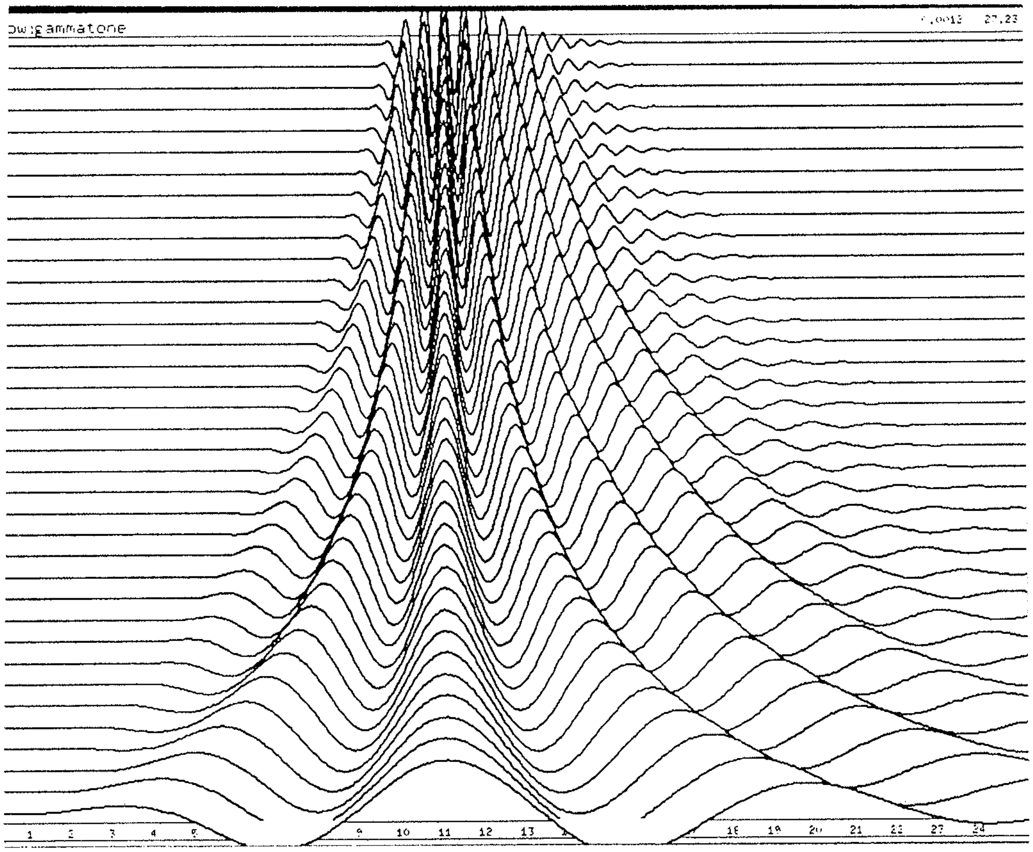
SVOS

Spiral
Voice Operated Switch



Phase I, Final Report, August 1988

The Auditory Filterbank

SPIRAL VOS FINAL REPORT

PART A

THE AUDITORY FILTERBANK

Executive Summary

The spiral voice operated switch (VOS) is essentially a multi-channel pitch extractor. The spiral VOS research programme has concentrated on the filterbank that creates the multi-channel representation of the sound and the spiral periodicity extractor that forms the basis of the spiral processor. In this paper we present a brief demonstration of the advantages of a multi-channel spiral VOS and then concentrate on the auditory filterbank research. Specifically, we describe the development of an auditory filterbank that meets the stringent requirements of the speech and hearing communities, and our subsequent development of a recursive version of this "gammatone" filterbank. Finally we assess the feasibility of producing a hardware version of the filterbank that operates in real-time, and conclude that a 32-channel version of the filterbank could be implemented on a single DSP chip, with the expectation that it would run in real-time at sampling rates up to 10 kHz.

The details of the research are described in three Annexes to this report: The first Annex describes our preliminary study of the feasibility of a hardware auditory filterbank and our survey of the forms of spectral analysis used in the hearing and speech communities. The second Annex provides an extended comparison of the gammatone filterbank and its predecessor, the roex filterbank, and the development of a gammatone filterbank tuned to human parameter values. The final Annex describes the procedure whereby one can implement a recursive version of the gammatone filterbank. The main body of the report provides an overview of the research programme with illustrations of the analysis technology and the output of the gammatone filterbank.

SPIRAL VOS FINAL REPORT

PART A

THE AUDITORY FILTERBANK

ROY PATTERSON
IAN NIMMO-SMITH

Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge
CB2 2EF

JOHN HOLDSWORTH
PETER RICE

Cambridge Electronic Design
Science Park
Milton Road
Cambridge

AUGUST 1988

Submitted to:

MR. A.W. HADDEN
Ministry Of Defense (PE) ARAD 13
St. Giles Court
St. Giles High Street
London WC2H 8LD

DR. G. ROOD
Mission Management 4c
Royal Aircraft Establishment
Farnborough
Hants

SPIRAL VOS FINAL REPORT - PART A - THE AUDITORY FILTERBANK

CONTENTS

ANNEX A - A PRELIMINARY STUDY OF THE FEASIBILITY OF A HARDWARE VERSION OF THE AUDITORY FILTERBANK

ANNEX B - AN EFFICIENT AUDITORY FILTERBANK BASED ON THE GAMMATONE FUNCTION

ANNEX C - IMPLEMENTING A GAMMATONE FILTERBANK

ABSTRACT

The spiral voice operated switch (VOS) is essentially a multi-channel pitch extractor. The spiral VOS research programme has concentrated on the filterbank that creates the multi-channel representation of the sound and the spiral periodicity extractor that forms the basis of the spiral processor. In this paper we present a brief demonstration of the advantages of a multi-channel spiral VOS and then concentrate on the auditory filterbank research. Specifically, we describe the development of an auditory filterbank that meets the stringent requirements of the speech and hearing communities, and our subsequent development of a recursive version of this "gammatone" filterbank. Finally we assess the feasibility of producing a hardware version of the filterbank that operates in real-time, and conclude that a 32-channel version of the filterbank could be implemented on a single DSP chip, with the expectation that it would run in real-time at sampling rates up to 10 kHz.

The details of the research are described in three Annexes to this report: The first Annex describes our preliminary study of the feasibility of a hardware auditory filterbank and our survey of the forms of spectral analysis used in the hearing and speech communities. The second Annex provides an extended comparison of the gammatone filterbank and its predecessor, the roex filterbank, and the development of a gammatone filterbank tuned to human parameter values. The final Annex describes the procedure whereby one can implement a recursive version of the gammatone filterbank. The main body of the report provides an overview of the research programme with illustrations of the analysis technology and the output of the gammatone filterbank.

INTRODUCTION

In September of 1986, ARAD 13 of MOD PE contracted Cambridge Electronic Design (CED) to determine whether the performance of the standard voice operated switch (VOS) could be improved by the inclusion of a pitch extractor -- in particular, the Spiral Pitch Processor developed by the Applied Psychology Unit of the Medical Research Council. In a traditional VOS, the criterion for turning the switch on is simply an increase in ambient energy above a specified level, independent of the form of the energy. In practice, when the criterion is set to a moderate level, transient noises cause false triggering at a rate that pilots report is unacceptably high. Consequently, the pilots typically set the trigger to a high level, with the result that when they wish to turn the switch on they find that they have to shout. Most of the energy in speech is contained in the vowels and so it will typically be a vowel that triggers the standard VOS. Vowels are quasi-periodic sounds which suggests that it should be possible to improve the performance of a VOS by restricting it to trigger only on periodic, or quasi-periodic sounds. Since the periodicity in question is the pitch of speech, an auditory pitch mechanism was chosen, namely the spiral pitch processor.

The spiral processor is one stage of an auditory model designed to a) simulate the operation of the human cochlea and b) transform the output of this physiological simulation into a representation that is more like the sensations humans hear. A block diagram of the model appears in Figure 1. The cochlea is simulated by two stages of processing: The first is an auditory filterbank which simulates the Spectral analysis performed by the basilar membrane. The second stage is a set of haircell simulators that convert the output of each channel into a stream of pulses intended to represent the neural firing pattern in that channel of the auditory nerve. The array of haircell outputs provides a detailed representation of the neural firing pattern flowing up the auditory nerve in response to a complex sound like speech. The pattern is referred to as a pulse ribbon for convenience. The focus of the pulse ribbon model is auditory neural processing which is performed in three stages as shown in the large dashed rectangle of Figure 1. The three stages perform transformations intended to capture the salient characteristics of human phase perception, pitch perception, and timbre perception, respectively. The spiral processor is the fourth stage and it determines whether the neural firing pattern is periodic. The research effort in the Spiral VOS Project is concentrated on two of the five stages; namely, the auditory filterbank, Stage 1, and the spiral processor, Stage 4.

A. Extracting Pitch in a Noise Environment

An assessment of the operation of a 27-channel version of the first 4 stages of the model is presented in Figure 2. In this case, both the signal and the background noise are wideband stimuli with energy in the region 100-4,000 Hz. The signal is a pulse train presented in the temporal centre of a burst of noise. There are four conditions in the experiment represented by the four rows of the figure. The signal power is held constant across the four conditions. The signal-to-noise ratio was varied from +12 to -6 dB by progressively increasing the noise level. The lefthand column of sub-figures shows the waveforms for the four conditions; as the level of the background noise rises it obscures the signal. The signal-to-noise ratio is indicated by the first number in the window identifier. The righthand column of sub-figures shows the strength of the pitch estimate produced by the spiral processor as a function of time, in response to the stimulus in the lefthand column. The pitch value (125 Hz) is correctly detected by the spiral processor for all but the lowest signal-to-noise ratio. A more traditional pitch extractor performed well at 12 dB signal-to-noise ratio but failed at 6 dB signal-to-noise ratio. A detailed review of the spiral processor is presented in a companion report entitled "Spiral VOS Final Report: Part B, the Pitch Extractor". The purpose of the current report, Part A, is to present the research done on the auditory filterbank in Stage 1 of the model.

