CHARGE-COUPLED DEVICES FOR SPEECH TRANSFORM CODING

K.K.Patel*, A.P.H.McCabe**, D.J.Burt**, and K.V.Lever***

ABSTRACT

This paper describes the design and performance of chirp correlators used in a feasibility study of speech transform coding. Two chips have been evaluated, a 50-point Block CZT processor and a 100-point sliding CZT processor. The relative merits of the two systems are compared in the context of data-compression of speech signals using analysis/synthesis coding techniques. Particular emphasis is placed on alternative methods of obtaining 100% duty cycle, on multiplexing the Block CZT processor, and on establishing the feasibility of the inverse sliding CZT.

INTRODUCTION

Transform Coding is a technique for the data-compression of speech signals [1] which promises to provide a reasonable quality of transmission for a moderate data rate in the 4-16 kbits/s region intermediate between waveform coding and vocoder-based techniques. Much of the work done so far has addressed fundamental theoretical issues such as the design of efficient spectral coding algorithms, and has been appropriately supported by slow-time computer simulation studies[2]. The next stage would appear to be to examine the feasibility of real-time implementation. Despite the fact that high-speed signal processing chips are under development which will advance the state-of-the-art of digital FFT hardware,[3] the entirely digital approach does not yet seem to be feasible. What is required in the short term is a single-board low-cost low-power processor, and in the long term a single-chip processor to perform real-time Fourier analysis. Present state-of-the-art CCD's are already close to providing this capability. The CCD Chirp-Z Transform (CZT) using discrete chirp correlator chips is well-established [4,5,6,7] and a single-chip version has recently been reported.[8]

The objective of this study is neither the development of transform coding algorithms, nor the fabrication of an LSI CCD spectrum analyser chip. Instead, we have concentrated on an issue which supplements both these activities: to determine whether current CCD technology is adequate for application to speech transform coding. We have made extensive use of analytical models and software simulation, but for the purposes of this presentation we shall emphasise the results we have obtained by means of prototype hardware systems.

Fig. 1 depicts a speech communication link employing transform coding and decoding. Redundancy inherent in the speech waveform is more obviously manifest in the frequency domain than the time domain as vacancies in the spectrum. Unimportant spectral coefficient can be digitised with small word-lengths or omitted altogether, reducing the total number of digits required to represent the original signal. Various quantisation schemes have been suggested, and those based on Rate-Distortion Theory [9] have the least distortion for a prescribed

* *Chelsea College, University of London.*
** *G.E.C. Hirst Research Centre.*
*** *Brunel University.*

average channel bit-rate. In addition coding algorithms which are adaptive[2] on a batch-to-batch basis show the greatest data reduction, provided the batch length is well matched to the time interval over which correlation between signal samples persists. The data-reduction efficiency also depends on the nature of the orthogonal transform used. The Karhunen-Loeve Transform is optimal in that spectral coefficient are totally decorrelated and data-compression maximised, but has the disadvantage of being too complicated to implement in real-time. The Discrete Cosine Transform (DCT) is close to optimal and the Discrete Fourier Transform (DFT) somewhat less so.

The CCD CZT is, of course, an implementation of the DFT. Although we have considered the use of a CCD DCT employing the Chirp-Z architecture, we have, for reasons that will emerge, concentrated on the sub-optimal DFT, for ease of implementation.

A complete analyser/synthesiser system utilises two CZT processors in cascade, one for forward and one for inverse transformation. Usefully, the design of the inverse processor is only trivially different from that of its forward transform counterpart. In addition, there is a useful simplification in that the chirp postmultiplication in the forward transform and chirp premultiplication in the inverse may be omitted, since - being unit-modulus complex conjugate operations - they would only cancel out if they were included. The analyser/synthesiser subsystem is shown in Fig.2.

