ANNEX B OF THE SVOS FINAL REPORT
(Part A: The Auditory Filterbank)

AN EFFICIENT AUDITORY FILTERBANK BASED ON THE GAMMATONE FUNCTION

Roy Patterson and Ian Nimmo-Smith

MRC Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF

John Holdsworth and Peter Rice

Cambridge Electronic Design
Science Park
Milton Road
Cambridge

December 1987

This Annex is a revised version of a paper presented at a Speech-Group meeting of the Institute of Acoustics on Auditory Modelling, which was held at RSRE, Malvern, 14-15 December 1987.

Abstract

This paper describes the development of an auditory filterbank to perform the initial frequency analysis in models of human hearing and speech perception. It is based on the gammatone function used by physiologists to summarise 'revcor' measurements of the impulse response of the auditory filter in small mammals. The first section shows that the amplitude characteristic of the gammatone function is very similar to that of the roex filter shape of Patterson, Nimmo-Smith, Weber and Milroy (1982), which is known to predict human masking data well (Patterson & Moore, 1986). The second section argues that the minimum-phase characteristic of the gammatone filterbank is the preferred alternative for an auditory filterbank, and it introduces a method of compensating for the strong skew on the output of the auditory filterbank. The last section presents a recursive implementation of the gammatone filter that is both accurate and efficient.

The result is an auditory filterbank with a unique combination of advantages: It is based on physiological data and modelling. It predicts human masking data accurately. It is almost as fast as the simplified filterbanks currently being used as front-ended processors for automatic speech recognition.

INTRODUCTION

In February of 1987, speech and hearing groups met at the Institute of Hearing Research (IHR) in Nottingham to discuss the use of auditory filterbanks in speech and hearing research, and to determine the extent to which we could agree on the characteristics of a 'standard' filterbank. Several groups described auditory filterbanks based on the 'roex' filter suggested by Patterson, Nimmo-Smith, Weber and Milroy (1982). The roex filter characteristic was derived from psychophysical data, and it has been shown to predict auditory masking in a wide variety of conditions (for a review see Patterson & Moore, 1986). Although the psychophysical experiments provide data that leads to a detailed description of the amplitude characteristic of the auditory filter, the assumptions made in deriving the filter shape preclude the derivation of the phase characteristic and thus the impulse response of the filter.

In order to generate a dynamic roex filterbank, one has to make some assumptions about the phase characteristic of the auditory filter, and each group described their assumptions and the advantages and disadvantages of the systems that those assumptions generated. There were two main contenders for the roex phase characteristic, namely 'linear phase' and 'minimum phase'. A brief overview of the systems and issues is presented in the latter half of this Introduction. Briefly, the minimum-phase assumption has the advantage that it leads to realistic, asymmetric impulse responses. It has the disadvantage, in common with most current filterbanks, that the output of the filterbank has a strong rightward skew in the low-frequency channels. The linear-phase filterbank has the disadvantage of producing symmetric impulse responses. It does, however, provide a straightforward method for compensating for the rightward skew in the output of the filterbank. At the same meeting, Martin Cooke suggested an auditory filterbank based on the gammatone function used by physiologists to fit 'revcor' data, and he pointed out that Schofield (1985) had demonstrated that the amplitude characteristic of the gammatone function provides a good fit to the human filter shapes in Patterson (1976). The phase characteristic was not discussed.

In this paper we extend the work of Schofield (1985) concerning the amplitude characteristic of the gammatone, and then consider its phase characteristic. Briefly, with regard to phase, the gammatone filterbank appears to offer the best of both worlds; it has a minimum-phase characteristic and so produces a realistic, asymmetric impulse response; at the same time there is a straightforward method for compensating for the skew in the low frequency channels of the output. We also discovered a recursive implementation of the gammatone filter that makes it much more efficient than its roex equivalent. Accordingly, we conclude that the gammatone filterbank is the preferred option whenever a symmetric amplitude characteristic is acceptable.

A. The Roex Filterbank with a Linear Phase Characteristic

The advantages of the roex, linear-phase filterbank are threefold: 1. The filter shape is known to predict simultaneous masking well. 2. The linear-phase assumption permits the derivation of impulse responses from asymmetric, as well as symmetric amplitude characteristics. 3. There is a simple procedure for introducing phase compensation to remove the skew from the cochleogram when required. In the current software, the user specifies the frequency range of the filterbank and the spacing of the filters in equivalent rectangular bandwidths (ERB's). The program then generates a filter centred at 1.0 kHz and proceeds up and down from this point in frequency calculating filters centred at the required frequencies. The amplitude characteristics of three roex(p) filters centred at 0.427, 1.0 and 2.09 kHz are presented in Figure 1. The function relating bandwidth to centre frequency is that suggested by Moore and Glasberg (1983). In the figure, the lower and upper filters are six ERB's below and above the 1.0 kHz filter, respectively.

The impulse responses for a set of 24 filters whose centre frequencies span the range 100 to 4,000 Hz are shown in Figure 2. They are finite impulse responses (FIR) filters which were derived by specifying the amplitude characteristic and assuming a linear phase characteristic. The linear phase assumption was used for two reasons: Firstly, we wanted to ensure that the amplitude characteristic was accurate; there is a well-known Fourier technique for generating an impulse response from an amplitude characteristic given the linear phase assumption. Secondly, it is a simple matter to align the impulse responses of linear-phase filters in time (as shown in Figure 2), and so remove the strong rightward sweep in the output of the filterbank. Patterson (1987b) recently published a set of psychophysical experiments on timbre perception which indicate that the phase lag in the low-frequency channels of the auditory system does not affect the perception of a sound. The result was interpreted to mean that the auditory system knows its own phase response and measures temporal information, relative to its own phase lag.

The phase compensation issue is illustrated in Figures 3 and 4 which show the response of the filterbank to the vowel in 'mat' without compensation and with compensation, respectively. The output of the filterbank is referred to as a cochleogram and the surface of the cochleogram is intended to represent the motion of the basilar partition as a function of time. In this case, the filterbank contained 189 channels spread across the frequency range 100 to 4,000 Hz. In Figure 3, there is no compensation and so the cochleogram has a strong rightward skew in the low-frequency channels. In the upper half of the figure, where the channels are naturally aligned, we can see formants 2, 3, and 4 showing clearly in each cycle of the stimulus.

It is not particularly clear, however, where the first formant lies. When the same stimulus is analysed by a filterbank that compensates for the auditory phase lag, the resulting cochleogram is as shown in Figure 4. The upper 3 formants are clearly defined, as in the previous case, but now it becomes more apparent that the first formant is centred around the fourth and fifth harmonics; one can readily count the number of peaks per cycle for the low harmonics. (The cochleograms in Figures 3 and 4 were actually produced by a gammatone, rather than a roex, filterbank).

With regard to machine processing of voiced speech sounds, we would expect phase compensation to improve the performance of pitch synchronous feature extractors, inasmuch as it makes the pitch frames more rectangular; that is, it brings the information associated with one glottal pulse together into one pitch frame.

B. The Roex Filterbank with a Minimum Phase Characteristic

The roex filterbank with a minimum phase characteristic also has three advantages, two of which are the same as for the linear-phase version -- the filter shape predicts masking well and the minimum-phase assumption can be used with asymmetric amplitude characteristics. The minimum-phase roex has the additional advantage of producing more realistic impulse responses with faster onsets and slower offsets. It has the minor drawback that it is less obvious as to how to compensate for the skew in the cochleogram. Accordingly, we began to implement a minimum-phase, roex filterbank at APU.

At the same time we began experimenting with the gammatone filter and during the course of analysing its amplitude characteristic, John Holdsworth recognised that one could implement a recursive filter that was a very close approximation to the gammatone filter. It was clear that this 'recursive gammatone' would lighten the computational load significantly and so bring a real-time filterbank closer to reality. As a result, we chose to pursue the recursive gammatone filterbank.

The gammatone filter has one notable disadvantage; the amplitude characteristic is virtually symmetric for orders equal to or greater than two, and there is no obvious way to introduce asymmetry. The tails of the auditory filter become asymmetric as level increases with the lower skirt shallower than the upper skirt. There is far less asymmetry in the passband, however, even at high levels (Lutfi & Patterson, 1984), and it is the passband that determines masking in the vast majority of practical situations (Lower et al., 1986). Thus, we allowed the speed advantage to prevail and pursued the recursive gammatone option rather than the minimum-phase roex or the linear-phase roex.

II SPECIFICATION OF AN AUDITORY GAMMATONE FILTERBANK

To begin with, to avoid confusion, we should distinguish between the revcor function and the gammatone function which are often used interchangeably. The revcor function is a continuous representation of a set of data points obtained from an experiment in which the firings of a primary auditory fibre are correlated with the waveform at the input to the ear. Theoretically, the result of this reverse correlation procedure is an estimate of the impulse response of the fibre concerned (de Boer & Kuyper, 1968). The gammatone function is an equation that some physiologists use when they require an analytic expression for the revcor function (Johannesma, 1972). The equation for the gammatone function is

$$gt(t) \propto t^{n-1}\exp(-2\pi bt)\cos(2\pi f_o t+\phi) \quad (t \geq 0) \qquad (1)$$

where n is the order, b is a bandwidth parameter, f_o is the filter centre frequency and ϕ is the phase of the finestructure of the impulse response. Johannesma (1972) used this function to summarize revcor data, although he did not refer to it as the gammatone function, and the function was not fitted to revcor data. The name appears to have been adopted by de Boer and de Jongh (1978). The name refers to the fact that the expression before the cosine term is the gamma function from statistics, and the cosine term is a tone when the frequency is in the auditory range. Thus, the name draws our attention to the fact that, we can think of the impulse response of this filter as a burst of the centre frequency of the filter enclosed in a gamma-shaped envelope.

An array of gamma envelopes, for the case where the order is fixed at 4 and the bandwidth parameter is the equivalent rectangular bandwidth (ERB) of the auditory filter is presented in the upper part of Figure 5. The envelope rises and falls more slowly in the low-frequency channels where the filter is narrow. The lower half of the figure shows the corresponding gammatone impulse responses where it can be seen that the low-frequency impulse responses are much longer than the high-frequency impulse responses. The set of impulse responses can then be convolved channel-by-channel with the signal to produce a cochleogram, the surface of which provides a representation of the motion of the basilar partition in response to the stimulus. The cochleograms of the vowel in mat (Figures 3 and 4) were produced by a gammatone filterbank of this type. The remainder of this Section is concerned with the specification of an auditory gammatone filterbank; that is, the preferred parameter values for a filterbank that best represents human hearing as we know it.

