Proceedings of The Institute of Acoustics

A FILTERBANK MODEL OF THE COCHLEA

N.D. BLACK & R. LINGGARD

DEPT. OF ELECTRICAL & ELECTRONIC ENGINEERING,
THE QUEEN'S UNVIERSITY OF BELFAST.

1. INTRODUCTION

A plausible strategy for the machine recognition of speech, is to use a set of processing algorithms which emulate the various aspects of human audition. The most important of these is the spectral analysis which attempts to model cochlear response. This paper presents a filterbank model of cochlear action which has 145 bandpass filters. Centre frequencies are on a logarithmic scale, with bandwidths assigned according to critical band theory. The output of each filter is rectified, smoothed, and logged and is assumed to be proportional to the signal which, in the cochlea, excites the perpherial auditory neurons.

Two other algorithms, intended to follow the cochlea analysis, are presented. The first of these is designed to enhance the harmonic content of the spectra. The second, uses the peaks found in the enhanced spectra to detect harmonic patterns corresponding to various fundamental frequencies (pitch).

2. FILTERBANK DESIGN

It is well known that the mechanical response of the Basilar membrane is low pass, whereas the neural output from the cochlea appears to be caused by a band-passed half wave rectified signal. The filterbank proposed here, uses bandpass filters, and the unknown mechanism which converts the Basilar membrane response to neural signals, is assumed to be included in the model.

The filter functions used are based on the 4th-order analogue filter:- (C.F. Flanagan).

$$H_{(s)} = \frac{b^2 S^2}{(S^2+bs+\omega_o)(S^2+bs+\omega_o)}$$

where ω_o is the filter centre frequency and b the bandwidth.

This is transformed using the Bilinear transform to give the Z-plane transfer function

$$H_{(z)} = (1-2Z^{-2}+Z^{-4}) / (1+\alpha Z^{-1}+\beta Z^{-2}+\eta Z^{-3}+\delta Z^{-4})$$

The actual shapes of the digital frequency responses are warped by the bilinear transform so that they become progressively less symmetrical as their centre frequencies approach the half sampling frequency, as described in Fig. 1. Fortunately, similar assymmetry occurs in frequency responses of the basilar membrane, so that the filterbank gives quite a good fit to actual cochlea response.

The filterbank is designed to cover the frequency range 100 Hz to 6400 Hz in six octaves with 24 filters per octave giving (6x24)+1 = 145 channels, spaced on a logarithmic scale. The two end filters are specified by

$$S(1) = \ln(100) = 4.6052$$

$$S(145) = \ln(6400) = 8.7641$$

The logarithmic spacing between filters is then

$$\frac{S(145) - S(1)}{144} = 0.0288$$

and the centre frequencies of the filter of channel n is given by

$$f(n) = \exp.\ (4.6052 + 0.0288(n-1))$$

The bandwidths are assigned in accordance with critical band theory and are shown in Fig. 2. A 'sharpness' function K is used to multiply the filter bandwidths, so that the resoultion of the filterbank as a whole may be increased or decreased as K is varied.

3. SPECTRAL ENHANCEMENT

Using bandwidths consistent with critical band theory, gives filterbank Q values in the range 1 (low frequencies) to 8 (high frequencies). An example of an output spectrum for a three component harmonic series is shown in Fig. 3(a). It can be seen that the harmonics are not fully resolved. Obviously increasing the 'sharpness' of the filterbank would reveal the harmonics structure, but this would require Q-values unrealistically high. It would also impair the time response of the filterbank.

The enchancement algorithm simply subtracts from each channel a small fraction of the outputs of the two adjacent channels. This fraction is necessarily small but by iterating the process, the enhancement effect is spread over many more than two channels. If Y(n) is the output of channel n of the filterbank, then one iteration of the enhancement process gives the output

$$Y_1(n) = Y(n) + \left[\alpha Y(n) - \beta\{Y(n-1) + Y(n+1)\}\right]$$

Fig. 3(a) and 3 (b) show the before and after enhancement spectra, with the enhanced spectra exhibiting fully resolved harmonics.

