NEUVIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 16 au 20 MAI 1983

DATA COMPRESSION of SPEECH SIGNALS by VARIABLE RATE SAMPLING

Bülent Sankur[x] , W.Steenaart

University of Ottawa. Dept. Elec. Eng., Ottawa, Ontario, CANADA

RESUME

Dans ce travail, les quantificateurs sequentiels avec taux d'échantillonage dépendent du signal d'entreé ont été étudiés. Les propriétés d'un modulateur à delta asynchrone pour le codage efficace de la parole ont été analysées en particulier. Le signal d'entrée est échantilloné à taux non-uniforme et dans une manière adaptative. En consequence ce signal est converti dans un processus d'intervalles interbit, lequel est comprimé, mis dans un mêmoire tampon et transmis par multiplexage avec la séquence des polarités. Nous avons aussi étudié differentes stratégies d'adaptation. La performance de ce scheme est characterisée par le rapport signal-bruit et par un facteur de suréchantillonage. Cette étude est basé surtout sur des travaux de simulation extensives en utilisant des processus aléatoires dont la form d'onde ressemble à celle de la parole naturelle.

SUMMARY

Sequential quantizers with signal dependent sampling rate are investigated. In particular the signal coding properties of the asynchronous delta modulator are analyzed. In this encoder the input signal is adaptively sampled, converted into an interbit interval process, which is further compressed, buffered and multiplexed into the channel with the bit polarity sequence. Various adaptation strategies are also considered. The performance of this scheme as measured by both the signal to noise ratio and the oversampling factor is characterized with respect to the encoder and adaptation parameters. This study is based on extensive simulation results with speech-like waveform processes.

INTRODUCTION

Source coding techniques of the delta modulator (DM) or DPCM type have a narrow dynamic range within which satisfactory performance is obtained. Various improvements and adaptation schemes purport to match the encoder dynamics to the local (instantaneous or block) characteristics of the input signal. Schemes such as adaptive forward or backward quantization, fixed or adaptive linear prediction and several other adaptation strategies have brough about significant improvements in the peak signal to noise ratio (SNR) as well as have augmented the robustness of the encoders.

In this study we investigate a delta modulator where the sampling instances are nonuniform (hence, asynchronous delta modulator) and they occur in time in a signal adaptive fashion. This way the nonstationary dynamic behaviour of the input speech signal is reflected into the instantaneous sampling rate of the DM. The block digram of a general asynchronous delta modulator (ASDM) is shown in Fig. 1.

The support from the Naturel Sciences and Engineering Research Council under Grant A 8572 and Strategic Grant G 0665 is gratefully acknowledged.

[x] On leave from Boğaziçi University, Bebek, Istanbul, Turkey.

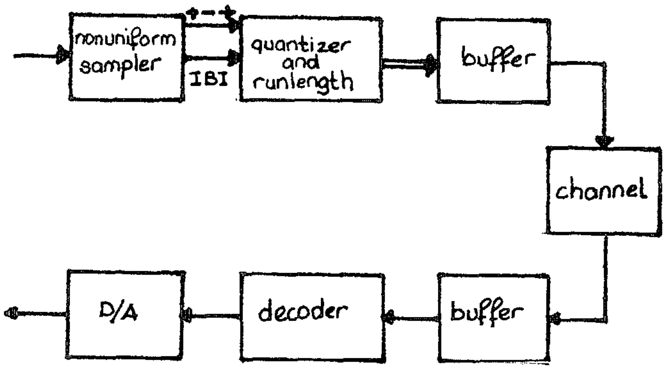

Fig.1. Block diagram of an encoder with variable rate sampling.

The input signal is being nonuniformly sampled according to some scheme (section 4) and consequently the signal information is mapped into both an amplitude sequence (typically a polarity sequence as in the delta modulator) and an interbit interval sequence (IBI), i.e., the time intervals between the successive samples. In the second stage both the polarity sequence and the IBI process are further compressed. For example, one can envision runlength encoding of the polarity sequence and delta modulation with adaptive forward quantization of the IBI process. Finally the buffer has the function of absorbing the excess or slack of activity between the source and channel sides. At the receiver side one has the decoding buffer, runlength decoder and d/a converter. A simple but effective nonuniform sampler is shown in Fig.2.

