A FUNCTIONAL MODEL OF NEURAL ACTIVITY PATTERNS AND AUDITORY IMAGES

Roy D. Patterson and John Holdsworth

OUTLINE

1. Introduction..548

2. The Cochlea Simulation 548
   2.1 Spectral Analysis and the Gammatone Auditory Filterbank 549
   2.2 Feature Enhancement and Two-Dimensional Adaptation 551

3. The Auditory Image 557
   3.1 The Temporal Integration Problem in Audition 557
   3.2 Quantised Temporal Integration 558
   3.3 The Trigger Mechanism 560
   3.4 The Auditory Image of /ae/ 561

This version of the paper has been remastered from the original text and scanned figures to provide a version for electronic distribution.

1

[1] Advances in Speech, Hearing and Language Processing
Volume 3, Part B, pages 547-563

ISBN: 1-55938-430-1

 548

1. Introduction

A comprehensive, but compact, model of auditory processing has been developed to analyse everyday sounds like music and speech and the sound environments in which they occur. The model transforms a complex sound into a multi-channel activity pattern like that observed in the auditory nerve, and then it converts this 'neural' activity pattern into an 'auditory image' that is intended to represent our initial impression of the sound.

Since we desire an accurate description of human perception, the initial stage that generates the neural activity pattern is implemented with a model whose architecture relects auditory anatomy and whose processing algorithms reflect auditory physiology -- a cochlea simulation. At the same time, since we are primarily interested in the neural activity patterns that it produces rather than the physiology *per se*, and since the speed of the model is very important, we have chosen to implement a *functional* model of the cochlea. Thus, for example, the initial spectral analysis is performed by a bank of recursive auditory filters rather than a comprehensive model of cochlear mechanics. The first half of this paper presents the functional cochlea and our representation of the neural activity pattern (NAP) arriving at the input to the cochlear nucleus.

The NAP is converted into an auditory image using a bank of temporal integration units, one per channel of the NAP. Each unit monitors the activity in its channel looking for peaks, which are used to trigger the transfer of a section of the NAP to the corresponding channel of an image buffer. The NAP sample is added, point for point, with what is already there in the buffer, and the multi-channel result is the auditory image. This 'triggered' temporal integration converts the NAP of a periodic sound into a static auditory image, and the NAP of a dynamic sound into a flowing image in which the motion occurs at the rate that we hear it in the sound. The mechanism and the resultant auditory image are presented in the second half of the paper. The anatomy and physiology of auditory temporal integration (TI) are largely unknown and so our model of image construction is, by its nature, purely functional. However, it is likely to occur earlier rather than later in the processing chain, and so we are inclined to look for the trigger mechanism, at least, in the cochlear nucleus.

2. THE COCHLEA SIMULATION

The cochlea simulation is composed of two processing modules: (A) a *gammatone* auditory filterbank which performs a spectral analysis and converts the acoustic wave into a multi-channel representation of basilar membrane motion, and (B) a two-dimensional adaptation mechanism that 'transduces' the membrane motion and converts it into a multi-channel representation of the neural activity pattern arriving at the cochlear nucleus. The transduction process includes compression and suppression, as well as adaptation.

2.1 Spectral Analysis and the Gammatone Auditory Filterbank

There are three parts to the argument for choosing the gammatone function as an auditory filter, one physiological, one psychological and one practical:

1. *Physiological*: The impulse response of the gammatone filter function provides an excellent fit to the impulse responses obtained from cats with the revcor technique (de Boer and de Jongh, 1978; Carney and Yin, 1988). The gammatone filter is defined in the time domain by its impulse response

$$gt(t) = a\, t^{(n-1)} \exp(-2\mathrm{p}bt) \cos(2\mathrm{p}\, f_o t + Ø) \quad (t>0) \qquad (1)$$

The name *gammatone* comes from the fact that the envelope of the impulse response is the familiar *gamma* function from statistics, and the fine structure of the impulse response is a *tone* at the centre frequency of the filter, fo. (de Boer and de Jongh, 1978). The primary parameters of the gammatone (GT) filter are b and n: b largely determines the duration of the impulse response and thus,the bandwidth of the GT filter; n is the order of the filter and it largely determines the tuning, or Q, of the filter. Carney and Yin (1988) fitted the GT function to a large number of revcor impulse responses and showed that the fit was surprisingly good.

2. *Psychological*: The properties of frequency selectivity measured physiologically in the cochlea and those measured psychophysically in humans are converging. Firstly, the shape of the magnitude characteristic of the GT filter with order 4 is very similar to that of the roex(p) function commonly used to represent the magnitude characteristic of the human auditory filter (Schofield, 1985; Patterson and Moore, 1986; Patterson, Holdsworth, Nimmo-Smith and Rice, 1988). Secondly, the latest analysis by Glasberg and Moore (1990)

indicates that the bandwidth of the filter corresponds to a fixed distance on the basilar membrane as suggested originally by Greenwood (1961). Specifically, Glasberg and Moore now recommend Equation 2 for the Equivalent Rectangular Bandwidth (ERB) of the human auditory filter.

$$\text{ERB} = 24.7(4.37fo/1000 + 1) \qquad (2)$$

This is Greenwood's equation with constants derived from human filter-shape data. The traditional 'critical band' function of (Zwicker, 1961) overestimates the bandwidth at low centre frequencies.

When fo/b is large, as it is in the auditory case, the bandwidth of the GT filter is proportional to b, and the proportionality constant only depends on the filter order, n (Holdsworth, Nimmo-Smith, Patterson and Rice, 1988). When the order is 4, b is 1.019 ERB. The 3-dB bandwidth of the GT filter is 0.887 times the ERB.