BLOCK CZT PROCESSOR

The first processor we examined was a 50-point Block CZT processor based on the GEC MA344 chip shown in Fig.3. Although we assessed this system in some detail space limitations preclude all but a brief mention of our findings. The most obvious defect of a Block CZT processor is the 50% duty cycle. If a single-chip processor were available, it would be feasible[8] to overcome this by duplication, working two parallel processors alternatively (i.e. two in the forward and two in the inverse transform). We, however, have examined what may be a more economical solution. A CCD delay line double-buffering system using GEC CD 256 delay lines was constructed, and worked much as expected: see Fig.4. The use of time-compression techniques of this type is not new [10], but there is one feature of our version that is interesting. We used off-the-shelf 256 element delay lines (50-element devices were not available), which are somewhat more than twice as long as is necessary to restore the 100% duty cycle. As a result we were able to utilise the CZT processor idle time to process an extra signal channel. Fig.5 shows the processor working in the multiplex mode. If four double length buffer delay lines were to be integrated onto the CCD chip this technique would permit a single-chip processor to be used in the 100% duty cycle mode for both forward and inverse processing. Potentially of some interest, this technique was not pursued. Our assessment of the prototype was that the technique is not viable using discrete CZT filters with off-chip output circuitry. The dynamic range would be inadequate due to the accumulation of noise and distortion in the cascade of input buffer, forward CZT, inverse CZT and output buffer.

SLIDING CZT PROCESSOR

The infeasibility of the Block CZT approach was due to the extra processing required to restore 100% duty cycle. The Sliding CZT (SCZT) overcomes this problem in a different way. No block structure is imposed, and for 100-element transversal filters, 100 spectral components are obtained. There are, however, some penalties attached to this

benefit. Unless the signal is periodic or stationary the SCZT phase spectrum is inaccurate, so that in an analysis/synthesis system severe phase distortion would result. Furthermore, Quatieri[11] has indicated that the inverse SCZT is not a true inverse but only an approximation, implying further distortion. He has proposed a number of remedies, but we have not attempted to implement them. Instead, guided by the observation that phase information is often discarded with impunity in speech processing, we have simply constructed a forward/inverse SCZT system and have performed objective and subjective assessments to determine the degree of distortion involved. The objective of transform coding is to selectively degrade the speech quality to achieve a required data rate, and we believe that the analyser/synthesiser performance is sufficiently good to ensure that the overall performance of the data-compression scheme we intend to implement will be determined not by the CCD analyser/synthesiser, but by the digital encoding/decoding.

An unwindowed 100-point SCZT chip (GEC MA344C3) shown in Fig.6 has been fabricated for this study, and has been used to implement the demonstration system described in the next section. This chip has tap-weights designed to allow it to function also as a Block DCT processor, but in view of the inadequacy of our 100% duty cycle restoration circuitry, we have not pursued its evaluation.

DEMONSTRATION SYSTEM

The demonstration system consists of forward and inverse SCZT processors in cascade. Fig.7 shows a 1 kHz input sinewave, its real and imaginary spectral components, and the resynthesised output signal. Notice the considerable loss of signal energy in the forward transformation process. The input sinewave amplitude is at the largest value compatible with acceptable distortion, and the peak of the output spectrum is at roughtly the same level, despite signal compression. Unity gain is a necessary design feature, as the spectral components have to be transmitted over the same channel as the original signal samples would have been. To make a fair comparison between coded and uncoded systems (as regards the effect of digital transmission errors for instance) it is necessary that the same channel peak amplitude constraint be applied to both systems. This equal peak amplitude criterion results in a scaling down of the spectrum which violates the classical Parseval equality for the Fourier transform : compare the sparse output spectrum with the ubiquitous input sinewave. This has an important ramification : loss of dynamic range when the signal is reconstructed from its spectral coefficients. The observed dynamic range of the spectrum is about 40 dB and it is only this large by virtue of the forward CZT filters being "full up" with sinewave samples. When the spectral signal, which is virtually "empty", is retransformed the output signal level, with respect to the background noise in the output circuitry is lower by a factor corresponding to the spectral occupancy. It is not possible to overcome this problem by amplifying the spectrum before inverse transformation as this would cause the signal-handling capacity of the CCD inputs to be exceeded. In the case shown the spectral occupancy factor is about 50:1 implying a loss of 17 dB in dynamic range. The resynthesised signal is amplified back up to a level comparable to that of the original signal, but this of course does not restore the dynamic range. This loss mechanism appears to be fundamental, and applies equally to any implementation analogue or digital or whatever. It would seem to imply that if a resynthesised output SNR of 40 dB were required, then the dynamic range of the forward processor would have to be about 20 dB better than this, requiring a more stringent design of the off-chip sense circuitry. Fortunately, as subjective tests show, a requirement of

40 dB output SNR is excessive.