A. A Comparison of Roex and Gammatone Amplitude Spectra

Schofield (1985) has recently demonstrated that a gammatone filter with order 4 provides a good fit to the average auditory filters presented in Patterson (1976). Patterson used a five parameter, rounded-exponential function to fit his filter shapes in order to ensure that the derived filters were not unduly constrained by the fitting process; he was not concerned with the efficiency of the fitting process or the parsimony of the expression used to represent the filter shape. He measured the filter shape at three centre frequencies, 0.5, 1.0 and 2.0 kHz, but he was unable to find a satisfactory expression relating the filter parameters to filter centre frequency. In short, the desire to measure the filter shape as accurately as possible is incompatible with the desire to summarise the resulting parameter values in a simple, smooth function.

Subsequently, as more filter shape data were gathered, Patterson and Nimmo-Smith (1986) and Patterson et al (1982) developed a family of rounded exponential, or roex, filters that enable one to choose the number of parameters in the fit in accordance with the accuracy required. The simplest member of the family, has only one parameter, p, which determines both the width and shape of the filter passband. Despite its simplicity, the roex(p) filter has proven quite successful in predicting auditory masking; see Patterson and Moore (1986) for a review. It is also the shape that Assman of IHR used in their filterbank. In the first part of this section, we compare the roex(p) filter shape to the amplitude characteristic of the gammatone filter, in an effort to extend Schofield's (1985) observations from the five-parameter filter of Patterson (1976) to the one-parameter filter of Patterson et al (1982). In the latter part of this section we consider the relationship between the gammatone and the roex(p,w,t) filter shapes. Within the roex family, the roex(p,w,t) member provides the best alternative when it is important to approximate not only the passband but also the tails of the filter outside the passband.

1. Comparison of the Roex(p) and Gammatone(n,b) Filters

Schofield (1985) used a gammatone function with a fixed order, 4, and varied the bandwidth parameter, b, to fit the three average filter shapes in Patterson (1976). The figure he presented shows that an excellent fit was obtained over the first 35 dB of the filter's dynamic range. We began by comparing the roex(p) filter at three centre frequencies, 0.43, 1.0 and 2.09 kHz with the amplitude characteristic of the gammatone(4,b) function. For a given order, n, there is a fixed linear relationship between the ERB of the gammatone function and b; for order 4, the gammatone ERB is 1.02b. When the ERB of the gammatone function was set equal to the ERB of the roex(p) filter, the fits were found to be extremely good (Figure 6) over the first 40 dB of the dynamic range (Figure 6). Beyond this, the gammatone function falls a little more

slowly than the roex function, and at the lowest centre frequency, they diverge by as much as two decibels when the attenuation characteristic reaches about 60 dB down.

In choosing to match the ERB of the two filters, we have chosen to minimise the area between the attenuation characteristics of the two functions in linear terms -- a criterion which ensures that they will predict the same threshold for a tone presented in a broadband noise. When the masker has a sharp edge some distance from the centre frequency of the relevant auditory filter, the criterion of matched ERBs leads to a one or two decibel discrepancy between the threshold predictions of the two filters. Since the discrepancy is progressive, the average decibel difference can be reduced considerably by the simple expedient of increasing the bandwidth parameter of the gammatone filter by 10 percent as shown in Figure 7. In either case, it is clear that the gammatone(4,b) function provides an extremely good fit to the roex(p) filter shape over a dynamic range of 60 dB.

We also compared gammatone functions of order 3 and order 5 with the roex(p) filter shape. The fit is not quite as good in either case. The gammatone(3,b) attenuation characteristic overestimates the roex(p) filter shape a little in the region 20-30 dB down, and underestimates the filter shape a little in the region beyond 40 dB down; the gammatone (5,b) does the reverse, underestimating the attenuation characteristic of the roex(p) filter slightly in the region 0-16 dB down and overestimating the attenuation characteristic slightly in the region beyond 30 dB down. Nevertheless, all three of these gammatone filter functions provide excellent fits to the roex(p) filter shape.

2. Comparison of the Roex(p,w,t) and Gammatone(2,b) Filters

In the case of the human auditory filter, in the region beyond 35 dB down from the peak of sensitivity, the sharp skirts that define the passband of the filter give way to rather shallower tails; the slope of the filter characteristic drops from around 100-150 dB/octave on the skirts to around 30-50 dB/octave in the tails. In the tail region, both the roex(p) and gammatone(4,b) filters fall much faster than the auditory filter. In many cases, the discrepancy is not important; for example, the prediction of threshold in broadband noise. Nevertheless, we extended the comparison of the roex and gammatone filters because a) there are situations where one wants a more accurate filter representation (e.g. in patient studies), and b) there was reason to believe that the gammatone(2,b) filter would provide a better approximation to the auditory filter shape, and the gammatone(2,b) filter is considerably faster than the gammatone(4,b) filter.

A two stage fitting process was used in this case: Three roex(p,w,t) filters with centre frequencies of 0.43, 1.0 and 2.09 kHz were calculated. The p values were taken from the ERB-rate function of Moore and Glasberg (1983), as before. The w and t values were derived from Patterson et al (1982); in particular, w was set to 0.0025 and t

was set to 0.2p. Since the tails of this filter are still relatively steep, and since the weighting factor for the tails is very small, the ERB of this filter is virtually identical to that of the roex(p) filter with the same p value. Accordingly, we began by fixing the order of the gammatone at 2, and varying the scalar applied to the bandwidth of the gammatone filter to match the passband of the gammatone(2,b) attenuation characteristic to that of the roex(p, 0.0025, 0.2p) filter. When the scalar is 0.6, the fit of the gammatone and roex passbands is very good at all three centre frequencies, as shown in Figure 8. There was no important reason to take the fitting process further. However, it was clear that the discrepancy could be largely removed by broadening the tails of the roex filter and moving them up a little, so that they take over from the passband a little earlier. Accordingly, we reversed the process at this point and fitted the tails of the roex filter to those of the gammatone(2,b) filter. We found that excellent fits were obtained for a variety of combinations of w and t; for example, 0.01, 0.28p; 0.007, 0.25; and 0.005, 0.24p. This range of w and t values is compatible with that observed in the psychophysical data. A comparison of gammatone(2,b) and roex(p, 0.005, 0.24p) filters is shown in Figure 9.

In conclusion, the gammatone(2,b) filter has an attenuation characteristic whose bandwidth can be adjusted to provide an even better approximation to the human auditory filter than the gammatone(4,b) or roex(p) filters. It should be noted, however, that this filter responds a little more slowly to changes in the stimulus envelope and as such it may be less representative of human hearing.

B. Phase Compensation in the Gammatone Filterbank

The impulse responses for the set of filters in earlier versions of the pulse ribbon model (Patterson, 1987a, 1987b) are shown in the upper righthand quadrant of Figure 10. From the point of view of auditory perception, this set of impulse responses has an advantage and a disadvantage. The disadvantage is that the envelopes of the individual impulse responses are symmetric, whereas those of the auditory filter are symmetric with the onsets steeper than the offsets, and we might, one day, expect to find that this infidelity is reflected in psychophysical data -- for example, differential masking of short signals presented just before and after a large impulse.

The advantage of the linear-phase filterbank is that it produces a cochleogram with little or no skew in the low-frequency channels. This occurs because the peaks of the envelopes coincide with a peak in the fine structure of the impulse response, and the envelope peaks are aligned across channels. The data from our phase experiments (Patterson, 1987b) indicated that the phase lag of the cochlea does not affect timbre perception, and so we would argue that the normalized cochleogram is a good representation on which to base models of auditory perception. In this section, we

experiments (Patterson, 1987b) indicated that the phase lag of the cochlea does not affect timbre perception, and so we would argue that the normalized cochleogram is a good representation on which to base models of auditory perception. In this section, we show how to implement the asymmetric impulse response of the gammatone filter, while at the same time enabling one to compensate for the phase lag and remove the cochleogram skew when appropriate.

1. The Gammatone Impulse Response Without Phase Compensation

The impulse responses for a 24-channel, gammatone(4,b) filterbank comparable to the roex filterbank in Figure 2, are presented in Figure 10a. As noted earlier when this filterbank is applied to a signal it produces a cochleogram with a strong skew to the right in the low-frequency channels. This skew is a natural property of cochlear processing, and there is no doubt that the phase lag that it imparts is present in the firing pattern that flows up the auditory nerve. The phase lag does not, however, appear to play a role in timbre perception (Patterson, 1987b), and so we wanted to develop a phase compensation that would remove the skew from the cochleogram.

The method that we employed was suggested by the form of the gammatone function where the envelope terms occur separately from the centre frequency term. The gamma envelopes appear on their own in the upper section of Figure 5. As a first approximation, we aligned the gammatone impulse responses so that the envelope peaks occurred at the same time for all channels. The resulting set of impulse responses is shown in Figure 10b. There is far less skew in the low-frequency channels of the cochleograms produced by this filterbank, but some does remain. A comparison of the filterbanks in Figures 10b and 2 suggests a further refinement of the alignment process. The filterbank in Figure 2 is more orderly because in each case a peak in the oscillating component of the impulse response coincides with the peak of the envelope of the impulse response. This is only rarely the case in Figure 10b. Accordingly, we adjusted the phase of the oscillating component until a fine structure peak coincided with the envelope peak. The results are shown in Figure 10c. Both de Boer (1976); and Buunen (1976) have shown that shifting the fine structure relative to the envelope does not affect the perception of a sound and so it is unlikely that this adjustment to the gammatone filterbank will adversely affect a model's ability to predict perceptual changes.

The difference between envelope phase compensation on its own and envelope-plus-peak compensation can be seen in Figure 11 which shows the cochleograms produced when the two filterbanks are stimulated by the pulse train shown in the upper section of the figure. The centre section of the figure shows the

case for envelope compensation on its own. Although the majority of the skew of the cochleogram has been removed by this process there remains a slow curvature; the peaks of the fine structure drift slowly left or right relative to the envelope peak. The cochleogram in the lower part of the figure has both envelope and fine structure compensation. The resulting cochleogram is more rectangular and it is somewhat easier to read; for example, the transition from one, low harmonic to the next is more obvious in this form.

III A RECURSIVE GAMMATONE FILTER

In the previous section, we indicated that the cochleogram was calculated by convolving the signal with the set of gammatone impulse responses that define the filterbank and, indeed, at the start of our work with the gammatone, we did use the convolution method of filtering. Although this method is accurate and the filters are reliably stable, it is expensive in terms of computation. To begin with, we were more concerned with accuracy than efficiency and so the convolution method was appropriate at that point. However, it is our intention to produce a software filterbank that can be used as a frontend in speech and hearing research, and ultimately, to produce a hardware filterbank for automatic speech recognition. As a result, we were aware of the computational load that the convolution method imparts, and concerned to find a recursive filter that would improve the efficiency.

