AN ACTIVE COCHLEAR MODEL FOR SPEECH RECOGNITION

E. Ambikairajah (*) and E. Jones (**)

(*) Department of Electronic Engineering, Regional Technical College, Athlone, Ireland
(**) Department of Electronic Engineering,University College Galway, Ireland

ABSTRACT - A model of the cochlea which includes both passive and active elements is described in this paper. The cochlear model consists of a cascade of 128 digital filters, of which 60 fall within the speech bandwidth of 250 Hz to 4 KHz. The model presented in this paper is suitable for implementation using a digital signal processor, and could act as a front-end processor for a speech recognition system.

INTRODUCTION

Previous research has assumed that the cochlea is a passive system (Lyon, 1982; Ambikairajah et al., 1989; Linggard & Ambikairajah, 1986). However, recent research indicates that the cochlea contains both passive and active elements (Neely, 1989; Davis, 1983). According to Davis (1983), the passive system is operational at normal stimulus amplitudes, while the active system comes into play when the ear is presented with a low-amplitude stimulus. Several active basilar membrane models have previously been proposed, e.g. Neely & Kim (1985), de Boer (1983); however, these models require considerable computation. Some models achieve greater efficiency by using analogue circuitry, e.g. Zwicker (1986). A computational active cochlear model is presented in this paper. Weaker formants are enhanced in the output of this model. This behaviour is similar to the behaviour of the adaptive Q model proposed by Hirahara & Komakine (1989). An implementation of this model on a TMS320C25 is described in a companion paper, also presented at this conference.

DEVELOPMENT OF THE AUDITORY MODEL

The passive cochlea

A block diagram of the overall auditory model is shown in Figure 1. The basic model used for the passive cochlea is that developed by Ambikairajah et al. (1989), in which the basilar membrane is modelled as a cascade of 128 digital filters, each of which has a different resonant frequency in the auditory spectrum. A stimulus, representing sound pressure, is applied at the input of the model, and travels down along the cascade of digital filters. The fluid pressure at the input to each section of the model is converted into mechanical displacement of that section. The pressure transfer function in the s-domain, for a single section, is given by (Linggard & Ambikairajah, 1986) :

$$\frac{V_o(s)}{V_i(s)} = K\left(\frac{a}{s+a}\right)\left(\frac{\omega_p^2}{s^2+B_ps+\omega_p^2}\right)\left(\frac{s^2+B_zs+\omega_z^2}{\omega_z^2}\right) \qquad (1)$$

where V_i is the pressure input to the section, V_o is the pressure output from the section, a is the low-pass filter pole frequency, ω_p is the resonant pole frequency, ω_z is the resonant zero frequency, B_p is the resonant pole bandwidth, B_z is the resonant zero bandwidth and K is a gain factor. The membrane displacement transfer function is given by :

$$\frac{V_m(s)}{V_i(s)} = K\left(\frac{a}{s+a}\right)\left(\frac{\omega_p^2}{s^2+B_ps+\omega_p^2}\right) \qquad (2)$$

where V_m is the membrane displacement.

The capacitor model of the inner hair cell

The transduction of membrane displacement to electrical energy takes place in the inner hair cells. The model of the inner hair cell used in the present work is a capacitor model, in which the input voltage corresponds to the spatially differentiated membrane displacement output of the auditory model. According to Shamma & Morrish (1987), spatial differentiation of the membrane displacement represents coupling between the cilia of the inner hair cells, through the fluid in the subtectorial space (high-pass filter effect). The remainder of the model consists of a half wave rectifier and a low pass RC filter.

The active cochlear model

Different explanations for the existence of the active cochlea have been proposed, however, the approach taken in this paper is that the basilar membrane is normally in a passive state (low-Q), but upon stimulation by a frequency of low amplitude, the section of the basilar membrane corresponding to that frequency is switched to an active state (high-Q). This is in agreement with the suggestion of Davis (1983). In this state, the experimentally-observed increased sensitivity is provided by some active mechanism feeding energy into the basilar membrane. Davis proposed that the mechanism for the active cochlea has its origin in the outer hair cells (see Figure 1, in which the ith section incorporates a closed loop). This is supported by physiological evidence (Zenner, 1988). Figure 2 shows an expanded view of the feedback loop used to model the active cochlea, for the ith section. The open loop membrane displacement is fed through a switching mechanism, S, and a piezoelectric network (the outer hair cell) to form a closed loop. The transfer function of a piezoelectric network in the s-domain, P(s), is given by:

$$P(s) = G\,\frac{s^2 + B_N s + \omega_N^2}{s^2 + B_N s + \omega_D^2} \qquad (3)$$