The loss mechanism is manifested in the following way. For the case of a sinewave input the unamplified reconstructed output signal is even lower than the spectral output, as the inverse processor is performing pulse expension rather than sinewave compression. At the inverse SCZT CCD outputs the signal level is about 0.15 V superimposed upon a 5 V dc offset. This is processed to remove the dc offset and the signal amplified to 25 V with a residual dc offset of 1.5 V.
These composite signals are presented to the MDACs for multiplication by the postchirps. As a result the output signal is contaminated by a replica of the postchirp waveform which is only 25 dB down. Thus the dynamic range of our system is limited at the small signal end not by noise but by chirp breakthrough. This can be seen in Fig.8, which shows the spectrum of a 1kHz input sinewave (lower trace) and that of the re-synthesised output. Note the structured nature of the background unwanted signal. This is chirp breakthrough, at worst 25 dB down on the main signal.

A second feature is noticable in Fig.7: the reconstructed sinewave is deteriorated by about 3 dB amplitude modulation having a ripple frequency equal to the batch length (100 samples). Though not extremely deleterious in a speech system this is large enough to require explanation. Its origin can be traced in part, at least, to charge transfer inefficiency (CTI). This gives rise to a trailing charge packet adjacent to the main peaks (upper and lower sidebands) in the output of the forward SCZT processor, having a measured amplitude of about 6%. This implies a value for CTI of 2×10^{-4}, which is typical for these 3-phase surface-channel polysilicon CCD transversal filters. As far as spectral analysis is concerned, this is probably innocuous, as it affects spectral resolution and dynamic range in only a minor way. But the "echoes" in the frequency domain show up as a ripple in the time domain of the reconstructed signal. Analysing a simple model for this deterioration predicts about 1 dB ripple at the batch frequency, so that there is a substantial component of the ripple that is unaccounted for. It may be that we are seeing here a manifestation of the approximate nature of the inverse SCZT.

CONCLUSIONS

The performance of our prototype SCZT analyser/synthesiser system is limited in several ways: by a fundamental loss mechanism imposed by the channel peak amplitude constraint, by signal fluctuations caused by CTI and perhaps the inherently inaccurate nature of the inverse SCZT, and by chirp waveform breakthrough at the output of the MDACs in the inverse processor. All these mechanisms contribute to a measured SNR which is rather poor, and yet subjectively the reconstructed speech quality is judged to be good enough for transform coding purposes.

The feasibility of the Slding CZT approach and the superiority of the Cosine Transform for speech coding leads us to ventrue that a combination of the two concepts would be worth exploring. As far as we can judge by the published literature the feasibility of a Sliding Cosine Transform has not yet been examined.

Other topics than those reported here have been studied, notably the improvement of off-chip sense circuitry, use of $2\frac{1}{2}$-phase clocking, the role of signal windowing techniques and the implementation of encoding and decoding: but space limitation precludes their discussion here. The work reported is also being continued, particularly with a view to establishing a more thorough understanding of the deterioration mechanisms and their amelioration.

REFERENCES

1. J.L.Flanagan et al, "Speech Coding", IEEE Trans. Commun., Vol.COM-27, No.4 April 1979 pp.710-737.
2. R.Zelinski and P.Noll, "Adaptive Transform Coding of Speech Signals", IEEE Trans. Acoustics, Speech and Signal Processing, Vol.ASSP-25 No.4 August 1977 pp.299-309.
3. S.S.Magar, "Integrated Circuits for Fast Fourier Transform", Proceedings, European Solid State Integrated Circuit Conf., Ulm, W.Germany 1977.
4. E.Bull and B.Wardrop, "A Discrete Fourier Transform Processor using Charge-Coupled Devices", Marconi Review, Vol.XL 1977 pp.3-40.
5. K.V.Lever, "Signal Processing with CCD Transversal Filters", Microelectronics and Reliability, Vol.16, 1977 pp.479-486.
6. Reticon R5601-1/2 Quad chirped transversal filter, Sigma Ltd. Pubn.No.3801 1978.
7. R.W.Broderson, C.R.Hewes and D.D.Buss, "Spectral Filtering and Fourier Analysis using CCDs", Proceedings IEEE international Symposium on Circuits and Systems, Newton, Mass. April 1975 pp.43-68.
8. W.L.Eversole et al, "A Completely Integrated 32-point Chirp-Z Transform", IEEE Journal of Solid State Circuits, Vol.SC-13, December 1978 pp.822-831.
9. T.Berger, "Rate Distortion Theory", Englewood Cliffs NJ., Prentice Hall 1971.
10. D.B.G.Roberts, "CCD and SAW : Friends or Foes?", IEE Specialist Seminar on the Impact of New Technologies in Signal Processing, Conf.Pubn.No.144 1976 pp.53-60.
11. T.E.Quatieri Jr., "Short-Time Spectral Analysis with the Conventional and Sliding CZT", IEEE Trans. Acoustics, Speech and Signal Processing, Vol. ASSP-26, No.6 December 1978 pp.561-566.