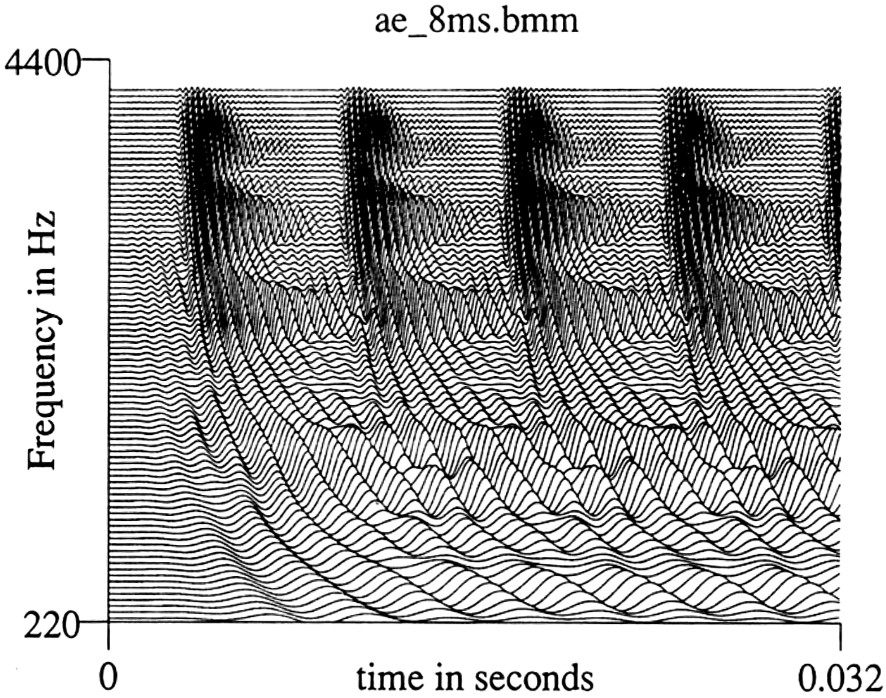

Figure 1. A simulation of the basilar membrane motion produced in response to four cycles of the vowel /ae/ in 'past'. The simulation was produced with a gammatone auditory filterbank and each line shows the output of one auditory filter.

Finally, with respect to phase, the GT is a minimum-phase filter and, although the phase characteristic of the human auditory filter is not known, it seems reasonable to assume that it will be close to minimum phase.

3. *Practical*: An analysis of the GT filter in the frequency domain (Holdsworth, Nimmo-Smith, Patterson and Rice, 1988) revealed that an nth-order GT filter can be approximated by a cascade of n, identical, first-order GT filters, and that the first-order GT filter has a recursive, digital implementation that is particularly efficient. For a given filter, the input wave is frequency shifted by -fo, low-pass filtered n times, and then frequency shifted back to fo.

The resultant gammatone auditory filterbank provides a good tradeoff between the accuracy with which it simulates cochlear filtering and the computational load it imposes for applications where the level of the stimulus is moderate. The response of a 49-channel gammatone filterbank to four cycles of the vowel in the word 'past' is presented in Figure 1. Each line shows the output of one filter and the surface defined by the set of lines is a simulation of basilar membrane motion as a function of time. At the top of the figure, where the filters are broad, the glottal pulses of the speech sound generate impulses responses in the filters. At the bottom of the figure, where the filters are narrow, the output is a harmonic of the glottal pulse rate. Resonances in the vocal tract determine the position and shape of the energy concentrations referred to as 'formants'.

The GT auditory filterbank is particularly appropriate for simulating the cochlear filtering of broadband sounds like speech and music when the sound level is in the broad middle range of hearing. The GT is a linear filter and the magnitude characteristic is approximately symmetric on a linear frequency scale. The auditory

 550

filter is roughly symmetric on the same scale when the sound level is moderate; however, at high levels the highpass skirt becomes shallower and the lowpass skirt becomes sharper. For broadband sounds, the effect of filter asymmetry is only noticeable at the very highest levels (over 85 dBA). For narrowband sounds, the precise details of the filter shape can become important. For example, when a tonal signal is presented with a narrowband masker some distance away in frequency, the accuracy of the simulation will deteriorate as the frequency separation increases.

2.2 Feature Enhancement and Two-Dimensional Adaptation.

The inner hair cells in the cochlea convert the motion of the membrane and outer hair cells into neural transmitter, and the concentration of the neural transmitter determines the probability that the sensory neurons attached to the hair cells will fire. Data from individual inner hair cells indicate that they *compress*, *rectify* and *lowpass filter* their input, although not necessarily in that order. They *adapt* to level changes rapidly, and there is a lateral interaction in the frequency dimension of the cochlea which results in areas of intense activity *suppressing* areas of lesser activity. *Adaptation* and *suppression* enhance features that arise in basilar membrane motion, which indicates that the array of inner hair cells is more than just a transduction mechanism. Rather, it should be regarded as a sophisticated signal processing unit designed to remove the smearing introduced by the filtering and compression, and to emphasise the information concerning the energy concentrations in the stimulus.

In the cochlea simulation, this processing is performed by a functional model of the inner hair cells and primary nerve fibers. It includes a bank of logarithmic compressors and a bank of adaptation units that apply adaptation to the membrane motion simultaneously in time and in frequency (Holdsworth, 1990) and so it is referred to as *two-dimensional adaptation*. It converts the surface that represents our approximation to basilar membrane motion into another surface that represents our approximation to the pattern of neural activity that flows from the cochlea up the auditory nerve to the cochlear nucleus -- the NAP. In the process, the adaptation introduces compression, rectification, adaptation and suppression and lowpass filtering.