where G is a gain factor, ω_N is the zero frequency of the piezoelectric network, B_N is the bandwidth, and ω_D is the pole frequency of the piezoelectric network. The closed loop transfer function (Figure 2) is given by

$$\frac{V_m(s)}{V_i(s)} = \frac{\dfrac{a}{s+a}\,\dfrac{\omega_p^2}{s^2 + B_p s + \omega_p^2}}{1 + G\,\dfrac{\omega_p^2}{s^2 + B_p s + \omega_p^2}\,\dfrac{s^2 + B_N s + \omega_N^2}{s^2 + B_N s + \omega_D^2}} \qquad (4)$$

where $V'_m(s)$ is the membrane displacement of the closed loop system. Equation (4) will yield a fifth-order system which can be factored to give

$$\frac{V'_m(s)}{V_i(s)} = \frac{K\,(s^2 + B_N s + \omega_N^2)}{(s + a_1)(s^2 + B_{p1} s + \omega_{p1}^2)(s^2 + B_{p2} s + \omega_{p2}^2)} \qquad (5)$$

In order to make the problem more tractable, it is assumed that appropriate choice of the parameters ω_N, ω_D and B_N would yield $\omega_N = \omega_{p2}$ and $B_N = B_{p2}$. This would give an approximation to (5) as

$$\frac{V'_m(s)}{V_i(s)} = \frac{K}{(s + a_1)(s^2 + B_{p1} s + \omega_{p1}^2)} \qquad (6)$$

where ω_{p1} is the resonant pole frequency of the closed loop system, and B_{p1} is the pole bandwidth of the closed loop system.

Equation (6) is of the same form as the displacement transfer function of the passive system (equation (2)). However, consider the resonant frequency of the closed loop system, ω_{p1}, which is not the same as that of the open loop system, ω_p. This will give a slight shift in the resonant frequency when the section switches to the active state, as stated by Davis (1983) and Hirahara & Komakine (1989). The degree of this frequency change is not known exactly, so, for simplicity, this model assumes that the shift is negligible, hence $\omega_{p1} = \omega_p$. In this model, the active system is modelled by using second order resonant pole sections with Q factors which are much higher than those in the passive system. This can be achieved by choosing B_{p1} to be smaller than B_p.

Implementation of the digital equivalent of equations (1) and (2) for the active system requires no increase in calculation over the passive model, the only change being that different digital filter coefficients are used. Ideally, a different set of filter coefficients should be calculated for each level of increased selectivity; however, in the interests of efficiency, only two sets of coefficients were calculated. Switching a section of the basilar membrane to the active state is carried out by inserting the appropriate coefficients into the digital equations for that section.

Smoothing of the inner hair cell output

When the stimulus is a speech signal, the outputs of only the sixty filters spanning the speech spectrum (filters numbered 47 to 106, spanning the range 250 Hz to 4 KHz) are actually examined. To enable the formant content of the speech to be discerned, each inner hair cell output is processed according to the following equation:

$$y(i) = w(\frac{N}{2})\, x(i) + \sum_{K=1}^{\frac{N}{2}} [\, w(\frac{N}{2} - k)\, x(i-k) + w(\frac{N}{2} + k)\, x(i+k)] \qquad (7)$$

where

$x(i)$ = output of inner hair cell i before smoothing;
$y(i)$ = output of inner hair cell i after smoothing;
$w(k)$ are weights which determine the degree of smoothing.

Switching between the passive and active states

The basilar membrane is initially in the passive (low-Q) state until a frequency component has been identified, and has been recognised as being of low amplitude. The decision to switch is based on the amplitude of the inner hair cell output. If the output of a section is consistently less than a certain threshold, then that section of the model is switched to the active state by switching the digital filter coefficients. An increase of sensitivity of 20 dB at the characteristic frequency of the appropriate digital filter was chosen. Once a section of the cochlea has been switched to the high-Q state, it is switched back to the low-Q state after a certain period of time.

RESULTS OF THE SIMULATION OF THE ACTIVE MODEL

Response of the active cochlea to a low amplitude sinusoidal stimulus

A stimulus containing a high-amplitude component at 2 KHz and a low-amplitude component at 8 KHz was applied to the model. If the model operates correctly, the filter corresponding to 2 KHz (filter 65) should remain in the passive state, whereas the filter corresponding to 8 KHz (filter 26) should initially be in the passive state but, upon detection by the model of a peak of low amplitude, should be switched to the active state. Figure 3(a) indicates the inner hair cell output of the cochlea after application of the stimulus. At this point, both filters are in the passive state. Figure 3(b) shows the response some time later. Now, the response to the frequency component at 2 KHz is unaltered because the corresponding filter has remained in the passive state. However, filter 26 has been switched to the active state and is providing much increased amplification to the 8 KHz component. The model was thus found to correctly identify and amplify low amplitude frequency components of a stimulus.

