Proceedings of the Institute of Acoustics

A MODEL OF THE PROCESSING OF VOICED PLOSIVES IN THE AUDITORY NERVE AND COCHLEAR NUCLEUS

N. Blackwood, G. Meyer and W.A. Ainsworth

University of Keele, Department of Communication and Neuroscience, Keele, Staffordshire ST5 5BG.

1. INTRODUCTION

Models of the peripheral auditory system are of interest to both physiologists and speech researchers. In the case of speech research, it is hoped that the performance of a speech-processing or recognition system might be improved by the incorporation of some of the features of the human auditory system. This approach could prove especially fruitful in solving the problems with noise and sex-difference, where the performance of the auditory system is considerably better than current speech-recognisers.

Physiologically-based models have until recently extended only as far as the auditory nerve level. These models can be considered simply as a filter-bank followed by a series of non-linear stages, which convert the acoustic stimulus into action potentials in the auditory nerve. Output from such models can be considered to be a representation of the neural activity in the auditory nerve, and shows similarities to a spectrogram representation.

The next level of processing in the auditory system takes place in the cochlear nucleus. The working of the cochlear nucleus as a whole is not yet fully understood. However the responses of different cell types found in the cochlear nucleus are known (fig 1.1). The non-homogeneous responses of these cells makes it likely that they are used to extract different features for further processing at higher levels. Several attempts at modelling the cochlear nucleus have been made recently [1], [2].

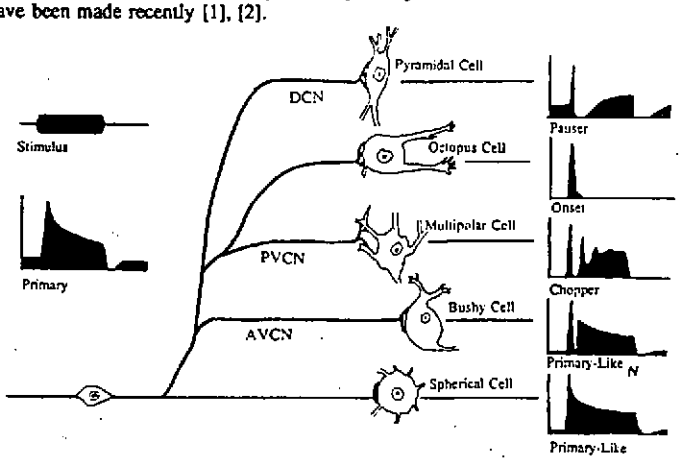

Fig. 1.1 Idealised temporal responses in the auditory nerve and cochlear nucleus.

This paper proposes a model of the peripheral auditory system up to the auditory nerve, and a set of units which have similar response patterns to the cells of the cochlear nucleus. These units connected in parallel could form a model for cochlear nucleus processing.

The responses of the auditory nerve (AN) and cochlear nucleus (CN) models to voiced plosives are also described. This is a first step in investigating the encoding of speech with the hope of discovering the important cues which are extracted or enhanced, and which may be important in speech recognition.

2. AUDITORY NERVE MODEL

The AN model simulates the cochlea, hair cells and the auditory nerve itself. In the cochlea, sounds are transformed into frequency dependent localised vibrations of the cochlear partition. The inner hair cells transform the motion of the cochlear partition into spike trains which are transmitted via the auditory nerve to the cochlear nucleus. The auditory nerve consists of about 30000 individual fibres in man, each of which is finely tuned to a specific frequency, designated the characteristic frequency. See [3] for a review of auditory nerve anatomy and physiology.

The present software model is a multichannel extension of a hardware model of a single AN fibre [4]. Fig 2.1 shows a block diagram of a single channel of the AN model. The input is an auditory stimulus such as speech or simpler test stimuli, ie clicks, noise, or pure and complex tones. The output is a train of spikes.