1. *Adaptation in the Time Domain:* Consider the problem that arises when one attempts to distinguish two related signals as they appear at the output of one individual member of the auditory filterbank. The signals are an impulse and an impulse which has been passed through a resonance; they are presented as a composite signal with the second beginning 20 ms after the first in Figure 2a. When this pair of signals is passed through the auditory filter centred at 1.0 kHz the result is as shown in Figure 2b. Whereas the two signals appear very different in Figure 2a, they are much less different in Figure 2b because they occur in conjunction with the impulse response of the filter. This is an unavoidable by-product of the process of spectral decomposition performed by an auditory filterbank. It is important to be able to distinguish the impulse from an impulse that has been passed through a resonator because the individual formants of speech can be thought of as impulses that have been passed through a resonance.

The difference between the filterbank response to the two stimuli is more clearly seen when we look at the output of the logarithmic compressor, as shown in Figure 2c. The rate of decay of the impulse response of the gammatone filter is a negative exponential. Since the compressor is a log function, the filter decay function is a straight line with a negative slope on these co-ordinates. The damped sinusoid causes the filterbank output to decay more slowly as seen both in Figures 2b and 2c; it is this slower rate of decay that discriminates the damped sinusoid from the impulse and in the compressed output the difference between the two is more marked.

In general terms, this suggests that one should measure the output of the filter relative to the filter's impulse response. In order to implement this property we use adaptive thresholding and the dotted line in Figure 2c provides an example of the adaptive threshold. In this particular mechanism there is a contrast between the rate at which the threshold is permitted to rise and the rate at which it is permitted to decay. There is no restriction on the rate of rise and so the adaptive threshold follows the leading edge of the signal in the upwards direction. In the downwards direction, however, the rate of descent is limited to a rate slightly less than that of the impulse response of the filter so that cycles of the compressed filterbank output which are simply due to the ringing of the filter fall just below the adaptive threshold. Figure 2d shows the compressed output measured relative to the adaptive threshold and this is the output of the mechanism. It shows a dramatic difference between the response of the impulse and the damped sinusoid. The details of adaptive thresholding and a circuit diagram for the process are presented in Holdsworth (1990).

2. *Adaptation in the Frequency Domain:* An analogous procedure can be used to implement adaptive thresholding in the frequency domain where the results appear as suppression. Once again we use a composite

signal to demonstrate the operation of the mechanism. In this case the signal is composed of two sinusoidal components one at 1000 Hz and the other at 2300 Hz; the latter component is 24 dB weaker than the former.

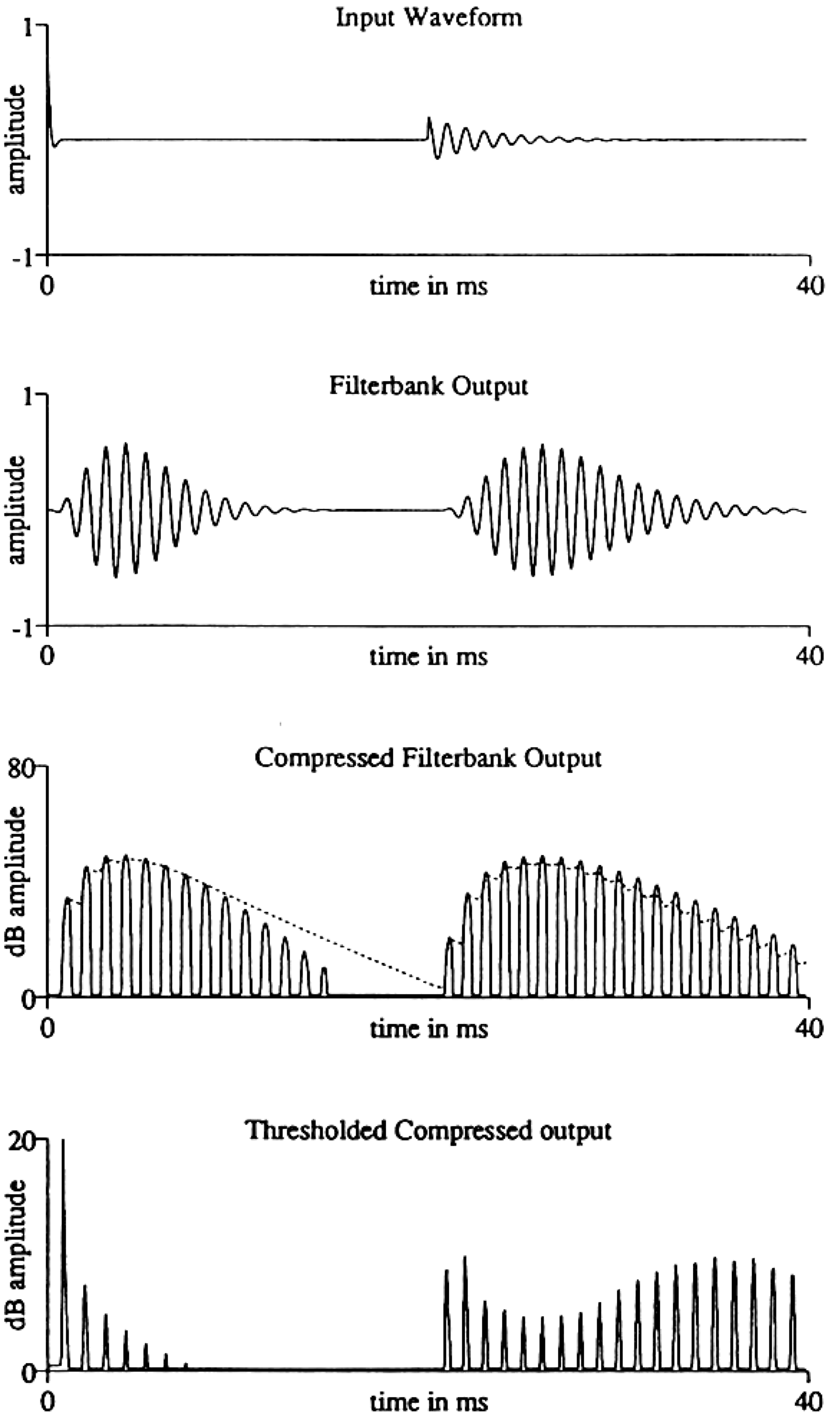

Figure 2. A schematic representation of the operation of two-dimensional adaptation in the time domain.