4. PITCH DETECTION BY HARMONIC MATCHING

The pitch detection algorithm operates in the frequency domain, and conforms to the 'place theory'. The process examines the enhanced spectra for peaks and each spectrum is reduced to a binary pattern so that channels are labelled one (peak) and zero (no peak). This pattern is then compared with a set of template patterns, each of which represents a perfect set of harmonics for a particular pitch frequency. For example pitch at 100 Hz will have harmonic peaks at frequencies 100 Hz, 200 Hz etc., and hence have ones in channel numbers 1,25,39,49 ... etc. All other channels will be zero. For pitch of 103 Hz the harmonic template will have ones at 2,26,50 etc. This is the same pattern for 100 Hz shifted by one channel. In fact because of the logarithmic scale, all the harmonic templates are shifted versions of each other. To account for pitch less than 100 Hz hypothetical channels having negative channel numbers are used to represent frequencies less than 100 Hz. Thus the harmonic template for 50 Hz has ones at channel no -23,1,15,25 etc.

The harmonic-pattern matching algorithm compares the peak pattern for an incoming spectrum with each of the harmonic templates. The actual comparison is an exclusive - Or operation and the output is the number of ones in the result. This is a distance measure in that a perfect match results in zero output. A typical 'distance spectrum' is shown in Fig. 4 and the minimum of the graph indicates true pitch.

5. CONCLUSIONS

The cochlea model presented here is designed to be a good compromise between complexity and convenience. It is complex (145 channels) but convenient and quick to implement. Its main deficiency is that it does not model the cochlea phase response accurately nor does it include non-linearaties.

Spectral enhancement is a plausible mechanism to explain the frequency sensitivity of the ear, although the algorithm presented is intended to model the effect rather than the mechanism itself. Likewise the harmonic pattern matching algorithm is a plausible explanation of human pitch detection, though how this could be accomplished in neural networks is still uncertain.

REFERENCES

1. FLANAGAN, J.L. 1972 p.p. 110 ' Speech Analysis Synthesis And Perception' 2nd Ed. Pub. Springer-Verlag.

2. GREENWOOD, D.D. 1961 J. Acoust. Soc. Am. 33, 1344-1356, Critical Bandwidth And The Frequency Coordinates of The Basilar Membrane.

3. HAWKING, J.E. and STEPHENS S.S. 1950 J. Acoust. Soc. Am., 22, 6-13, The Masking of Pure Tones And of Speech By White Noise.

4. SCHARF B. 1970 p.p. 159-202 Foundations of Modern Auditory Theory, Ed. J.V. Tobias, Acd. Press.

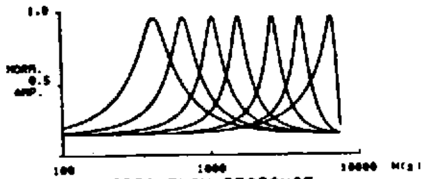

FIG.1 FREQUENCY RESPONSE

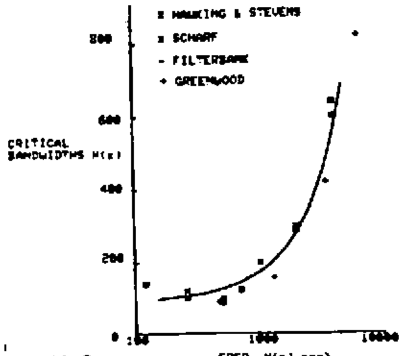

FIG.2 FILTERBANK BANDWIDTHS

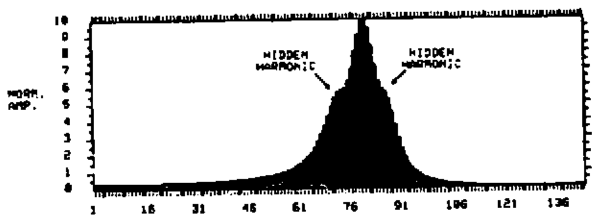

FIG.3(a) SPECTRUM OF MODULATED TONE

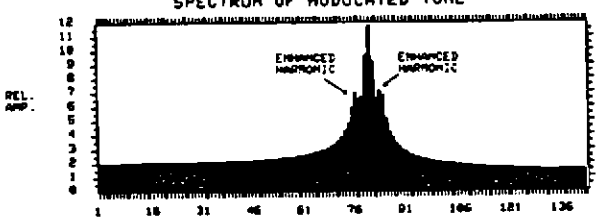

FIG.3(b) ENHANCED SPECTRA

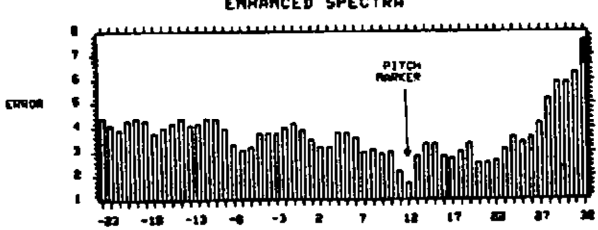

FIG.4 MINIMUM ERROR PITCH PLOT