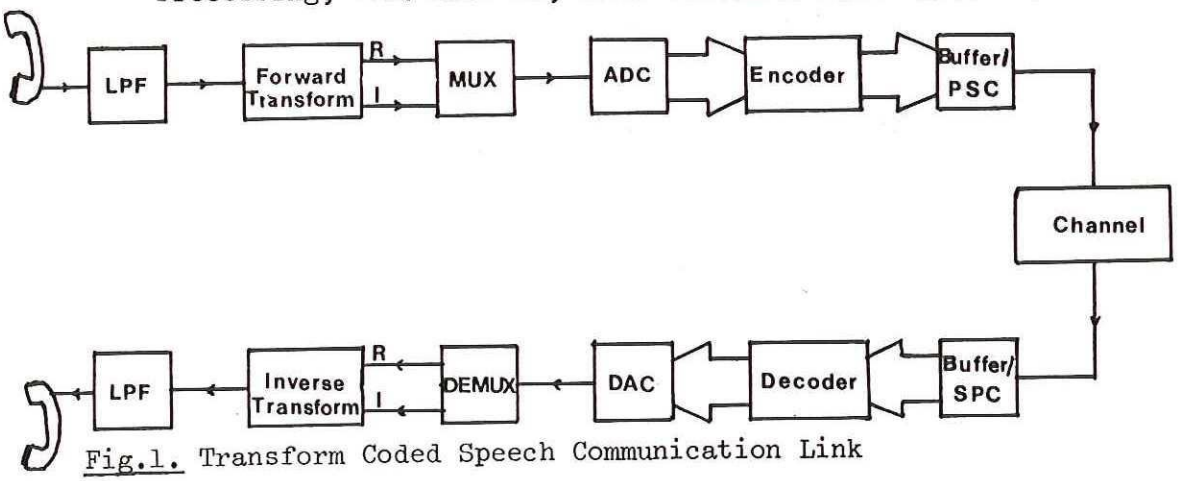

Fig.1. Transform Coded Speech Communication Link

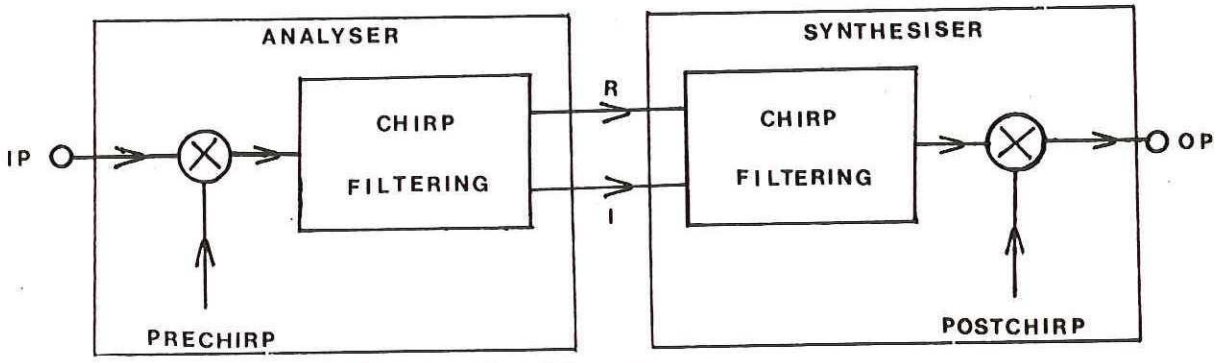

Fig.2. Analysis/Synthesis Subsystem

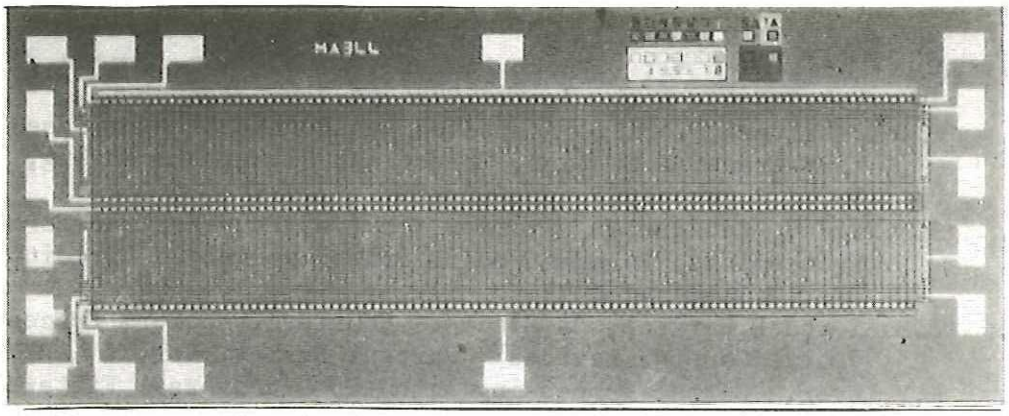

