from auditory nerve output that is of considerably higher frequency resolution than that from a standard auditory peripheral model. The neural circuits used are called Lateral Inhibitory Networks. They have long been known to be responsible for early spatio-temporal processing in the visual system. Here we investigate the use of such networks for early auditory processing. We describe the analytical basis, problems with various variants of the model, and show some initial results yielded by the research.

1. Introduction

The use of neural network models of the auditory system has been finding increasing popularity in auditory circles. The neural oscillator models of Brown *et al.* [3] show that neural networks can be used to perform auditory grouping as described by Auditory Scene Analysis [1]. Neural networks offer the modeller an alternative to the largely ad-hoc 'representational' approach that more is commonly used (e.g.[2]).

The cochlea can be considered a filter bank array. Much of the temporal fine structure in each channel is retained, so frequency information is stored in both spatial and temporal form. The spatial sorting of frequency alone cannot account for the observed frequency resolution of our hearing and temporal fine structure is used to give more accurate pitch. However, neurological investigations have shown that the auditory system organises its mechanisms tonotopically, suggesting spatial processing of frequency.

Lateral Inhibitory Networks (LINs) are known to perform image enhancement in the visual system. The work of Hartline [4] on the eye of the Limulus Crab showed how simply interconnected banks of neurons were performing spatio-temporal processing of the photoreceptor cell output. Adjacent cells 'inhibit' one another in order to sharpen the image. Shamma [6] has shown that such networks can also sharpen the frequency resolution of models of Auditory Nerve output.

Lateral Inhibitory Networks are non-adaptive networks with a single layer of neurons. Inputs are arranged tonotopically. Neurons receive excitatory input from the their preferred freqency channel and inhibitory input from neighbouring neurons' inputs or outputs. LINs can often be approximated by linear systems. This makes analytical treatment possible, unlike the case of symbolic auditory processing models. In such cases a network reduces to a 2-dimensional filter. A limitation of previous work on auditory LIN processing has been the lack of a firm analytical methods for network synthesis. Exact parameters for a network are chosen quite arbitrarily (see [4]), resulting in non-optimal processing. In the following sections, we shall look at the need, the possiblities and problems of proper design of LINs and their use for auditory modelling.

2. The LIN models and their analysis

2.1. General

LINs are configured as either recurrent or non-recurrent networks (Fig 1). A non-recurrent network neuron takes all its inputs from the actual inputs to the network. Each input is weighted differently, normally with excitatory central connections and inhibitory surround, connection strength falling off with distance. In a recurrent network, inhibitory inputs are the outputs from adjacent neurons. Input to the recurrent network is a single input to each neuron.

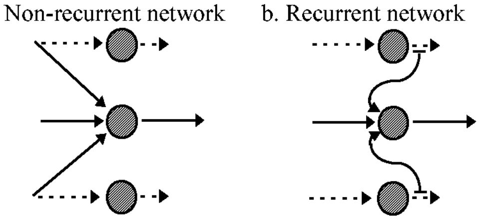


Fig 1. Neuron connectivities

2.2. A Non-recurrent network

Fig 2 shows a generic neuron. The non-recurrent network neuron has three parts: i. weighted sum of inputs; ii. low-pass filter; iii. sigmoidal non-linearity. In discrete time form, it can be described by the difference equations:

$$y_i(t) = \sum_{n=-N/2}^{N/2} c_n x_{(i+n)}(t) \qquad \text{(i)}$$

$$y_i'(t) = \sum_{p=0}^{P} b_p y_i(t-p) - \sum_{q=1}^{Q} a_q y_i'(t-q) \qquad \text{(ii)}$$

$$y_i''(t) = \frac{y_{\max}}{1+e^{-b(y_i'(t)-y_0)}} \qquad \text{(iii)}$$

where $x_i(t)$ is the i^{th} input at time t; N is the number of inputs to the neuron and c_n its weights; P and Q are the number of terms in the low pass filter and a_q and b_p its coefficients; y_{max}, b, and y_0 are the parameters controlling the non-linearity.

i/p sum feedback sum non-linearity

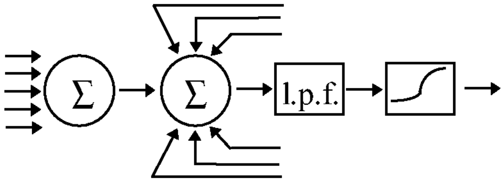


Fig 2. Generic Neuron Model

Ignoring the sigmoidal non-linearity, we have a linear system amenable to linear systems theory. Thus we can express the network in the two dimensional frequency domain. We refer to the transforms as being in *temporal* and *spatial* frequency.

