An Active Model of the Auditory Periphery with Realistic Temporal and Spectral Characteristics

Ambikairajah, E., McDonagh, B.

Speech Research Group
Department of Electronic Engineering
Regional Technical College, Athlone
Republic of Ireland

ABSTRACT

This paper describes three auditory models which attempt to reproduce accurate temporal and spectral behaviour in a manner which is not computationally excessive. A particular problem in modelling the cochlea is to compress the input Sound Pressure Level (SPL) range into a smaller range of values (the compression ratio for a human is approximately 2.5:1), while maintaining realistic cochlear bandwidths and latencies. This paper describes a transmission line model which aims to incorporate such a compression ratio into the model, and compares its performance with that of two alternative parallel filterbank models.

INTRODUCTION

The first model presented is a modification of the transmission line model of Ambikairajah *et al*[1]. The filters are arranged in cascade, with the output of each filter being fed to a second filter which is tuned to the same frequency. The outputs of the second filters are sent to a peak detect mechanism which controls the Q values of the appropriate filters. This arrangement enables an almost uniform 60 dB of compression over most of the cochlear filters to be obtained. This high compression ratio is possible because the signal at any given tap has been boosted by all the preceding filter gains. The low Q values used also mean that realistic latencies are achieved (between 10 and 20 ms for the 1 kHz channel).

For comparative purposes another recently proposed model is considered, that of Carney [2], in which a single section of the basilar membrane is modelled by a gammatone filter. The bandwidth is controlled by a level-dependent non-linear saturating feedback, which simulates the activity of the Outer Hair Cells, and the parameters are chosen to most closely match the responses of cat auditory nerve fibers. In this paper, a number of these gammatone filters are arranged in a parallel filterbank formation, with each filter tuned to a different frequency of the auditory spectrum. Using the parameter values specified by Carney[2] for all filters, the tuning curves resulting from this model seem to be excessively broad as compared with those of the first model, and the signal compression is less than that achievable with the modified transmission line technique.

As a third model, an alternative parallel filterbank is presented, which reduces the computational effort involved, while maintaining the compression ratio and achieving greater frequency selectivity. In this model each section of the membrane consists of a second order bandpass filter. The output of this filter is fed to a second filter tuned to the same centre frequency. The second filter has a variable Q, which is dependent on the signal level. It is controlled by a capacitor hair cell model. By varying the value of Q, the broadening of the cochlear bandwidth and 20 dB of compression can be realised in a more efficient fashion to that of Carney.

For the three models the output of the cochlear filters are input to the synapse model proposed by Carney [2]. This consists of an inner hair cell, (a saturating nonlinearity followed by two lowpass filters), and an inner hair cell-auditory nerve synapse. The three models are compared under the three main criteria of (a) temporal response, (b) neural tuning curves, (c) level dependence.

The design procedure for the implementation of the three models is described. All three models operate at a sampling frequency of 20 kHz and cover the frequency spectrum from 75 Hz to 3.5 kHz. The principal intended use for these models is as front-end processors for speech recognition systems.

MODEL 1

The transmission line model of the cochlea of Ambikairajah *et al* [1], uses 128 cascaded filters to represent adjacent sections of the basilar membrane. Each filter is tuned to an individual frequency of the auditory spectrum, which ranges from 3.5 kHz to 75 Hz. As a pressure signal propagates along the cascade high frequency components are filtered out by the 2nd order lowpass filters. The pressure transfer function for a single section of the membrane, and it's digital filter realisation are given in Figure 1.