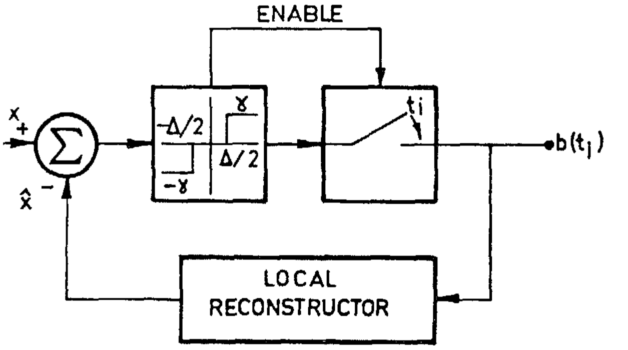

Fig.2. Corridor type asynchronous delta modulator.

In this scheme one envisions a sensivity corridor around the signal as represented by the-possibly-variable or adaptive deadzone in the 1-bit quantizer. Here a sample is transmitted only if the error signal crosses one of the corridor walls. The size of the corridor itself as well as the stepsize could be varying adaptively. The adaptation logic for both is available at the receiver side since the adaptation rule is typically derived from the bit polarity and/or IBI process itself. In the seqel, the signal coding properties of this type asynchronous delta modulator are considered.

2. PREVIOUS WORK

The idea of variable rate sampling with a view to exploit for the purpose of bit rate reduction the alternation of quiescent and relatively more active sections of the input signal has appeared directly or indirectly in various source coding algorithms. These algorithms have in common:

i) Speech is an intermittent and nonstationary process. Hence source coding rate and channel transmission rate should not be equal all the time.

ii) A buffer is needed to absorb the difference between source and channel activities.

iii) The sampling rate or instances are determined in a signal adaptive fashion.

Sankur (1), Hawkes and Simonpieri (2), Mark and Todd (3), Murakami (4) have investigated various nonuniform sampling and asynchronous delta modulation algorithms. The intermittent nature of speech, more explicity, the fact that conversational pauses and micropauses account up to 60-70% of the time, has been used in the TASI and DSI systems, (5,6) as well as in some packet transmission systems (7). Dubnowski and Crochiere (8) have considered a generalization of the TASI process; in this scheme the short-time signal variance is computed and the sample bit assignment is made using a rate distortion theory based algorithm. Significant improvements are obtained only when several source signals (speech channels) are concatenated and compressed together. An interesting technique analyzed by LoCicero and Prezas (9) provides double adaptation, i.e., both for the stepsize and the sampling rate. While the stepsize adaptation is based on the Song mode adaptive DM (ADM), the sampling rate is adjusted on a block basis considering the slope statistics. Jayant and Christensen (10) have considered variable aperture coding of speech where IBI's are allowed to be only multiples of Nyquist intervals. An ADM with runlength coding has been considered by Barba et al. (11). Other examples can be found in (12) and (13).

3. PRELIMINARIES

The investigation of ASDM is based on extensive simulation studies where an artificial model for speech signals was used as shown in Fig. 3.

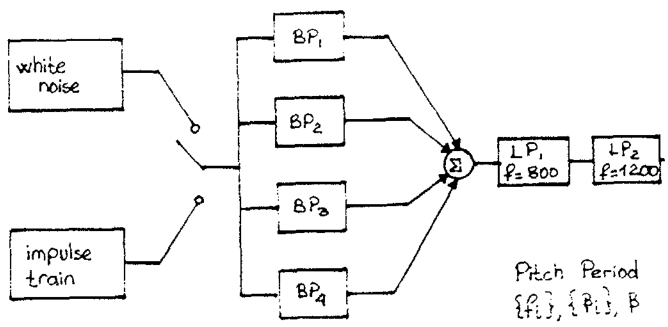

Fig. 3 Block diagram for the generation of the artificial speech.

For each speech segment (usually 8192 samples) one specifies the pitch period, the voiced/unvoiced decision, the bandwidths (B_i) and center frequencies (f_i) of the four formant filters. For each segment these quantities are selected randomly, (B_i) and (f_i) being uniformly distributed in:

$400 \leqslant f_1 \leqslant 700$ $\qquad 50 \leqslant B_i \leqslant 110$
$700 \leqslant f_2 \leqslant 1800$ $\qquad 50 \leqslant B_i \leqslant 110$
$1800 \leqslant f_3 \leqslant 2900$ $\qquad 100 \leqslant B_i \leqslant 400$
$2900 \leqslant f_4 \leqslant 3900$ $\qquad 100 \leqslant B_i \leqslant 400$

