Synopsis

Introduction :

The advantages of transmitting speech signals in digital form are well known, and several systems for the conversion of the analogue speech into the digital form have been described in the literature (J-4, S-12). There are areas of communication, however, where the high transmission rate associated with the digitized speech is considered undesirable. Thus it may be necessary to digitize speech at sufficiently low bit rates to enable commercially available data modems to be used and a satisfactory signal transmitted over audio lines. All the advantages of digital communication including secrecy can thus be made available over commonly used facilities. Also the available transmission channel bandwidth requirements are reduced considerably with these low bit rate systems and many more channels can be accommodated in an existing communication system - the only limitations being the acceptability of the reproduced speech and the cost involved.

The low bit rate digitization techniques can be subdivided in two main classes of speech encoders. In one class, the speech signal is analyzed to extract some significant parameters which define it and from which the speech can be synthesized back. These parameters are then digitized and transmitted to the receiver. At the receiver the speech is synthesized from the received parameters, which are now an approximation to the original values because of the quantization involved in the transmission. Systems using this 'analysis-synthesis' approach include the digital forms of the channel vocoder (D-8, B-10), the formant vocoder (S-13), the adaptive predictive coder (A-3) and the linear predictive coder (A-4) etc.

In the second class of encoders, the basic approach is not to analyze the speech signal for its significant parameters but to digitize in such a way as to reproduce the waveform faithfully. Such encoders include the Pulse Code Modulation [PCM] (R-1), the Differential Pulse Code Modulation [DPCM] (C-2) and the Delta Modulation [DM] coders (D-1, D-2) etc. Several systems which are a hybrid of the two classes of encoders have also been reported in the literature (D-6, Q-1, C-13). It now seems certain that the 'analysis-synthesis' approach encoders are potentially capable of obtaining an acceptable quality of speech reproduction at the minimum possible data rates. These systems are in general very complex, though this fact has lost much of its significance in the light of contemporary advances in integrated circuit technology. At the present level of technology, several of these systems are also not yet able to achieve this digitization on a real time basis.

In the class of waveform digitizers, PCM was the earliest technique developed where the waveform was sampled, quantized and finally coded in a binary format. Communication links using PCM have been in operation for a long time. The differential PCM, invented later, involved the coding of only the incremental values of the waveform and thus took advantage of the statistics of the signal. At about the same time, the Delta Modulation was proposed which, though conceived independently of DPCM, is a special case of differential encoding - being essentially a one bit oversampled DPCM system. The early DM failed to evoke much interest due to the severe limitations in its dynamic range and consequently poor speech reproduction. Matters improved with the introduction of the adaptive DM (ADM) schemes. Amongst the many factors which contributed to the vigorous development of the ADM system was the simplicity of its implementation. In addition to its simplicity, the ADM system has the promise of a better quality of reproduction at low bit rates than is possible with PCM/DPCM.

Most of the ADM coders described in the literature operate at the medium bit rates of 16 to 64 Kb/s approximately and are now commercially available as integrated chips (F-14). Above these bit rates, the ADM is not popular because of the better performance of PCM type coders. At bit rates below 16 Kb/s, the existing ADM coder strategies are not able to cope well with some of the problems which arise in the encoding process. Straight forward encoding of speech using PCM/DPCM at low bit rates is not suitable because of the excessive quantizing noise generated. The problems faced in the ADM encoding can be broadly classified into two main categories. In the first place, the ratio of the sampling frequency to the input signal frequency is small and hence the number of steps taken by the coder during one cycle of the input signal becomes rather small. Consequently a relatively large step size has to be used to handle a given signal amplitude. This implies a large quantization noise which is subjectively irritating to the listener. Another implication of the same restriction is that the higher input signal frequencies, which are small in amplitude, are badly coded and result in a poor speech quality. The other source of trouble is the fact that, at these bit rates, the correlation amongst the output bits becomes small and the feedback control of the step size adaptation from the output digital stream becomes difficult. On the other hand a forward control of the adaptation from the input signal leads to difficulties in tracking at the receiver since the input signal is not available at the receiver.