An indication of the magnitude of the computational load imposed by real-time, FIR filterbanks is provided in Figure 12. It shows the number of millions of operations that have to be performed per second, as a function of three variables -- the number of channels in the filterbank (the ordinate), the number of coefficients in the impulse response (the abscissa) and the sampling rate (the depth dimension). A modest FIR filterbank with only eight channels, and 32 points per impulse response, operating at a sampling rate of 10 kHz, requires on the order of 2.5 million operations per second (MOPS). A large FIR filterbank with 128 channels and 256 points per impulse response, operating at a rate of 25 kHz requires 800 MOPS! Currently, digital-signal-processing chips only perform on the order of 10 MOPS, and so the large filterbank would require on the order of 80 of these DSP chips to run in real time.

Looking at the dimensions in turn, only one would appear to offer any opportunity for achieving significant improvements in speed, namely the filter-length dimension. With regard to sampling rate, the audio bandwidth is about 0.4 times the sampling rate, and frequencies up to 8 kHz are required to distinguish some fricatives. Accordingly a competitive filterbank would have to be able to run at a minimum of 20 kHz. With regard to channels, there is some discussion as to the number that are actually required for a competent speech recognition machine; however, few people think that the number could be less than about 32. The reason is that the bandwidths of the filters are such that it requires around 32 to cover the range 50-8,000 Hz. At filter densities less than this, components in the speech wave can fall between filters. Thus, the only real hope of improving the speed is to reduce the number of coefficients per filter -- that is to find a recursive filter that provides an adequate approximation to the auditory filter.

There are several problems with recursive filters: They require very accurate coefficients and so they usually require floating point computations. They also exhibit

stability problems when the filter is relatively narrow and the centre frequency is a small proportion of the sampling rate -- conditions which exist for all of the low-frequency filters in the first three octaves of the filterbank. These problems can be alleviated by down sampling the signal for the low-frequency channels, and thereby raising the ratio of the centre-frequency to the sampling-rate. But this in turn necessitates the use of anti-aliasing filters at each down-sampling point, and some method of correcting for the phase shifts that these extra filters impart. Fortunately, John Holdsworth discovered that a gammatone filter of order n, could be very accurately approximated by a cascade of frequency-shifted, lowpass filters, for which a recursive implementation was available. Some care had to be taken when implementing the digital form of the recursive filter in order to avoid small phase shifts -- particularly at the higher frequencies. However, once this problem was overcome, he was able to demonstrate that when the recursive and non-recursive filterbanks were applied to a wideband noise, the difference between corresponding filter outputs was negligible. The details of the recursive filter derivation and its implementation are described in a separate technical report (Holdsworth, Nimmo-Smith, Patterson, & Rice, 1988) which appears as Annex C of the Spiral VOS Final Report: Part A..

A comparison was made of the relative efficiency of the recursive and non-recursive gammatone(4,b) filterbanks. The results for sampling rates of 10 and 20 kHz are shown by broken and solid lines, respectively, in Figure 13. The ordinate is the number of seconds required to filter a one-second sample of sound, that is, the number of 'times real time', on a standard MicroVAX II computer. The sloping curves show the results for the convolution method with the FIR filters. In the lowest channels the process takes 80 to 320 times real time depending on the sampling rate; in the highest channels this drops to between 10 and 40 times real time. The filtering times increase by a factor of four when the sampling rate is doubled, because the number of points in the wave doubles and the number of coefficients in the impulse response also doubles.

The horizontal lines near the bottom of the figure show the results for the recursive filters. When the sampling rate is 10 kHz, the process takes 6.8 times real time except in the highest frequency filters where oversampling is required to preserve the accuracy -- in which case the process takes 13.6 times real time. When the sampling rate is 20 kHz, the processing time doubles for the lower frequency channels. It doubles, rather than quadrupling, because we do not have to double the number of coefficients in this case. Furthermore, once the sampling rate is well above the centre frequency of the highest filter, there is no need for oversampling. Across the entire filterbank, the recursive filter is about five times faster than the non-recursive filter at 10 kHz, and ten times faster when the sampling rate is 20 kHz.

With regard to overall performance, a recursive, 32-channel gammatone(4,b)

filterbank, operating at 20 kHz, runs at 438 times real time on a MicroVAX II. Although this may seem like rather poor performance, it does enable one to process sufficient sound to do hearing research with complex sounds, and to investigate speech perception at the level of individual syllables or words. Fortunately, DSP chips operate at higher speeds and they are optimised to perform operations like those involved in digital filtering. A recursive gammatone(4,b) filterbank with 32 channels and the equivalent of 16 coefficients per channel, running at 20 kHz, requires about 10 MOPS. There are now floating point DSP chips which claim performance in this range which indicates that it should now be possible to produce a real-time auditory filterbank that runs on one DSP chip. If the gammatone(2,b) filterbank proves acceptable, there would seem every possibility of implementing it on one DSP chip.

ACKNOWLEDGEMENTS

We would like to thank Egbert de Boer for helpful discussions concerning the phase characteristic of the gammatone filter, and in particular for pointing out that it is a minimum-phase filter. This work has been carried out with the support of Procurement Executive, Ministry of Defence.

REFERENCES

Buunen, T.J.F. (1976). On the perception of phase differences in acoustic signals. Doctoral dissertation. University of Delft, Delft, The Netherlands.

de Boer, E. (1976). On the 'residue' and auditory pitch perception. In W.D. Keidel & W.D. Neff (Eds.) Handbook of Sensory Physiology, Vol. V, Springer-Verlag, Berlin, 479-583.

de Boer, E., & de Jongh, H.R. (1978). On cochlear encoding: Potentialities and limitations of the reverse correlation technique. Journal of the Acoustical Society of America, 63, 115-135.

de Boer, E., & Kuyper, P. (1968). "Triggered correlation". IEEE Transactions on Biological & Medical Engineering, BME 15-3, 169-179.

Holdsworth, J., Nimmo-Smith, I., Patterson, R., & Rice, P. (1988) . Implementating a gammatone filterbank. APU Technical Note.

Johannesma, P.I.M. (1972). The pre-response stimulus ensemble of neurons in the cochlear nucleus. Proceedings of the Symposium on Hearing Theory, pp.58-69. IPO, Eindhoven, The Netherlands.

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.

Lutfi, R.A. & Patterson, R.D. (1984). On the growth of masking asymmetry with stimulus intensity. Journal of the Acoustical Society of America, 76, 739-745.

Moore, B.C.J., & Glasberg, B.R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," Journal of the Acoustical Society of America, 74, 750-753.

Patterson, R.D. (1976). Auditory filter shapes derived with noise stimuli. Journal of the Acoustical Society of America, 67, 229-245.

Patterson, R.D. (1987a). A pulse ribbon model of peripheral auditory processing. In William A. Yost & Charles, S. Watson, (Eds.), Auditory Processing of Complex Sounds. Hillsdale, N.J.: Erlbaum, 167-179.

Patterson, R.D. (1987b). A pulse ribbon model of monaural phase perception. Journal of the Acoustical Society of America, 82, (5), 1560-1586.

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), pp. 299-307.

Patterson, R.D. Nimmo-Smith, I. Weber, D.L., & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. Journal of the Acoustical Society of America, 72, 1788-1803.

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

FIGURE LEGENDS

Figure 1. 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 2. An array of 24 impulse responses for roex(p) filters whose centre frequencies range from 100 to 4,000 Hz. The linear-phase assumption leads to symmetric impulse responses which have been aligned at their temporal mid-points.

Figure 3. 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 4. 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 3. Note that the strong rightward skew produced by the phase lags of the low-frequency filters has now been removed.

Figure 5. An array of gamma impulse responses for a 24-channel auditory filterbank (lower portion), and the equivalent array of gamma envelopes (upper portion). The range of the abscissa is 25 ms; the filter centre frequencies range from 100 to 4,000 Hz.

Figure 6. A comparison of the gammatone(4,b) and roex(p) filters at three centre frequencies, 0.43, 1.00 and 2.09 kHz. In this case, the gammatone filter has been matched to the roex filter by equating the ERB, thus minimising the difference in the area under the curves. The range of the ordinate is 60 dB, the abscissa ranges from an octave below to an octave above the centre frequency in each case.

Figure 7. The comparison of the gammatone(4,b) and roex(p) filters at three centre frequencies (0.43, 1.00 and 2.09 kHz). In this case, the bandwidth of the gammatone filter has been increased by 10% to minimise the decibel difference between it and the roex filter. The range of the ordinate is 60 dB, and the abscissa ranges from an octave below to an octave above the centre frequency of the filter in each case.

Figure 8. A comparison of the gammatone(2,b) and the roex(p,w,t) filters at three centre frequencies (0.43, 1.00 and 2.09 kHz). The parameters for the roex filter are taken from Patterson et al (1982). The gammatone filter has been fitted to the roex

by equating their ERBs. The range of the ordinate is 50 dB, and the abscissa ranges from an octave below to an octave above the centre frequency, in each case.

Figure 9. A comparison of the gammatone(2,b) and the roex(p,w,t) filters at three centre frequencies (0.43, 1.00 and 2.09 kHz). In this case the roex parameters, w and t, have been adjusted to improve the fit to the gammatone(2,b) filter to show that the discrepancy can easily be minimised. The range of the ordinate is 50 dB, the abscissa shows a range from an octave below to an octave above the centre frequency of the filter in each case.

Figure 10a. The impulse responses for a gammatone auditory filterbank without phase compensation. The filterbank has 37 channels covering the frequency range 100 to 5,000 Hz. The range of the abscissa is 25 ms.

Figure 10b. The impulse responses for a gammatone auditory filterbank with envelope phase-compensation; that is, the peaks of the impulse-response envelopes have been aligned vertically. The filterbank contains 37 channels ranging from 100 to 5,000 Hz. The range of the abscissa is 25 ms.

Figure 10c. The impulse responses for a gammatone auditory filterbank with envelope and fine-structure phase-compensation; that is, the envelope peaks have been aligned and then a fine-structure peak has been aligned with the envelope peak. The filterbank contains 37 channels ranging from 100 to 5,000 Hz. The range of the abscissa is 25 ms.

Figure 11. A comparison of the output of a gammatone filterbank with envelope phase compensation only (middle panel), and envelope plus fine-structure compensation (lower panel).

Figure 12. 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 13. A comparison of the speed of FIR and recursive gammatone filters (open and filled symbols, respectively).