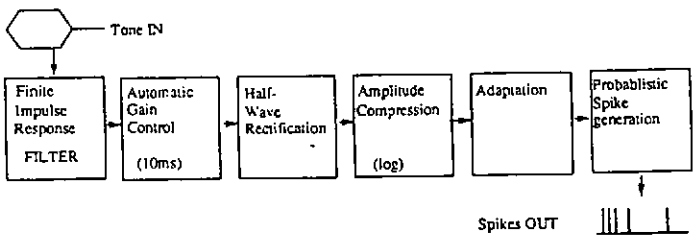

Fig. 2.1 Single channel of auditory nerve model.

Each channel of the model consists of three stages:

2.1 Cochlear filtering.
The filtering processes of the cochlea is simulated by a set of 32 FIR filters whose centre frequencies are spaced 0.5 bark apart in the range 100Hz to 4700Hz. The impulse responses of the filters are based on the impulse responses of cat AN fibres which are derived by reverse correlation [5]. A time-delayed gamma-tone function has been shown to have a good correspondence with the revcor function [6], and has been used in the present version of the model.

2.2 Hair Cell Transduction.
Hair cell transduction and synaptic activation are modelled by a series of non-linear stages:

2.2.1 Automatic Gain Control (AGC). An AGC mechanism controls the spike discharge rate for stimuli of greater than 40dB above spike activation threshold.

2.2.2 Half-wave Rectification. The predominately unidirectional hair cell and synaptic response is approximated by half-wave rectification.

2.2.3 Logarithmic Compression. This is intended to mimic the approximately linear relation of the spike discharge rate and stimulus level in dB between threshold and saturation. The value of the spike activation threshold for each channel is derived from human behavioural threshold data.

2.2.4 Adaptation. Short-term adaptation is simply modelled by an exponential decay in spike discharge rate with time. Off-suppression is produced by the same mechanism.

2.3 Probabilistic Spike Generation.
The spike generator mechanism consists of the product of two quadrant multiplication between low-pass filtered white noise and the output of the previous stages of the model, undergoing level detection. If the threshold level is exceeded, a spike is generated, followed by a refractory period of 2ms.

A more detailed description of the AN model is given elsewhere [7].

3. COCHLEAR NUCLEUS MODEL

The auditory nerve divides into a number of tonotopically ordered branches which project onto the cochlear nucleus. The cochlear nucleus can be spilt into the ventral (VCN) and dorsal (DCN) cochlear nucleus. The VCN can be further spilt into the posterior (PVCN) and the anterior (AVCN) ventral cochlear nucleus. The tonotopic organisation of the cochlea is preserved in each of these regions.

In contrast to the homogeneous responses to single frequency tone bursts found in the auditory nerve, the cochlear nucleus has at least five different response types [8]. They correspond to different cell types found in the cochlear nucleus. Fig 1.1 shows idealised versions of such temporal responses and corresponding cell types.

The cochlear nucleus units proposed here are based on a simple neuronal model (fig 3.1). Each particular CN unit is modelled by adjusting certain parameters to approximate the observed physiological characteristics of CN neurons.

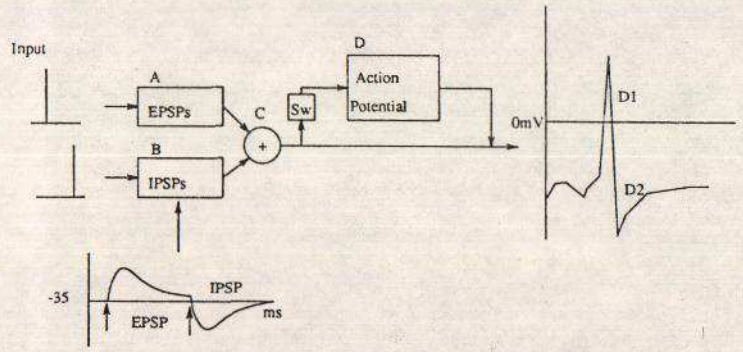

Fig. 3.1 Block Diagram of processing in the neuronal model.

Excitatory (EPSP) and inhibitory post synaptic potentials (IPSP) are generated (A & B) by low pass filtering input spikes. If the integrated intracellular potentials (C) exceed a given threshold, an action potential (AP) is generated (D). The AP consists of two separate processes, a large fast depolarisation (D1) followed by a more gradual hyperpolarisation (D2).The AP generator is disabled at the peak of the AP and is only reprimed when the intracellular potential falls below a predefined voltage. Normally this happens during the hyperpolarisation.