B. The Advantage of a Multi-Channel Pitch Extractor

The value of the filterbank is illustrated by two extensions of the analysis outlined above. The noise spectrum in a helicopter is far from flat; there is a great deal of noise in the frequency region below 1,000 Hz, but comparatively little in the region above 1,000 Hz. In the first extension, we approximated the conditions in a helicopter by lowpass filtering the background noise used to test the spiral processor. Then we reran the four conditions presented in Figure 2. The results are presented in Figure 3: The stimulus waveforms in the lefthand column of subfigures are similar to those in Figure 2 in the sense that the signal becomes less clear as the noise level rises. The traditional, single-channel pitch extractor, which operates on the raw waveform, shows only a marginal improvement in performance as would be expected. The pitch-strength contours in the righthand column of Figure 3 show that the performance of the spiral processor improved in all four conditions, and when the signal-to-noise ratio is -6 dB performance is virtually as good as in wideband noise. In the second extension, we showed that the spiral processor is not limited to detecting pitch in high-frequency channels (that is channels above 1,000 Hz). Figure 4 shows the results of the same performance test when the noise is highpass filtered at 1,000 Hz rather than lowpass filtered at 1,000 Hz. Once again, performance improves in all four conditions, and performance is well above threshold when the signal-to-noise ratio is -6 dB.

This simple demonstration shows the advantage of a multi-channel system based on an auditory filterbank. In short, since the spiral processor detects periodicity on an individual channel basis, and since it restricts the calculation of pitch to those channels with positive periodicity readings, it can effectively combine the periodicity information from any channels that have a reasonable signal-to-noise ratio, and at the same time reject those channels that have a poor signal-to-noise ratio.

C. Optimising the Filterbank

The roex filterbank that was originally used to demonstrate the feasibility of a multi-channel spiral processor is restricted to stationary sounds. Furthermore, the filtering algorithm is far too slow to support a real-time Spiral VOS. As a result we initiated a two-part research programme to develop a dynamic filterbank and to find a filtering algorithm that was more efficient. This report describes the research effort and the resulting 'gammatone' filterbank. One channel of this filterbank operates in near real-time on a Micro VAX II computer when the input is speech digitised at a 10 kHz sampling rate. Current estimates indicate that a high performance DSP chip like the TMS320C-30 could support a real-time, 16-channel auditory filterbank. This should be sufficient for a Spiral VOS operating in a helicopter environment and so we conclude that a one-chip, real-time filterbank is now feasible.

I DEVELOPMENT OF A TIME-DOMAIN AUDITORY FILTERBANK

Physiologists, psychoacousticians, and speech scientists all agree that the initial stage of auditory processing is a spectral analysis that can be simulated with reasonable accuracy using a bank of linear bandpass filters, but they use widely differing forms of spectral analysis. In July of 1986, a small meeting was convened at the Applied Psychology Unit in Cambridge to determine whether the three fields agreed on the general characteristics of the filterbank as it pertains human auditory filtering. The physiologists were represented by Professor Evans of the Department of Communication and Neuroscience at Keele University, the psychoacousticians by Dr. Patterson of the Applied Psychology Unit of the Medical Research Council, and the speech community by Dr. Moore from the Speech Research Unit (SRU) of the Royal Signals and Radar Establishment (RSRE). The meeting concluded that it might actually be possible to get agreement, not only on the general characteristics of the filterbank, but also on a set of parameter values that could be used as a basis for an 'informal standard' filterbank. A standardised filterbank would enable us to establish a common representation of the frequency selectivity available to the higher centres of the auditory system. The meeting agreed, that for purposes of studying auditory perception and speech perception, the best representation of the amplitude characteristic of the auditory filtering process was provided by the roex filter shape of Patterson and Nimmo-Smith (1982), and that the best specification of the parameter values of the filterbank was provided by Patterson and Moore (1986).

The meeting was also attended by representatives of the Institute of Sound and Vibration Research (ISVR). Subsequently, at the behest of RAE Farnborough, the ISVR and APU submitted a proposal to MOD PE to determine the feasibility of producing a real-time hardware version of the auditory filterbank to be used in analysing the auditory environments in helicopters.

The Initial Time-Domain Roex Filterbank

The first problem was to demonstrate that a time-domain version of the roex filterbank could be produced, and to examine the output of such a filterbank. The basic problem is that the psychophysical technique used to derive the roex filter shape reveals the amplitude characteristic of the filter but not the phase characteristic, and so it was not possible to specify the impulse response of the filter uniquely. This is a fairly common engineering problem and several techniques exist for deriving the impulse response from the amplitude characteristic by making some general assumptions about the phase characteristic. In this particular case the ISVR used an FIR technique to derive the impulse response of the auditory filter from its amplitude characteristic by assuming that the phase characteristic was linear in the frequency region of the filter's passband. The result was the roex linear-phase filterbank tuned to human auditory parameter values. The filterbank was integrated into a sound editing programme on a Micro VAX II computer and used to produce illustrations of the filterbank operating on a selection of sounds including auditory warnings and speech waveforms. Examples are shown in Figures 5 and 6.

The results were discussed with speech groups including the Institute of Hearing Research in Nottingham. They suggested holding a second filterbank meeting in an attempt to establish if there was interest in the speech community at large for a standardised auditory filterbank. In preparation for the meeting, APU prepared an interim report of the work done to that point on the 'hardware filterbank' project. A revised version of the interim report is presented as Annex. A. It has two main sections: The first outlines the basic filterbank concepts and suggests parameter values for an initial, time-domain, auditory filterbank for use in helicopters (see Table 1). The second section reviews the various spectral analysis systems used in speech and hearing research. It is concluded that a generalised version of the initial filterbank might provide the basis for a 'standard' filterbank for hearing and speech research (see Table 2).

The 'standard filterbank' meeting was organised by IHR and held in Nottingham in February of 1987, where a number of scientists presented their approach to the problem of specifying a practical auditory filterbank. Dr. Patterson presented the roex linear-phase filterbank developed for RAE by ISVR and APU. Dr. Darwin from Sussex University presented a roex minimum-phase filterbank developed at IHR by Dr. Assmann. The minimum-phase filterbank has a more realistic impulse response than the linear-phase filter characteristic. Dr. Cooke from the Engineering Department at Sheffield University presented a transmission line filterbank similar to that proposed by Lyon (1982). The meeting concluded that the roex filter shape combined with the parameter values provided by Patterson and Moore (1986) would provide a useful 'standard' filterbank provided that a suitable phase characteristic, and thus an appropriate impulse response, could be established.

As a result of the meeting, the APU in conjunction with ISVR began to study the implications of implementing a roex, minimum-phase filterbank like that suggested by IHR on the Micro VAX II computer. At the same time, APU and CED began to follow up Schofield's (1985) observation of the similarity between the amplitude characteristics of the revcor filter used in physiology and the roex filters used in psychophysics. It soon became clear that the revcor approach was preferable, whenever one can assume that the filter is roughly symmetric on a linear frequency scale.

II THE GAMMATONE AUDITORY FILTERBANK

In the summer and autumn of 1987, John Holdsworth of CED programmed a gammatone auditory filterbank on the Micro VAX II and integrated it with the sound editor Camsed. This enabled us to compare the amplitude characteristic, or shape, of the gammatone filter with that of the roex filter across the frequency range of speech. This is the same roex filter as that employed in the spectral filterbank programmed by ISVR for RAE Farnborough -- a filter shape which is known to predict auditory masking in helicopters with a high degree of accuracy (Lower et al, 1986). The gammatone filter shape, with order 4, was found to provide an extraordinarily close approximation to the roex filter shape, from which we can conclude that the gammatone filter will predict masking with the same accuracy as the roex filter. A comparison for three filters with centre frequencies near 500, 1000 and 2000 Hz is shown in Figure 7. Patterson & Moore (1986) have shown that the roex filter can predict masking in a wide range of situations so long as the masker is a stationary sound. The roex is limited to stationary sounds because of our lack of information concerning the phase characteristic of the filter shape. The gammatone filter is derived as an impulse response, and so it has a complete phase characteristic as well as an amplitude characteristic. Although the data were gathered in experiments on small mammals, it seems reasonable to extrapolate to humans as it only involves a scaling of the bandwidths, which would appear to be better than assuming a linear-phase characteristic as previously. As a result, APU and CED pursued the gammatone filter option, while ISVR pursued the minimum-phase option.

During the course of the research on the gammatone filter, John Holdsworth, discovered a recursive filter algorithm for calculating the filter output. It was clear that the recursive filter would be much more efficient than its FIR equivalent and so the recursive version was implemented and refined. It is this filter which produces the performance described at the end of the Introduction to this paper; namely, that a single filter running on a Micro Vax II, or a SUN work station, operates in near-real time on a speech signal digitised at 10 kHz.

The Recursive Gammatone Filterbank

The discovery of the recursive gammatone filterbank proved to be a significant breakthrough. It brought us to a position in advance of that which the original group of scientists meeting in July of 1986 had thought could be achieved. For here, was a filterbank that met the main requirements of not only the psychoacousticians, but also the physiologists and the speech scientists. Specifically, the gammatone filter was acceptable to physiologists as a function to represent cochlea filtering, inasmuch as they actually discovered the function and fitted it to physiological data. The same filter shape in the form of the roex filter, was known to be able to predict a wide variety of human auditory masking data and so it was acceptable to psychophysicists. Finally, it was also acceptable to speech scientists because, in its recursive form, it was almost as fast as the optimised filters that they were then using as frontend processors for speech recognition.