552

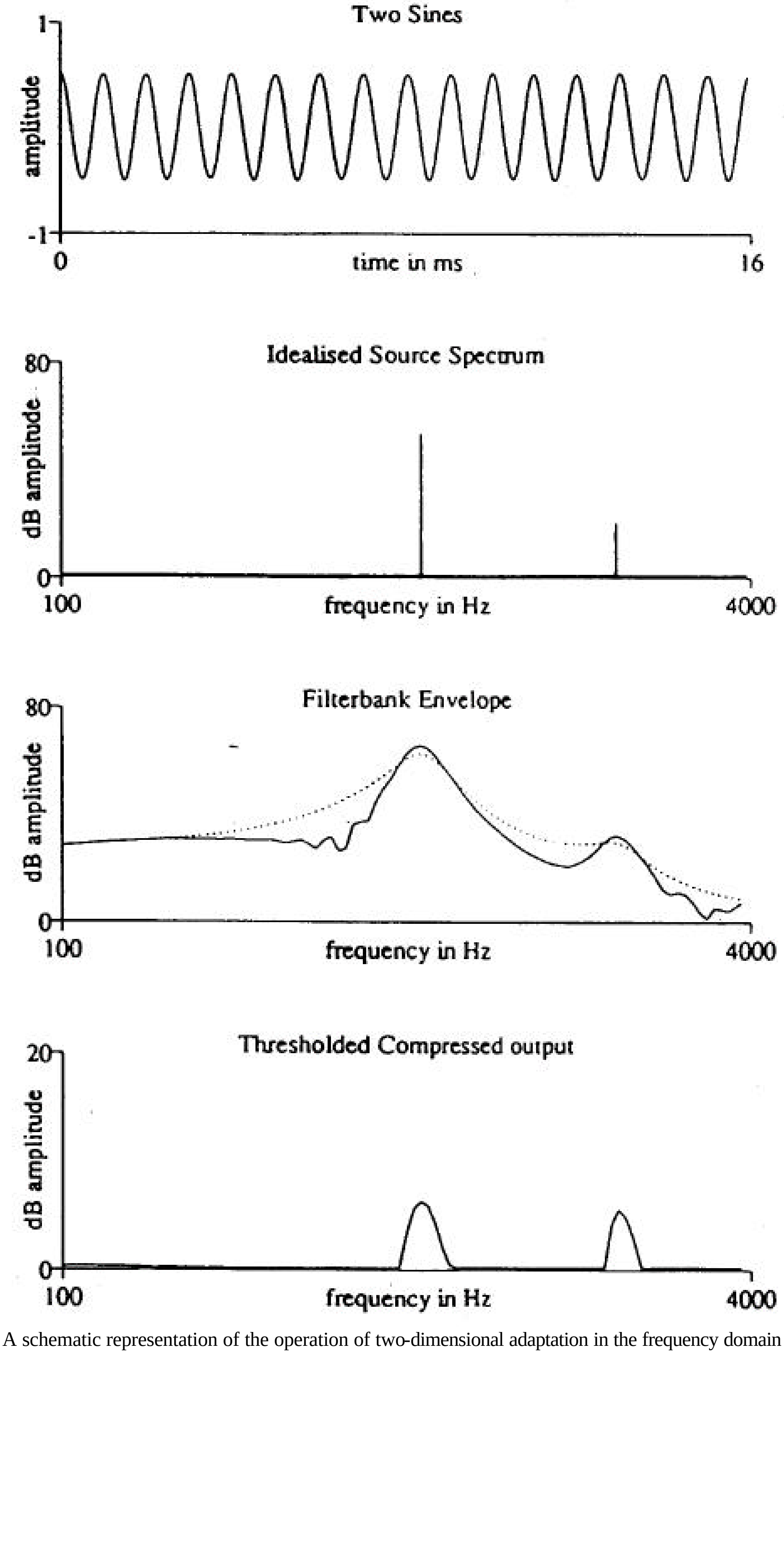

Figure 3. A schematic representation of the operation of two-dimensional adaptation in the frequency domain

The composite waveform is shown in Figure 3a; it is essentially a 1 kHz sine wave since the high frequency component is so small. The long-term power spectrum of the composite signal is shown in Figure 3b. The response of the filterbank to this composite signal is shown in Figure 3c. In point of fact, the solid line in Figure 3c shows the instantaneous value of the sustaining envelope of the filterbank output; that is, the solid line shows the contour produced by joining the maxima of all of the individual filter envelopes at this instant in time. Since the skirts of the individual filters are not infinitely steep, all of the filters in the region of the 1 kHz tone
respond to the tone. The degree of response decreases as the difference between the filter centre frequency at 1 kHz increases, but the spread of activity is significant. This spreading is an unavoidable property of any filterbank that has a reasonable temporal response and which cannot integrate forever. A device which monitors the filterbank envelope for signals, and which detects a local maximum in the envelope, should measure activity in that region relative to the smearing function; that is, the detection process should take note of the fact that activity in that region can only fall away at a certain rate given the properties of the filterbank.