Several authors have described systems in which these problems have been overcome or bypassed in various ways. These systems include the two channel delta modulation (G-7) in which the speech is digitized by an ADM system operating at various sampling rates in the range of 6.8 to 13 Kb/s. The step size control information is transmitted to the receiver through an auxillary coder using 400 bits/s sampling rate, and multiplexed with the main digital stream. This system reproduces the signal with an SNR of 18 dB when operating at 13.4 Kb/s. Dunn's system (D-6) uses a two coefficient adaptive prediction filter followed by an amplitude controlled delta sigma modulator in an open loop configuration. The main coder operates at 8 Kb/s while the predictor coefficients and amplitude information are transmitted at 1.6 Kb/s, to give an overall transmission rate of 9.6 Kb/s. In both the systems described above, the step size information has to be multiplexed into the digital stream. Moye's system (M-3) avoids this by using a self adaptive filter as an adaptive deconvolver to remove redundancy from the input, and the residue is coded by a 2 bit companded PCM for transmission. Quereshy & Forney (Q-1) have used Dunn's ideas, as modified by Melsa et al, (G-8) along with variable length self synchronizing codes to reduce the output bit rate. In this system an SNR of about 14 dB is obtained at 9.6 Kb/s. Crochiere et al (C-13) have used an entirely different approach to achieve a low bit rate coding. They have divided the speech into a suitable number of sub-bands and coded them preferentially - the important bands being coded in more detail while the other bands are coded in less detail. They have reported fairly good results at 9.6 Kb/s and lower.

The aim of the present investigations has been the design and study of encoders suitable for the waveform encoding of speech at bit rates below 16 Kb/s or more precisely in the range of 7-10 Kb/s. The DM format only has been used for these investigations and Adaptive Quantization with Backward Estimation [AQB] has been chosen to make the system adaptive. In the course of this work two different adaptation strategies of AQB were developed. In the first strategy, the control of the adaptation of the step size, though obtained from the output digital stream, is really proportional to the amplitude of the input signal, while in the second strategy, the control signal takes into account the slope of the input signal also.

Basically in a DM coder the error signal, obtained as a difference of the input and the reconstructed feedback signal is quantized into two levels, ± 1, and the consequent quantizing error is large. Obviously in such a situation a three level quantizer using '±1' and '0' levels will introduce lesser quantizing noise. Therefore for a given sampling frequency and thus, the channel band width, it will be possible to obtain a better SNR using this three level or ternary quantizer over the two level or binary quantizer. The idea of a ternary encoder coupled with a second level run length encoding seems to be very attractive. Using the AQB strategies described above, the author has successfully developed adaptive Binary and adaptive Ternary low bit rate DM systems.

The final bit rate of transmission in all waveform digitizers naturally depends upon the message bandwidth and the quality of signal reproduction desired at the receiver. It seems therefore natural to question the rationale of the message bandwidth being fixed in all such encoders. Thus, if the message bandwidth can first be reduced through a reversible process and subsequently digitized using waveform coders then a given quality of signal reproduction can be achieved at significantly lower bit rates. Some efforts have been made in this direction (S-17, M-8) but not many satisfactory systems have been reported in the literature. The author has developed a bandwidth compression- expansion scheme for speech signals where a bandwidth reduction by a factor of two or of four is obtained and a subsequent expansion to the original bandwidth is made at the receiver. Although this system does not fall into the category of an analysis - synthesis type vocoder, it is expected that this band compressed speech can be digitized at 4.8 Kb/s using APCM techniques.

The Adaptive Binary Delta Modulation Systems :

The twin objectives of a high SNR and a good stability in the Adaptive Binary DM Coder at low bit rates has been realized through the use of a two loop feedback coder configuration. A double integrator is used in one of the loops which is adaptive and active at high input levels and results in a high SNR. The other loop, using a single integrator is non adaptive and dominates at low input levels, giving rise to good stability and low idle channel noise.

The amplitude of the feedback pulse in the ADM coder - and thus its step size - is adapted by a control signal developed from the output pulse stream of the coder in the usual fashion (C-8). The N successive digits of the single integrator DM coder output are stored in an N bit shift register, and then compared by a coincidence detector. The detector gives an output whenever the N digits are identical - high or low. The output of the detector is then smoothened by a syllabic filter to give the control signal. This control signal is proportional to the amplitude as well as the rms slope of the input signal. It has been seen that if the sampling rate is less than 16 Kb/s, the control signal is best developed when N = 2. The control signal changes the amplitude of the feedback pulse through the use of an FET as a voltage dependent resistance. The amplitude of the feedback pulse thus becomes approximately proportional to the amplitude and slope of the input signal. The reconstructed feedback signal is obtained by integrating this pulse in a double integrator, and after suitable scaling forms the outer (adaptive) loop. The inner (non-adaptive) loop consists of a single integrator as mentioned earlier.

