# Spectral Enhancement of Sounds by the Stellate Microcircuit of the Ventral Cochlear Nucleus

*Timothy Esler and David B. Grayden*

NeuroEngineering Laboratory, Dept. of Electrical & Electronic Engineering, The University of Melbourne, Victoria 3010, Australia
Centre for Neural Engineering, The University of Melbourne, Victoria 3010, Australia
tim.esler@gmail.com, grayden@unimelb.edu.au

## Abstract

T Stellate (TS) cells in the auditory cochlear nucleus have been shown experimentally to accurately encode sound over a range of input sound pressure levels (SPL) in the presence of significant noise. One function postulated for the stellate cell circuit is to increase the dynamic range and signal-to-noise ratio (SNR) of the incoming signal, more accurately highlighting significant frequencies. A network model utilising leaky integrate-and-fire neurons, wide-band lateral inhibition, and selective processing of inputs is developed to reproduce the stellate microcircuit function. Model simulation demonstrates enhancement of peak identification and dynamic range with tone and vowel stimuli.

**Index Terms**: auditory brainstem, cochlear nucleus, lateral inhibition, selective processing, speech enhancement

## 1. Introduction

The cochlear nucleus is the first processing centre for ascending auditory signals from the cochlea. A number of distinct neuronal circuits have been identified that form parallel representations of incoming information from auditory nerve fibres (ANFs). The focus of this research is the T Stellate cell circuit of the ventral cochlear nucleus.

The output of the T Stellate (TS) cell (also known as type I multipolar or chopper cell) population has been shown experimentally to accurately encode important frequency components in the input signal over a wide range of input sound pressure levels (SPL) and in the presence of significant signal noise [1, 2, 3]. Significantly, the TS cell population has been shown to have a greater level of fidelity over a wider dynamic range than either of the low spontaneous rate (LSR) or high spontaneous rate (HSR) ANF populations in isolation [1, 2]. Based on these observations, it is thought that one purpose of the stellate cell circuit is to accurately highlight the most significant oscillation frequencies, thereby increasing the signal-to-noise ratio (SNR) of the incoming signal [4]. A clear application of this in speech processing is the identification of vowels, which requires the first two or three formant frequencies to be identified.

The TS circuit is thought to consist of an interconnected network of T Stellate, D Stellate (DS) and tuberculoventral (TV) cells, with the latter two providing glycinergic inhibition to TS cells [5]. A number of mechanisms have been developed to explain the function of the TS circuit. Due to their wide-band ANF input, DS cells effectively capture the average signal over a range of frequencies around the current characteristic frequency (CF); through inhibition of TS cells over a range of frequencies, important spectral peaks can be highlighted by rejection of the background signal [3]. This kind of noise suppression is known as “lateral inhibition”, as it acts to reduce noise located between spectral frequency peaks, increasing the SNR.

Another mechanism, dubbed “selective processing”, occurs on the single cell level, and enables TS cells to respond to a much wider range of input levels than the ANFs that innervate them [2]. Selective processing refers to the ability of a cell to differentiate between inputs based on the level of the input sound, enabling greater dynamic range. For TS cells, this means that for quiet sounds, the response of the cell is dominated by inputs from high spontaneous rate ANFs and, for loud sounds, by low spontaneous rate ANFs.

A phenomenological neural network model of the stellate microcircuit of the ventral cochlear nucleus is developed by combining common neural modelling techniques with experimental data. This work is inspired by earlier detailed neural modelling research [4, 6]. Each cell in the network is approximated using a deterministic leaky integrate-and-fire neuron model, which allows for the dynamics of individual cells to be adjusted to match observed physiological responses. In addition, the model allows different inputs to be processed independently by the cell models, implementing a selective processing mechanism. The number and type of cell inputs, the network connectivity, and the time constants associated with cellular dynamics agree, where possible, with experimental observations. In addition, the performance of the model in response to a variety of stimuli is adjusted to achieve the desired signal processing functions regarding noise reduction, peak identification and dynamic range.