The time course and amplitude of the PSPs and AP are controlled by a set of parameters. Other parameters defining the response pattern are the firing threshold and the repriming level.

The ability to vary virtually all cell parameters means that the response characteristics of single cells can be controlled to match original cell recordings.

- EPSPs and IPSPs can be low-pass filtered to a degree which reflects the position of their generation;
- the duration of hyperpolarisation can be adjusted to set the relative refractory period;
- the firing threshold can be set to adjust or prevent spontaneous activity; and finally
- the action potential repriming voltage can be set to provide a very powerful action potential blocking mechanism.

In this way, units having the response characteristics of Primary-like, Chopper, $Onset_I$, $Onset_{IN}$(type II) and Pauser(type IV) cells can be simulated [9]. Of particular interest in the processing of plosives are $Onset_I$ units.

3.1 $Onset_I$ units

$Onset_I$ responses are recorded in octopus cells, which occur in the PVCN in most mammals [10]. The cell bodies lie at the edge of the octopus cell area with their dendrites passing over a large number of auditory nerve fibres, which explains the large receptive fields found in octopus cells. Intracellular recordings show that spikes are

followed by no or very little hyperpolarisation.

Onset$_1$ units are modelled using the action potential blocking mechanism which is invoked after every action potential. The blocking mechanism is set at the onset of an AP and reset when the intracellular potential falls below a set voltage. In all other units this happens immediately with the onset of hyperpolarisation. In onset$_1$ units hyperpolarisation is suppressed and the repriming threshold is set to a low value so that for sustained stimuli only the onset spike is generated after which the spike generator is disabled.

To produce reliable onset responses, even at low stimulus amplitudes, the equivalent of 80 auditory nerve fibres are connected to the onset unit so that a steady depolarisation is produced. The need for a large number of afferent fibres is enhanced by the use of short duration EPSPs which provide relatively little temporal integration. The afferent nerve fibres are spread over a large range of centre frequencies (CF +- 2 Bark) to simulate the wide receptive field of Onset units. The firing threshold is set high (-38mV) to suppress spontaneous activity.

4. RESPONSES TO VOICED PLOSIVES

Both the auditory nerve and cochlear nucleus models have been tested with simple stimuli [7]. In this paper the responses of the models to VCV utterances are investigated.

4.1 Stimuli

27 Vowel-Consonant-Vowel (VCV) utterances by a single speaker were recorded. The utterances consisted of the voiced plosives /b/, /d/ and /g/ preceded and followed by the vowels /a/, /i/ and /u/. The stimuli were sampled at 10kHz.

4.2 AN Model

Each stimulus was presented for 10 repetitions and the output spikes collected into post stimulus histograms (PSTH) for each channel of the model. Fig 4.1 shows the PSTHs produced by the AN model in response to the utterance /aba/. These PSTHs were analysed using a fixed, short-time analysis window in two different ways, giving average discharge rate and synchronised discharge rate representations of the model responses, reflecting the type of analysis used by auditory physiologists [3]. A fixed-window length of 12.8ms was chosen. The window was shifted by increments of 6.4ms through the PSTH data, to construct a spectrogram representation.

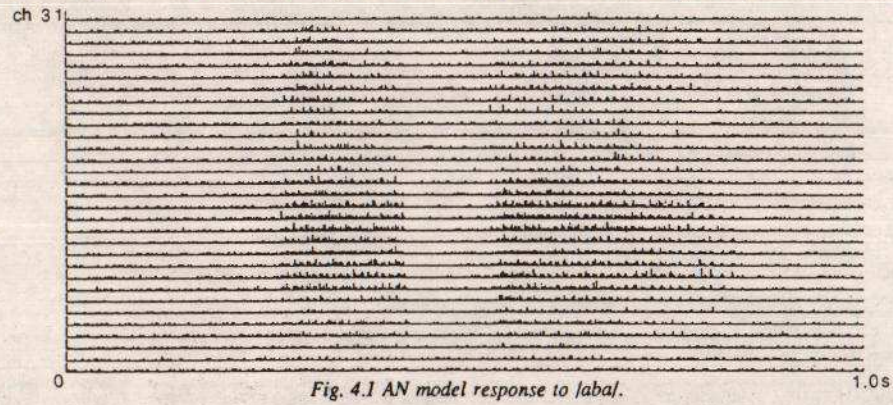

Fig. 4.1 AN model response to /aba/.