The system has been analyzed with sinusoidal inputs, and it is seen that for large signal levels, the SNR is of the form :

$$\mathrm{SNR}\Big|_{dB} = 20 \log k \frac{f_r^{3/2}}{f_m \; f_a^{\frac{1}{2}}} \qquad \ldots (1)$$

where f_r is the sampling rate

f_m is the signal frequency

f_a is the message bandwidth

and k is a constant dependent on the integrator breakpoints, and the gain of the feedback loops.

At small signal levels, the SNR is given by :

$$\left.\mathrm{SNR}\right|_{dB} = 20 \log k' \frac{V_m \, f_r^{3/2}}{f_a^{\frac{1}{2}}} \qquad \ldots (2)$$

where V_m is the signal amplitude and

k' is another constant.

It is seen from eq. (1) and (2) that different conditions prevail at the large and small input signal levels resulting in these two expressions of the SNR. Thus for the large signal levels the coder adapts the step size in such a way as to remain near the slope overload conditions and hence the SNR does not depend on the signal level. It, however, depends on the signal frequency because the signal frequency determines the step size leading to slope overload conditions at a given input level. At lower signal levels, on the other hand, the step size is almost constant and hence the SNR depends upon the signal level.

The step size in the coder is decided by the corner frequencies of the various integrators - single and double - used in the feedback loops, as well as by the scaling factors of the two loops. The double integrator allows the step size to increase much more rapidly in comparison to the output of a single integrator. Similarly, the frequency response, i.e. the output amplitude at the decoder for various input frequencies is very much dependent on the breakpoints of the integrators. Increasing the breakpoints improves the frequency response, but the quantization becomes poorer, resulting in a lowering of the SNR and vice-versa. Thus an optimization of the system requires a proper selection of the breakpoints of the integrators. This optimization has been done on a subjective basis and a suitable balance between the quantizing noise versus the frequency response has been arrived at. The optimized breakpoint for the single integrator is 800 Hz, and for the double integrator the breakpoints are 200 Hz and 800 Hz with a zero at 1.6 KHz.

With these optimized poles and zeros of the feedback loops, the maximum SNR has been theoretically calculated to be 20 dB for a sinusoidal input signal of 400 Hz at a bit rate of 10 Kb/s. Furthermore the dynamic range, i.e., the input signal range over which the SNR is above (say) 15 dB, has been calculated to be 20 dB for the particular scaling factors used in the feedback loops. It has been experimentally seen on the hardware wired system that at a sampling rate of 9.6 Kb/s, an input signal of 400 Hz frequency is coded with a maximum SNR of 19.5 dB, and the SNR remains above 15 dB for an input range of 17.5 dB below the overload conditions. The experimentally obtained values of the maximum SNR increase by about 9 dB for an octave increase in the sampling frequency. It has thus been seen that the values experimentally measured correspond fairly well with the theoretically calculated values and the present system compares well with other such systems reported earlier.

While the SNR and the dynamic range figures discussed above give a fair indication of the performance of a speech encoding system, they are not really representative figures, particularly at low coding rates. In view of this, the performance of the system has been subjectively evaluated using the "goodness rating test" (I-3) for the evaluation of the quality of the reproduced speech. While evaluating on this basis several different listening filters have also been tried out to obtain the best subjective performance. It was found that an optional notch filter just above the first formant used at the receiver improves the performance. When operating at 9.6 Kb/s the system has been rated as being "good" - the fourth grade on a five point grading scale. At 7.2 Kb/s, the system has been rated as "fair".

The subjective evaluation gives an overall idea of speech quality, including intelligibility, granular noise, speaker recognition and pleasentness of listening. Of these several subjective characteristics of speech, perhaps the single most important factor is the intelligibility. The intelligibility of the reproduced speech has also been measured for word intelligibility using the "modified rhyme test" (H-3). At 9.6 Kb/s the mean intelligibility score is 88%, while at 7.2 Kb/s, the mean value is 74% (L-3).