Once again we use the adaptive thresholding technique to prevent the mechanism from producing output that is simply a by-product of the operation of the filterbank itself. We construct a threshold that is suspended by the peak of a local maximum and which can only drop away from that local maximum at a limited rate as shown by the dotted line in Figure 3c. The output of the device is measured relative to the adaptive threshold and it is shown for the composite signal in Figure 3d. Note that the features in Figure 3d associated with the two signal components are considerably sharpened relative to their representation in Figure 3c.

There is one further aspect of the mechanism that should be described at this point. In both the time domain and the frequency domain, the adaptive threshold falls away from the local maximum at a rate that is not quite as fast as the filterbank would permit. The reduction of slope means that small features will be suppressed in the region of a large feature. In general, this is a useful property when attempting to separate signals from noise, and it is a property observed in auditory nerve fibers.

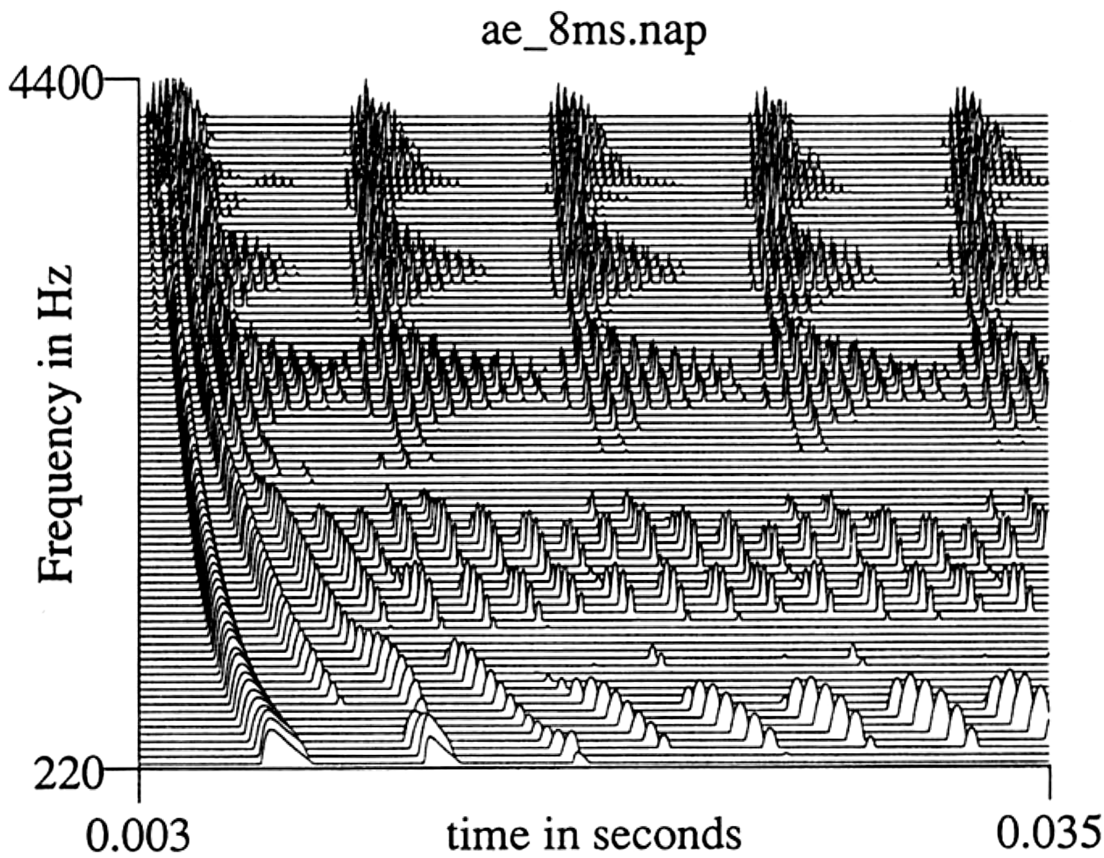

Figure 4. A simulation of the neural activity pattern arriving at the cochlear nucleus in response to four cycles of the vowel /ae/ in ‘past’. The simulation was produced with a two-dimensional adaptation mechanism and each line shows the neural activity associated with one auditory filter.

 554

The NAP produced by the full cochlea simulation in response to four cycles of the vowel /ae/ is presented in Figure 4. The formants have been sharpened and much of the filterbank ringing has been removed. Furthermore, the individual half-cycles of the filtered waves have been sharped in time which improves the temporal accuracy of the NAP.

As a model, two-dimensional adaptation is a functional representation of inner haircell and primary fiber processing, rather than a collection of units, each of which is intended to simulate a single inner hair cell. Specifically, there is only one cascade of rectifier, compressor, adaptation unit and lowpass filter for each auditory filter, and together, that single cascade is intended to represent the activity of all of the inner hair cells and primary fibers associated with that auditory filter. This functional model is more physiological than the auditory filterbank, in the sense that the processes in the module are motivated by the mechanisms observed in the cochlea. At the same time, however, the NAP is probably a less accurate representation of neural firing patterns than the filterbank output is of basilar membrane motion because there is relatively less information available on the aggregate behaviour of the inner hair cells and primary fibers.

3. THE AUDITORY IMAGE

3.1 The Temporal Integration Problem in Audition

When the input to the cochlea is a periodic sound, like a vowel or a musical note, the NAP of the sound oscillates. In contrast, the sensation produced by such a sound does not flutter or flicker; indeed, periodic sounds produce the most stable auditory images. Traditional models suggest that we integrate the NAP over time using a sliding temporal window and so smooth out the rapid oscillations of the periodic sound. If we have an integrator for each channel and the 3-dB duration of each temporal window is long with respect to the period of the sound, the result is a stable central spectrum that could form the basis of our stable auditory image. Unfortunately, this simple model of TI does not work for the auditory system. If the output is to be stable, the integrator must integrate over 10 or more cycles of the sound. We hear stable images for pitches as low as 50 cycles per second, and so the integration time would have to be 200 ms at the minimum. Such an integrator would cause far more smearing of auditory information than we know occurs. For example, phase shifts that produce small changes half way through the period of a pulse train are often audible (see Patterson, 1987, for a review). Small changes of this sort would be obscured by lengthy temporal integration. Thus, the problem in modelling auditory temporal integration is to determine how the auditory system can integrate information to form a stable auditory image without losing the fine-grain temporal information within the individual cycles of periodic sounds.