In December of 1987, SRU convened another informal meeting of speech and hearing groups interested in establishing a 'standard' auditory filterbank. The meeting was held at RSRE and in preparation for the meeting, APU prepared a written paper describing our research on the gammatone filterbank and its performance. The paper was presented by Roy Patterson on the first day of the meeting, and it formed the basis of a discussion session on the second day. A revised version of the paper is presented as Annexe B of this report.

The Introduction to the paper outlines the choices involved in choosing a phase characteristic for the auditory filter and it presents the alternative filterbanks that were available to us at the time. It also illustrates the advantage of phase compensation, that is, shifting the channel outputs in time to compensate for the longer phase lags occurring in the narrow, low-frequency channels.

The second section introduces the gammatone filter function and sets out the scientific basis for choosing the gammatone. Specifically, it presents a comparison of the roex and gammatone amplitude spectra. It shows that the fourth order gammatone filter provides the best approximation to the roex(p) filter -- that is the filter shape used in the ISVR programme for predicting auditory masking. It is important to note that in so doing we are fitting a function that is used to approximate physiological impulse responses (the gammatone) to a function (the roex) which is used by psychologists to approximate auditory filter data. In point of fact, we know that the 'true' auditory filter has somewhat shallower tails than the roex(p) filter outside the passband, and that the roex(p,w,t) filter provides a better approximation to the human auditory filter. Accordingly, the paper goes on to compare gammatones of different order to the more complicated roex filter, and it is shown that a second order gammatone provides an even better approximation to the roex(p,w,t) filter. That is, a gammatone filter with fewer stages, and requiring less computation, actually provides a better fit to the amplitude characteristic of the auditory filter!

The final part of the second section compares several methods of phase compensation and discusses the motivation for the different forms. There are two main arguments for phase compensation: firstly, it appears that the auditory system knows its own phase characteristic; secondly phase compensation tends to rearrange the channels in such a way as to bring together in time, those sections of the filter output associated with a particular instant in the input wave. Examples of filterbank output for the vowel in 'mat' are shown with and without phase compensation in Figures 8 and 9, respectively.

The third and final section is concerned with the recursive version of the gammatone filterbank. It begins with an analysis of the computational load implied by the speech community's desire for filterbanks with as many as 128 channels, operating at sampling frequencies up to 25 kHz, with FIR filters involving as many as 256 taps. The analysis shows that this kind of device would require on the order of 800 million operations per second which is simply not feasible in the foreseeable future even with the fastest DSP chips. A summary of the analysis is presented in Figure 10. The speed of the recursive gammatone is compared with that of an FIR gammatone from which it is concluded that the recursive gammatone is roughly equivalent to an FIR filter with between 12 and 16 coefficients. The paper concludes that a hardware, recursive gammatone filterbank with 32 channels is feasible and that it might be expected to run at sample rates up to 20 kHz on one of the new floating-point DSP chips.

III DOCUMENTATION AND DISTRIBUTION OF A GAMMATONE FILTERBANK PROTOTYPE

There was considerable discussion at the RSRE meeting concerning the theoretical and practical advantages of the gammatone filterbank, and following the discussion, a number of groups expressed interest in acquiring a software version of the filterbank that they could use for research purposes in their own laboratories. In an effort to support this interest, and in preparation for publication of our research, we prepared a document on the implementation of the gammatone filterbank. We also prepared some portable computer modules for calculating and applying the gammatone filterbank. The implementation document appears as Appendix C of this report.

The first section describes the gammatone filter in the time-domain. The second section describes the gammatone filter in the frequency domain, and reviews the argument that one of the terms in the frequency domain expression is negligible -- an approximation that is necessary for the derivation of the recursive gammatone filter. The third section derives the equivalent rectangular bandwidth of the gammatone filter and shows how to match filters of any order to the equivalent rectangular bandwidth suggested for humans in Patterson & Moore (1986). The fourth section explains our method of phase compensation. The fifth and final section describes the digital implementation of the gammatone filterbank. Briefly, the recursive gammatone is a cascade of frequency-shifted lowpass filters. The section describes our method of frequency shifting and the computation of the lowpass filter.

At the same time, we prepared the computer modules for calculating and applying a gammatone filterbank. There are two primary modules: the first sets up the general characteristics of the filter such as its order and the form of phase compensation; the second is used at run time to specify the lower and upper limits of the filterbank and the filter density, that is, the number of filters per ERB. At the time of writing, the filter modules had been successfully transferred to eight speech and hearing groups.

IV AUDITORY PRE-PROCESSING AND RECOGNITION OF SPEECH

By far the largest "market" for auditory research tools is not the hearing community, but rather the speech community. Until recently much of the speech community was content to use Fourier analysis or LPC analysis as a substitute for auditory analysis. Speech recognition machines have not made nearly as much progress as the speech community had anticipated, and many speech scientists now feel that they need to improve the resolution of their frontend processors. A portion of the community argue the best way to do this is to implement a full auditory model as a pre-processor for speech recognition. In an attempt to prompt collaboration in this area, the APU have prepared a chapter for a European volume on Cognitive Science Research Directions. In the chapter, Patterson and Cutler (1988) describe the advantages of an auditory cognitive approach to speech recognition. The most important sections for current purposes are the Introduction and Section I on Auditory Pre-Processing. In the Introduction, we compare the spectrographic representation of four vowels with the cochleogram representation, that is, the output of a multi-channel filterbank. In short, we argue that the traditional spectrogram simply does not have sufficient resolution to show the shapes of formants as they exist in the auditory system. We also propose a new basis for feature extraction in which the frames of the analysis are not determined by the Fourier transform, but rather by the pitch periods in the stimulus.

The first section of the main text deals with auditory pre-processing. It comprises the first published description of the gammatone filterbank and shows how it can be combined with the haircell model of Meddis (1986) to produce a full cochlea simulation. The rest of the section then describes the three stages of neural auditory processing that we believe are required to prepare the initial auditory image, that is, the initial sensation that a sound produces. This is a stabilised image which changes only when we hear a change in the sound, not when the waveform changes. It is this image that we feel is the product of the peripheral auditory system and the input to the speech system. The remaining sections of the paper outline the speech recognition problem from the cognitive psychological point of view and suggest that perhaps the best way to proceed currently is to develop a three stage connectionist model of the recognition process to follow the auditory pre-processor. A schematic representation of the auditory/connectionist model is presented in Figure 11.

The publication of the paper comprised the final work associated with the current project.

ACKNOWLEDGEMENTS

This work has been carried out with the support of Procurement Executive, Ministry of Defence.

REFERENCES

Lower, M.C., Patterson, R.D., Rood, G., Edworthy, J., Shailer, M.J., Milroy, R., Chillery, J., & Wheeler, P.D. (1986). The design and production of auditory warnings for helicopters 1: the Sea King. Institute of Sound and Vibration Research Report AC527A.

Lyon, R.F. (1982). A computational model of filtering, detection, and compression in the cochlea. Proceedings, IEEE ICASSP, Paris, 1282-1285.

Meddis, R. (1986). Simulation of mechanical to neural transduction in the auditory receptor. Journal of the Acoustical Society of America, 79, 702-711.

Patterson R.D., & Cutler A. (1988). Auditory preprocessing and recognition of speech. In A.D. Baddeley and N.O. Bernsen (Eds.) Research Directions in Cognitive Science, Vol. 1, Cognitive Psychology. London: Erlbaum (in press).

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

Patterson R.D., & Nimmo-Smith, I. (1986). Thinning periodicity detectors for modulated pulse streams. In B.C.J. Moore & R.D. Patterson (Eds.) Auditory Frequency Selectivity. Plenum: New York, 299-307.

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

FIGURE LEGENDS

Figure 1. A schematic representation of peripheral auditory processing as represented in the pulse ribbon model of hearing. The filterbank (32) together with the bank of pulse stream generators (33) represents the processing performed by the human cochlea. The processing modules shown in the large dashed rectangle in the centre of the figure (34-40) represent the auditory neural processing that takes place prior to the formation of the main auditory image. (Figure reprinted from UK Patent Application No. 8531871).

Figure 2. The performance of the multi-channel spiral processor operating on a broadband periodic signal embedded in broadband background noise. The signal level is fixed. The different rows of the figure shows the results of the analysis as the background noise rises from -12 to +6 dB as shown by the first number in parentheses in the window label. In each case, the subfigure in the lefthand column shows the stimulus, and the subfigure in the righthand column shows the strength of the pitch estimate as a function of time. As the background noise rises the strength of the pitch estimate decreases.

Figure 3. The performance of the multi-channel spiral processor operating on a broadband periodic signal embedded in lowpass background noise. In each case, the subfigure in the lefthand column shows the stimulus, and the subfigure in the righthand column shows the strength of the pitch estimate as a function of time. The signal level is fixed. As the background noise rises the strength of the pitch estimate decreases; but performance remains well above that in the broadband case even at the highest noise level.

Figure 4. The performance of the multi-channel spiral processor operating on a broadband periodic signal embedded in highpass background noise. In each case, the subfigure in the lefthand column shows the stimulus, and the subfigure in the righthand column shows the strength of the pitch estimate as a function of time. The signal level is fixed. As the background noise rises the strength of the pitch estimate decreases; but performance remains well above that in the broadband case even at the highest noise level.