$$\frac{V_o(s)}{V_i(s)} = \frac{\omega_p^2}{s^2 + B_p s + \omega_p^2} \, \frac{s^2 + B_z s + \omega_z^2}{\omega_z^2}$$

$$Q_P = \frac{\omega_P}{B_P},$$

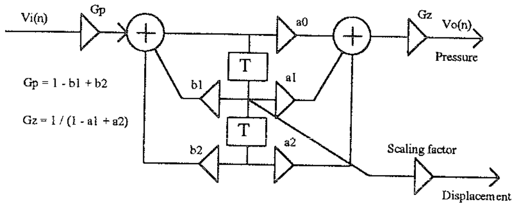

Figure 1. A single section of the digital filter model of the basilar membrane

In the modified transmission line model, a second filter is used at each tap to enhance the gain and selectivity of the model. This filter is tuned to the same frequency as the corresponding filter in the cascade structure (Figure 2). Displacement outputs from the first filters undergo two spatial differentiation's to accentuate the tuning characteristics before being input to the second filters. This differentiation represents the mechanical coupling of adjacent sections of the membrane. The transfer function for the second filter is the same as the pressure transfer function of the cascaded filters.

The second filter output is to be used to control the filter bandwidths and gain. That is, at low signal levels the gain is high and the bandwidth is low, and at high signal levels the gain is low and the bandwidth is broader. This dynamic in the gain and bandwidth of the filters is obtained by adjusting the Q's of each filter (Kates[3]) as a function of the peak output from the second filters. The maximum Q values (Q_{p1}) of the cascaded filters vary from 1.2 to 8.0 from the low to high frequency ends, on a linear cochlear distance scale. The relationship between the zero and pole parameters are: $\omega_{z1}(i) = 1.15\omega_{P1}(i-2)$ (i = filter number, ω_{z1} = zero resonant frequency, ω_{p1} = pole resonant frequency), and $Q_{z1} = 1.50 Q_{P1}$. For the *second filter* the maximum Q values (Q_{p2}) are also frequency dependent, determined by the relationship $Q_{P2} = 1.5(1+f)$ with f in kHz, and the zero/pole parameters vary according to $\omega_{z2} = \omega_{P2}/2$ and $Q_{z2} = 2.0 Q_{P2}$.

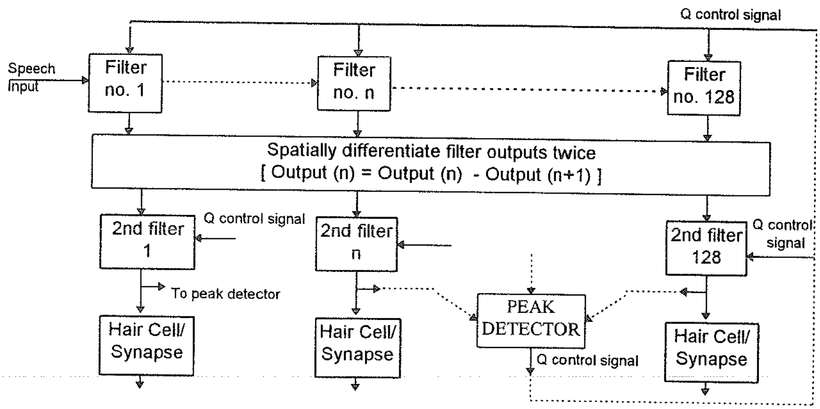

Figure 2. Transmission Line Model (Model 1)

Feedback Control

The feedback control mechanism works as follows: the outputs of each of the second filters are sent to a peak detect mechanism. This detects the peak output over a span of 8 filters either side of the one whose Q is to be controlled. This peak output is then used to determine the ratio of peak output to maximum output ('r'), which in turn controls the Q setting of each filter. The cascaded filter Q's are given by $Qp_1 = (0.3+0.7r)*Q_{1max}$ and the second filter Q's are given by $Qp_2 = (0.1+0.9r)*Q_{2max}$. Filter outputs should lie between 40 and 80 dB, to ensure that the 40 dB range of filter outputs matches the 40 dB dynamic range of the inner hair cell/synapse model. This is why the frequency responses are scaled to peak at -20 dB at low Q.

Figure 3 shows the frequency responses of every 6th filter when all the Q's (Q_{p1} and Q_{p2})are fixed in both the low (Fig. 3a) and high (Fig. 3b) states. A scaling factor is used to give all the responses a maximum gain of -20 dB at low Q. In the high Q state, both the low and high frequency responses do not have a full 60 dB gain. This is due to the fact that for the high frequency filters there are not enough preceding filters for the gain to have accumulated to 60 dB, while for the low frequency case it is because the overall gain at these frequencies does not sum to 60 dB.