Response of the system to speech input

The results presented in this section are those obtained when the model was tested with the utterance "one". The three parts of Figure 4 show the smoothed inner hair cell output at the end of three different 16 ms frames. For comparison, Figure 4(b) includes the unsmoothed inner hair cell output for the same time frame as the smoothed output. From this plot, the harmonic content of the speech signal is clearly visible. In Figure 4(a), all sections of the cochlear model are in the passive state; however, the third formant (F_3 = 3.2 KHz) is of low amplitude, and the corresponding section of the model is switched to the active state. The position 16 ms after switching is indicated in Figure 4(b). Note the increased amplification of F_3. As stated above, when a digital filter is switched to the high-Q state, it is switched back to the low-Q state after a certain period. Figure 4(c) shows the situation after the section of the model corresponding to the third formant has been switched back to the passive state. Similar behaviour was observed when the auditory model was tested with a fricative sound. A comparison of the auditory model output and the results of an LPC analysis of the same input data was also carried out. It was found that the peaks in the auditory model output were in approximately the same places as those in the LPC results.

Thus, the model is seen to be effective, not only in detecting the formants present in a speech signal, but also in enhancing any formants which may be of low amplitude by switching the appropriate sections of the model to the active state.

CONCLUSIONS

A computational model of the cochlea, incorporating both passive and active elements, has been presented in this paper. Results obtained when the model was excited with both a sinusoidal stimulus and a speech signal were presented. A method for smoothing the inner hair cell outputs, when the model was presented with a speech signal, was described. The filtering action of this operation enabled the formant content of the speech signal to be observed. A method for switching between the passive and active systems, based on the amplitude of the frequency components, was also presented. Results show that the model is capable of providing increased amplification to a low-amplitude formant, by switching the appropriate section to the active state.

ACKNOWLEDGEMENTS

This research was funded by a grant from EOLAS, the Irish Science and Technology Agency, under the Strategic Research Programme, 1987-1989.

REFERENCES

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

de Boer, E. (1985) *On active and passive cochlear models - towards a generalised analysis*, J. Acoust. Soc. Am. **73**, 574-576.

Davis, H. (1983) *An active process in cochlear mechanics.* Hearing Research **9**, 79-90.

Hirahara, T. & Komakine, T. (1989) *A computational cochlear nonlinear preprocessing model with adaptive Q circuits.* Proc. ICASSP '89, 496-499.

Linggard, R. & Ambikairajah, E. (1986) *A computational model of the basilar membrane.* Proceedings of the Conference on Speech Science and Technology (SST-86), Canberra, 286-291.

Lyon, R. F. (1982) *A computational model of filtering, compression and detection in the cochlea.* Proc. IEEE, ICASSP, Paris.

Neely, S. T. (1989) *A model for bidirectional transduction in outer hair cells.* Cochlear Mechanisms (J. P. Wilson & D. T. Kemp, eds.), 75-82, Plenum Press, New York.

Neely, S. T. & Kim, D. O. (1985) *An active cochlear model showing sharp tuning and high sensitivity*, Hearing Research **9**, 123-130.

Shamma, S. & Morrish, K. (1987) *Synchrony suppression in complex stimulus responses of a biophysical model of the cochlea.* J. Acoust. Soc. Am. **81**, 1486-1498.

Zenner, H. P. (1988) *The secret behind the sense of hearing.* German Research, Jan. 1988, 10-11.

Zwicker, E. (1986) *A hardware cochlear nonlinear preprocessing model with active feedback*, J. Acoust. Soc. Am. **80**, 146-153.

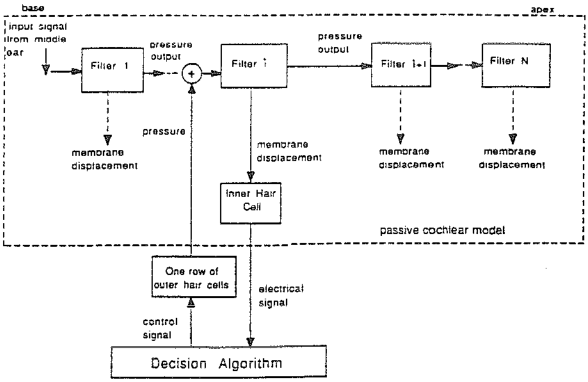


Figure 1. Overall block diagram of the cochlear model (the part enclosed within the dotted line is the passive model, and the ith section contains both passive and active elements'