Figure 5. The output of a 379-channel gammatone filterbank operating on an experimental auditory warning. The range of the abscissa is 50 ms; the filter centre frequencies range from 100 to 5,000 Hz. The stimulus was created by summing about 10 discrete frequency components all of which had fixed amplitudes and phases. The frequency spacing was essentially random and so, as the figure shows, the instantaneous frequency of the components of the filterbank output varies as a function of time.

Figure 6. The waveform of the word "lot" (top panel) and the output of a 379-channel gammatone filterbank in response to the word. The range of the abscissa is 300 ms. The filter centre frequencies range from 100 to 5,000 Hz. The dark areas in the figure show the tracks of the formants as the word proceeds.

Figure 7. The amplitude characteristics of three roex(p) filters centred at 0.43, 1.00 and 2.09 kHz. The lower and upper filters are centred 6 ERBs below and above the 1 kHz filter respectively. In each case, the range of the abscissa extends from an octave below to an octave above the centre frequency of the filter, on a linear frequency scale. The range of the ordinate is 40 dB.

Figure 8. A cochleogram of four cycles of the [ae] in "past" produced by a gammatone filterbank without phase compensation. The triangular objects are the upper three formants of the vowel. The duration of each period is 8 ms. The ordinate is filter centre frequency on an ERB scale. The centre frequencies range from 100 to 4,000 Hz, and the 1,000-Hz filter occurs about half way up the figure. Note the strong rightward skew induced by the phase lags of the low-frequency filters in the lower half of the figure.

Figure 9. A cochleogram of four cycles of the [ae] in "past" produced by a gammatone filterbank with phase compensation. The coordinates are the same as for Figure 8. Note that the strong rightward skew produced by the phase lags of the low-frequency filters has now been removed.

Figure 10. The computer speed required to support a real-time auditory filter bank based on FIR filters and digital convolution. The figure shows that as the number of channels rises from 8 to 128 (the ordinate), and as the number of filter coefficients increases from 32 to 256 (the abscissa), the number of Mops increases from 2.5 to 320. If the sampling rate is increased from 10 kHz to 25 kHz (depth), the Mop rate rises from 320 to 800.

Figure 11. A comparison of existing (upper row) and proposed (lower row) methods of word recognition using an auditory/connectionist approach. The spectrogram in the upper row is replaced by a full cochlea simulation and a pulse ribbon model of auditory neural processing in the lower row. The monolithic connectionist model in the upper row is replaced by a psychological, staged model in the lower row, wherein features are extracted from the auditory image and converted into a sublexical form of phonology before the phonology is assembled into word candidates. (Figure reprinted from Patterson & Cutler, 1988).

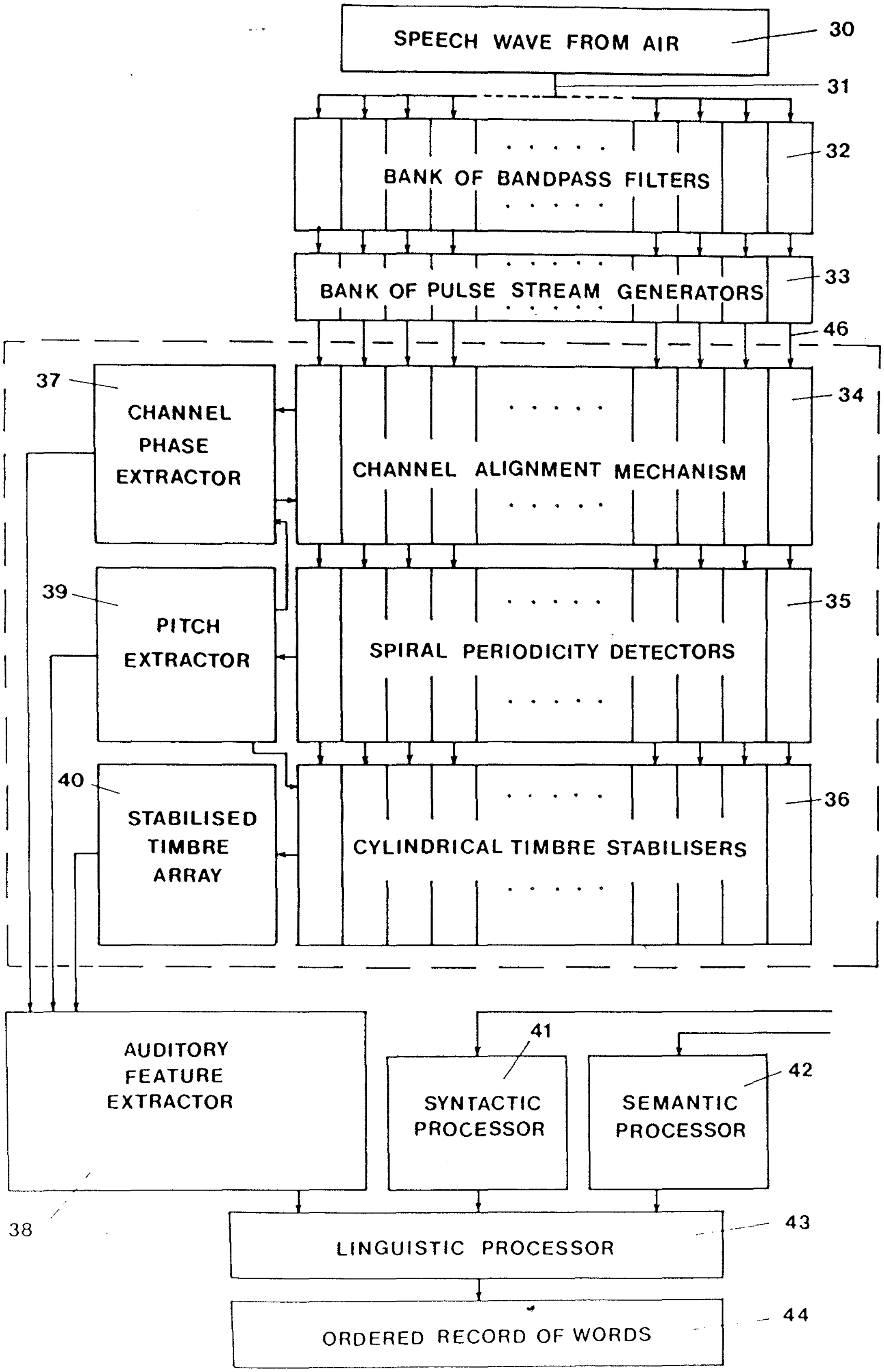

SPEECH WAVE FROM AIR
30
31
BANK OF BANDPASS FILTERS
32
BANK OF PULSE STREAM GENERATORS
33
46
37
CHANNEL PHASE EXTRACTOR
CHANNEL ALIGNMENT MECHANISM
34
39
PITCH EXTRACTOR
SPIRAL PERIODICITY DETECTORS
35
40
STABILISED TIMBRE ARRAY
CYLINDRICAL TIMBRE STABILISERS
36
AUDITORY FEATURE EXTRACTOR
38
41
SYNTACTIC PROCESSOR
SEMANTIC PROCESSOR
42
LINGUISTIC PROCESSOR
43
ORDERED RECORD OF WORDS
44