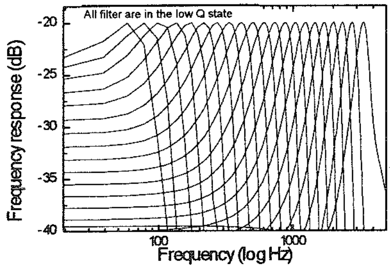

Figure 3(a)

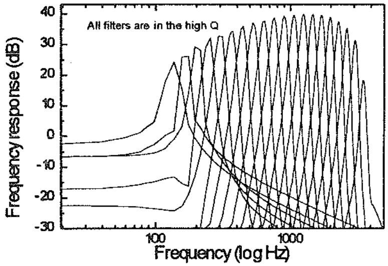

Figure 3(b)

Frequency responses for model 1.

Carney's hair cell model

For the purpose of comparing the three cochlear models the same inner hair cell and synapse model is used in all three cases.

Figure 4 shows the sustained firing rate of the synapse for an intermediate frequency channel. The firing rate is computed (as in [2]) from the average output of the synapse over the duration of the input tone. The first 5 ms of the response is discarded so as not to include the initial burst of activity. The firing rate is almost constant at low signal input levels. Then at about 40 dB SPL, the firing rate increases steadily until it eventually saturates.

The sustained rate has a dynamic range of about 40 dB. It is because of this physiological limitation in the ability of the auditory nerve to encode signals over a large dynamic level that the cochlea must compress the enormous variation of sound inputs into a more manageable range. Since the transmission line model discussed in the preceding section has a compression ratio of 2.5:1 over an input range of 100 dB, the auditory synapse model is able to successfully encode this range of signal inputs.

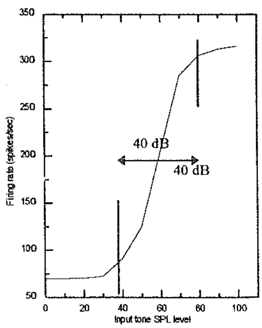

Figure 4. Rate Level Curve for the synapse model.

MODEL 2

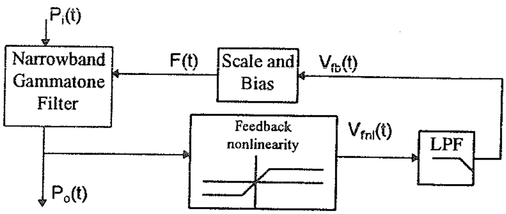

Figure 5. Gammatone filter (Model 2)

The narrowband filter used to characterise the tuning of one section of the basilar membrane in the auditory model proposed by Carney[2] is as shown in Figure 5. The filter itself is a linear revcor filter which is modelled by the gammatone function

$$g(t) = [(t-\alpha)/\tau]^{\gamma-1} e^{-(t-\alpha)/\tau} \cos[\omega_{cf}(t-\alpha)] \qquad t \geq \alpha$$
$$g(t) = 0 \qquad t < \alpha$$

In this paper 128 such gammatone filters are arranged in a filterbank formation. Figure 6 shows the frequency responses of every 6th filter in the filterbank. These responses are obtained by taking an FFT of the impulse response of each filter with the parameters set to those specified by Carney.

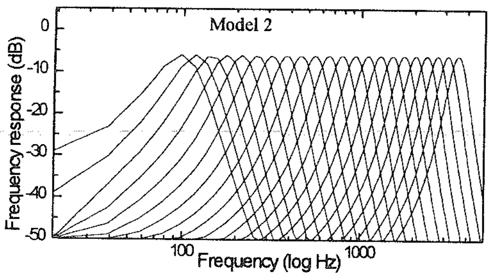

Figure 6. Frequency responses of model 2.