Fig.3. 50-point Block CZT Chip GEC MA 344

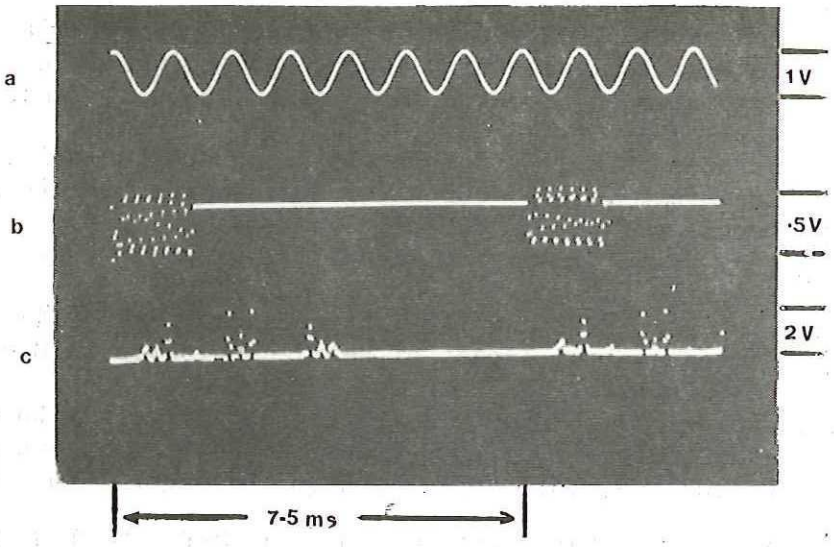

Fig.4. 100% Duty Cycle (a) 1kHz Sinewave Input (b) Time Compressed Buffer Output (c) Block CZT Spectral Output