Figure 1

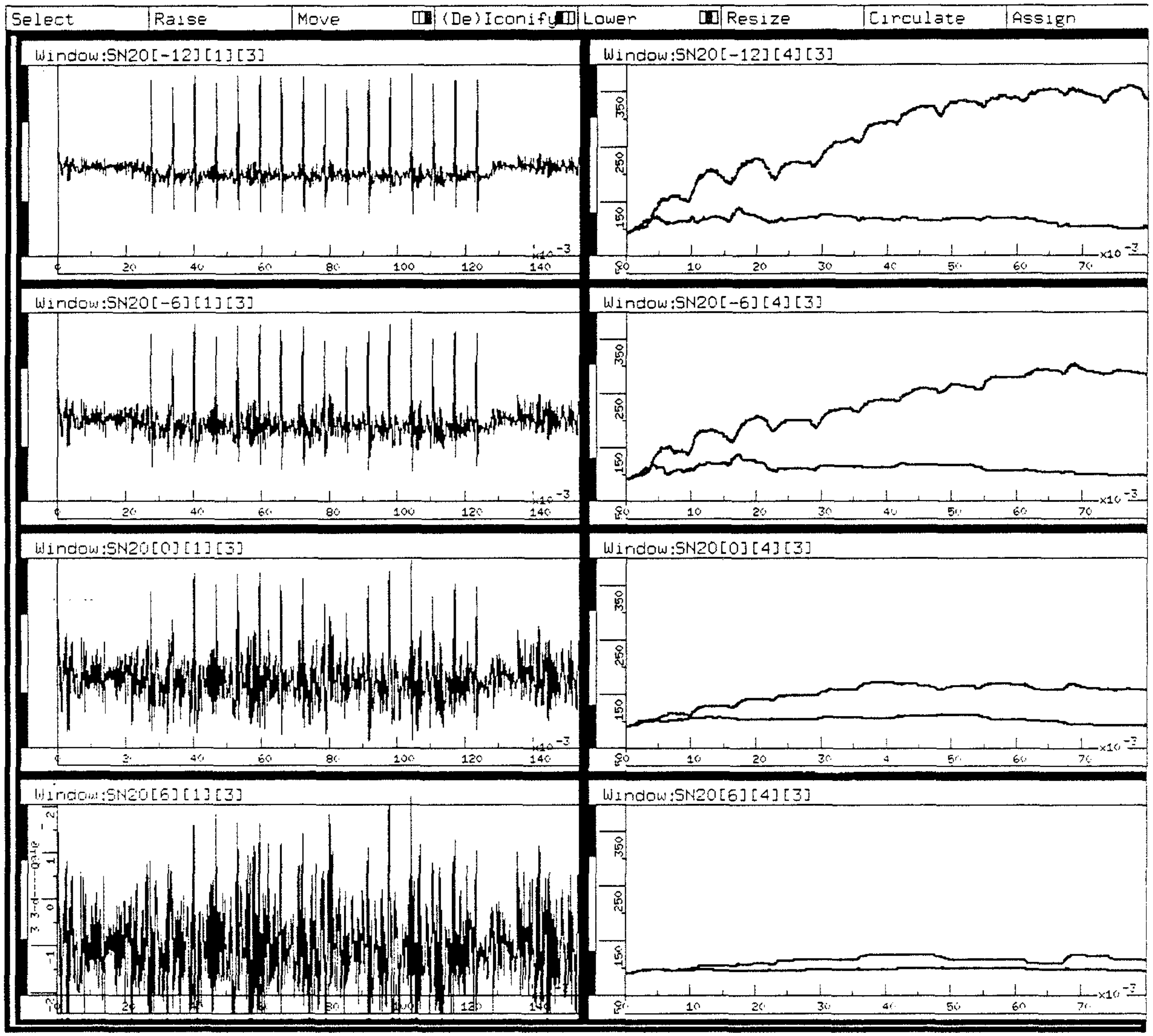

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:SN20[-12][1][3]
Window:SN20[-12][4][3]
Window:SN20[-6][1][3]
Window:SN20[-6][4][3]
Window:SN20[0][1][3]
Window:SN20[0][4][3]
Window:SN20[6][1][3]
Window:SN20[6][4][3]
0 20 40 60 80 100 120 140
x10 -3
50 10 20 30 40 50 60 70
x10 -3
150 250 350
-2 -1 0 1 2

Figure 2

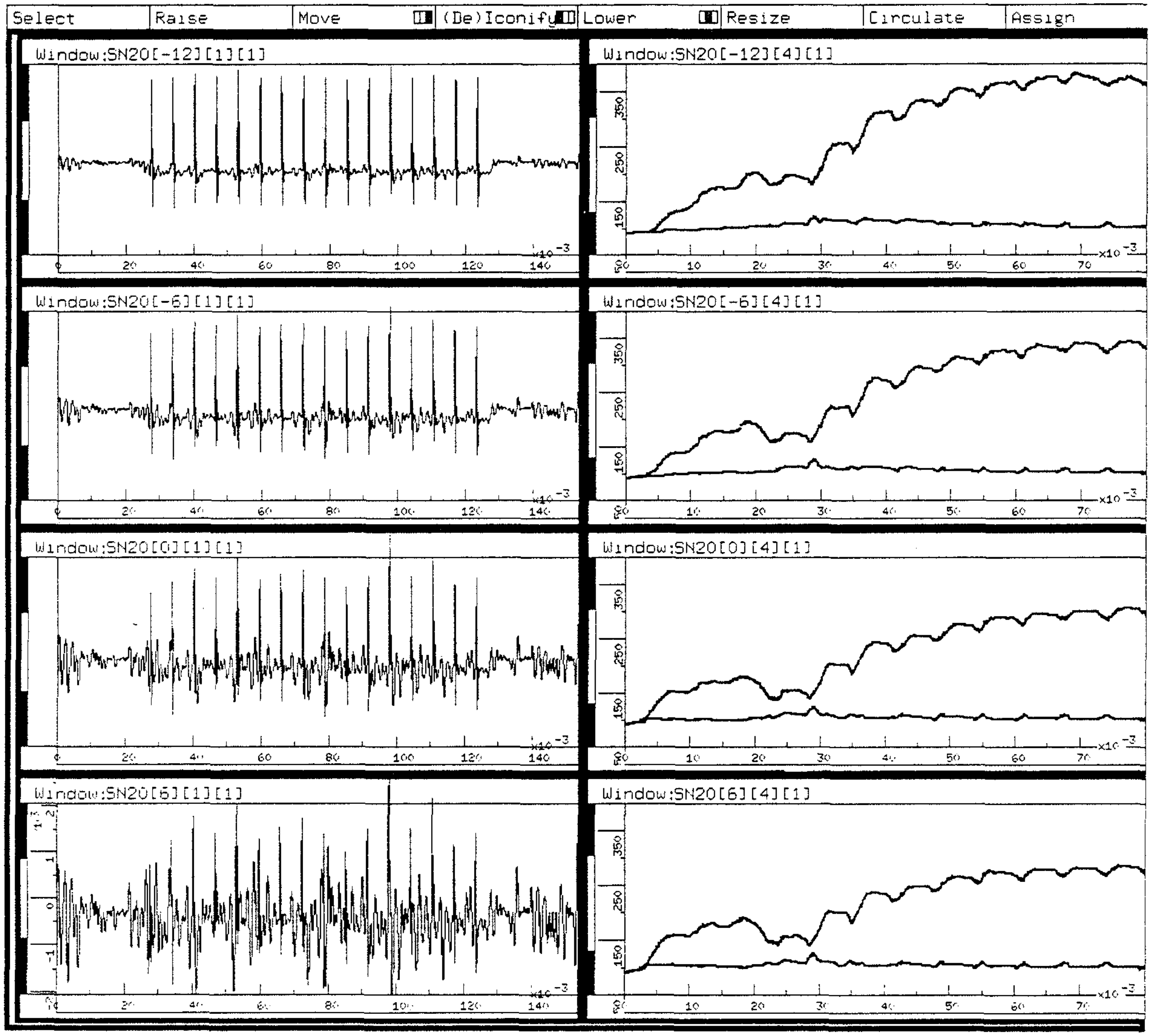

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:SN20[-12][1][1]
20 40 60 80 100 120 140 x10^-3
Window:SN20[-12][4][1]
150 250 350
0 10 20 30 40 50 60 70 x10^-3
Window:SN20[-6][1][1]
20 40 60 80 100 120 140 x10^-3
Window:SN20[-6][4][1]
150 250 350
0 10 20 30 40 50 60 70 x10^-3
Window:SN20[0][1][1]
20 40 60 80 100 120 140 x10^-3
Window:SN20[0][4][1]
150 250 350
0 10 20 30 40 50 60 70 x10^-3
Window:SN20[6][1][1]
-1 0 1 2
20 40 60 80 100 120 140 x10^-3
Window:SN20[6][4][1]
150 250 350
0 10 20 30 40 50 60 70 x10^-3