## 2. Cell Models

### 2.1. ANF Model and Stimulus Types

In this study, a “humanised” version of the auditory nerve model developed by Zilany and Bruce [7] is employed to generate inputs into the stellate network model. Spikes from both low spontaneous rate (0.1 spikes/s) and high spontaneous rate (60 spikes/s) ANFs in equal proportion are generated as inputs at 300 equally spaced positions along the cochlea with CFs ranging from 80 to 16000 Hz. Position is mapped to frequency based on the Greenwood function [8],

$$f = 165.4(10^{2.1x} - 1), \quad (1)$$

where $f$ is CF and $x$ is relative position along the cochlea from the apex. At each sampled frequency, 10 high spontaneous rate and 10 low spontaneous rate ANF outputs were generated.

A range of input sounds are used to stimulate the ANF model, including ramped pure tones and synthesised vowel sounds, to evaluate more complex performance features of the model. Each of these input types is masked with varying levels

of speech-weighted noise and at a range of SPLs to examine the breadth of model responses. The noise is generated by applying a first-order Butterworth filter with cut-off of 500 Hz to white Gaussian noise.

### 2.2. D Stellate (DS) Cell Model

DS cells receive wide-band input from ANFs, leading to a strong response to wide-band stimuli [5]. DS cell physiological responses to tones are classified as onset chopping responses, responding most reliably at stimuli onset. In the current model, the input into a given DS cell is the sum of ANF outputs from two octaves either side of the DS cell's CF (approx. 2000 ANF synapses). Due to the large number of inputs, it is assumed synaptic weights are small and the neuron has a short membrane time constant to prevent saturation.

The starting point for the DS cell model is the onset model of Kalluri and Delgutte [9], a leaky integrate-and-fire point neuron. Inputs are modelled as exponential functions of current with time constant, $\tau_s$,

$$I(t) = u(t)\frac{Ca}{\tau_s}e^{-\frac{t}{\tau_s}}, \quad (2)$$

where $u(t)$ is the Heaviside step function, $C$ is the membrane capacitance, and $a$ scales the current. The post-synaptic membrane potential is modelled as an equivalent resistor-capacitor (RC) electrical circuit,

$$\tau_m \frac{dV(t)}{dt} + V(t) = RI(t). \quad (3)$$

At each time step, the membrane voltage is compared to a threshold value and an output spike is generated if the threshold is exceeded. Following spiking, the membrane voltage is reset to the resting membrane voltage (0V is used for all models for simplicity) and a fixed refractory time is enforced before subsequent spike generation is permitted. In order to align with experimental data, model time constants and synaptic weights were small.

The Kalluri and Delgutte model [9] relies on precisely chosen input weights to concentrate spiking at onset. As a result, increasing input levels cause cells to respond more strongly, causing the neuron to cease providing only an onset response. To extend the onset chopping response to a wider range of input SPLs, the model is extended to include a threshold adaption property, which responds to output spikes by temporarily raising the threshold. In line with existing model dynamics, threshold adaption, $T$, is modelled as an RC circuit, driven by an exponentially-decaying adaption parameter, $T'$,

$$T'(t) = u(t)\frac{C_{\mathrm{th}}a_{\mathrm{th}}}{\tau_{\mathrm{th,s}}}e^{-\frac{t}{\tau_{\mathrm{th,}}}}, \quad (4)$$

$$\tau_{\mathrm{th,m}}\frac{dT(t)}{dt} + T(t) = R_{\mathrm{th}}T'(t). \quad (5)$$

The effect of adaption is to prevent excessive spiking after onset, achieving more robust onset behaviour and a more physiologically realistic response to varying input SPL.

Fig. 1A shows post-stimulus time histogram and rate-level functions for the DS cell model. In line with experimental data, the simulated DS cell responds more strongly to stimuli such as noise, as can be expected due to the wide-band input.