Adaptive Ternary Delta Modulation :

While the above system produces a fairly acceptable quality of speech reproduction at 9.6 Kb/s it was felt than an improvement of the quality may be obtained by the use of a three level quantization instead of the two level scheme used so far. The nature of this improvement can be understood from the following considerations. In the adaptive binary DM, the coder takes a step, either in the positive or in the negative direction at each sampling instant. Thus if the error just prior to the sampling instant is small, the step taken increases the error, irrespective of the direction of the step. To avoid this increase in error, the coder may be allowed a third alternative - as in the ternary quantizer - which is distinct from the taking of a positive or a negative step. This alternative is to take no step at all, thus retaining the reconstructed signal at the same level as prior to the sampling instant. In this case the error is kept small instead of being increased as in the binary DM, and thus the reconstruction of the signal is improved. Delta Modulators built on this premise, and having a three level output (-1, 0 or +1) are called ternary delta modulators [TDM] (F-2).

It has been shown in the present study that the quantizing noise generated in a TDM system is given as

$$N = \frac{\Delta^2}{3}\,(1 - 3t + 3t^2) \qquad \ldots (3)$$

where Δ is the step size

and t is the threshold as a fraction of the step size .

If the error signal ϵ_K is within $\pm\, t\Delta$, the coder output p_k , is 'zero' otherwise it is +1 or -1 depending on the sign of the error as shown below.

$$
\left.\begin{aligned}
p_k &= 1 \text{ if } \epsilon_k > t\Delta \\
p_k &= 0 \text{ if } |\epsilon_k| \leqslant t\Delta \\
p_k &= -1 \text{ if } \epsilon_k < -t\Delta
\end{aligned}\right\} \quad \ldots (4)
$$

It is seen from eqn. (3) that with everything else remaining the same, the noise is minimum when the threshold is adjusted to $t = 1/2$. In other words the noise power is 6 dB lower in the optimum ternary case in comparison to the binary case (when $t = 0$). In an experimental study, a linear TDM system has been built and tested for various values of the threshold 't'. The three level comparator really consists of two binary comparators and an adder. The feedback network is an RC integrator with a corner frequency of 800 Hz. The maximum SNR for a sinusoidal input of 400 Hz is 25 dB for a sampling rate of 9.6 Ksamples/s. This value is 5.5 dB better than the corresponding value of 19.5 dB for two level quantization. The dynamic range of the linear TDM is quite small (10 dB, for SNR $\geqslant$ 15 dB) as expected but can be increased by the use of a compressor - expandor pair. However, the use of this companding, external to the feedback loops, causes a deterioration in the frequency response and as such the linear TDM with external companding is not quite suited to speech coding at these low sampling frequencies. The other alternative of using an AQB type of encoding like the binary system seems to be quite attractive. Adaptive strategies have been developed to change the step size in order to match the input changes, thereby obtaining the necessary dynamic range.

Before going into the details of the adaptive strategies developed for producing the variable step size, it should be noted that an additional factor of threshold size has been added by a three level adaptive quantizer. The step size could be changed with the threshold kept fixed or both the step size and the threshold could be changed together. Max (M-1) has suggested an optimum quantizer for the three level quantization, but unfortunately his results are not directly applicable here, since the probability density function at the input of the quantizer is itself affected by the quantizer. It however, appears that if the step size and the threshold are changed by the same control voltage, the system may yield an optimum performance. We shall first describe systems in which the adaptive step sizes have been produced without disturbing the threshold.

In the adaptive TDM also, therefore, the step size is controlled in the adaptive loop by changing the amplitude of the feedback pulse through a control signal and the stability etc. is assured through the non adaptive feedback loop. The control signal is developed from the coder output, by checking the pulse pattern at the output, using either one of two different schemes. In one scheme, the output of the coder is integrated, full wave rectified and filtered through a syllabic filter to obtain the step-size control signal. It has been shown that the control signal here depends only upon the amplitude of the input signal and the system is therefore refered to as the Amplitude Controlled Adaptive TDM. In the second scheme, named as the Slope Controlled Adaptive TDM, the generation of the control signal is similar to the scheme used in the adaptive BDM. The output of the coder is stored in a shift register and the occurance of N consecutive 'ones' or N consecutive 'minus ones' is detected by a coincidence detector whose output is filtered by a syllabic filter to obtain the control signal. N = 2 has been found to be optimum at the sampling frequencies used. The control signal so generated depends on the amplitude of the input signal and also on its rms slope. The second scheme performs slightly better than the first, presumably because the Delta Modulator is basically a slope coder.