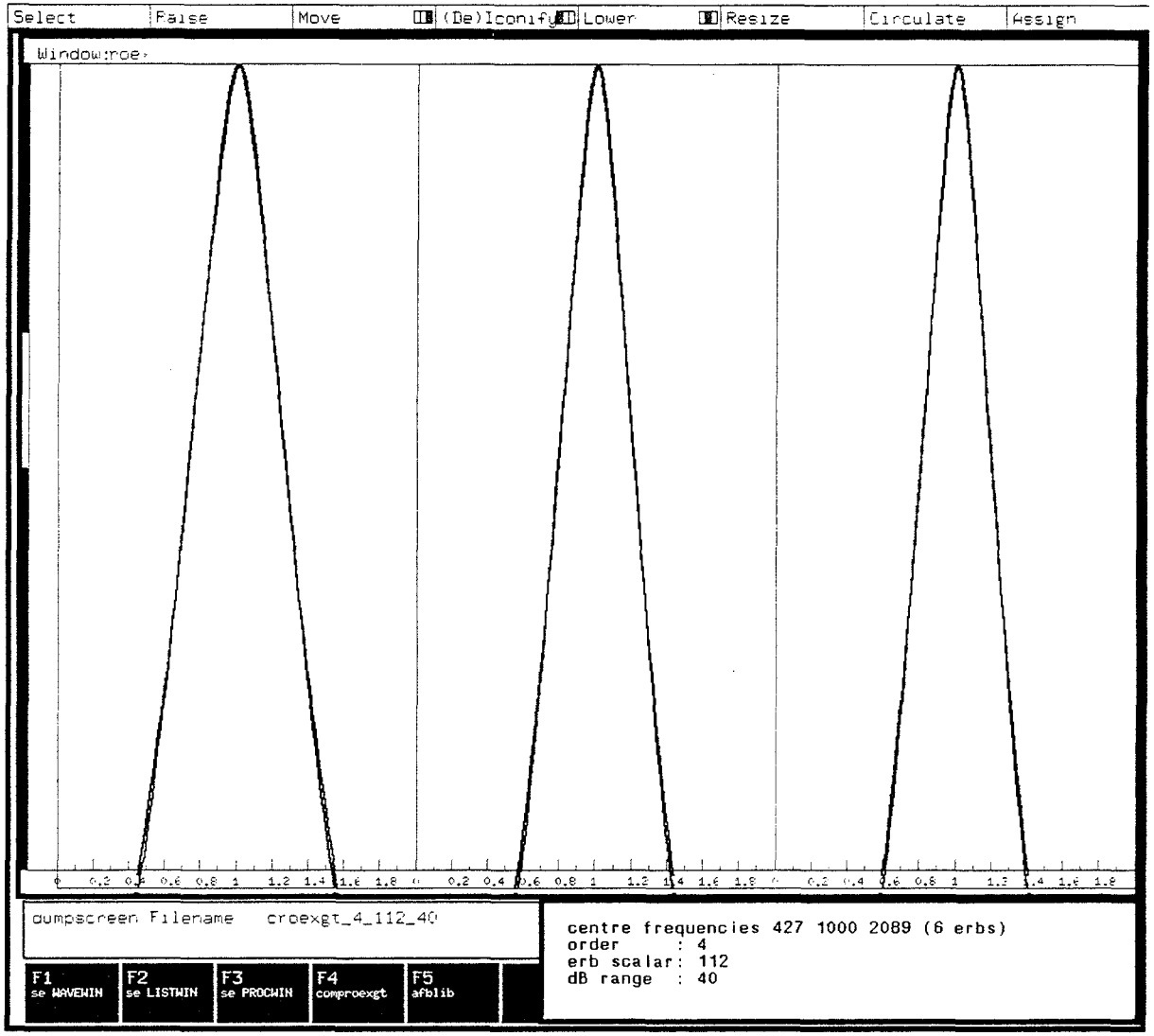

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:roe>
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8
dumpscreen Filename croexgt_4_112_40
F1 se WAVEWIN
F2 se LISTWIN
F3 se PROCWIN
F4 comproexgt
F5 afblib
centre frequencies 427 1000 2089 (6 erbs)
order : 4
erb scalar: 112
dB range : 40

Figure 1

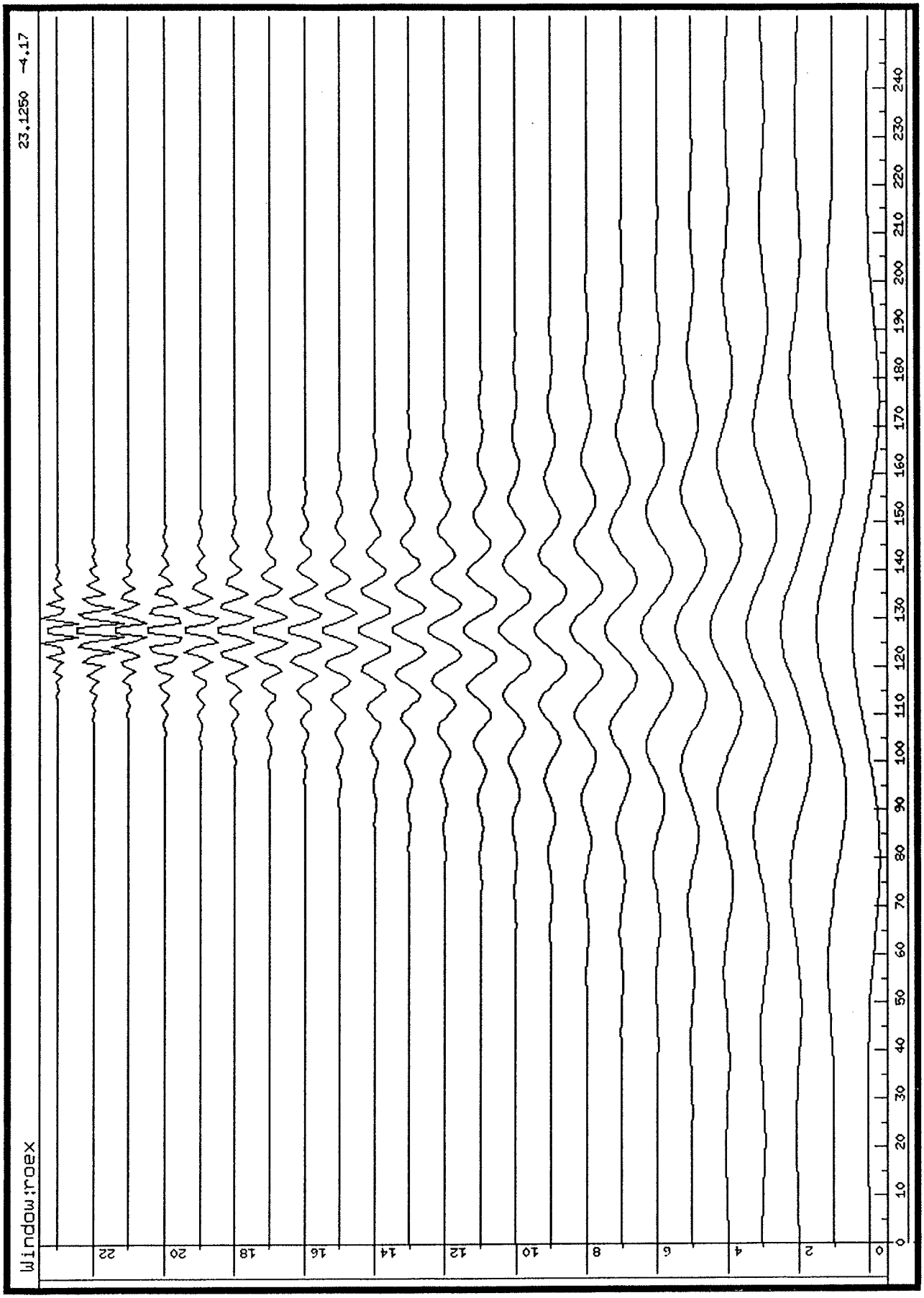

Window:roex
23.1250 -4.17
0
2
4
6
8
10
12
14
16
18
20
22
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240