$$H(\phi,\ \ ) = H(\phi)H(\ \ ) \qquad \textit{Overall response}$$

$$H(\phi) = c_0 + \sum_{n=1}^{N/2} 2c_n \cos(n\phi) \qquad \textit{Input sum}$$

$$H(\ \ ) = \frac{\sum_{p=0}^{P} b_p e^{-jp}}{1+\sum_{q=1}^{Q} a_q e^{-jq}} \qquad \textit{Low pass filter}$$

where ϕ and ω are spatial and temporal frequency.

2.3. A Recurrent Network

The recurrent network consists of the following stages: i. weighted sum of neuron outputs; ii. low pass filter; iii. sigmoidal non-linearity. Its discrete time formulation is as follows:

$$y_i(t) = x_i(t) + \sum_{n=-N/2}^{N/2} c_n y_{(i+n)}''(t-1) \qquad \text{(i)}$$

$$y_i'(t) = \sum_{p=0}^{P} b_p y_i(t-p) - \sum_{q=1}^{Q} a_q y_i'(t-q) \qquad \text{(ii)}$$

$$y_i''(t) = \frac{y_{\max}}{1+e^{-b(y_i'(t)-y_0)}} \qquad \text{(iii)}$$

If we again ignore the sigmoidal non-linearity, then *y''=y'*, and we end up with a linear system once more. The frequency response of this network can be expressed as

$$H(\phi,\omega) = \frac{B(\omega)}{1+A(\omega)-e^{-j\omega}B(\omega)C(\phi)}$$

where $$B(\omega) = \sum_{p=0}^{P} b_p e^{-jp\omega},\ A(\omega) = \sum_{q=1}^{Q} a_q e^{-jq\omega}$$

and $$C(\omega) = c_0 + \sum_{n=1}^{N/2} 2c_n \cos(n\phi)$$

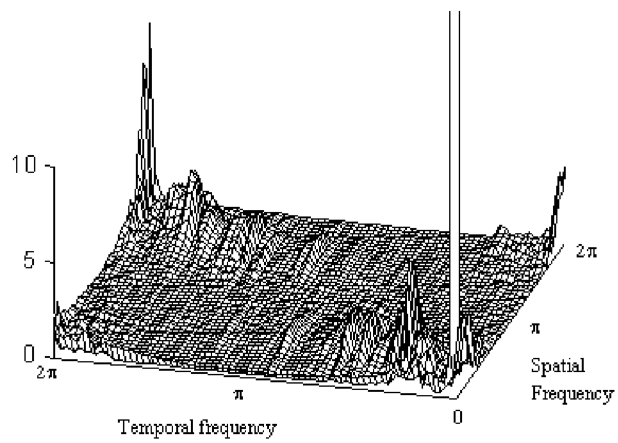
3. Network Synthesis and Examples

*3.1. Li*near *n*on-*re urr*en*t* ne*twork*

In a non-recurrent LIN, spatial and temporal filtering is decoupled, and synthesis of the network becomes synthesis of two separate 1-dimensional filters.

Fig 4 shows output from an auditory periphery model. The model consists of i. middle/outer ear bandpass filter; ii. bank of 100 8th-order gammatone filters; iii. Meddis '86 hair cells; iv. low pass filtering of each channel and decimation to 1kHz sampling rate.

From FFTs of the data, we can synthesize a network of the desired response. Fig 3 shows a 2-D transform of the output from the auditory periphery model. The filters are constructed as follows: The weighted sum inputs form a 15-term zero-phase FIR spatial bandpass filter, and the low pass filter is a 3rd order Chebychev. These can be easily synthesized using some filter design utility such as MATLAB. Fig 5 shows the result of processing using the network. This network was designed to produce an enhanced rate-place output. It makes clear components that were obscure in the peripheral output.



F*ig 3.* 2-dimensional FFT of output in Fig 4

*3.2. Li*near *r*e *urr*en*t mod*e*l*

We can linearize the recurrent model by taking $\mathbf{y}_i\ (t)$ as feedback. Here the spatial and temporal processing is coupled and so we must synthesize a two-dimensional filter. Work is still underway here, but filters are likely to be synthesized using approximation methods. The problem is further complicated by the need to ensure stability of the filter. Network stability has been found to be very sensitive to both initial conditions and data. Due to space restrictions, no example is given, but trial and error filter selection resulted in similar performance to the non-linearized model.