The average discharge rate is calculated by simply counting the number of spikes in each window. Fig 4.2a shows the average discharge rate spectrogram for /aba/. The synchronised discharge rate is a measure of the temporal response of the model. A Hamming window was applied to the windowed data, and a 128pt FFT calculated. The synchronised discharge rate for a given channel is the magnitude of the FFT component at the CF of that channel. Fig 4.2b shows the synchronised discharge rate spectrogram for /aba/. Fig 4.2b gives the better representation of the neural activity. Formant discrimination between F1 and F2 is clearer. This is because the

model exhibits synchrony suppression, but not rate suppression [3].

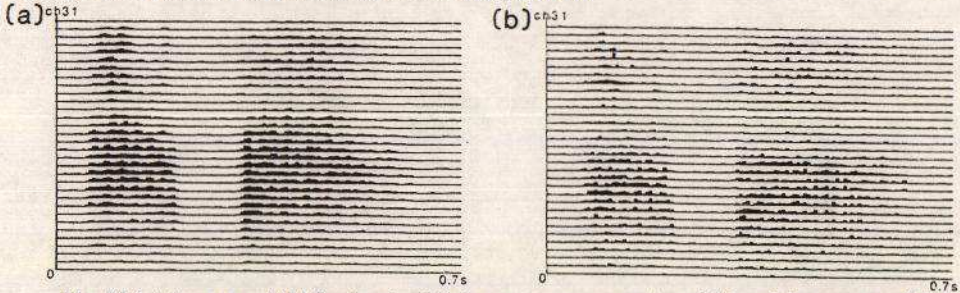

Fig. 4.2 (a) Average and (b) Synchronised discharge rate representation of AN model response to /aba/.

The synchronised discharge rate responses for the utterances /ada/ and /aga/ are shown in fig 4.3a & b. Visual inspection of the 'spectrograms' gives little indication of which plosive is present in the utterance. The formant transitions of the CV syllables are poorly encoded by the AN model.

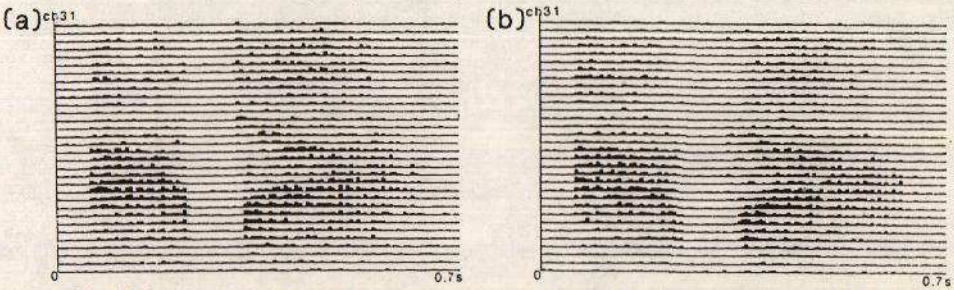

Fig. 4.3 Synchronised discharge rate representation of AN model response to (a) /ada/ and (b) /aga/.

4.3 CN Model

The PSTHs of the response of the AN model to the 27 VCV utterances were presented to arrays of a number of cochlear nucleus units: Primary-like, Chopper, $Onset_I$, $Onset_{IN}$(type II) and Pauser(type IV). The response of units with unilateral inhibition was also investigated. These units act as formant edge detectors in frequency space but have no direct physiological correlate in the CN. $Onset_I$ units seemed to be well suited for the encoding of plosives. The other units seemed more suited for the encoding of formant frequencies or the fundamental frequency, or in the case of type IV (pauser) units seemed unable to encode speech information in any useful way.