Figure 2

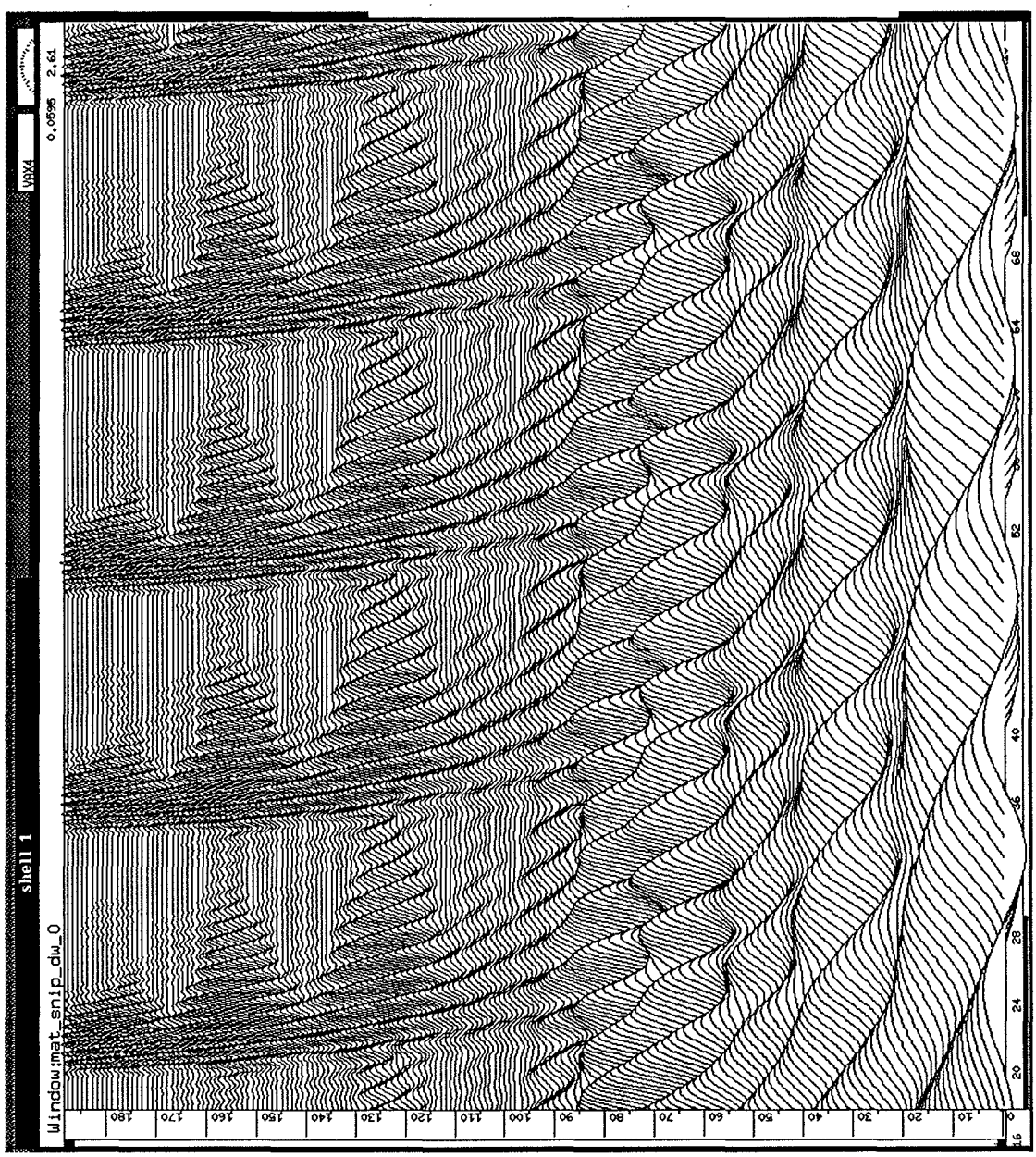


Figure 3

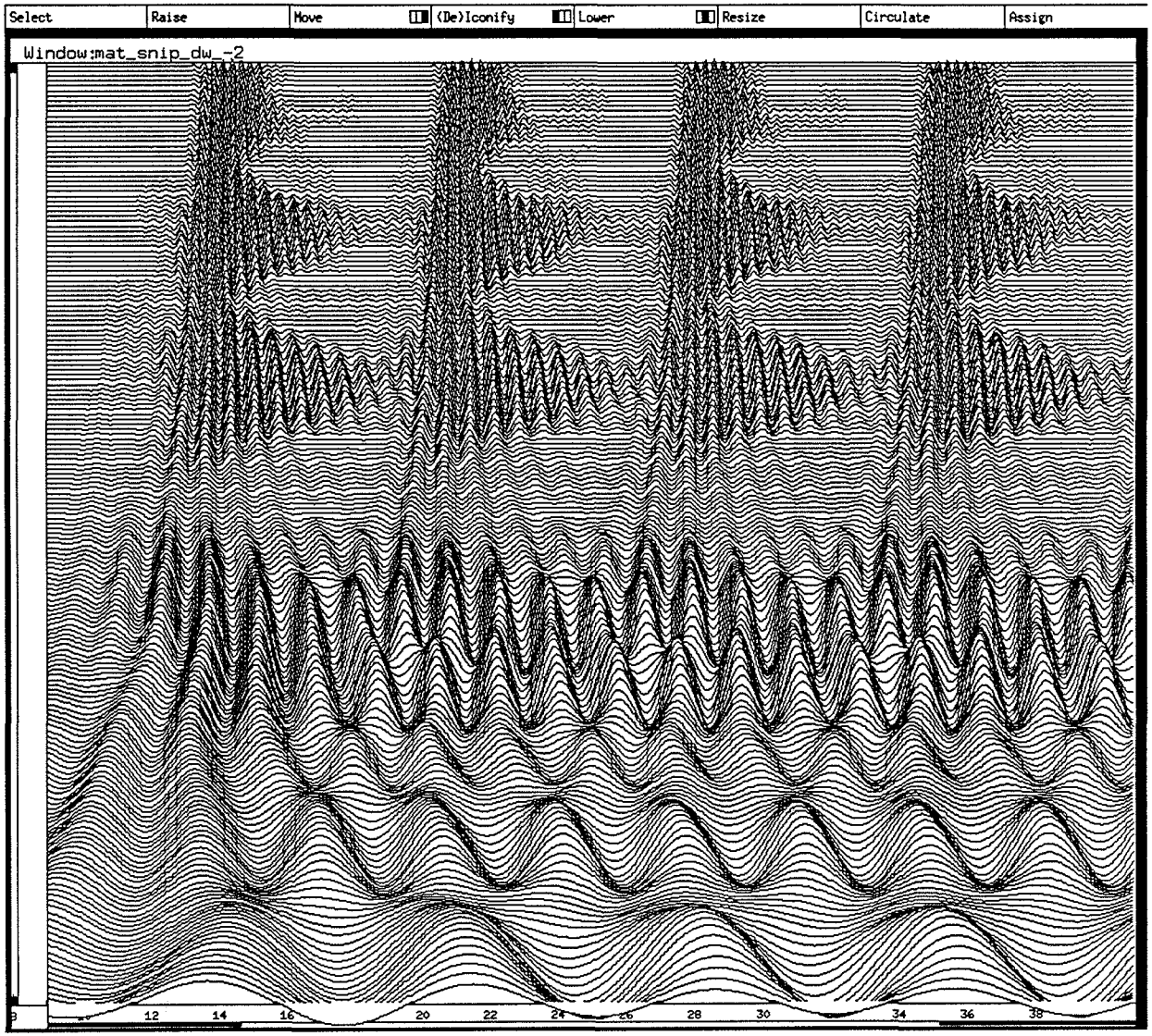

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:mat_snip_dw_-2
12
14
16
20
22
24
26
28
30
34
36
38

Figure 4

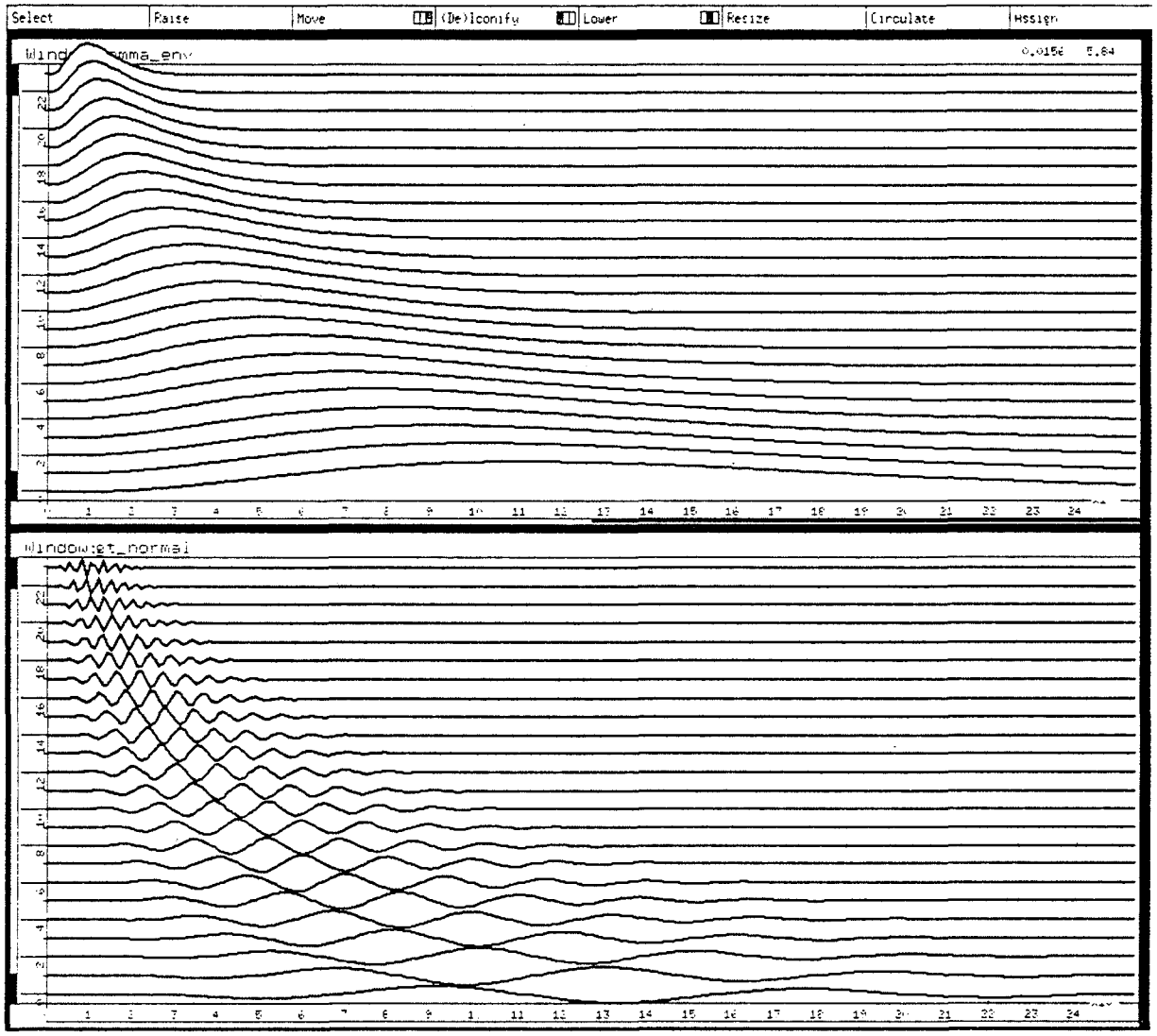

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Wind
mma_env
0.0156
5.64
window:gt_normal