Whilst linear recurrent and non-recurrent networks have broadly similar capabilities, they both have different limitations. The ease of design of non-recurrent networks is set against the high number of terms required for accurate spatial filtering.

3.3. Non-linear re urrent model

For the non-recurrent network, non-linearity performs nothing more than a selective compression/expansion of the signal. Adding non-linearity to the recurrent network changes the processing, as the signal being fedback is different. We cannot at present design these networks, but we know in general terms how they behave.

The sigmoid function has been observed to perform three functions i. Half-wave rectification; ii. Increase/ decrease feedback by dynamic expansion/compression; iii. Limit feedback. Increasing the feedback can cause what was a stable linearized system to become unstable and vice versa. The saturation can act to limit the degree of instability. At sub-saturation levels, a stable non-linear recurrent network exhibits linear-like behaviour. However, an unstable non-linear network can have its instability limited by the saturation of the sigmoid, leading to hysteresis i.e. output depends on current and past input.

Fig 6 shows the output from a stable non-linear recurrent network. The output is roughly similar to the non-recurrent network, although not quite as clear. The spatial filter has 7-terms. It is very difficult to find networks that exhibit desirable results on real signals when in saturation. Instability is as likely to generate unwanted oscillations as clear feature extraction.

4. Discussion

The relative success of the non-recurrent model demonstrates the importance of adequate design procedures in order to produce a successful network. Whilst non-linear recurrent networks hold promise of more interesting behaviour and lower filter orders, their unpredictability makes it very difficult to produce a practical network. It is clear that we need analytical methods for designing recurrent networks in order to realise their potential. These findings contrast with Shamma's approach, which emphasized the supposed invariance for a wide variation in inhibitory profiles (i.e. filter coefficients).

Evidence of the need for spatio-temporal processing in the auditory system is abound. The gradual decay of phase locking in Auditory Nerve signals makes a mechanism that gracefully handles the transition between average-rate and temporal processing desirable. The widespread occurance of tonotopic maps throughout the auditory system, the frequency resolution limitations of average-rate models, and the use of temporal fine structure, all strongly suggest the use of LINs in auditory processing.

Although recent models of the Ventral Cochlea Nucleus's often use isoloated neurons with Hogkin-Huxley dynamics, inhibition has been found in the Cochlea Nucleus (see [5]) and is thought important for certain observed response patterns. It also seems unlikely that such a simple mechanism would not be at work at other places in the auditory pathway.

Computational Auditory Scene Analysis models typically use ad-hoc methods for forming components from auditory nerve data. In LINs we have the beginnings of a biologically plausible mechanism for producing auditory representations based on both rate and fine structure of signals in the auditory nerve. In their current state of development however, they cannot compete with ad-hoc algorithms, such as the cross-correlation map of Brown and Cooke [2].

5. Conclusion and future work

We have described a highly biologically plausible neural network model for spatio-temporal processing of the auditory nerve signal. However, in all but the simplest case, design of these networks is a difficult task. Current research is now addressing these problems.

6. References

[1] Bregman A.S.(1990). *" uditory Scene nalysis"*, MIT press, London.

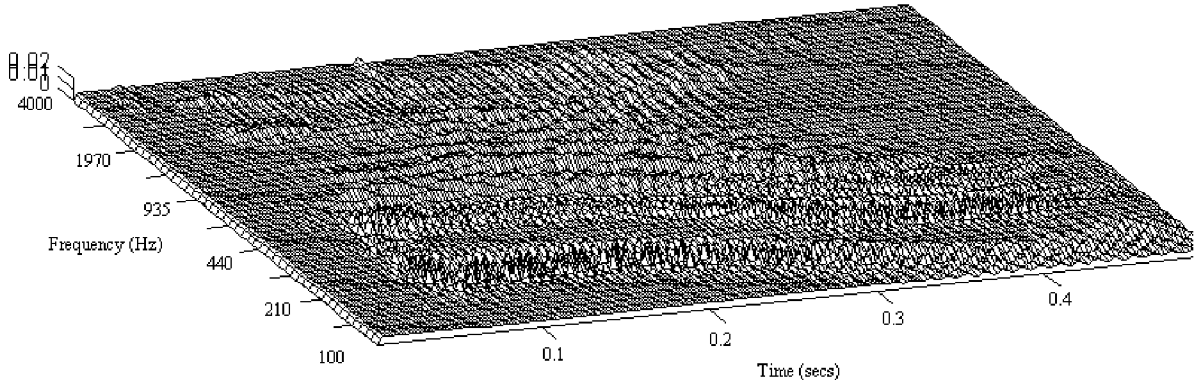
[2] Brown G.J., Cooke M.(1994), *"Computational uditory Scene nalysis"*, Comp. Spch. Lang., 8, P297-336.