4.3.1 Plosives in $onset_I$ units. $Onset_I$ units appear to act as edge detectors in time, with exactly one spike being generated at each tone onset. These units also produce a single spike at the plosive burst (fig 4.4,B). This suggests that the units may be used to encode one of the features involved in the perception of plosives. Ainsworth [11] showed that the conditional probability of a listener perceiving a synthetic stimulus as /b/ was greatest if a short noise burst was presented near the first fomat frequency, as /g/ if the noise was near the second formant frequency, and as /d/ if the noise was centered at a higher centre frequency. However, the experiments also demonstrated that the efficacy of this cue was dependent upon the vowel context. Fig 4.6a, b & c shows the responses of an array of $Onset_I$ units to a VCV stimuli, /aba/, /ada/ and /aga/. In all cases, both the formant onsets (A1, A2) and the plosive burst (B) are encoded by the unit. Another feature of the unit is its ability to phase lock into the glottal pulse period (C1 and C2).

In order to categorise the plosives, the position of the highest and lowest channel of the array activated by the plosive burst (ie. leading edge B, fig 4.6a,b&c) was plotted (fig 4.5). The graph shows a clear separation of /b/

AUDITORY PROCESSING OF VOICED PLOSIVES

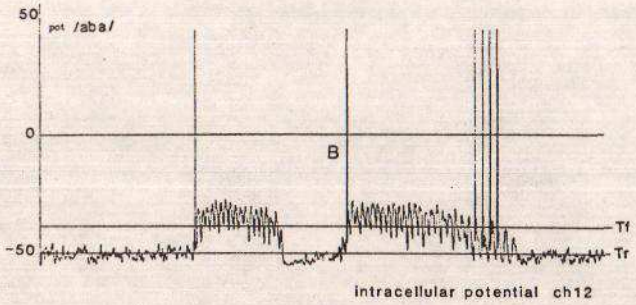

Fig. 4.4 Response of single Onset, unit to /aba/.

and /d/ sounds. The leading edge in /b/ virtually always starts at channel 0, ending in the middle to high frequency region. The plosive /d/ is characterised by a leading edge ranging from the highest frequency, 4700 Hz down to the middle and low frequency region. Finally /g/ sounds are grouped into two areas, the common feature is the spread of the leading edge. Leading edges of /g/ plosives are rarely spread over more than 15 channels. There is a large amount of overlap between /d/ and /g/ which show activity in the high frequency region. This may have been brought about by the frequency range of our model, maximum frequency of 4700Hz. All /d/ sounds have their leading edge reaching to this frequency region, and possibly higher.

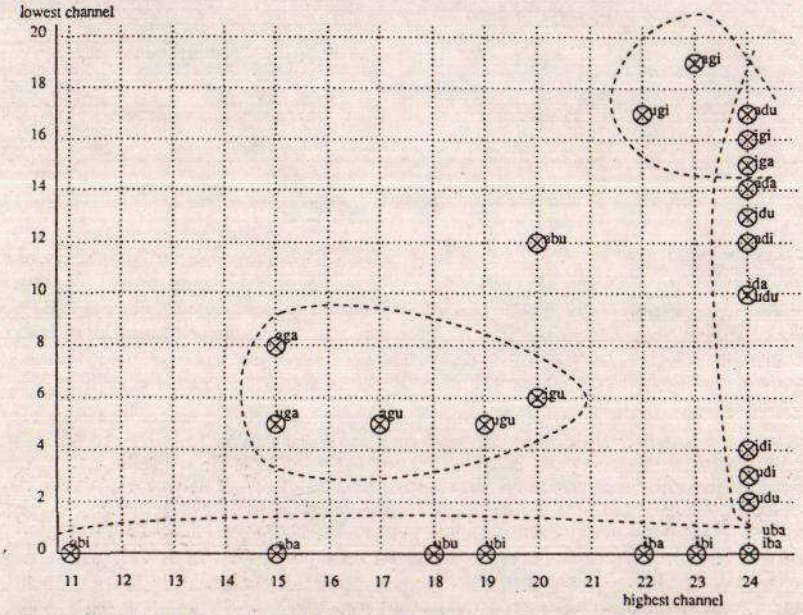

Fig. 4.5 Classification of voiced plosives in terms of burst region

The above results are in general agreement with psychophysical data [11]. Additionlly, it would be expected that the spread of the plosive burst is an important cue. For example, /g/ consistently has a small frequency range over which the plosive burst causes spike discharges. The discrimination of voiced plosives using only the