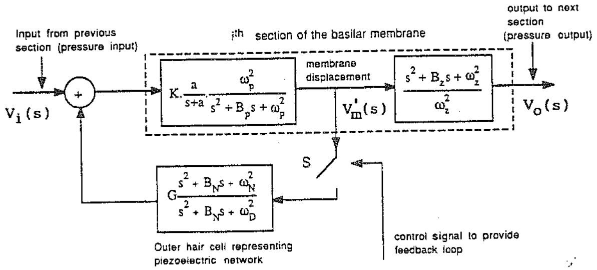
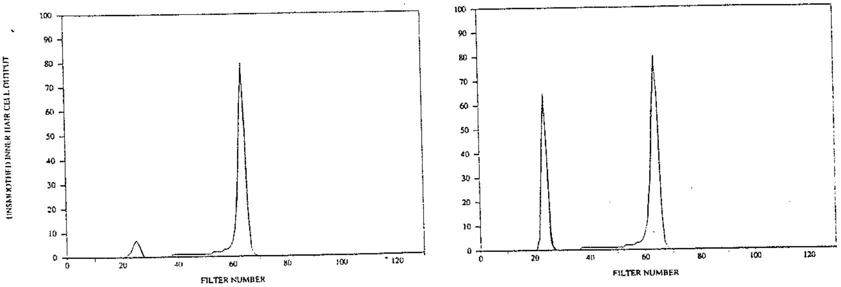

Input from previous section (pressure input)
ith section of the basilar membrane
output to next section (pressure output)
$V_i(s)$
+
$K.\frac{a}{s+a}.\frac{\omega_p^2}{s^2+B_ps+\omega_p^2}$
membrane displacement
$V_m'(s)$
$\frac{s^2+B_zs+\omega_z^2}{\omega_z^2}$
$V_o(s)$
S
$G\frac{s^2+B_Ns+\omega_N^2}{s^2+B_Ns+\omega_D^2}$
Outer hair cell representing piezoelectric network
control signal to provide feedback loop

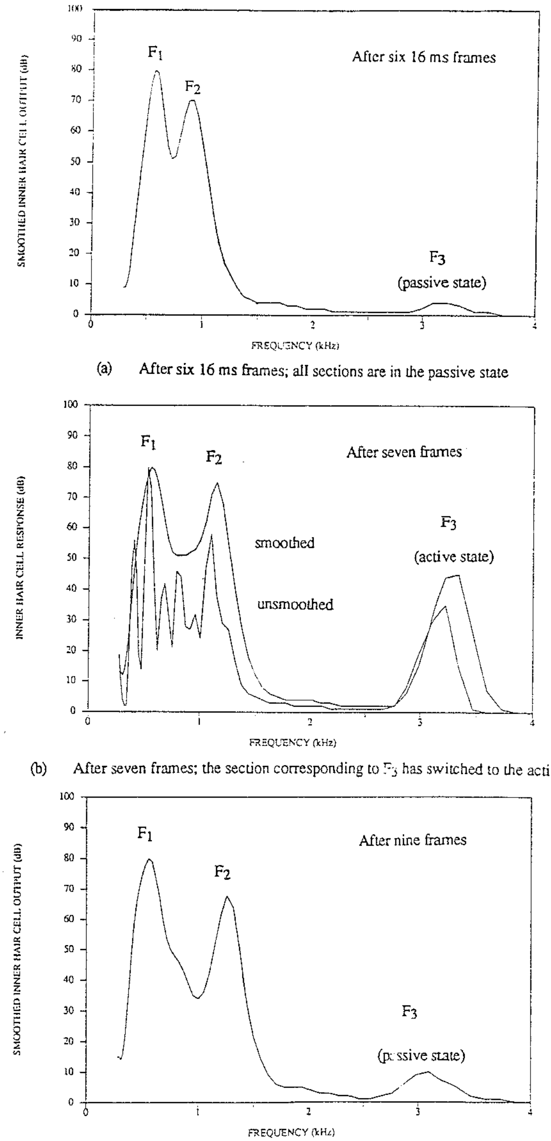
Figure 2. Block diagram of the ith section of the cochlear model, including the feedback loop responsible for the active cochlea



(a) All filters in the passive state

(b) Filter number 26 has switched to the active state

Figure 3. Response of the model to a sum of two sinusoids, one of high amplitude, the other of low amplitude



(a) After six 16 ms frames; all sections are in the passive state

(b) After seven frames; the section corresponding to F_3 has switched to the active state

(c) After nine frames; the section corresponding to F_3 has switched back to the passive state

Figure 4. Response of the model to the utterance "one"