In Fig.3. the β parameter specifies the sampling frequency normalized to the Nyquist rate. It was found (section 4) that β must be chosen in the range 8 to 16, corresponding to a sampling frequency range of 64 to 128 kHz.
To quantify the performance of the ASDM algorithm we have used the signal to noise ratio

$$SNR = 10 \log_{10} \frac{\sum x_i^2}{\sum (x_i - \hat{x}_i) LP}$$

as well as the oversampling factor, k, defined as

$$k = \frac{\# \text{samples in the segment}}{\# \text{Nyquist samples}}$$

Finally, we have found useful to express the parameters of ASDM as a percentage of the segmental variance; hence, e.g., AC 0.35 means that the corridor width is 35% of the segmental variance. The main parameters were: AC: corridor size
AP: stepsize

β sampling rate/ Nyquistrate

4. SIMULATION RESULTS

The SNR and oversampling factor k figures as an function of both the step and corridor sizes are shown in Tables I and II.

Table I: SNR figures for various AP and AC values.

SNR

AC \ AP	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	1.00
0.05	14.4	[illegible]	39.3	40.6	38.8	36.3	33.9	31.9	[illegible]	28.4	27.4	26.2	24.8	23.1	22.9	21.5	20.7	20.3	19.9
0.10	12.0	27.	34.4	38.8	35.4	36.6	36.9	36.	34.6	32.9	[illegible]	27.4	26.2	26.6	25.1	23.9	22.8	21.9	20.9
0.15	13.8	22.7	[illegible]	[illegible]	[illegible]	33.3	31.8	32.1	33.3	32.8	31.9	30.8	29.4	27.9	26.5	25.8	24.6	23.1	22.1
0.20	11.7	19.9	22.0	23.4	25.4	27.4	[illegible]	[illegible]	28.9	28.6	29.5	29.6	30.1	30.0	29.8	29.1	27.6	26.9	25.6
0.25	9.9	17.4	[illegible]	20.4	21.5	22.8	24.6	25.6	26.5	26.4	26.2	25.8	26.6	26.5	27.0	26.6	27.2	27.5	27.0
0.30	[illegible]	16.1	17.1	18.0	20.0	20.0	21.0	22.3	23.3	23.7	24.4	24.7	24.1	23.6	24.6	24.1	25.	25.0	
0.35	10.2	14.7	15.6	16.2	16.9	17.7	18.7	19.5	20.3	21.4	21.8	22.1	22.7	21.9	21.5	22.0	22.1	21.9	22.3
0.40	9.9	13.5	14.1	14.8	15.4	16.1	16.7	17.5	18.2	19.2	19.5	20.4	20.6	20.8	20.9	20.7	19.9	20.3	20.8
0.45	8.5	12.4	13.0	13.3	14.1	14.6	15.3	15.7	16.3	17.0	17.6	18.4	19.0	19.1	19.6	19.1	19.0	19.2	18.8
0.50	8.5	11.4	12.1	12.4	12.9	13.3	13.9	14.4	14.9	15.4	16.0	16.6	16.9	17.3	17.8	18.2	18.5	18.9	18.1
0.55	8.0	10.6	11.1	11.4	11.9	12.3	12.8	13.4	13.8	14.2	14.5	15.0	15.6	16.1	16.4	16.5	16.7	16.7	17.0
0.60	7.7	10.2	10.3	10.7	11.2	11.4	11.7	12.3	12.7	13.0	13.4	14.1	14.3	14.9	14.9	15.3	15.6	15.5	
0.65	7.7	9.2	9.6	10.1	10.3	10.7	11.1	11.4	11.6	12.0	12.5	12.8	13.0	13.4	13.8	14.1	14.7	16.8	14.8
0.70	7.1	8.8	9.0	9.3	9.6	10.0	10.4	10.5	11.1	11.2	11.6	11.9	12.2	22.7	12.7	13.1	13.5	13.8	
0.75	6.2	8.3	8.5	8.8	9.0	9.5	9.6	9.9	10.2	10.4	10.8	10.8	11.4	11.7	11.9	12.2	12.4	16.6	13.0
0.80	6.1	7.8	8.1	8.3	8.4	8.6	9.0	9.0	9.6	9.8	10.0	10.3	10.5	10.8	11.1	11.4	11.6		
0.85	6.1	7.4	7.6	7.8	8.0	8.2	8.5	8.6	9.1	9.2	9.4	9.8	9.8	10.0	10.4	10.6	10.7	[illegible]	11.3
0.90	5.7	7.0	7.1	7.2	7.6	7.7	7.9	8.3	8.4	8.7	8.7	9.1	9.3	9.5	90.0				
0.95	5.2	6.7	6.7	7.0	7.2	7.4	7.5	7.6	7.9	8.1	8.4	8.5	8.6	8.8	9.1	9.4	9.5	9.7	10.
1.00	5.3	5.9	6.4	6.6	6.7	6.9	7.1	7.4	7.5	7.7	7.7	7.8	8.3	8.4	8.5				