It is also useful to consider TI from an information processing perspective, and in particular, the problem of preserving formant information in speech. The shape of the NAP within the period provides information about the resonances of the vocal tract and thus the identity of the vowel, or other voiced part of speech. The information about the source arrives in packets whose duration is the period of the source. Many speech sounds have the property that the resonance information changes relatively slowly when compared with the repetition rate of the source wave (i.e. the pitch). Thus, one would like to combine source information from neighbouring packets, while at the same time taking care not to smear the source information contained within the individual packets. In short, one would like to perform *quantised* temporal integration, integrating over cycles but not within cycles of the sound.

Accordingly, a mechanism has been developed in which the larger peaks in the NAP are used to trigger a quantised TI process. The triggering mechanism is adaptive and for periodic sounds it tends to output one trigger pulse per period at the time of the largest peak in the period. The stream of trigger pulses can then be used to initiate period-synchronous integration which causes periodic information to accumulate and aperiodic information to die away. At the same time, the periodic information forms a *stabilised auditory image* (SAI) that provides a reasonable representation of the sensation that we hear.

3.2 Quantised temporal integration

The construction of one channel of the auditory image of a pulse train sound with an 8 ms period is illustrated in Figure 5. The centre frequency of the channel is 1.0 kHz and it is like a channel from the centre of the second formant of /ae/. The upper row of the figure (a) shows the NAP that would flow from the 1.0-kHz channel in response to 8 cycles of the sound; the flow proceeds from right to left as if produced by a chart recorder. It is assumed that there is a buffer store in the auditory system through which the neural activity flows in a first-in, first-out fashion, and as it flows, the level decays exponentially in time. The second row of the figure (b) shows the FIFO buffer with the neural activity pattern after 8 cycles of the sound. When a trigger peak occurs a copy of *all* the information in the FIFO buffer is transferred to the corresponding channel of the auditory image and summed point for point with the information that is currently there. The auditory image is a static buffer and so it is at this point in the system that the temporal code is converted to a spatial code. The strength of the auditory image decays exponentially in time but the position of the information does not change once it enters the image buffer.

The build up of the auditory image over the first 8 cycles of the sound is shown in the 8 rows of Figure 5c. The integration is triggered on the second peak of each impulse response in the NAP as indicated by the position of the cycle numbers in Figures 5a and 5b and by the position of the image relative to the righthand ordinate in the 8 rows of Figure 5c. The form of the auditory image after each transfer is shown for each cycle in successive rows of Figure 5c. Thus, the impulse response towards the righthand end of row c4, is the combination of 3 clusters of NAP pulses: the first from cycle one at the time of the second trigger pulse, the second from cycle two at the third trigger pulse, and the third from cycle three at the fourth trigger pulse.

After each cycle the auditory image becomes a cycle broader, but whereas the NAP flows across the FIFO buffer store, the auditory image simply emerges from the background in its final position. In general, then, the auditory image is stationary for periodic sounds because the NAP information for successive cycles is the same and successive cycles of the NAP accumulate into the same position in the auditory image. That is, the information for cycle n+1 of a periodic sound will be added in at the same point as the information from cycle n provided the trigger pulses are one period apart.

The column of impulse responses towards the righthand end of the image shows that the image grows over the first four to five cycles of the sound and then asymptotes to a fixed level as the summation and decay processes come into balance. Thus, this column shows quantised TI for samples of the NAP separated by one cycle of the sound. The second column from the righthand edge of the image shows TI for samples separated by two cycles, and in the third column from the right, TI for samples separated by three cycles. So, the width of the image reveals the duration over which the sound is periodic. The limit on the width of the auditory image for a strictly periodic sound is determined primarily by the decay rate in the NAP buffer.

3.3 The Trigger Mechanism

The trigger mechanism is relatively simple. A trigger threshold value is maintained for each channel and when a NAP pulse exceeds the threshold a trigger pulse is generated at the time associated with the maximum of the peak. The threshold value is then reset to a value somewhat above the height of the current NAP peak and the threshold value decays exponentially with time until another peak is encountered. The amount of threshold elevation is determined by the height of the most recent peak and the time between the last pair of trigger pulses in a way that causes the mechanism to emphasise the period of sound when it is periodic.

General purpose pitch mechanisms based on peak picking are notoriously difficult to design, and the trigger mechanism just described would not work well on an arbitrary acoustic waveform. The reason that this simple trigger mechanism is sufficient for the construction of the auditory image is that NAP functions are highly constrained. The microstructure invariably reveals a function that rises from zero to a local maximum smoothly and returns smoothly back to zero where it stays for more than half of a period of the centre frequency of that channel. On the longer time scale, the amplitude of successive peaks changes only relatively slowly with respect to time. As a result, for periodic sounds there tends to be one clear maximum per period in all but the lowest channels where there is an integer number of maxima per period. The simplicity of the NAP functions follows from the fact that the acoustic waveform has passed through a narrow band filter and so it has a limited number of degrees of freedom. In all but the highest frequency channels, the output of the auditory filter resembles a modulated sine wave whose frequency is near the centre frequency of the filter. Thus the NAP is largely restricted to a set of peaks which are modified versions of the positive halves of a sine wave, and the remaining degrees of freedom appear as relatively slow changes in peak amplitude and relatively small changes in peak time (or phase).