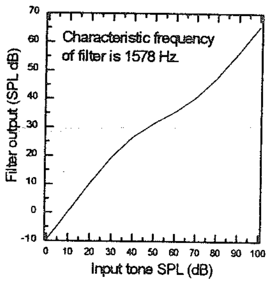

Figure 7. Filter I/O characteristics

Figure 7 shows the level of a single filter output in response to input tone level. This figure shows a compression ratio of about 1.33:1 - i.e. 100 dB input is compressed into 75 dB output. At high and low Sound Pressure Levels the gammatone filter behaves in a linear fashion. Compression due to the saturating feedback nonlinearity occurs over the range of approximately 40 to 90 dB. This is a difference in behaviour relative to the other two models presented here, in that in both other cases the compression takes place evenly over the entire input range.The various parameters of the gammatone filter and the feedback loop have been chosen by Carney to most closely approximate empirical data on cat auditory nerves, rather than on DSP considerations.

MODEL 3

For the purposes of speech recognition experiments the essential features of latency and cochlear tuning can be obtained using a much simpler, alternative parallel filterbank model, which is very much concerned with physiological accuracy with regard to the feedback control mechanism and the cochlear bandwidth.

This model uses a bandpass filter followed by a 2nd order lowpass filter, both of which are tuned to the same characteristic freqeuncy. The bandpass filter output is used to control the Q factor of the lowpass filter, which influences the gain and bandwidth characteristics of the filter (Figure 8). The feedback control mechanism consists of a simple capacitor outer hair cell model, the output of which is integrated over a short period to get the energy. This integrated value is compared with the pre-computed maximum and minimum values of OHC energy and used to change the lowpass filter coefficients.

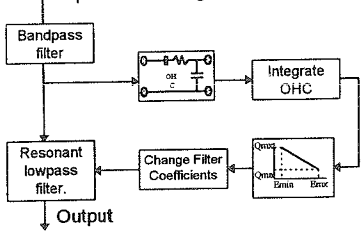

Figure 8. Model 3

Figure 9 shows the frequency response of the filterbank at high and low values of Q. There is a 20 dB difference in maximum gain between the two states.

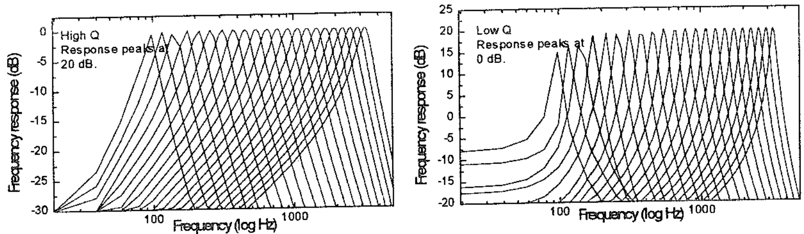

Figure 9 Frequency Response of model 3.

RESULTS

Figure 10 shows a comparison between the temporal responses of the three models in response to an input tone at CF at two different SPL's. The time taken for the synapse output to decay to it's resting rate after the signal has been removed indicates the latencies of the models. The three models all give reasonable latencies of less than 20 ms for an 80 dB signal, when the filter Q's are low. The effect of the superior compressive abilities of model 1 are evident in the synaptic response to a low level input.

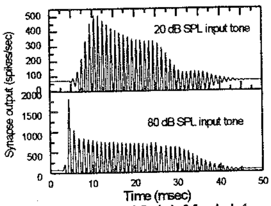

Figure 10 (a) Model 1.

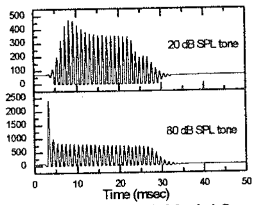

Figure 10 (b) Model 2.

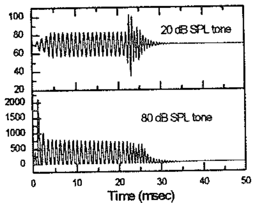

Figure 10 (c) Model 3.

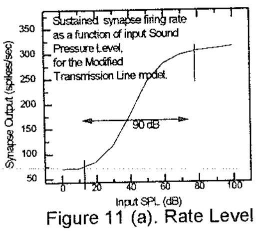

Figure 11 (a). Rate Level Curve for model 1.