A simulation study of the adaptive TDM encoder using the second scheme (L-4) has given a maximum SNR of 20.5 dB at a sampling frequency of 8 Ksamples/s, the input being RC shaped gaussian noise. The value of the SNR remains above 15 dB for an input dynamic range of 22 dB. A hardware wired adaptive TDM system using either of the two schemes for adaptation has been experimented with in all details for various input signals : sinusoidal, noise and speech. It has been expected and also experimentally verified that the maximum SNR is almost the same for both the schemes. The dynamic range over which the SNR is greater than 15 dB, however, is more in the case of the slope controlled ATDM, being 20 dB as compared with the 16 dB achieved by the amplitude controlled

ATDM. The maximum SNR of 23 dB obtained in the adaptive TDM is less than that of the linear TDM because of the fixed threshold as speculated earlier. The fall in SNR can be understood from eq. (3), where it is seen that the optimum threshold is a definite fraction of the step size. Thus, if the threshold is kept fixed while the step size is changed with the input, then clearly the threshold can be optimum for some inputs only while for the other inputs the SNR will be less than the optimum. In the two systems described so far the threshold and hence the SNR was optimum at the minimum step size and as such the maximum SNR achieved is less than the maximum expected. To obtain the optimum performance, therefore, it is necessary to vary the threshold size along with the step size.

A variable threshold ATDM has been designed where the control signal modifies the comparator threshold size along with the changes of the step size. It has been seen that for a sampling frequency of 9.6 Ksamples/s and an input signal of 400 Hz, the maximum SNR is 25 dB, which is roughly 2 dB better than the value obtained in the case of the fixed threshold schemes. However the dynamic range is not extended, since the variation of the threshold does not produce any significant change at the lower signal levels, where the fixed threshold itself had been close to the optimum. Extensive noise and speech input tests have been made and the results have been reported. The subjective quality of the variable threshold ATDM operating at 9.6 Ksamples/s has been rated as '4.7' on the 'goodness rating test'.

The effect of channel errors :

The effect of channel errors on the adaptive binary and ternary DM decoders and the degradation of the performance because of these errors has been studied in detail. It has been assumed that both for the binary and the ternary systems, the proper channel modems and equalizers are available in all the cases of baseband or radio frequency transmissions. The effect of the channel errors has been analyzed in terms of P_E , the probability of error, and it has been shown that the degradation in the received SNR i.e. the SNR at the output of the receiver, in the binary DM is given by the equation.

$$SNR\Big|_{dB} = -10 \log \left(1 + 24 \frac{f_r}{f_m} P_E \right) \qquad \ldots (5)$$

In the ternary case, on the other hand, one error may result in a single change of level, i.e. due to an error a + 1 changes to 0 for a - 1 changes to 0 etc. We shall assume the probability of such errors as P_1. It is also possible that one error causes a double change of level, i.e., due to an error a + 1 changes to a - 1 or vice-versa. Assuming the probability of such errors to be P_2 , the degradation of SNR has been calculated as

$$SNR\Big|_{dB} = -10 \log \left(1 + 24 \frac{f_r}{f_m} P_1 + 96 \frac{f_r}{f_m} P_2 \right) \ldots (6)$$

It has been shown that these formulae are valid both for the linear DM and the syllabically adaptive DM systems except for an additional effect of a marginal compression due to channel errors in the adaptive cases.