### 2.3. Tuberculoventral (TV) Cell Model

TV cells receive inputs from ANFs (excitatory) and DS cells (inhibitory). Studies have shown that ANF innervation derives from only a narrow frequency band, representing in a sharp tuning curve [5, 10]. Output spikes are suppressed for noise inputs and loud tones by strong inhibition by DS cells, resulting in a non-monotonic rate-level response with a single peak between 30 and 60 dB SPL [10]. Although not fully understood, the role of TV cells in the stellate circuit may be to enhance spectral peaks for loud input sounds. TV cells respond weakly at spectral peaks for loud inputs, but more strongly in adjacent, lower level areas. Hence, spectral peaks may be highlighted in TS cells through TV cell suppression of moderate level sidebands.

TV cell responses are characterised as onset-graded, with the response decaying exponentially from onset [10]. Due to the mix of inhibitory and excitatory inputs into TV cells, a modified version of the DS cell model discussed above is used. The TV cell model integrates inhibitory and excitatory inputs using independent RC circuits, accommodating dynamics for glutamatergic and glycinergic synapses separately. The output of each integration is summed to obtain the overall membrane voltage.

The inputs into each cell are 20 low and 20 high spontaneous rate ANFs from isofrequency fibres (i.e., two low spontaneous rate and two high spontaneous rate ANF model outputs) and a single isofrequency DS cell. To achieve the required onset graded response as opposed to the more abrupt onset chopping response, less severe threshold adaption is used. To achieve lasting inhibition from DS cells, a long glycinergic decay time constant is used. The TV model reproduces the non-monotonic rate response and onset-graded dynamics seen experimentally (Fig. 1B).

Table 1. *Model parameters for each neuron model in the network*

| Parameter | DS Cell Model | TV Cell Model | TS Cell Model |
|---|---|---|---|
| Synaptic Inputs | ± 2 octaves of ANFs | 20 HSR ANFs (isofrequency), 20 LSR ANFs (isofrequency), 1 DS cell (isofrequency) | 20 HSR ANFs (isofrequency), 20 LSR ANFs (isofrequency), ± 1 octave of DS cells |
| Excitatory input weight | $a$ = 0.12 V | $a$ = 0.7 V | $a_{LSR}$ = 5.2 V, $a_{HSR}$ = 0.5 V |
| Inhibitory input weight | N/A | $a$ = 2 V | $a_{DS}$ = 0.45 V, $a_{TV}$ = 0.5 V |
| Excitatory time constants | $\tau_m$ = 0.13 ms, $\tau_s$ = 0.1 ms | $\tau_m$ = 6 ms, $\tau_s$ = 2 ms | $\tau_m$ = 4 ms, $\tau_s$ = 2 ms |
| Inhibitory time constants | N/A | $\tau_m$ = 100 ms, $\tau_s$ = 2 ms | $\tau_m$ = 50 ms, $\tau_s$ = 3 ms |
| Threshold | $th$ = 3 V | $th$ = 4 V | $th$ = 4 V |
| Threshold adaption weight | $a_{th}$ = 2.5 V | $a_{th}$ = 1.5 V | N/A |
| Threshold adaption time constant | $\tau_{th,s}$ = 1 ms | $\tau_{th,s}$ = 10 ms | N/A |
| Threshold decay time constant | $\tau_{th,m}$ = 3 ms | $\tau_{th,m}$ = 100 ms | N/A |
| Refractory period | $t_{ref}$ = 0.5 ms | $t_{ref}$ = 1 ms | $t_{ref}$ = 0.7 ms |
| Input propagation delay | N/A | 2 ms | 2 ms |
| Shunting proportion | N/A | N/A | 0.7 |