Figure 3

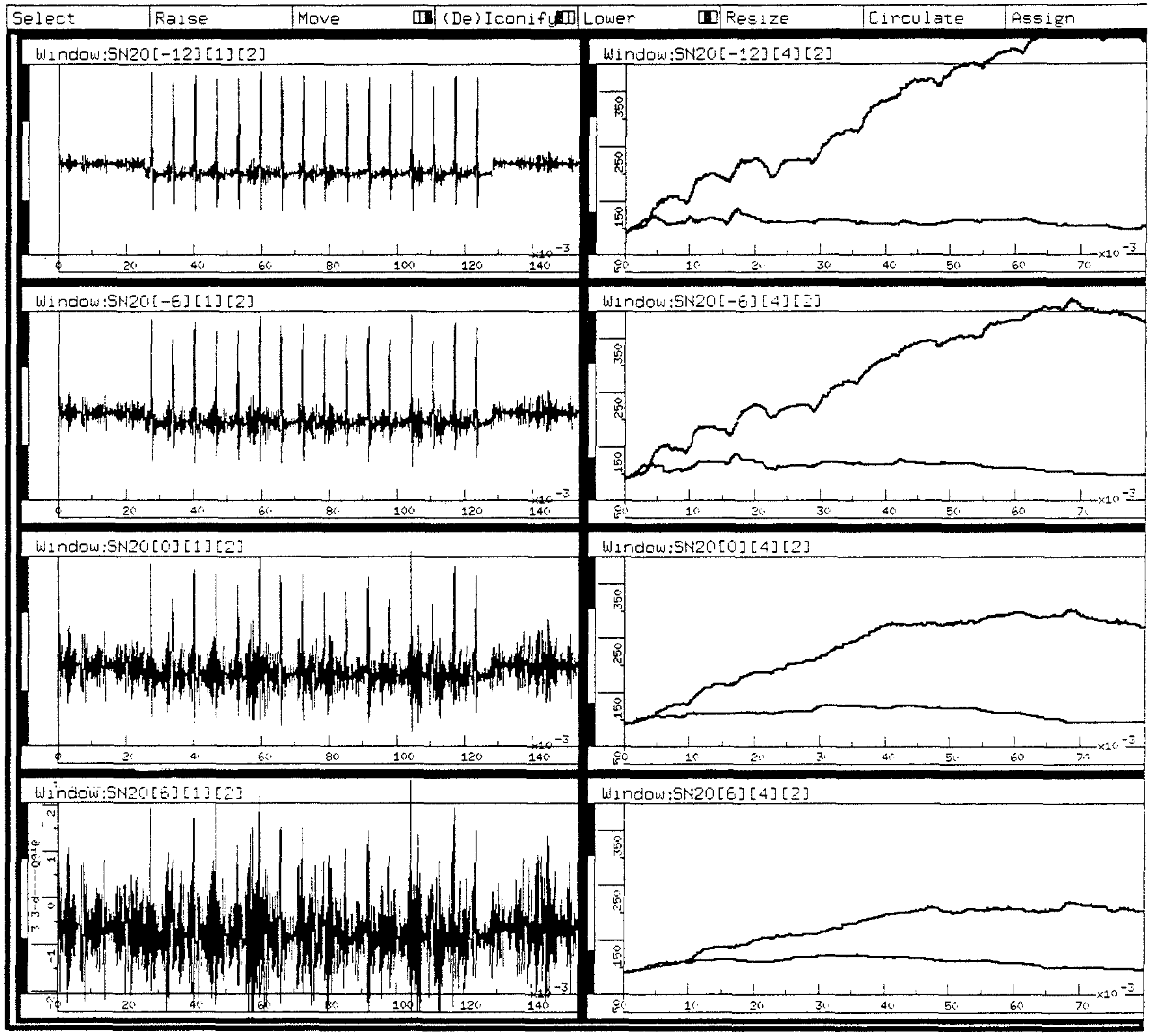
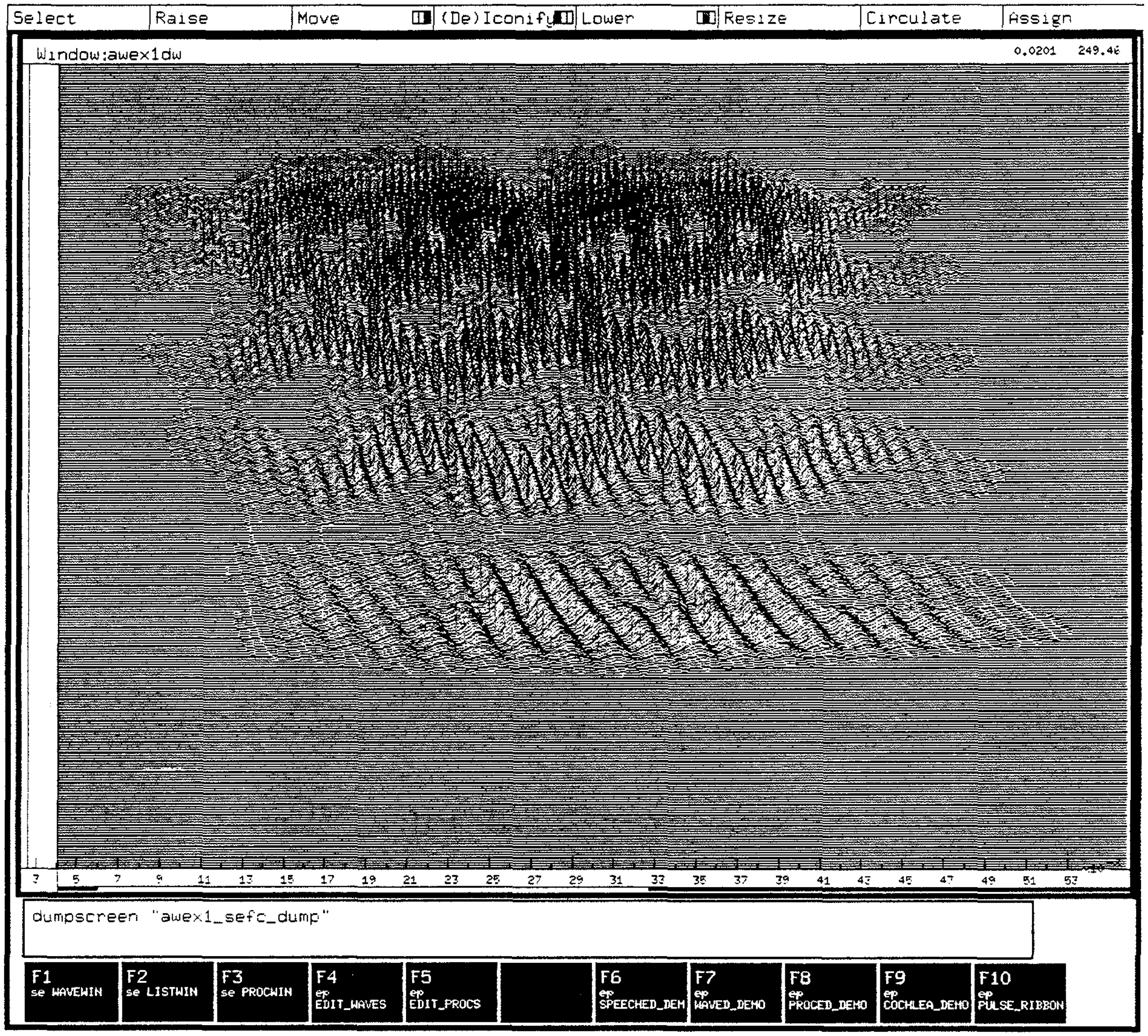

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:SN20[-12][1][2]
Window:SN20[-12][4][2]
Window:SN20[-6][1][2]
Window:SN20[-6][4][2]
Window:SN20[0][1][2]
Window:SN20[0][4][2]
Window:SN20[6][1][2]
Window:SN20[6][4][2]
0
20
40
60
80
100
120
140
x10^-3
50
150
250
350
10
20
30
40
50
60
70
x10^-3
-2
-1
0
1
2
x10^3