For the various modes of transmission like ASK, FSK, PSK etc., the requirements of channel band width have been linked with the sampling rate in the cases of the binary and the ternary systems. Similarly the channel errors have been computed for various channel SNR. These results have been used in conjunction with eq. (5) and (6) to compare the communication efficiency of the binary and the ternary systems. In this comparison it has been assumed that the channel bandwidth is fixed, and hence, with the use of suitable modems, the binary and ternary systems are operated at the same sampling rate. If there is no channel noise, then, as seen from eq. (3), the received SNR is 6 dB better in the ternary case than in the binary case. Because of the channel noise, however, the received SNR will fall in both cases - the fall in the ternary case being more. This is because of two effects. Firstly, the probability of error in the ternary case is more than that in the binary case, since the channel SNR has been assumed to be the same in both cases. Secondly, for the same probability of error the degradation of the received SNR is more in the ternary case as seen from eq. (5) and (6). For example, in the case of PSK with a total probability of error of 10^{-3}, the values of P_1 and P_2 are equal and are 5×10^{-4}. The degradation in the SNR in the ternary is calculated to be 0.96 dB while the degradation in the binary case is 0.40 dB showing that the received SNR in the ternary is still superior to that of the binary system. For the same channel SNR, therefore, we find that the received SNR of the ternary system, though superior at larger values of the channel SNR, is poorer than that of the binary system at values below a certain threshold. Thus for a given bandwidth, the ternary system would deliver a better signal quality provided an appropriate modulation technique is used. If, however, the ternary signals are converted to straight binary before transmission over a channel then the ternary system would have no justifiable advantage. On the other hand, a secondary coding of the ternary output into a binary code using run length codes etc., for example, may take advantage of the redundancy and the quiet passages in speech to reduce the effective bit rates to smaller values. Thus, either ternary modems and equalizers have to be used or proper secondary coding has to be used in order to take full advantage of the adaptive ternary DM proposed here.

Bandwidth Compression :

As discussed earlier it is feasible to reduce the bandwidth of the speech signals by taking advantage of the redundancy inherent in the speech. Of the several types of band reduction techniques the band reduction by frequency division is of particular interest in the context of waveform digitization. The speech bandwidth was therefore reduced by the technique based on analytical square rooting (S-10). It is well known that taking the square root $\sqrt{\sigma(t)}$ of an analytical signal $\sigma(t)$ reduces the bandwidth of the signal by half. The inverse process of squaring restores the original bandwidth though extra noise and distortion components get introduced in this reverse process. Both the square rooting and the squaring can be easily implemented if the original signal is of the form :

$$s(t) = a(t) \cos \theta(t) \qquad \ldots (7)$$

and the bandwidths of the amplitude component $a(t)$ and the phase component $\cos \theta(t)$ do not overlap. However, speech signals by themselves are not available in this form. Various approximations are, however, possible wherein the speech signals can be transformed into this format and the square rooting of the corresponding analytic signal can easily be done in an approximate way. These transformed signals are used to obtain the requisite band compression, by a separation of the amplitude and phase components.

As an example of this approach, if the speech is separated into its three prominent formants, then each formant by itself can be thought of as an analytic signal which can be processed. The individual formant signals are represented by eq. (7) and band compression by the square rooting seems to be quite attractive. However, in practice there are many difficulties associated with this process and the resulting system does not perform well. Even for a compression of two to one, the ratio of bandwidth to the centre frequency of the formant is quite large and a frequency division and squaring results in large amounts of spectrum overlap and irreversible fold over distortion. Interestingly, if the formant signals are translated to a higher frequency band then some of the fold over distortion component can be avoided and a suitable band pass filter can cut off the unwanted spectral components. The distortions introduced during the expansion also remain quite manageable at low compression ratios of two to one and four to one. Listening tests on the reproduced speech indicate a fairly good performance of '4.6' for a bandwidth reduction of two to one and rating of '4.2' for a reduction of four to one, using the goodness rating tests.

In another variation of the above scheme the whole speech frequency band is translated up to a higher frequency band without seperating it into formants. Obviously such a crude process results in a poorer performance and has only the merit of simplicity. It has been experimentally determined that in this case the four to one compression expansion scheme has a goodness rating of '4.1'. Extensive noise and sinusoidal signal tests have also been carried out on this scheme.

Conclusions :

Studies have been carried out on the digitization of speech using several configuration of binary and ternary DM type encoders operating at low bit rates. These studies have been made both in the software and in the hardware models. These studies involved both the experimental measurement of the systems objective performance, and a statistical evaluation of the systems' subjective quality.

The results of the experimental measurements, using sinewave inputs have been summarized in Table - I. From this table we see that the maximum SNR obtained by the binary systems is 19.5 dB if the input signal is a 400 Hz sinewave and the sampling rate is 9.6 Ksamples/s. With similar testing conditions the fixed threshold ternary systems produce an SNR of about 23 dB which improves to 25 dB for the variable threshold ternary system. The dynamic range of the slope controlled ternary systems is seen to be 25 dB, while that of the amplitude controlled ternary systems is 21.5 dB.