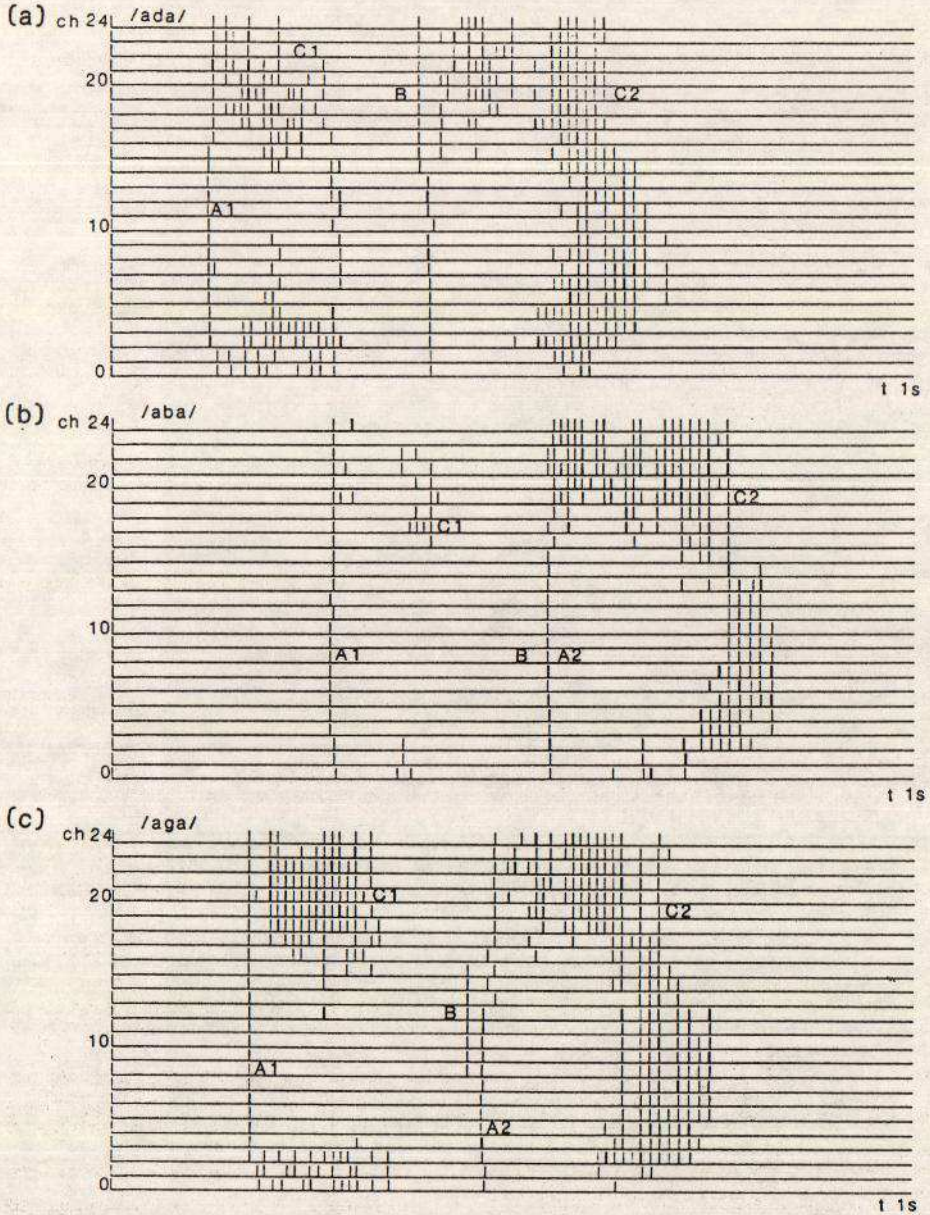

Fig. 4.6 Response of array of Onset$_1$ units to (a) /aba/, (b) /ada/ and (c) /aga/.

plosive burst frequencies as a cue is unlikely because of the large overlap of /g/ and /d/ in the high frequency region.