It is important to note that the triggering is done on a channel by channel basis and that it is asynchronous across channels, inasmuch as the major peaks in one channel occur at different times from the major peaks in

 556

other channels. It is this aspect of the triggering process that causes the alignment of the auditory image and which, in turn, accounts for the loss of phase information in the auditory system (Patterson, 1987).

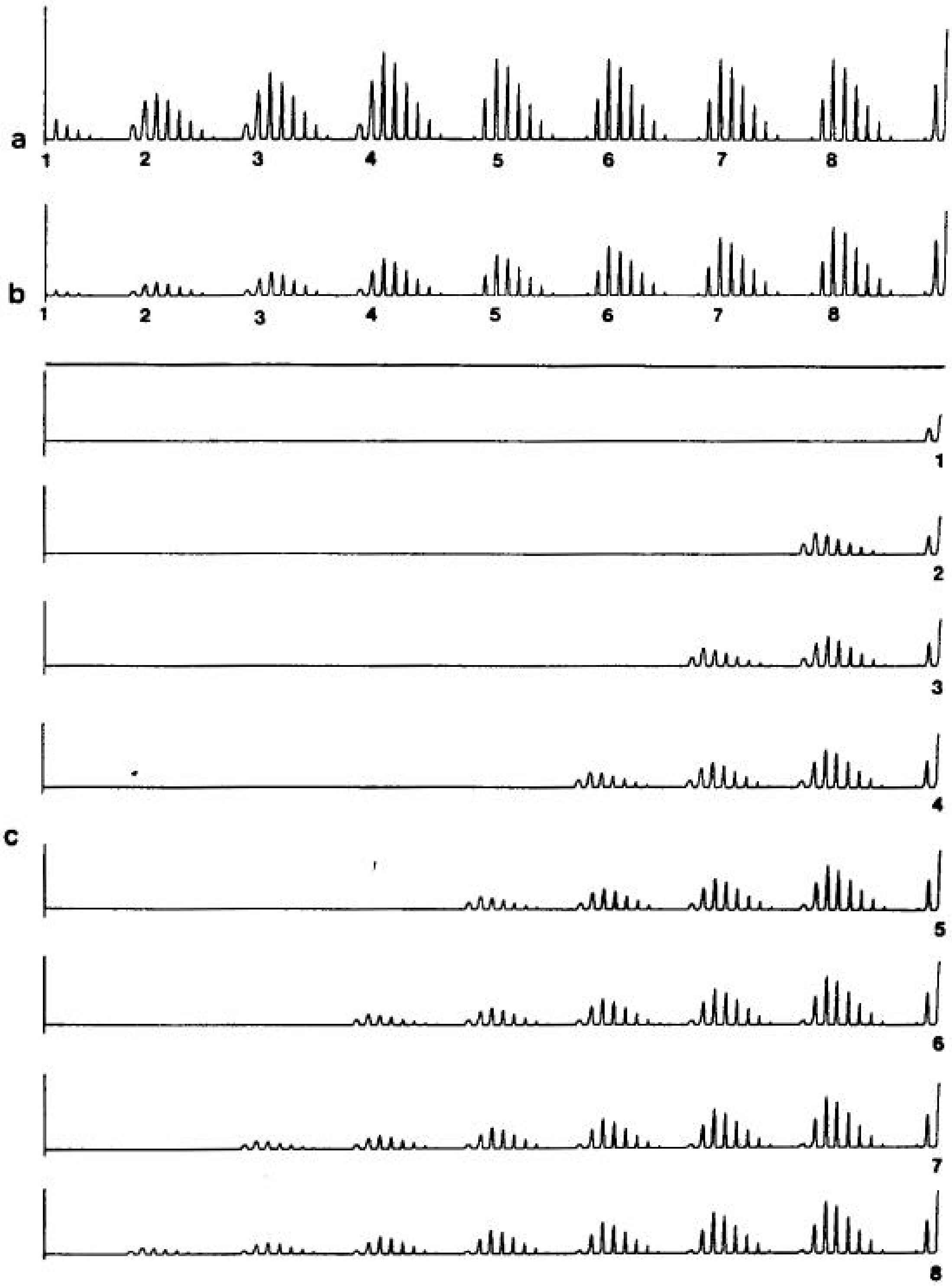

Figure 5. Illustration of the build-up of one channel of the auditory image produced by a pulse train with a period of 8ms. The center frequency of the channe is 1.0kHz. The upper row (a) shows the NAP produced by 8 cycles of the sound and the row below it (b) shows the decaying internal representation. Rows c1 – c8 show the auditory image just after each successive trigger pulse.

3.4 The Auditory Image of /ae/

A sample of the stabilised auditory image for the vowel /ae/ is shown in Figure 6. The auditory image has the same vertical dimension (filter centre frequency) as the neural activity pattern. The continuous time dimension of the neural activity pattern becomes a local time dimension in the auditory image; specifically, it is "the time since the trigger pulse". The trigger mechanism has phase aligned the low-frequency channels. Unlike the NAP, this image is stationary. When the vowel resonances change, the formants move vertically in the image, and as long as the change is not too rapid the motion is smooth and the formant details are preserved. Thus, this auditory image is a better representation of what we hear in the sound than the NAP.