In table - II, the various aspects of these coders have been compared on the basis of the results listed in Table - I. Thus it is seen that the ternary system has a better SNR than the binary system. Though the degree of companding is more in the binary case, the over all performance of the ternary system is better. The slope controlled adaptation is seen to perform better than the amplitude controlled adaptation in respect of the dynamic range, but the two are similar in respect of the SNR. The fixed threshold scheme is seen to be inferior to the variable threshold scheme as far as the SNR is concerned.

In table - III a comparison has been made of the subjective performance of the various schemes. It is seen that, subjectively assessed, the variable threshold adaptive ternary delta modulation is found to be the most pleasing system.

In table - IV, the performance of the systems developed by the author is compared with the performance of systems reported elsewhere. It is seen that the various systems perform similarly, with slight differences in the received SNR. Thus the author's adaptive TDM produces an SNR of 25 dB and the adaptive BDM produces 19.5 dB, comparing favourably with Greefkes' 2 channel ADM which produces only about 10 dB, and Dunn's system which produces a 16 dB SNR. It may be noted, however, that part of this improvement in SNR is due to the restriction of the speech bandwidth to 2400 Hz, which the author has employed.

The fact that all these different systems achieve only a comparable performance seems to imply that this is possibly close to the limit achievable by the DM systems at these rates. To improve the performance beyond these limits, some different ideas may have to be incorporated in the digitization scheme. One possible solution which has been offered here is to reduce the bandwidth of the speech signal by square rooting technique before digitization. That such a bandwidth reduction is feasible has also been experimentally demonstrated.

Thus we have established the feasibility of obtaining an acceptable quality of speech transmission by digital systems operating at low bit rates. Though the coders described here operate at bit rates of the order of 7 - 10 Kb/s, the use of bandwidth compression of the order of four to one coupled with

APCM coders of 3-4 bits/sample will lead to good quality waveform digitizers operating at around 4.8 Kb/s, and it is felt that such systems would be developed in the future.

While the ternary DM encoders have been developed in the present study, they cannot be usefully employed without proper modems and equalizers and hence it is necessary to design and study these modems and equalizers. Furthermore, to take advantage of the redundancy in the speech signals as well as of the silent portions of speech as encoded by the ternary systems, it may be useful to study the nature of the patterns which form in the output stream of the ternary coder. It is felt that with a knowledge of the distribution of these patterns, the ternary output could be encoded into binary patterns using run length codes.

Thus in conclusion we can say that using these simple hardware systems, as developed during the course of the present work, and the ideas of future work outlined above, an acceptable quality of digital speech can be obtained at bit rates as low as 7.2 Kb/s. Further, with the use of speech band compression and/or run length coding, this rate can be reduced to lower than 4.8 Kb/s, and can then be transmitted over telephone channels with suitable modems.

Table - 1

Summary of the performance of the systems investigated here

Property \ System	BADM	A.C. ATDM	S.C. ATDM	VT/A.C. ATDM	VT/S.C. ATDM
Peak SNR at 400 Hz f_r = 9.6 Ks/s	19.5 dB	23 dB	23.5dB	24.5dB	25 dB
Dynamic range at 400 Hz f_r = 9.6 Ks/s (range from +5 dB level to level where SNR 15 dB)	22 dB	21.5dB	25dB	21.5dB	25 dB
Peak SNR vs f_r	Increases at about 9 dB/octave				
Peak SNR vs f_m	Falls at about 6 dB/octave				
Frequency response at 0 dB (fall at 2400 Hz)	-10 dB	-10 dB	-10 dB	-10 dB	-10 dB
at -10 dB	Flat	Flat	Flat	Flat	Flat

Table - 2

Comparison of the systems investigated here

Property \ System	Binary vs Ternary	Amp. control vs. slope control.	Fixed vs variable threshold
Peak SNR	Ternary better by 6 dB	Same	Variable better by 2 dB
Degree of companding	Binary better	Slope control better	Same
Frequency response	Same	Same	Same
Speech Quality	Ternary better	Slope control better	Variable better

Table - 3(a)

Subjective Preference charts of the DM systems at f_r = 9.6 Kb/s

	BADM	CTDM	ATDM	VT.ATDM
BADM	-	1	0	0
CTDM	9	-	3	2
ATDM	10	7	-	4
VT ATDM	10	8	6	-

Table - 3(b)