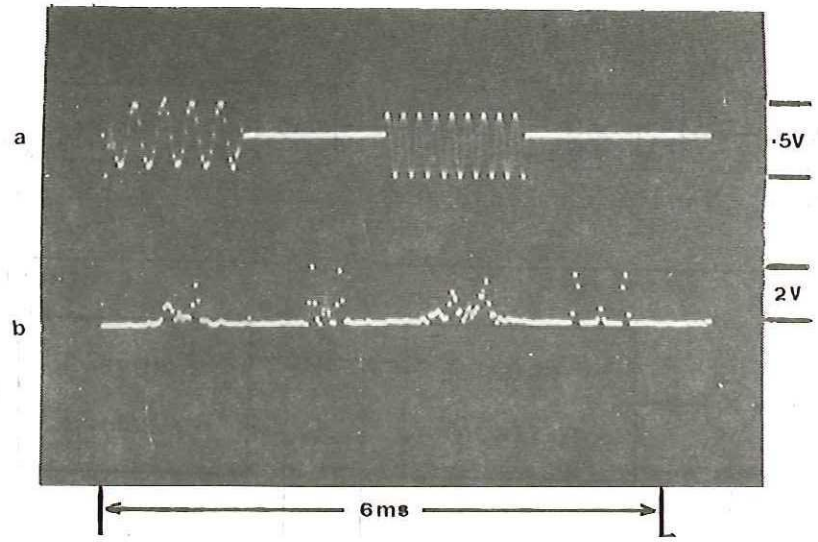

Fig.5. Multiplex Operation (a) Composite Input Signal (b) Output Spectrum

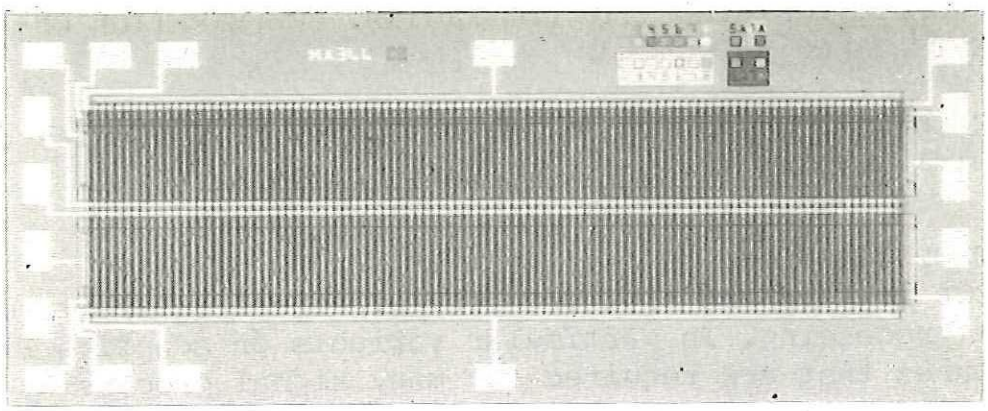

Fig.6. 100-point Sliding CZT Chip GEC MA 344C3

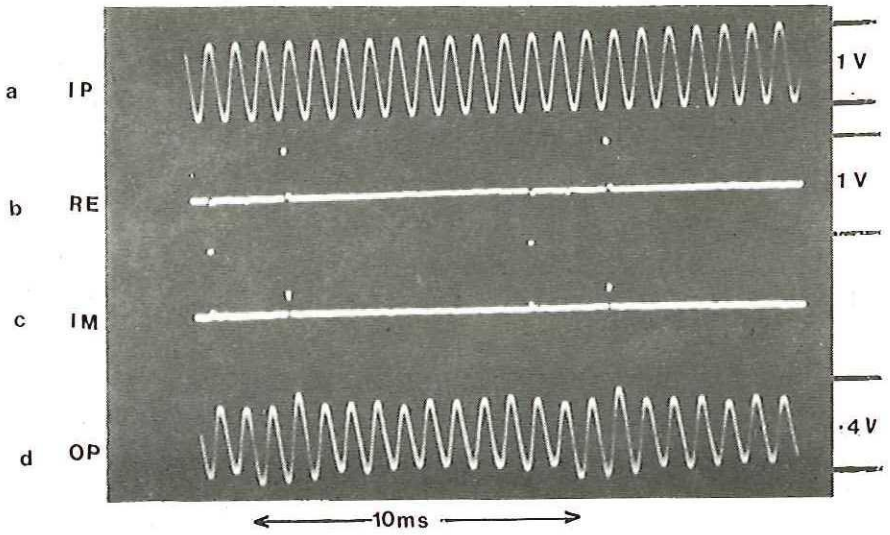

Fig.7. Analyser/Synthesiser Performance (a) 1kHz Sinewave Input (b) Real Spectral Output (c) Imaginary Spectral Output (d) Resynthesised Signal

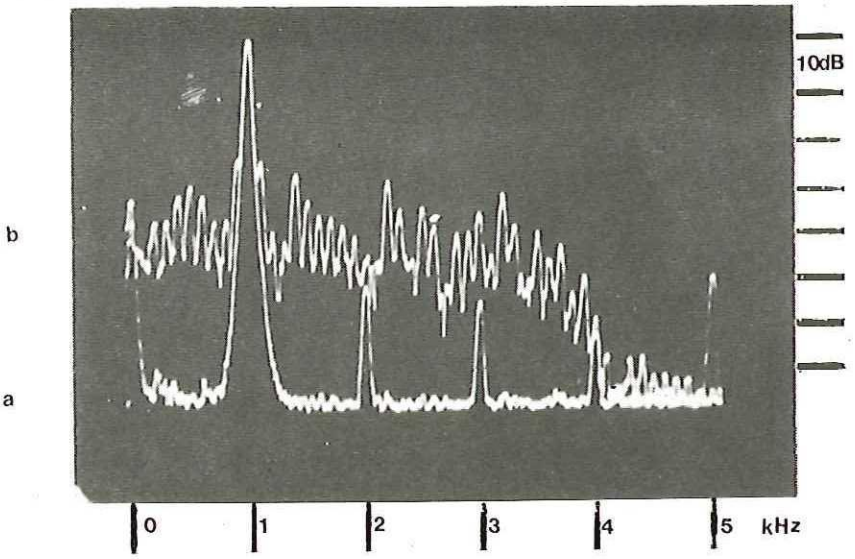

Fig.8. Spectrum of Resynthesised Output (HP3580A) (a) 1kHz Sinewave Input (b) Resynthesised Signal