Table.II: k figures for various AP and AC values.

k

	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
	9.8	5.4	6.5	9.1	10.8	11.5	12.1	12.8	13.1	13.3	13.5	13.9	14.0	14.	14.2	14.3	14.5	14.5	14.6
	9.4	5.2	5.7	3.0	4.1	6.2	7.7	8.7	9.5	10.0	11.1	11.5	11.8	11.7	12.0	12.4	12.5	12.8	13.0
	8.0	4.8	2.7	2.5	2.2	1.9	2.8	4.7	7.1	7.9	8.5	8.8	9.6	10.1	10.4	10.6	10.9	11.1	11.5
	8.0	4.5	2.9	2.4	2.2	1.6	1.7	1.3	2.3	3.7	4.9	5.7	6.5	7.3	8.0	7.9	8.5	9.3	9.5
	8.4	4.2	2.8	2.3	1.9	1.4	1.5	1.2	1.2	1.1	1.8	3.1	4.1	4.8	5.8	6.0	7.0	7.7	8.2
	7.8	3.9	2.9	2.1	1.6	1.6	1.3	1.2	1.1	1.0	.9	.9	1.6	2.7	3.6	4.4	4.9	5.9	6.2
	6.4	3.5	2.2	1.7	1.5	1.4	1.2	1.1	0.9	1.1	0.9	0.8	0.8	0.8	1.4	2.4	3.3	3.8	4.3
	5.8	3.5	2.3	1.9	1.4	1.2	1.1	1.2	1.0	1.1	.8	.8	.7	.7	.7	0.8	1.2	2.2	3.1
	6.4	3.1	2.0	2.0	1.4	1.5	1.2	1.0	.8	.7	.8	.9	.8	.7	.7	.6	.5	.6	1.1
	5.7	3.2	2.3	1.8	1.5	1.1	1.0	.9	.7	.7	.7	.7	.6	.6	.6	.6	.6	.7	.6
	5.4	3.0	2.1	1.8	1.2	1.2	1.0	.9	.8	.7	.7	.7	.7	.6	.6	.5	.5	.5	.5
	5.2	2.6	1.8	1.6	1.3	1.1	.9	.9	.8	.7	.6	.6	.6	.6	.5	.5	.5	.4	.4
	4.5	2.6	2.0	1.3	1.2	1.1	.9	.9	.7	.6	.7	.6	.6	.5	.5	.5	.6	.5	.5
	4.3	2.3	1.7	1.4	1.3	.9	.8	.7	.6	.7	.6	.5	.6	.5	.5	.5	.5	.5	.5
	4.7	2.3	1.7	1.2	.9	.9	.8	.8	.6	.6	.5	.5	.6	.5	.4	.5	.4	.4	.4
	4.3	2.3	1.5	1.1	1.0	.8	.7	.7	.6	.5	.5	.5	.4	.4	.5	.4	.4	.4	.4
	3.6	2.0	1.4	1.1	1.0	.9	.7	.7	.6	.6	.5	.5	.4	.5	.5	.4	.4	.4	.4
	3.6	1.8	1.3	1.	.9	.7	.6	.6	.5	.5	.5	.4	.4	.4	.4	.4	.4	.3	.4
	3.4	1.7	1.4	.9	.6	.7	.6	.6	.5	.6	.4	.6	.4	.4	.4	.4	.3	.3	.3
	1.3	1.7	1.2	1.	.7	.6	.4	.5	.5	.4	.4	.4	.4	.4	.3	.3	.4	.4	.4

We have delineated on these tables also roughly of SNR and k regions. Obriously there are several combinations of step and corridor sizes to attain a certain SNR and k standard. An interesting region occurs for $0.35 \leqslant AC \leqslant 0.70$ and $0.40 \leqslant AP \leqslant 0.80$ where one has SNR values corresponding roughly to the values attained by various mediumband speech coding techniques, i.e., $10 \leqslant SNR \leqslant 20$ and $0.5 \leqslant k \leqslant 1.2$.
The plots of the SNR and k factors are shown in Figs. 5 and 6 where we can observe that the choice of the AP parameter is much more critical for small corridor sizes as compared to relatively large corridor widths.
In the table below the dependences on the corridor width of the SNR-maximizing stepsize (AP^x) and k-minimizing stepsize (AP^{xx}) are indicated,

AC	0.01	0.1	0.3	0.5
SNR	40	30	18	12.5
AP^x	0.05	0.15	0.35	0.75
k	12	5	0.95	0.70
AP^{xx}	0.05	0.20	0.45	0.85

where it is seen that AP^x and AP^{xx} are often quite close to each other.