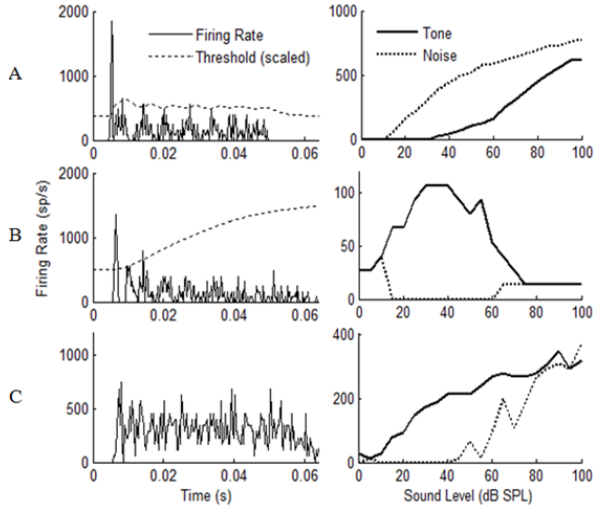


Figure 2: *Peri-stimulus time histograms and rate-level curves for each cell type in the microcircuit model. The DS cell model (A) exhibits onset chopping and a monotonic rate response, the TV cell model (B) exhibits onset-graded dynamics with a non-monotonic rate response, and the TS cell model (C) exhibits transient chopping behaviour with a large dynamic range.*

### 2.4. T Stellate (TS) Cell Model

TS cells receive excitatory input from low and high spontaneous rate ANFs, wide-band inhibitory input from DS cells and narrow-band inhibitory input from TV cells [5]. Investigations of synaptic physiology of TS cells reveal a minimum of five ANF inputs from a narrow frequency band [5]. Two TS cell subpopulations have been classified: transient and sustained chopper cells [11]. The distinction between these two subtypes is given by the duration of chopping evident in the cell's peri-stimulus time histogram (PSTH).

In the cell model, spiking regularity is achieved through temporal summation of inputs. 20 low spontaneous rate and 20 high spontaneous rate ANF inputs are input into each cell, taken from a narrow frequency band. Inhibitory input is taken from DS cells one octave to either side of the TS cell CF, and from one isofrequency TV cell. A delay of 2 ms is used to model spike propagation and integration times [11]. Threshold adaption is not used.

To enable selective processing of inputs, low spontaneous rate ANFs, high spontaneous rate ANFs, DS cells, and TV cells are integrated by the TS cell model independently [2, 6]. High spontaneous rate ANFs and DS cells both synapse distally, and low spontaneous rate ANFs synapse proximally. Due to this distribution, DS cell inhibition first acts to suppress high spontaneous rate inputs. For loud sounds, when DS cell inhibition is greater than high spontaneous rate excitation, the residual inhibitory graded potential diffuses toward proximal regions. Due to charge dispersion, the strength of inhibition decays to a lesser value by the time it reaches the proximal low spontaneous rate ANF synapses. Hence, low spontaneous rate inputs are only suppressed by this mechanism for very strong DS cell responses (such as for strong noise inputs). The membrane voltage from each input is summed at the distal and proximal points, then combined into a functional membrane potential for threshold comparison.

Due to this level-dependent processing mechanism, the developed TS cell model exhibits a robust representation of incoming signals for a large dynamic range, as shown in Fig. 1C. Compared to ANFs, the TS cell model has a rate response similar to that of high spontaneous fibres at low SPL and similar to that of low spontaneous rate fibres at high SPL, indicating an enhancement of dynamic range. Due to DS cell lateral inhibition, TS cells respond poorly to noise at low and moderate levels, which is a necessary property for the rejection of signal noise.

To classify the model's spiking regularity, the coefficient of variation (CV) of interspike intervals is calculated for successive 10 ms windows for the duration of the activity. Sustained chopper cells are have CV below 0.2, whereas transient chopper cells are have CV that is low initially (< 0.2) but increases for the remainder of a tone input (0.2 < CV < 0.5) [11]. This supports the classification of the current cell model as a transient chopper. Modelling transient-type TS cells is favourable for this analysis as they have been implicated as the primary cell responsible for spectral encoding of important features [12].

## 3. Results

Where possible, the pattern of synaptic connections made between the cells described above is based on experimental findings in the literature. The developed stellate network model follows a clear hierarchy of synaptic inputs, with DS cells receiving innervation only from ANFs, TV cells receiving innervation only from ANFs and DS cells, and TS cells receiving innervation from ANFs, DS cells, and TV cells. Due to this form of connectivity, the full response of each cell group can be calculated, beginning with DS cells, then used as input into remaining cell types.