Figure 5

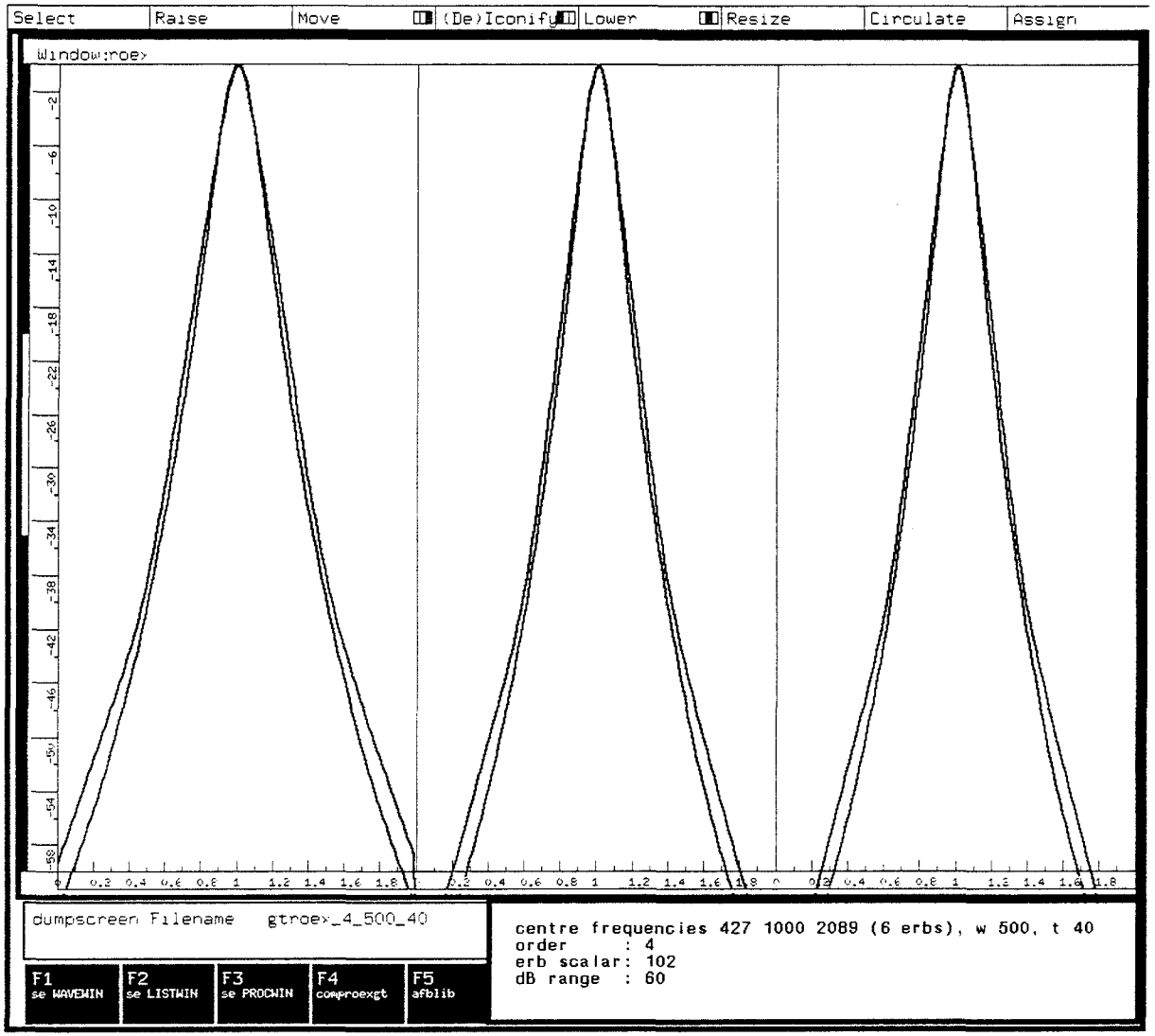

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:roex
dumpscreen Filename gtroex_4_500_40
centre frequencies 427 1000 2089 (6 erbs), w 500, t 40
order : 4
erb scalar: 102
dB range : 60
F1 se WAVEWIN
F2 se LISTWIN
F3 se PROCWIN
F4 comproexgt
F5 afblib

Figure 6

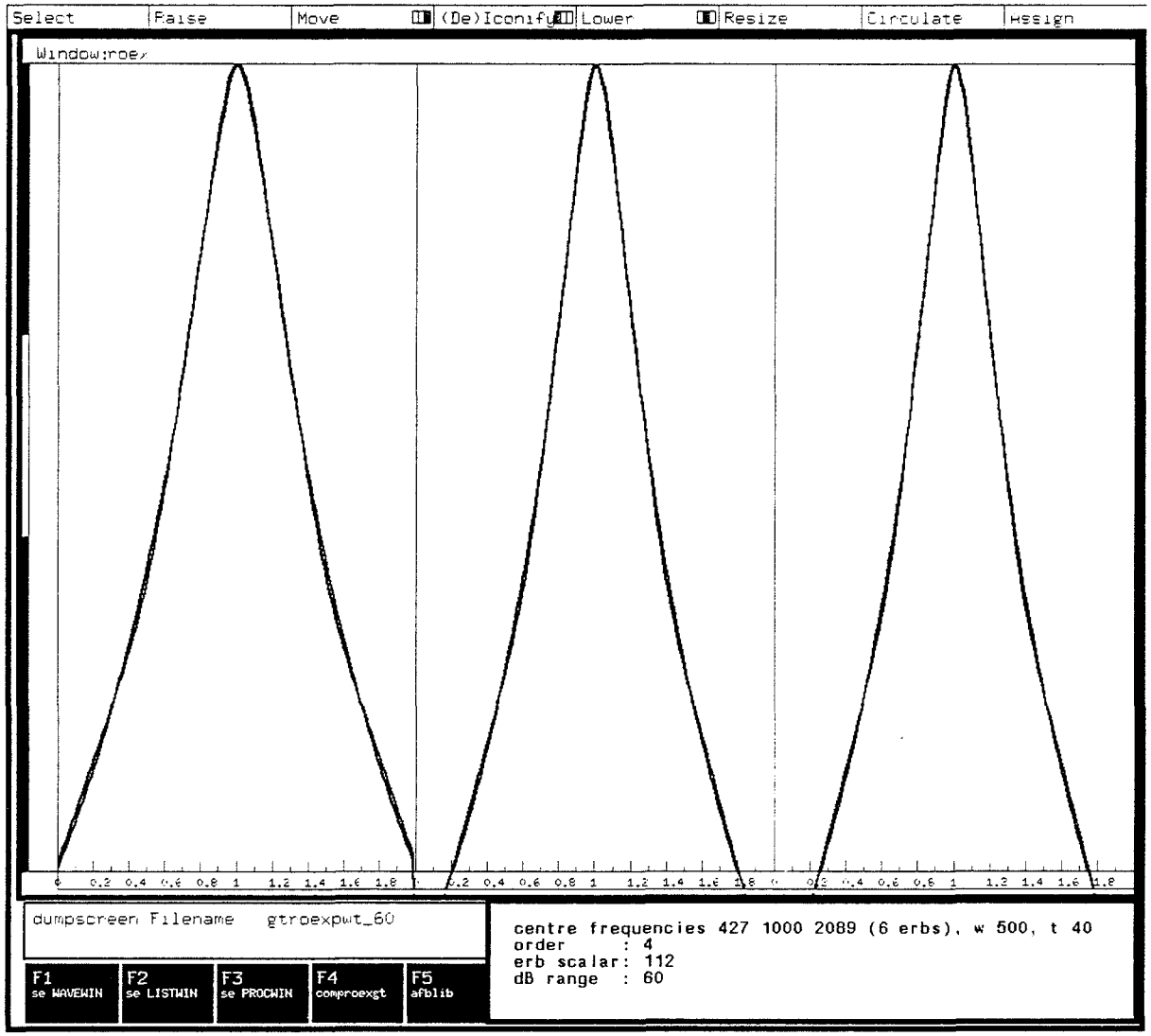

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:roex
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8
dumpscreen Filename gtroexpwt_60
F1 se WAVEWIN
F2 se LISTWIN
F3 se PROCWIN
F4 comproexgt
F5 afblib
centre frequencies 427 1000 2089 (6 erbs), w 500, t 40
order : 4
erb scalar: 112
dB range : 60

Figure 7

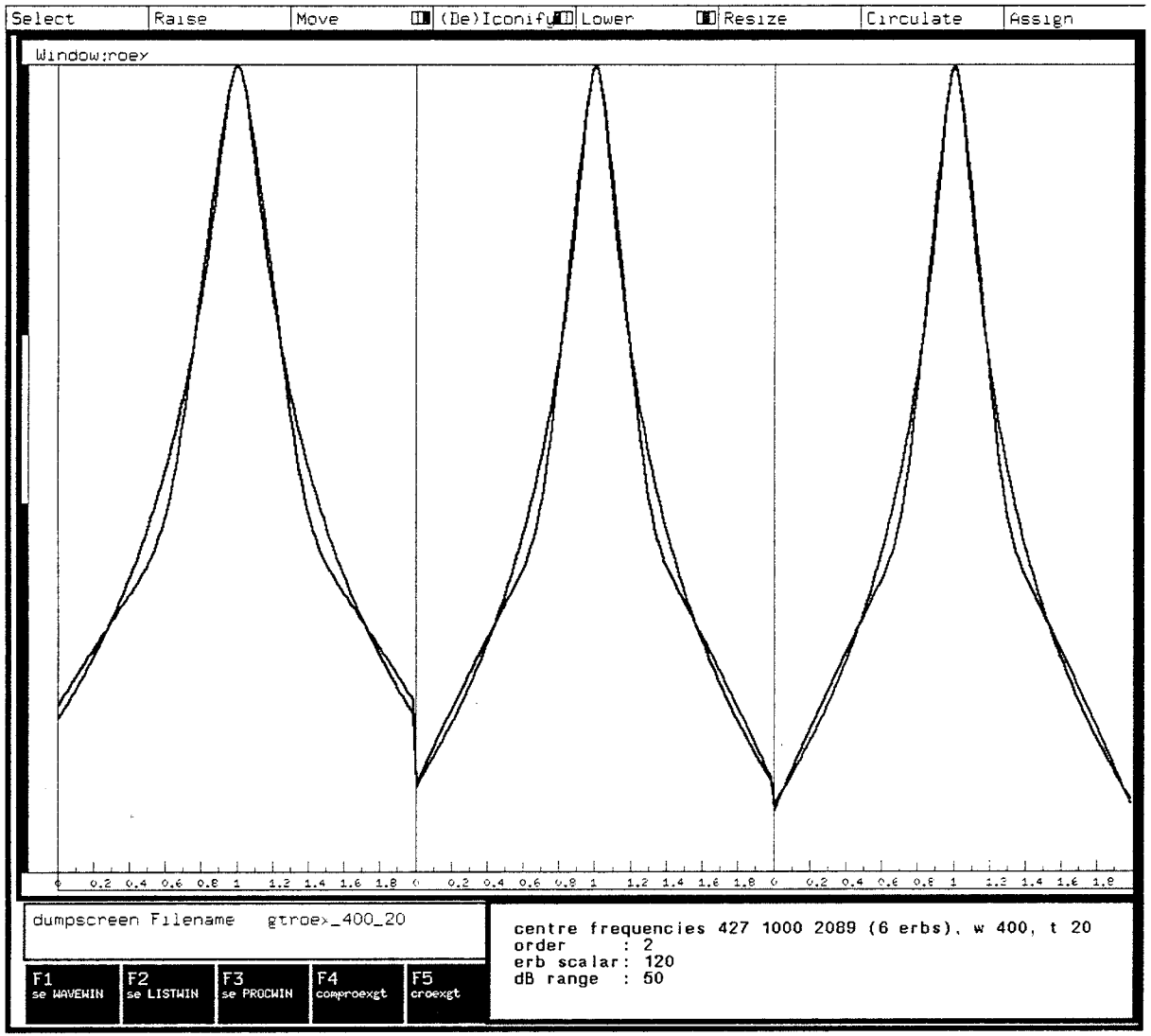

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:roex
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8
dumpscreen Filename gtroex_400_20
F1 se WAVEWIN
F2 se LISTWIN
F3 se PROCWIN
F4 comproexgt
F5 croexgt
centre frequencies 427 1000 2089 (6 erbs), w 400, t 20
order : 2
erb scalar: 120
dB range : 50