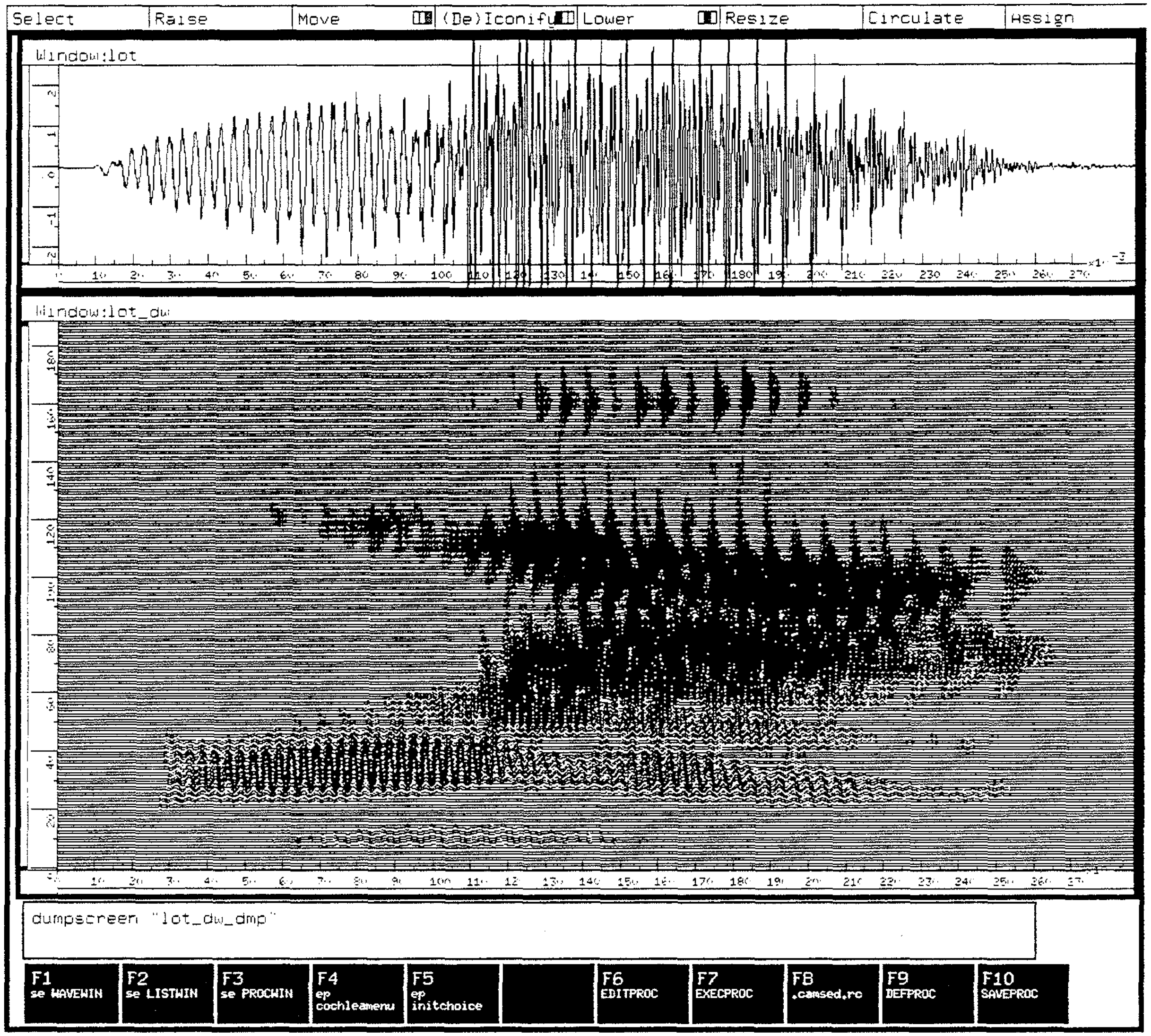
Figure 4



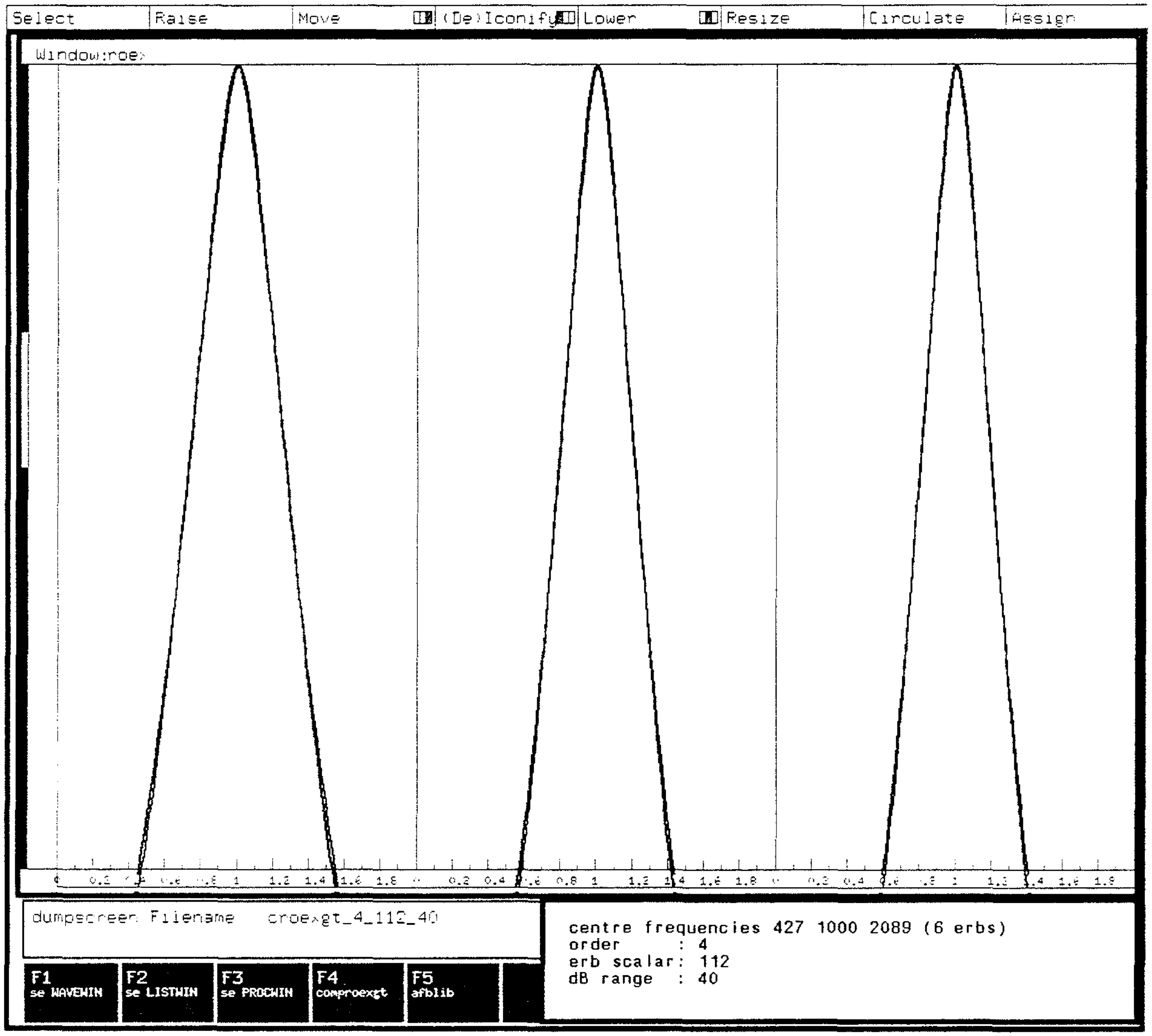
```
dumpscreen "awex1_sefc_dump"
```

Figure 5

Figure 6



```
dumpscreen "lot_dw_dmp"
```



```
dumpscreen Filename   croexgt_4_112_40
```

```
centre frequencies 427 1000 2089 (6 erbs)
order      : 4
erb scalar: 112
dB range   : 40
```

Figure 7

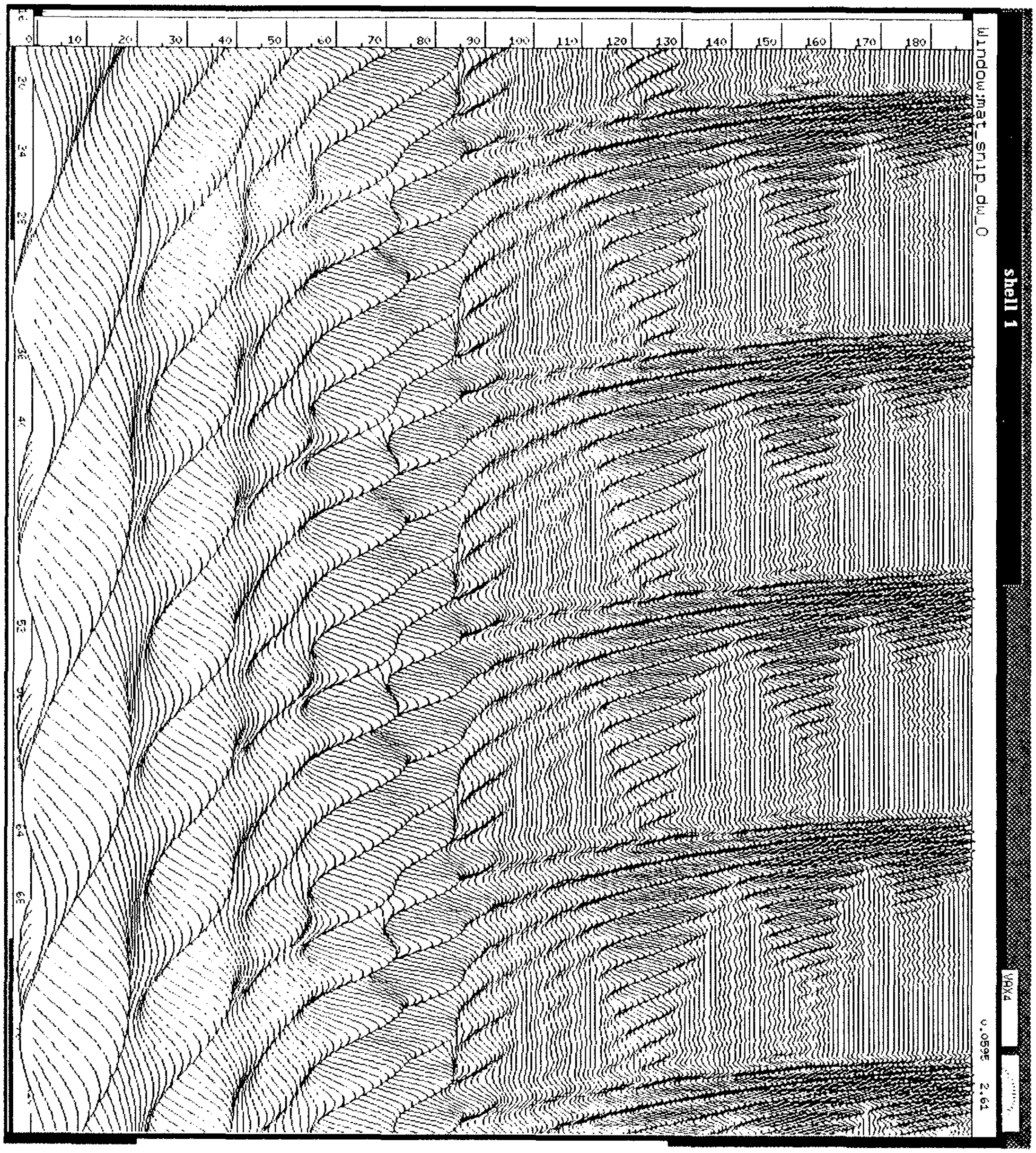

shell 1
Window:mat_snip_dw_0
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
VAX4
0.0595
2.61