At each frequency, both low spontaneous and high spontaneous rate ANF responses are simulated. Studies have shown that the ratio of TS to DS cell numbers in the ventral cochlear nucleus is approximately 15:1, so TS and TV cell outputs are calculated for every CF sampled in the auditory nerve, and DS cells at every 15$^{th}$ CF. To accommodate spectral edge regions for wide-band inputs, in which input bandwidth falls outside the range modelled, available inputs were scaled by the proportion of the octave range missing.

Spectral results are presented by calculating the average spiking rate for each cell in the network. Using both pure tones and the synthesised vowel /o:/ (as in "ball") as stimulus, the quality of spectral representation was analysed with

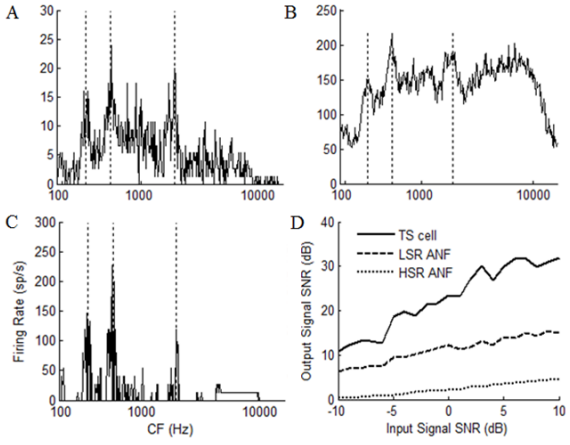


Figure 1: *Representation of the synthesised vowel /o:/ in noise with an SNR of -5 dB, from the (A) LSR ANF, (B) HSR ANF, and (C) TS cell model populations, showing the stellate microcircuit's function in enhancing signal contrast. (D) The ratio of the average output rate of cells within 1.5 critical bands of the known stimulus frequency to the average background output rate, over a range of input SNR values.*

respect to the SPL and the SNR of the stimulus. For vowel identification, the system should ideally highlight the first several formant frequencies, while reducing the level of the frequencies in between them.

As discussed earlier, a range of input sounds are masked using speech-weighted noise. Signal-to-noise ratios from -10 dB to +10 dB are used; -10 dB is tolerable with normal hearing, 5-10 dB for cochlear implant and hearing aid users. As can be seen in Fig. 2, even at low SNRs (<0 dB) the network suppresses background noise, highlighting masked spectral peaks; in Fig. 2C, the peaks are more clearly represented compared to auditory nerve input (Fig. 2A&B). For lower noise levels, the developed system removes almost all the background noise seen in the ANF spectral representations.

In order to quantify the level of noise reduction and spectral peak enhancement achieved by the model network, a signal-to-noise metric is used. This metric is calculated as the ratio of the average output rate of the cells within 1.5 critical bands of the known stimulus frequency with respect to the average background output rate. For vowel analysis, the metric is calculated for the first three vowel formants. As can be seen in Fig. 2, the TS cell representation is significantly higher than for either ANF population for the majority of input SNR levels.

## 4. Discussion

The computational model of the cochlear nucleus stellate cell microcircuit reproduces important physiological characteristics of the targeted cell types. Model simulation shows that both the temporal dynamics and rate responses of the DS, TV, and TS cells accurately replicate typical responses observed experimentally. Where possible, model parameters and time constants were made to align with those estimated from physiological experiments, giving the observed model performance greater validity. Furthermore, to verify the robustness of the results presented in this report, inputs tested in simulation were made to vary in both frequency and SNR. The model produces physiologically realistic output responses, and is able to perform functions relating to SNR improvement, peak identification and dynamic range enhancement. In addition, the use of selective processing and lateral inhibition in achieving these functions supports evidence indicating their role in feature extraction from ascending auditory signals.