Figure 8

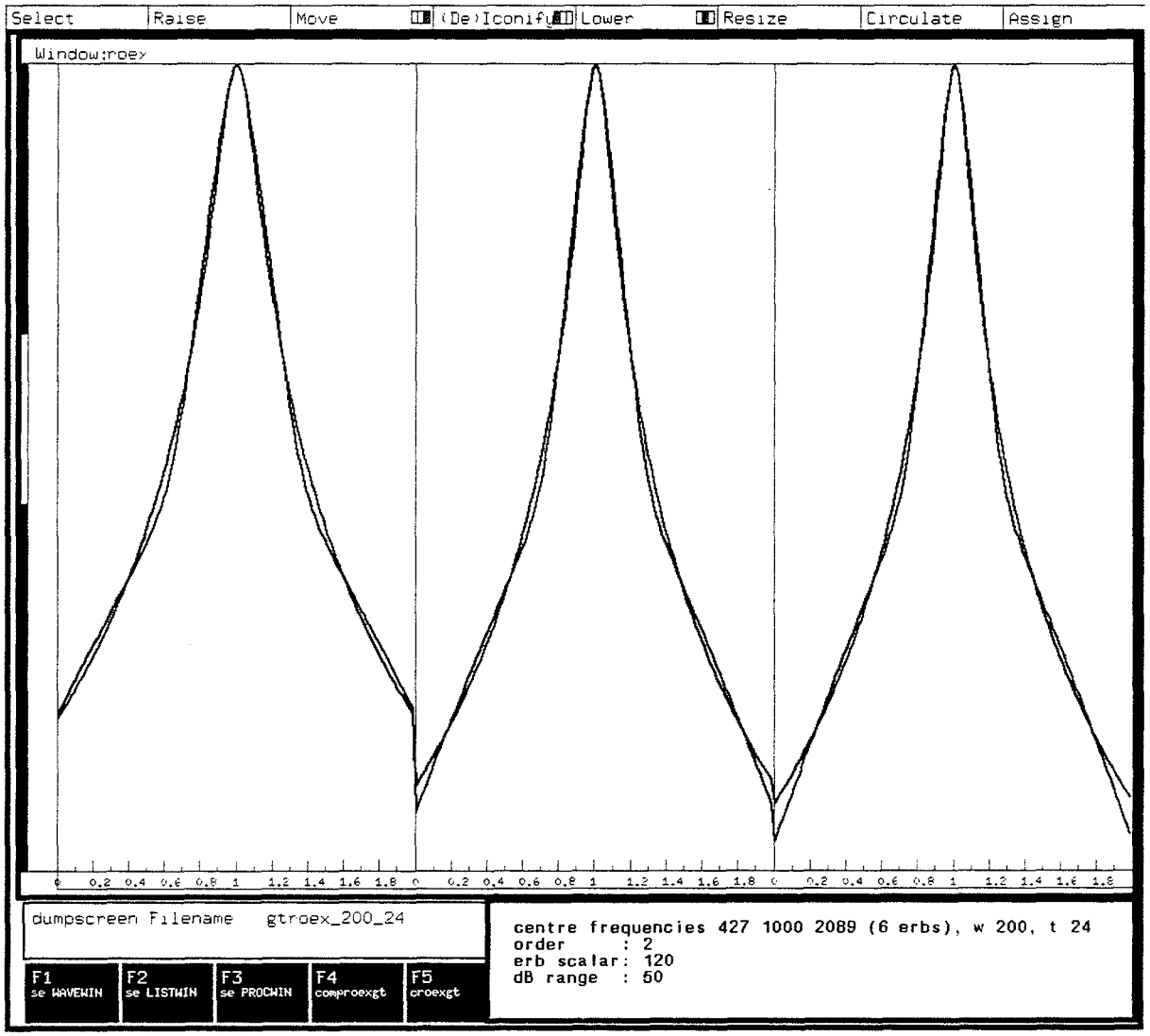

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:roex
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8
dumpscreen Filename gtroex_200_24
F1 se WAVEWIN
F2 se LISTWIN
F3 se PROCWIN
F4 comproexgt
F5 croexgt
centre frequencies 427 1000 2089 (6 erbs), w 200, t 24
order : 2
erb scalar: 120
dB range : 50

Figure 9

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
dumpscreen "gtir_0_dmp"
F1 se WAVEWIN
F2 se LISTWIN
F3 se PROCWIN
F4 setopics
F5 coslib
F6 ep johntest johntest
F7 ep spextest spextest
F8 ep PROCED_DEMO
F9 ep cochleamenu
F10 ep pulseribbon

Figure 10a

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:gammatone
dumpscreen "gtir_1_dmp"
F1 se MAKEWIN
F2 se LISTWIN
F3 se PROCWIN
F4 gatopics
F5 cosiib
F6 ep johntest johntest
F7 ep spextest spextest
F8 ep PROCED_DEMO
F9 ep cochleamenu
F10 ep pulsetrainboon

Figure 10b

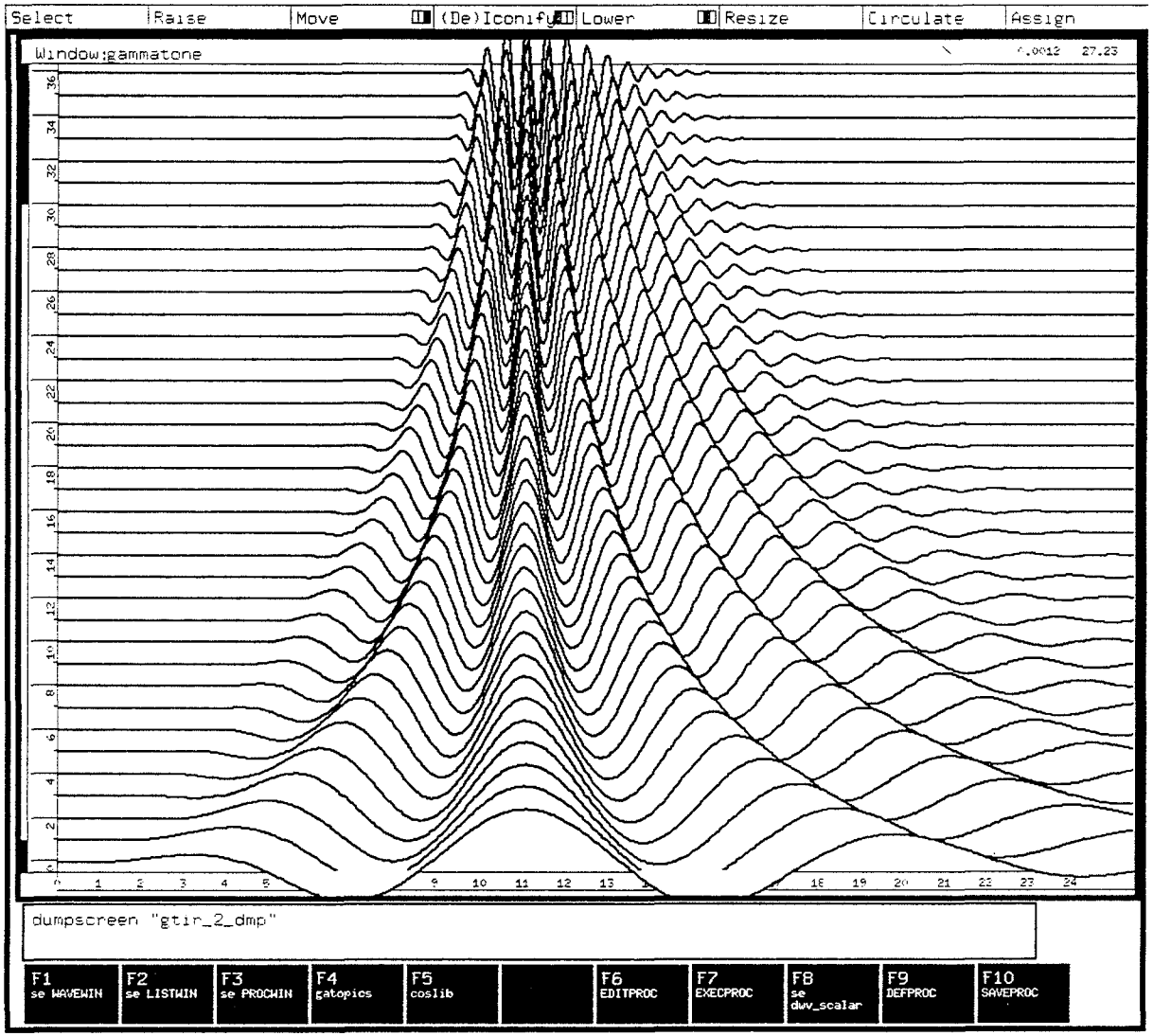

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:gammatone
dumpscreen "gtir_2_dmp"
F1 se WAVEWIN
F2 se LISTWIN
F3 se PROCWIN
F4 gatopics
F5 coslib
F6 EDITPROC
F7 EXECPROC
F8 se dwv_scalar
F9 DEFPROC
F10 SAVEPROC