Other important cues to the plosive type preceding a vowel are formant transitions, and there is evidence that some cochlear nucleus units exhibit a preference for frequency sweeps in one particular direction. However, a systematic encoding of frequency sweeps is more commonly associated with the central auditory system [12].

5. CONCLUSIONS

A physiologically-based model of the auditory nerve and cochlear nucleus has been produced. This has been used to process voiced plosives in VCV utterances. The results suggest that some of the units in the cochlear nucleus (Onset$_1$ units) may be responsible for detecting the onset burst frequency, a cue which has been shown psychophysically to be useful in distinguishing between voiced plosives. However the complexity of the system is such that much more work needs to be carried out before definite conclusions can be reached.

6. ACKNOWLEDGEMENTS

The work on the auditory nerve model is supported by SERC Image Interpretation Initiative Grant GR/E 4283.9, while the work on the cochlear nucleus model was supported by European Community contract SC1.0044.c(H).

7. REFERENCES

[1] M J PONT & R I DAMPER, 'Software for a Computational Model of Afferent Neural Activity from the Cochlea to the Dorsal Acoustic Stria', VSSP Technical Report, University of Southhampton (1989).

[2] D J SANDERS & G G R GREEN, 'Properties of modelled and networked cochlear nucleus neurons', Advances in Speech, Hearing and Language Processing, Vol 3 Cochlear Nucleus: Structure and Function in Relation to Modelling, JAI Press, (in press).

[3] E F EVANS, 'Cochlear Nerve and Cochlear Nucleus', Handbook of Sensory Physiology, V5(2), Springer Verlag, Berlin (1975).

[4] E F EVANS, 'An electonic analogue of single unit recording from the cochlear nerve for teaching and research', Journal of Physiology (London), 298: 6P-7P (1980).

[5] E de BOER, 'Reverse correlation II. Initiation of nerve impulses in the inner ear.', Proceedings Kon. Nederl. Akat. Wet., 72: 129-151 (1969).

[6] L H CARNEY & T C T YIN, 'Temporal Coding of Resonances by Low-frequency Auditory Nerve Fibers: Single-fiber responses and a Population Model', Journal of Neurophysiology, 60(5): 1653-1677 (1988).

[7] G F MEYER, N BLACKWOOD & W A AINSWORTH, 'A Computational Model of the Auditory Nerve and the Cochlear Nucleus', Proceedings of SCSI European Multiconference, Nuremberg, p573-578 (1990).

[8] E D YOUNG, W P SHOFNER, J A WHITE, J-M ROBERT & H F VOIGT, 'Response Properties of Cochlear Nucleus Neurons in Relationship to Physiological Mechanisms', Auditory Function, Annual Symposium of The Neurosciences Institute p277-312 (1988).

[9] G F MEYER & W A AINSWORTH, 'Modelling Response Patterns in the Cochlear Nucleus using Simple Units', Advances in Speech, Hearing and Language Processing, Vol 3 Cochlear Nucleus: Structure and Function in Relation to Modelling, JAI Press, (in press).

[10] C M HACKNEY, K.K. OSEN, J. KOLSTON, 'The Cochlear Nuclear Complex of Guinea Pig. Some Anatomical Observations', Anatomy and Embryology', 182 p123-149 (1990).

[11] W A AINSWORTH, 'Perception of Stop Consonants in Synthetic CV Syllables', Language and Speech, V11(3) p139-155 (1968).

[12] G J BROWN & M P COOKE, 'Modelling Modulation Maps in the Central Auditory System', Poster Communication at BSA short Papers Meeting Nottingham, (1989).