Stepsize adaptation: Consider an ASDM scheme where the stepsize is made to depend adaptively on the bit polarity pattern or the IBI process. Among various alternatives a simple scheme, whereby the stepsize is increased by a factor $APCM_+ > 1$ if ++ or -- occurs and it is reduced by a factor $APCM_- < 1$ if an alternating pattern + - or - + occurs, has been used. (see Fig. 4c). This algorithm is reminiscent of Jayant's adaptive PCM algorithm (14).
SNR and k results for various $APCM = APCM_+ / APCM_-$ pairs are shown in Fig.7. The stepsize adaptation didn't bring about any improvement in the peak SNR nor any further economy in the k factor. The only advantage was to widen the range of optimality of the AC-AP parameters.

Variable Corridor Size: A variation of ASDM is to change the corridor size depending on the size of the signal amplitude after each sampling instant. For example one can assign a larger corridor size if the signal amplitude is large and viceversa (Fig. 4d). With such a sceme one allows for rougher quantization at larger amplitudes, thus incurring into some SNR loss which would be traded for some reduction in the k factor. The corridor adjustment was carried out according to a Max-Lloyd quantizer steps, i.e., the corridor size was expanded by the same ratio as that of the corresponding stepsize in the quantizer to its minimum stepsize. However the advantages brough about by this scheme were also rather insignificant as shown in Fig. 8.

ASDM with Prediction: In the feedback loop in Fig.2 one can use a linear predictor of p'th order, where the predictor is used to predict succesive samples separated in time by T/β sec. (Fig. 4e). However the predictor, which was designed based on the statistics of the input process, is now being excited by an "unnatural" process, i.e., the reconstructed signal ($\hat{x}_k$) which may be a constant level waveform for several (e.g.,20) T/β subintervals. Consequently, the SNR gains when a linear predictor is used, are quite modest (%20) at the expense of a very large increase in the k factor.

Sampling time: In a practical realization the allowed values of the sampling intervals are also quantized. One can conjecture furthermore that as the sampling rate β decreases (minimum allowable sampling interval increases) the benefits of ASDM will also diminish. Indeed, especially at narrow corridor widths, the SNR performance decreases drastically with decreasing β. This sensivity, however, is much more subdued at larger (e.g. AC$=$ 0.35) corridor sizes.(Fig. 9-10). It appears that for speech-like signals studied in this work, a sampling rate of 8 to 16 times the Nyquist rate (β=8 to 16) suffices.

5. THE IBI PROCESS

The probability density function of the IBI process for various statistical speech models were analyzed in (1). The effects of quantizing the IBI process on the SNR valves and the total number of bits used are shown in Figs. 11 and 12. The quantization of the interbit intervals reflects to the signal reconstruction process as shown in Fig. 4 f. The following observations have been made:
i) The SNR performance reaches a saturation level soon with only 2-3 bit quantizers (actually 1-2 bit quantizer since the IBI valves are positive). ii) We have used two and optimum Gaussian quantizer. Similar performance results have been obtained with both. In particular the SNR valve is not critically dependent upon the μ-factor, though μ= 25 gave relatively better results. iii) For 10 to 16 k bit range one can obtain satisfactory performance (e.g. Ac$=$0.30, AP$=$0.45) vis-a-vis other source coding algorithms for medium band speech. (iv) There seems to be further room for SNR improvement in cosidering more complicated quantization schemes for the IBI process, especially at low bit rates.

6. CONCLUSION

A sequential quantizer with instantaneously variable sampling rate has been investigated. It has been demonstrated that the ASDM scheme is a viable and attractive alternative for medium band speech coding, i.e. in the 12 to 16 kb/s range. Furthermore this method is computationally very simple. The signal adaptive sampling that converts the input signal into an IBI process as well as a polarity sequence seems to exhaust most of the redundancy of the signal such that none of the adaptation strategies (i.e., stepsize adapatation) has brought about any significant improvements. There seems to be further room improvement, however, in the form of differential or block quantizing of the IBI process. Other questions that remain to be answered are the buffer statistics, spectral shaping of the quantization noise and effects of the channel errors. Finally an intriguing question is the effectiveness of tree codes with variable time quanta.