Several extensions on the network presented have been suggested as having an influence on TS cell processing, such as inhibitory Golgi cells of the ventral cochlear nucleus granule layer, or recurrent inputs between TS cell [5]. As the purpose of these inputs are not well understood, and the range of desired model responses can be achieved using a simpler model, these extensions were neglected in the current model. To enable selective processing, more commensurate input from low spontaneous and high spontaneous rate inputs were used, leading to transient chopper behaviour. The choice was made to focus on transient chopper TS cells rather than sustained choppers due to the assumed difference in inputs into each type. Sustained chopper cells are more likely to be dominated by high spontaneous rate ANF inputs, leading to a greater number of low weighted inputs, which lends itself to greater output firing regularity.

In future work, models of the stellate cell microcircuit should be extended to capture a number of other postulated functions, such as temporal coding of information in amplitude modulated signals. This is likely to require a more sophisticated network model. Beyond the addition of extra cell groups and more intricate intra-network connectivity, this can be achieved by moving to a more versatile, conductance-based model, and by using a finer frequency resolution for approximating neural inputs and outputs. In addition, more realistic inputs should be used rather than synthesised vowels.

## 5. Acknowledgements

Funding from Australian Research Council (DP140101520).

**References**

[1] X. Wang and M. Sachs, “Neural encoding of single-formant stimuli in the cat. II. Responses of anteroventral cochlear nucleus neurons,” *J. Neurophysiol.,* vol. 71, pp. 59-78, 1994.

[2] Y. Lai, R. Winslow and M. Sachs, “A Model of Selective Processing of Auditory-Nerve Inputs by Stellate Cells of the Antero-Ventral Cochlear Nucleus,” *J. Comp Neurosci.,* vol. 1, pp. 167-194, 1994.

[3] A. Recio and W. Rhode, “The Representation of Vowel Stimuli in the Ventral Cochlear Nucleus of the Chincilla,” *Hearing Research,* vol. 146, pp. 167-184, 2000.

[4] M. Eager, D. Grayden, H. Meffin and A. Burkitt, “Amplitude modulation in the stellate microcircuit of the cochlear nucleus,” in *Seventh International Conference on Intelligent Sensors, Sensor Network and Information Processing 2011*, Adelaide, Australia, 2011.

[5] M. Ferragamo, N. Golding and D. Oertel, “Synaptic Inputs to Stellate Cells in the Ventral Cochlear Nucleus,” *J. Neurophysiol.,* vol. 79, pp. 51-63, 1998.

[6] M. Eager, D. Grayden, A. Burkitt and H. Meffin, “A neural circuit model of the ventral cochlear nucleus,” in *Tenth Australian International Conference on Speech Science and Technology, 2004*, Sydney, Australia, 2004.

[7] M. Zilany and I. Bruce, “Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery,” *J. Acoust. Soc. Am.,* vol. 120, no. 3, pp. 1446-1466, 2006.

[8] D. Greenwood, “Critical bandwidth and the frequency coordinates of the basilar membrane,” *J. Acoust. Soc. Am.,* vol. 33, p. 1344, 1961.

[9] S. Kalluri and B. Delgutte, “Mathematical Models of Cochlear Nucleus Onset Neurons: I. Point neuron with many weak synaptic inputs,” *J. Comput. Neurosci.,* vol. 14, no. 1, pp. 71-90, 2003.

[10] W. Rhode, “Vertical Cell Responses to Sound in Cat Dorsal Cochlear Nucleus,” *J. Neurophysiol.,* vol. 82, pp. 1019-1032, 1999.

[11] A. Paolini, J. Clarey, K. Needham and G. Clark, “Fast Inhibition Alters First Spike Timing in Auditory Brainstem Neurons,” *J. Neurophysiol.,* vol. 92, pp. 2615-2621, 2004.

[12] M. Hewitt, R. Meddis and T. Shackleton, “A computer model of a cochlear-nucleus stellate cell: Responses to amplitude-modulated and pure-tone stimuli,” *J. Acoust. Soc. Am.,* vol. 91, no. 4, pp. 2096-2109, 1992.