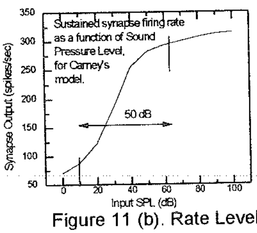

Figure 11 (b). Rate Level Curve for model 2.

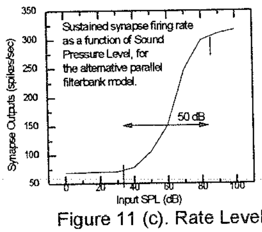

Figure 11 (c). Rate Level Curve for model 3.

Figure 11 shows the rate level curves for the three models. The synaptic output is plotted for a particular frequency channel, in response to tones at CF at varying SPL's. Models 2 and 3 both show a dynamic range of approximately 50 dB. Due to the high compression ratio in the case of model 1, response can take place over a much larger range of inputs (approx. 90 dB).

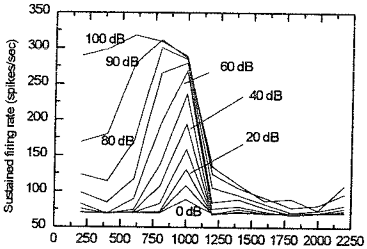

Figure 12 (a) Area rate curves model 1.

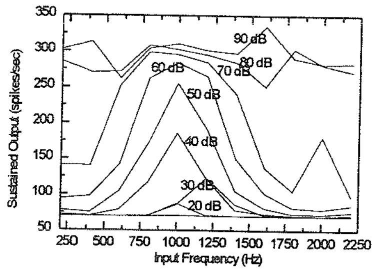

Figure 12 (b) Area rate curves for model 2.

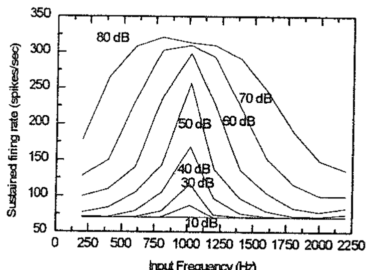

Figure 12 (c) Area rate curves for model 3.

Figure 12 shows the area rate curves for the three models. Model 1 (Fig 12a) shows a substantially superior bandwidth to the subsequent two models. The effect of the cascade of resonant lowpass filters is clearly visible in the steep high frequency slope of the curves. The effect of the 2.5:1 compression ratio is also visible, with both 0 dB and 90 dB signals clearly peaking at the filters centre frequency. Model 2 (Fig 12b) in contrast shows a relatively wide bandwidth. There is virtually no peak at low signal levels and the model shows no selectivity at higher SPL's. Model 3 (Fig 12c) also performs poorly at low SPL's, however it displays a superior bandwidth and has a better high signal level performance.

CONCLUSIONS

The Transmission Line model (Model 1) is evidently a superior cochlear model under the three criteria examined above. A realistic compression ratio is unattainable using a parallel filterbank formation (Models 2 and 3) of independent filters. The cascade structure model 1 also ensures a superior frequency selectivity in the channels of the auditory nerve, however the latencies of Model 1 are slightly longer than for the other two models.

ACKNOWLEDGEMENTS

This research has been carried with funding provided by EOLAS, the Irish Science and Technology Agency, to Regional Technical College, Athlone under the Applied Research Program 1993/94.

REFERENCES

[1] Ambikairajah, E., Black, N.D., & Linggard, R. (1989) *Digital filter simulation of the basilar membrane*, Computer Speech and Language, 3, pp 105-118.

[2] Carney, L.H. (1993) *A model for the responses of low-frequency auditory nerve fibers in cat*, J. Acoust. Soc. Am., 93(1), pp 401-417.

[3] Kates, J.M. (1993) *Accurate Tuning Curves in a Cochlear Model*, IEEE Trans. on Speech and Audio Processing, Vol. 1, No. 4, pp 453-462.