[3] Brown G.J., Cooke M.(1995), *"Temporal Synchronisation in a neural oscillator model of primitive auditory stream segregation"*, Proc. IJCAI Workshop on CASA, Montreal 40-47.

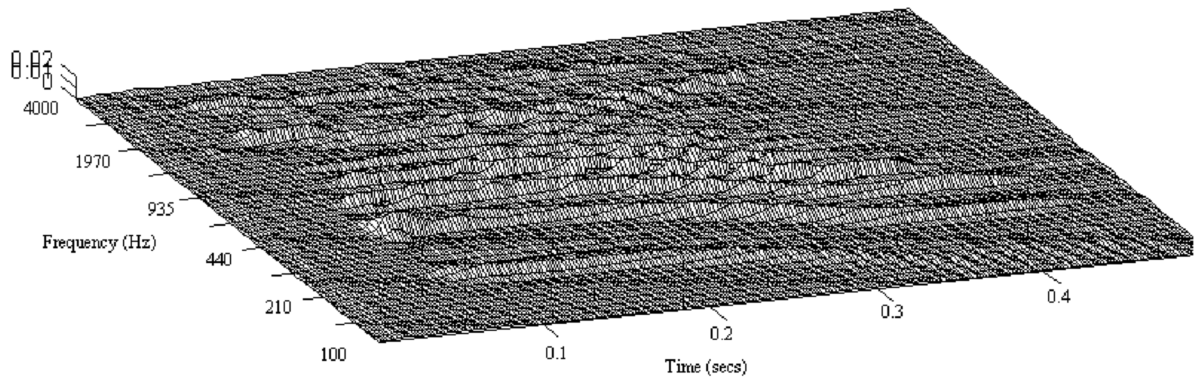
[4] Hartline H.K.(1974), *"Studies on Excitation and Inhi ition in the Retina"*, ed. Ed. Ratlif, The Rockefeller U.P., New York.

[5] Young E.D. et al.(1988), *"Repsonse Properties of Cochlea Nucleus Neurons"*, Auditory Function by G.M Edelman, Chapter 9, P277-312.

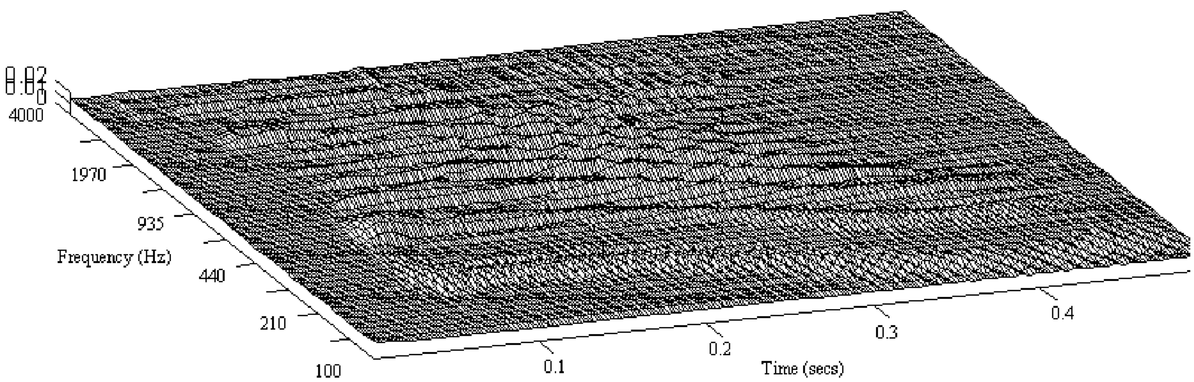
[6] Shamma S.(1989), *"Spatial and Temporal Processing in Central uditory Networks"*, Methods in Neuronal Modelling (eds. Koch, Segev). MIT press.



F*ig* **4**. Ouput from Auditory Peripheral for spoken word 'day'



F*ig* **5**. Results from processing using non-recurrent LIN



F*ig* **6**.. Results from processing using non-linear recurrent LIN