References

1-B.Sankur,"Signal coding Properties of Asynchronous Delta Modulator", Proc. Int. Conf. Acoust Speech Sig. Processing, 1704-1709, Paris (1982).

2-T.A. Hawkes, P.A. Simonpieri "Signal Coding Using Asynchronous Delta Modulation", IEEE Trans. Comm, COM-22, 346-348 (1974).

3-J.W.Mark, T.D.Todd, "A Non-uniform sampling Approach to Data Compression", IEEE Trans. Comm., COM-29, 24-33 (1981).

4-K.Murakami, et al., "Variable Sampling Rate Coder" Tech. Rep., 26, Univ. Osaka, 499-505 (1976).

5-S.J.Campanella, "Digital Speech Interpolation Techniques", Proc. IEEE Nat. Telecomm. Conf., 14.1.1-14.1.5 (1978).

6-J.P.Adoul, F.Daaboul,"Digital TASI Generation with Voiced/Unvoiced Discrimination for Tripling T1 Carrier Capacity", Proc. IEEE Int. Comm. Conf., 32.5.310-32.6.314, Chicago (1974).

7-S.A.Webber, C.J.Harris, J.L.Flanagan,"Use of variable Quality Coding and Time-interval Modification in Packet Transmission of Speech", Bell Sys. Tech. Journal, 56, 1569-1573 (1977).

8-J.J. Dubnowski, R.E. Crochiere, "Variable Rate Coding of Speech", Bell Sys. Tech. J., 58, 577-599(1974).

9-J.L.Cicero, P.P.Prezas,"Doubly Adaptive Delta Modulation for Bandwidth Compression of Speech Signals", Proc. Nat. Telecomm. Conf.,36.5.1-36.5.5 (1980).

10-N.S.Jayant, S.W. Christensen, "Adaptive Aperture Coding for Speech Waveforms", Bell Sys. Tech. J., 58, 1631-1645 (1979).

11-J.Barba et al.,"A Modified Adaptive Delta Modulator", IEEE Trans. Comm., COM-29, 1767-1785 (1981).

12-A.g.Tescher, V. Cox, "Image Coding: Variable-Rate DPCM Through fixed Rate Channel", Proc. Soc. Photo-Optical Inst. Eng.,119, 147-154 (1977).

13-H.Meyr, H.C. Rosdolsky, T.S. Huang, "Optimum Run Length Codes", IEEE Trans. Comm., COM-22,826-833 (1974).

14-N.S.Jayant, "Digital Coding of Speech Waveforms: PCM, DPCM and DM Quantizers", Proc. IEEE, 62, 611-632 (1974).

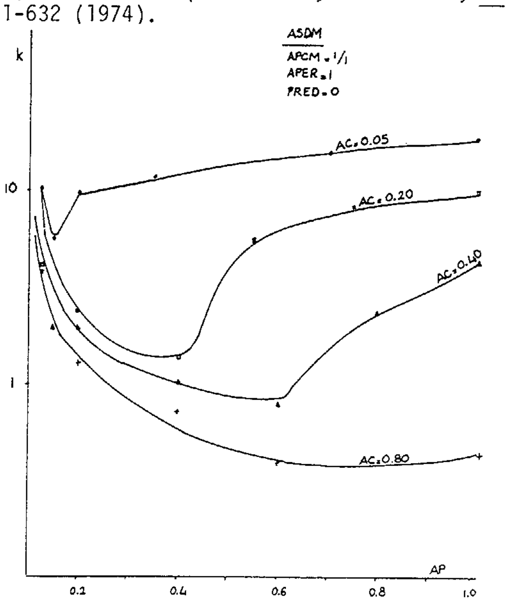

Fig.6. Plot of the oversampling factor vs AP.