Figure 10c

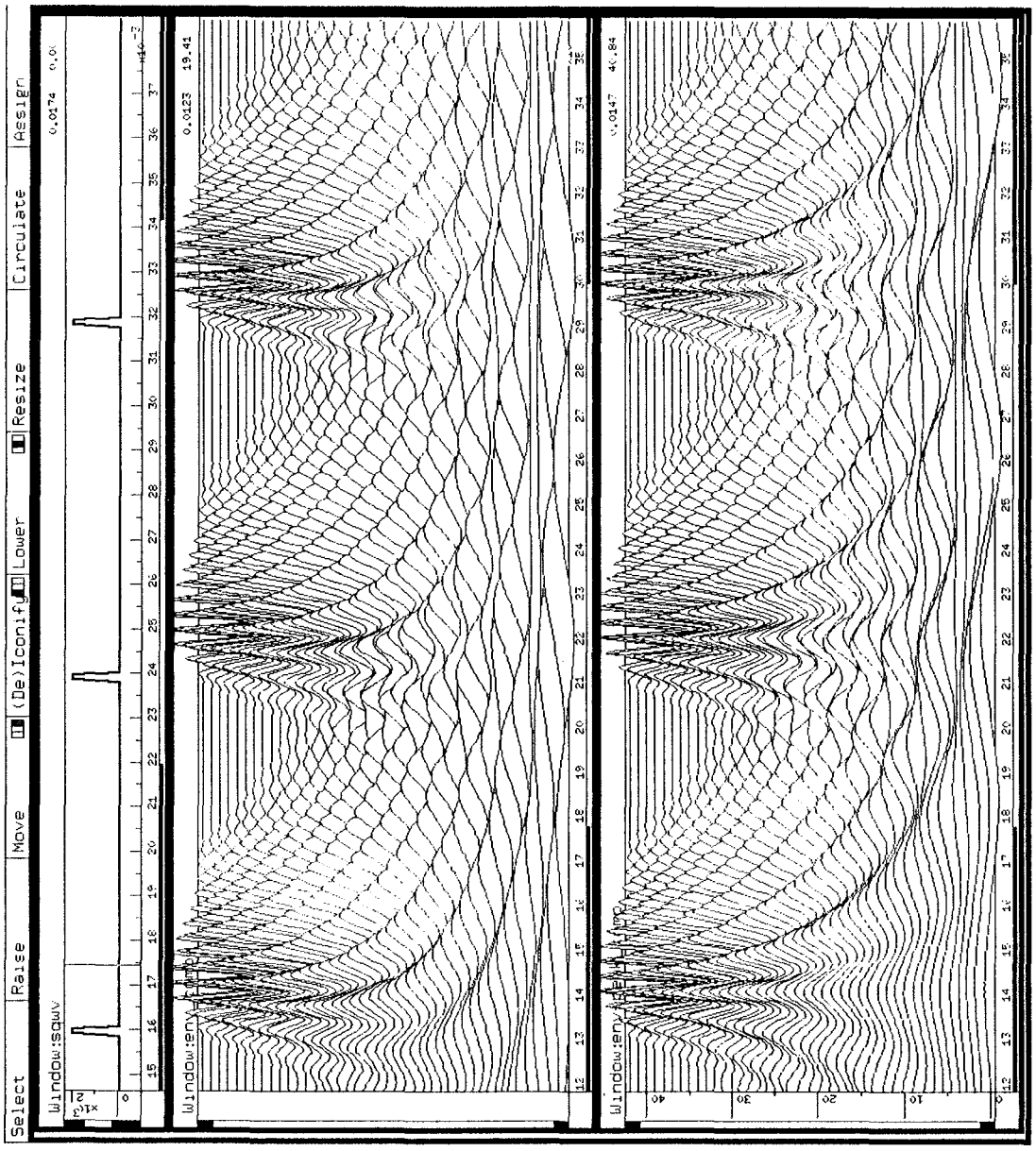

Select
Raise
Move
(De)Iconify
Lower
Resize
Circulate
Assign
Window:sqwv
Window:env
Window:env

Figure 11

F.I.R. filter processing speed

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

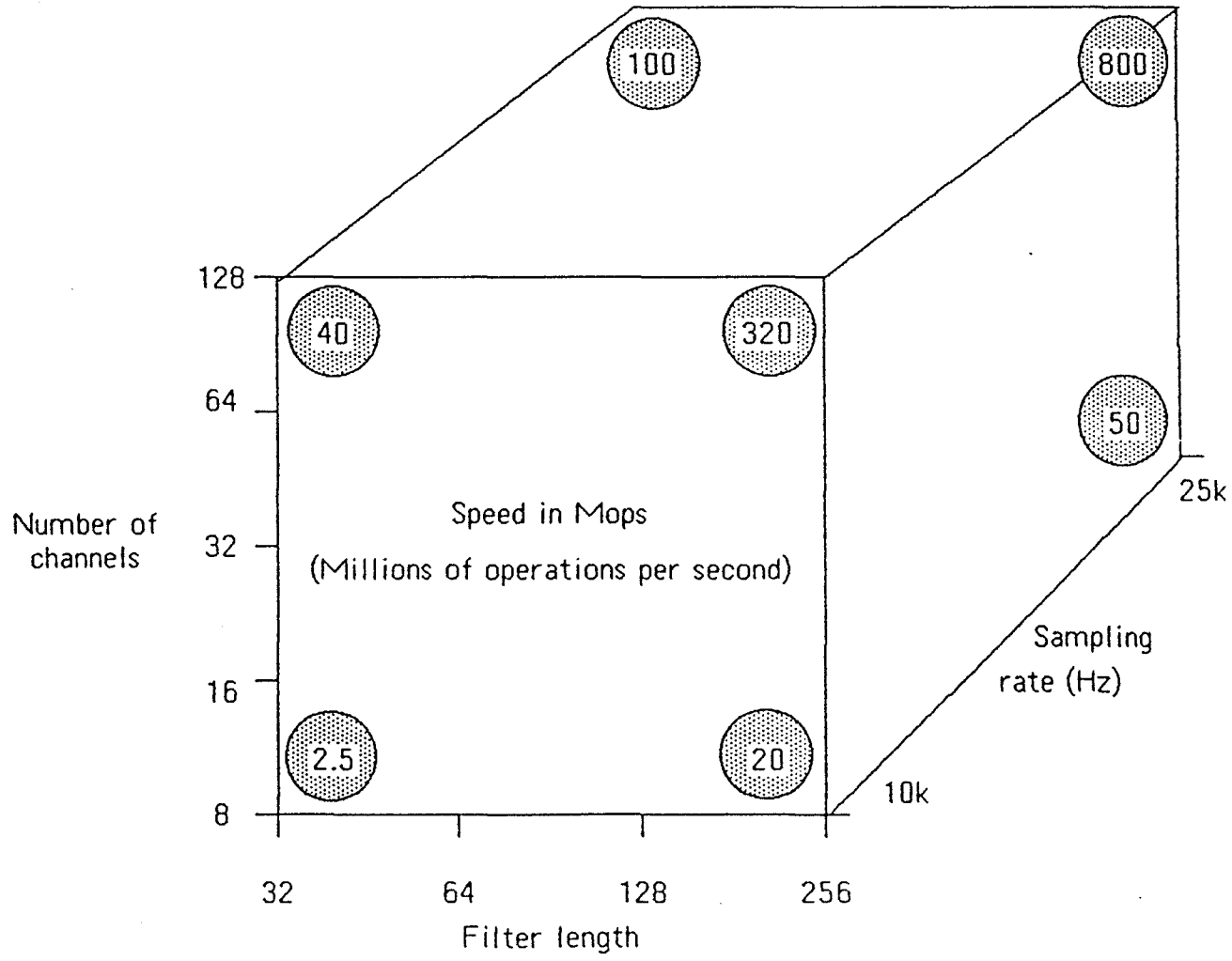
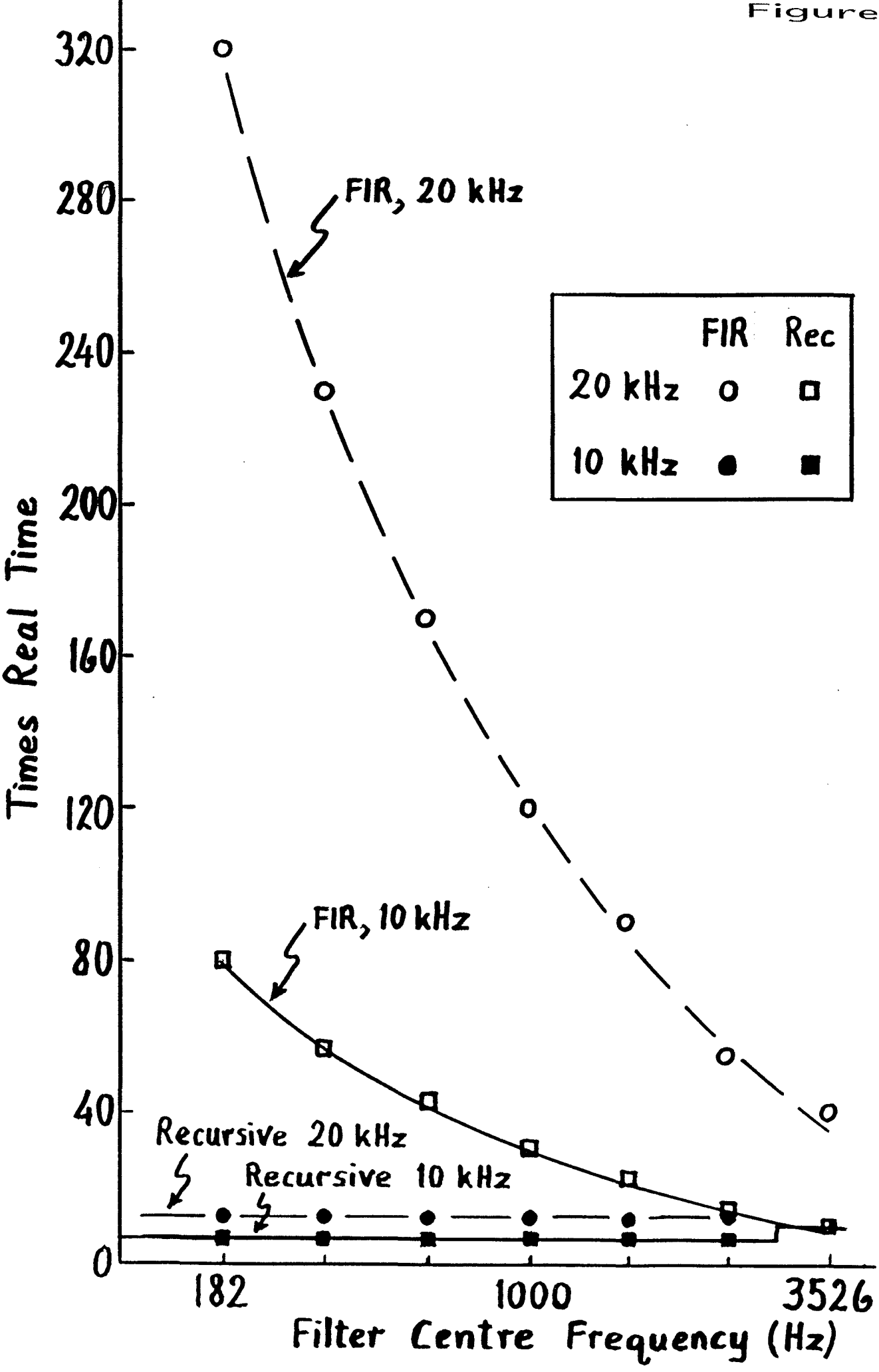


Figure 12

Figure 13


320
280
240
200
160
120
80
40
0
Times Real Time
FIR, 20 kHz
FIR, 10 kHz
Recursive 20 kHz
Recursive 10 kHz
FIR Rec
20 kHz
10 kHz
182
1000
3526
Filter Centre Frequency (Hz)