Subjective comparison of the control signal generators

	Amp. Controlled ATDM	A.C. V.T. ATDM
Slope controlled ATDM	6;4	-
S.C. V.T. ATDM	-	6;4

Table - 4

Comparison with other low bit rate coders

System	BW	Trans-mission rate	Speech Quality	SNR	Remarks
1) Binary, 2 loop ADM developed by the author	2400 Hz	7.2 Kb/s	3	17.5 dB	1) Hardware model built
		9.6 Kb/s	4	19.5 dB	
2) Ternary ADM developed by the author.	2400 Hz	6.4 Ks/s	Better than corres-ponding B,ADM	19 dB	1) Hardware model built
		7.2 Ks/s		23.5 dB	
		9.6 Ks/s		25 dB	
3) 2 channel delta modulator	300 Hz to 3 KHz	6.4 Kb/s to 13.4 Kb/s	1.3	6 dB	1) Hardware model.
			4	18 dB	2) Requires multiplex-ing of 2 channels.
					3) SNR below overload.
4) Dunn's system	3400 Hz	9.6 Kb/s	= 4 bit log PCM	= 16 dB	1) Requires multiplex-ing of prediction coefficients.
					2) Very complicated.
					3) Hardware model built.

Table - 4 (Contd.)

System	BW	Transmission rate	Speech Quality	SNR	Remarks
5) Moye's system	2400 Hz	9.6 Kb/s	Adequate for Military purposes		
6) Quereshi & Forney's system	3000 Hz	9.6 Kb/s	=16 Kb/s CVSD	13.74 dB	1) Requires long buffers
		16 Kb/s	=32 Kb/s CVSD	17.63 dB	2) Only simulated
7) Crochiere's Sub band coders : fixed sub bands		4.8 to 16 Kb/s	For fixed bands 7.2 Kb/s = 17.2 Kb/s ADM		1) Very complicated
Variable sub-bands			For variable sub bands : 4.8 Kb/s = 17.2 Kb/s ADM		2) Only simulation results available.

List of Symbols and Abbreviations

a	a constant
A	variable gain ; threshold
ADM	adaptive delta modulation
ADPCM	adaptive differential pulse code modulation
ADΣM	adaptive delta sigma modulation
AI	articulation index
APCM	adaptive pulse code modulation
ATDM	adaptive ternary delta modulation
b	a constant
BDM	binary delta modulation
C	crest factor; capacitance
CDM	continuous delta modulation
DCDM	digitally controlled delta modulation
DM	delta modulation
DPCM	differential pulse code modulation
D M	delta sigma modulation
$e(t)$	excitation function
E	error
E'	error just after sampling
E_k, $E(t)$	error at kth sampling instant or at time t
$E(f)$	amplitude spectrum of excitation function
f	frequency; input frequency
f_a	input bandwidth
f_m	input frequency

f_r	sampling frequency
G_1, G_2	loop gain
$h(t)$	impulse response
$H(f), H(s)$	transfer function
i, j	index variable
k, K	constants
$L_n, L(t)$	digital output at nth clock pulse or time t
LDM	linear delta modulation
LPF	low pass filter
LTDM	linear ternary delta modulation
N	noise; limit of index variable.
N_f	noise spectral power density
N_q	quantizing noise
NRZ	non return to zero
p_i	probability of run of length i
$p(t)$	pulse output at time t; p.d.f. of t
P	probability
P_E	probability of error
PCM	pulse code modulation
r	resistance (variable)
R	resistance (fixed)
RZ	return to zero
$S(t)$	signal
$S_c(t)$	frequency translated signal
$\hat{S}$	predicted or estimated value of S ; Hilbert transform of S
S	signal power

S_f	signal power spectral density
SAF	self adaptive filter
SNR	signal to noise ratio
SNR_c	channel SNR
t	time; threshold
T	clock interval
T_1, T_2	loop transfer function
TDM	ternary delta modulation
v	voltage (variable)
V	voltage (constant)
V_m	signal voltage
x_r	input sample r
$x(t)$	input at time t
Y, Z	variables
α	a constant
$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \gamma$	predictor coefficient
Δ	step size
Δ_r	step size at sample r
Δ-ΔM	delta-delta modulation
ϵ	error
ϵ_k	error at sample k
$\epsilon(t)$	error at time t
μ	compression ratio
σ	standard deviation
τ	pulse duration, integration time constant
ω	angular frequency.

CONTENTS

Page No