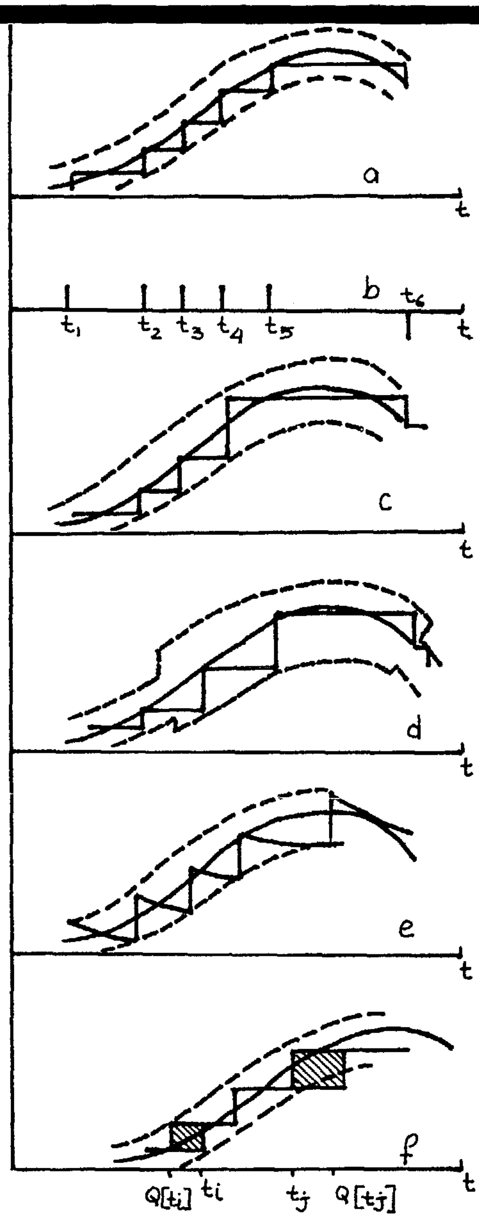

Fig.4 Waveforms: a) ASDM b)sampling instances c) ASDM with adaptive stepsize d)ASDM with variable corridor e)ASDM with prediction f)Time quantization error

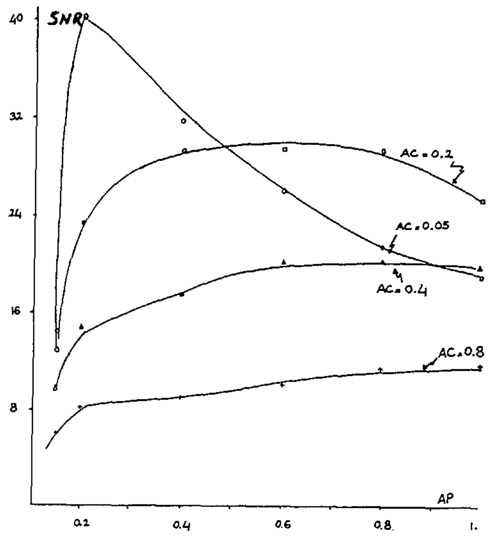

Fig. 6.Plot of SNR and k vs corridor aperture.

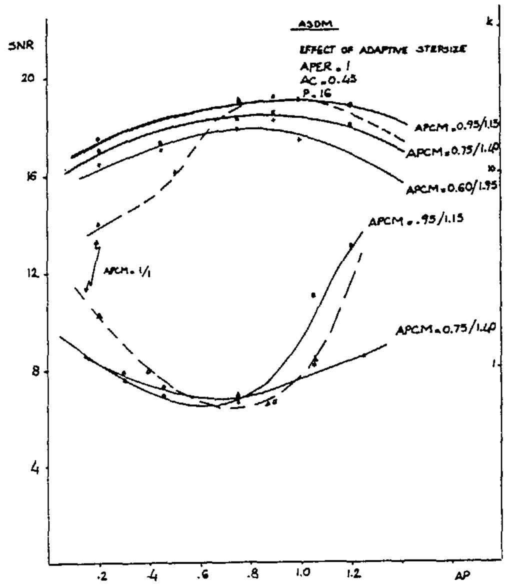

Fig.7. Plots of SNR and k vs stepsize adaptation parameter.

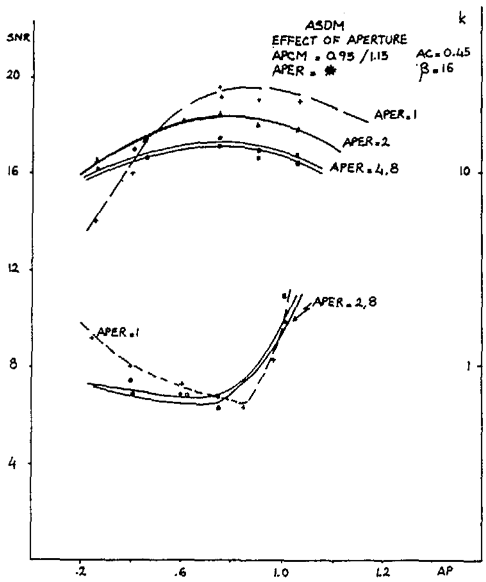

Fig.8 Plot of SNR and k vs corridor aperture.