Figure 8

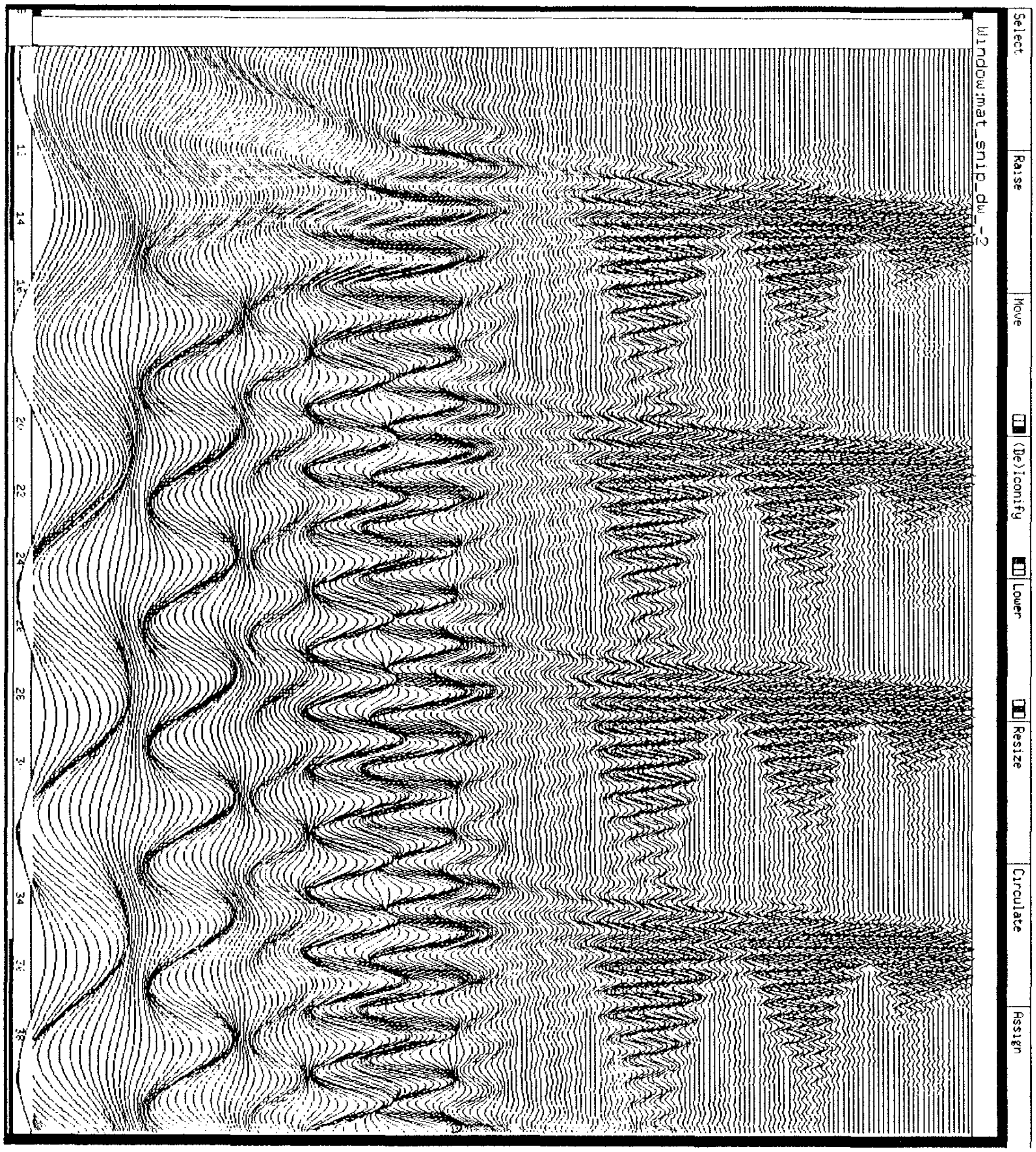

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:mat_snip_dw_2
12
14
16
20
22
24
26
28
30
34
36
38

Figure 9

F.I.R. filter processing speed

Length of filter	32 – 256 taps
Number of channels	8 – 128
Sampling rate	10kHz – 25kHz

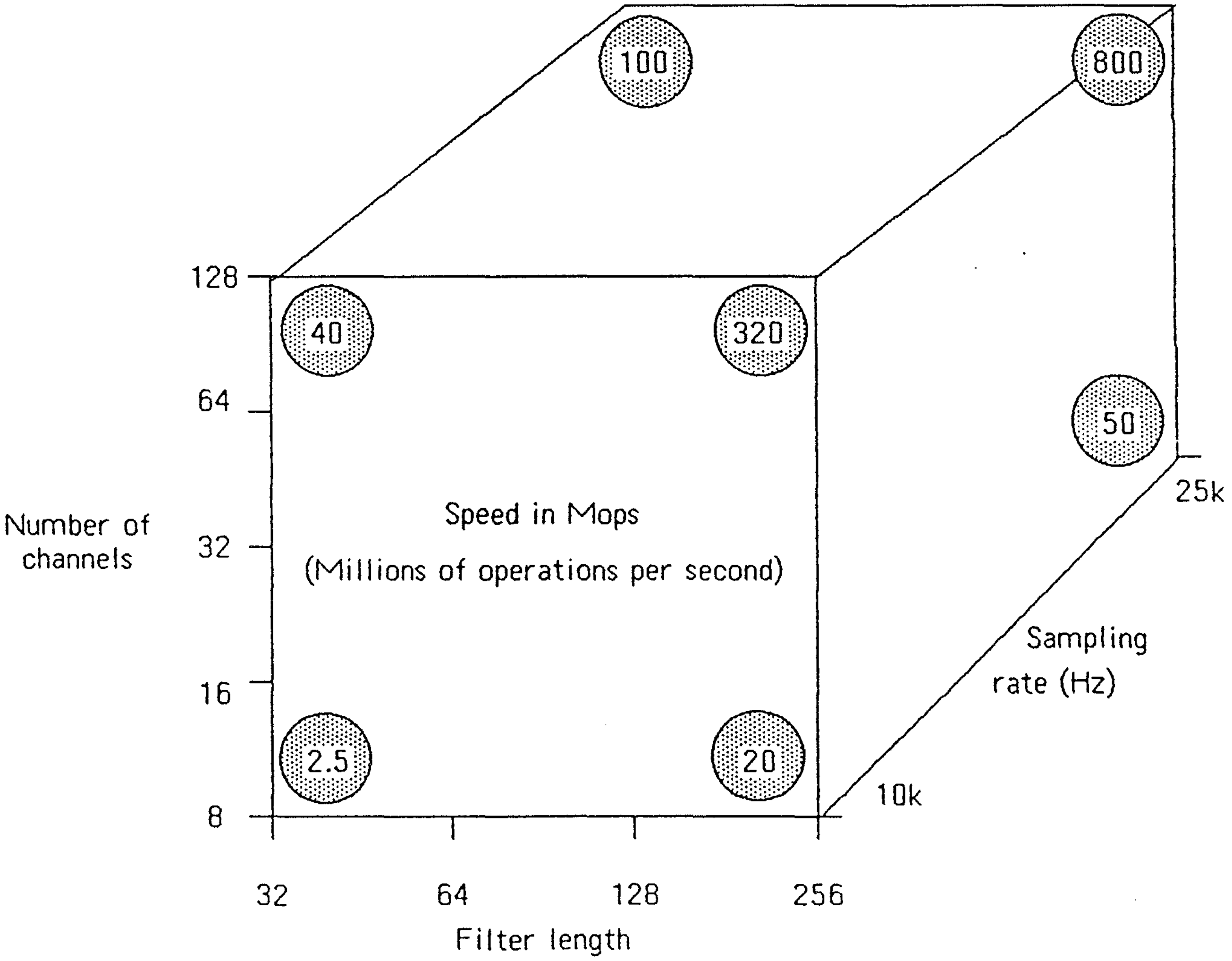
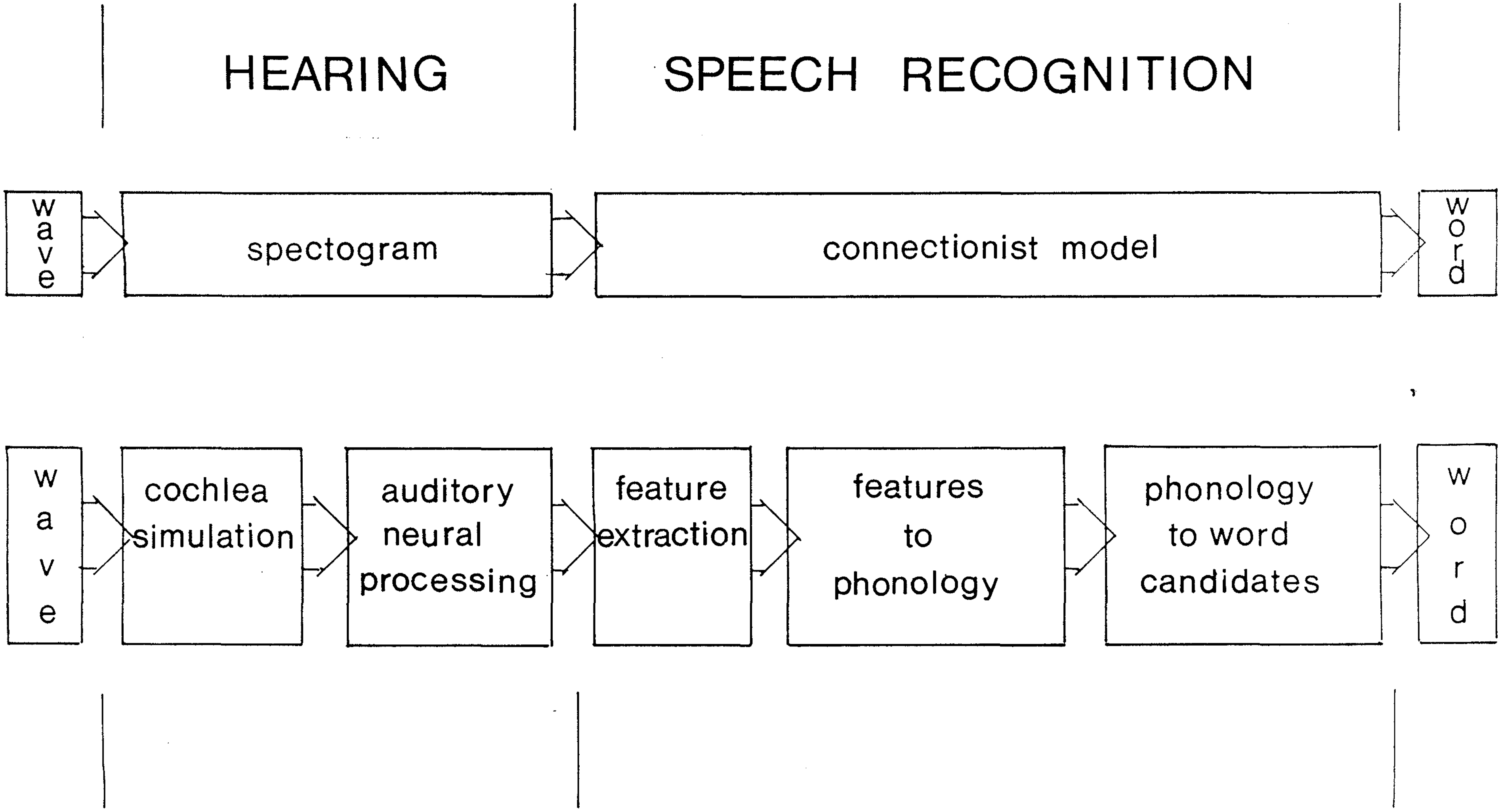


Figure 10

Figure 11



Auditory/Connectionist Speech Recognition