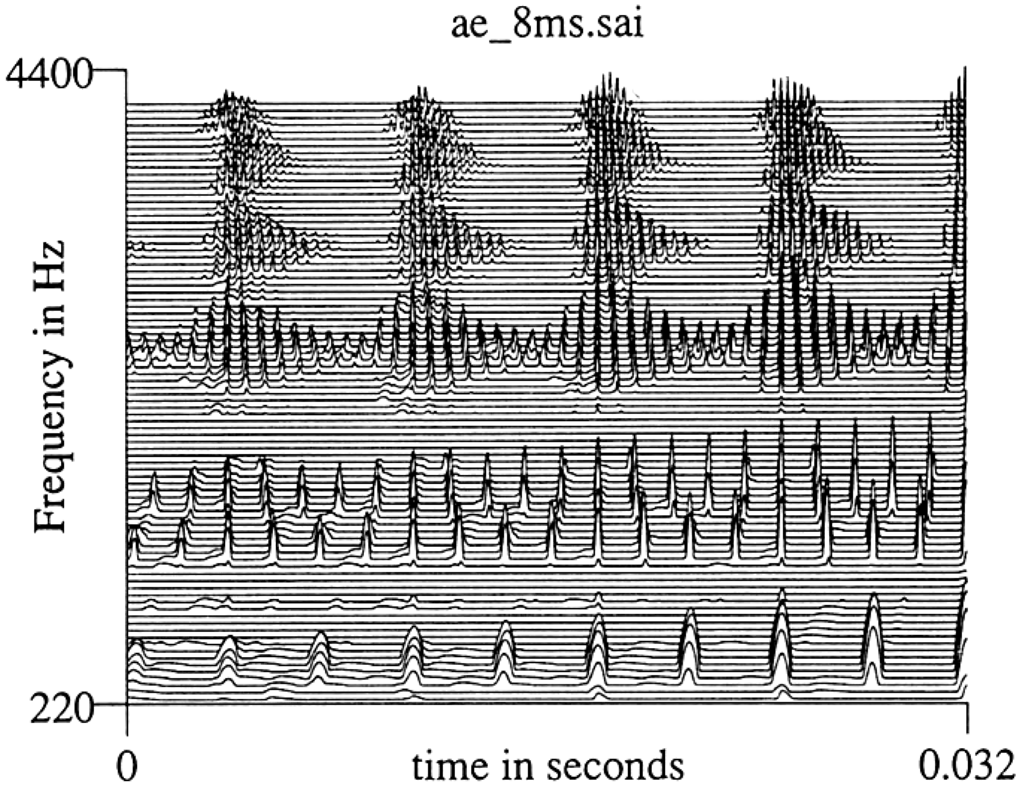

Figure 6. A simulation of the auditory image produced by the vowel /ae/ in 'past'. The auditory image is produced by quantised temporal integration of the neural activity pattern.

The motivation for developing a functional model of quantised TI is essentially the same as that for developing the cochlea simulation: to enable us to examine the results of auditory processing as we understand it. Nevertheless, the character of the modelling is quite different. The cochlea simulation attempts to provide a convenient summary of physiology as we currently know it; quantised TI attempts to provide links between peripheral physiology and human perception -- links where the physiology is either unknown or not sufficiently well understood to specify the central stages.

As a psychological model, quantised TI arises from the desire to explain certain aspects of auditory perception in humans. It is a functional model in the sense that it attempts to perform one of the transformations that we know is required in an auditory model; namely, the production of stabilised images from oscillating neural activity patterns. But there is a major distinction between the psychological, functional model represented by quantised temporal integration, and the functional model provided by the auditory filterbank. Whereas the auditory filterbank provides a convenient approximation to a process whose

 558

physiology is relatively well understood and whose site is not in doubt, quantised TI is a conjecture put forward to explain how sensations described by humans could be derived from a seemingly incompatible representation produced by a physiological model of an earlier stage: the cochlea.

Acknowledgements

The authors would like to thank Mike Allerhand for his assistance with the programming of the auditory image and its trigger mechanism. The work was supported by the MRC and two grants: MOD PE XTR/2239 and Esprit BRA 3207.

References

L.H. Carney and C.T. Yin (1988) 'Temporal coding of resonances by low-frequency auditory nerve fibers: Single fibre responses and a population model', J.Neurophysiology, 60, 1653-1677.

E. de Boer and H.R. de Jongh (1978) 'On cochlear encoding: potentialities and limitations of the reverse-correlation technique', J. Acoust. Soc. Am. 63, 115-135

B.R. Glasberg and B.C.J. Moore (1990) 'Derivation of auditory filter shapes from notched-noise data', Hearing Research, 47, 103-138.

D.D. Greenwood (1961) 'Critical bandwidth and the frequency coordinates of the basilar membrane', J. Acoust. Soc. Am. 33, 1344-1356.

J. Holdsworth (1990) 'Two-dimensional adaptive thresholding.' Annex 4 of APU AAM-HAP Report 1

J. Holdsworth, I. Nimmo-Smith, R.D. Patterson and P. Rice (1988) Annex C of 'Spiral Vos Final Report, Part A: The Auditory Filterbank,' APU report 2341.

R.D. Patterson (1987b) 'A pulse ribbon model of monaural phase perception', J. Acoust. Soc. Am. 82, 1560-1586.

R.D. Patterson and J. Holdsworth, J. (1990) 'An introduction to auditory sensation processing', Annex 1 of APU AAM-HAP Report 1

R.D. Patterson and B.C.J. Moore (1986) 'Auditory filters and excitation patterns as representations of frequency resolution', in B.C.J. Moore (ed), Frequency Selectivity in Hearing, pp.123-177. London: Academic Press Ltd.

R.D. Patterson, J. Holdsworth, I. Nimmo-Smith and P. Rice (1988) 'SVOS Final Report: The Auditory Filterbank', APU report 2341

D. Schofield (1985) 'Visualisations of speech based on a model of the peripheral auditory system', NPL Report DITC 62/85.

E. Zwicker (1961) Subdivision of the audible frequency range into critical bands (frequenzgruppen). J. Acoust. Soc. Am. 33, 248.