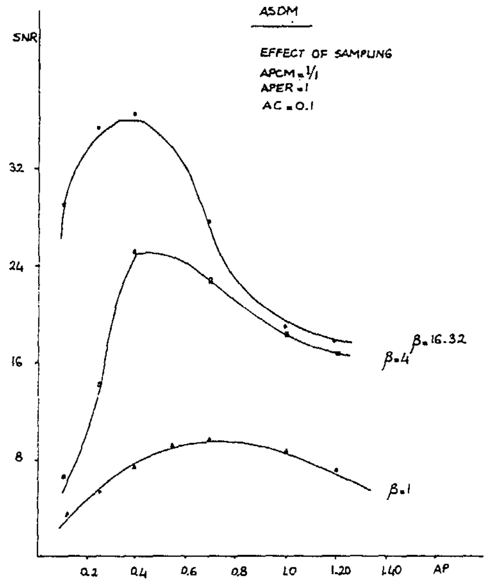

Fig.9 Effect of Sampling; SNR vs AP

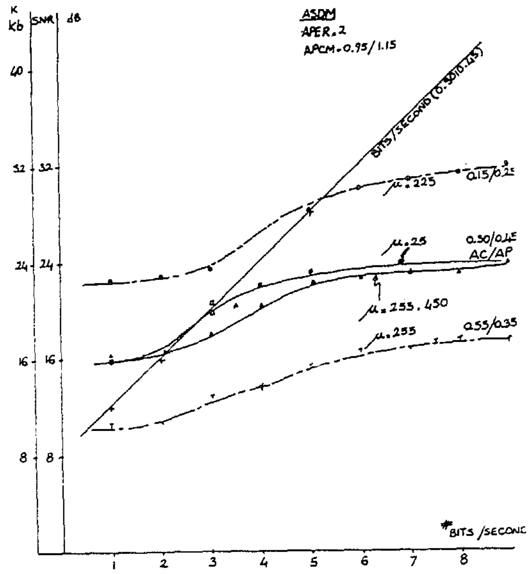

Fig.11 SNR vs interbit quantization (logarithmic companding).

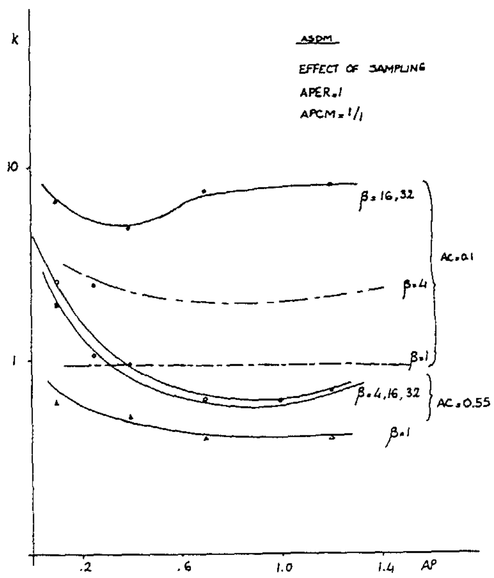

Fig. 10 Effect of sampling ; k vs AP

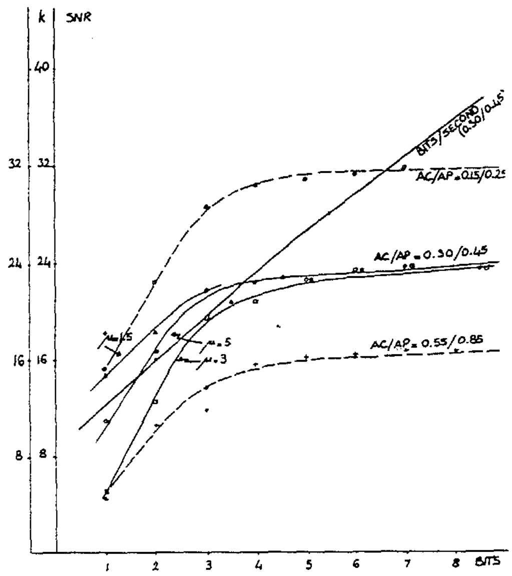

Fig. 12 SNR vs interbit interval quantization (Gaussian quantizer).