THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 3, March 1978
Printed in U.S.A.

Gradient Encoding for Low-Bit-Rate Stored Speech Applications

By S. V. AHAMED

(Manuscript received October 7, 1977)

In stored speech applications, the waveform of the message is completely specified and can be effectively used to reduce the bit rate at which the message may be synthesized. In gradient encoding, we propose to match the gradient of the output wave of the differential decoder with the required gradient between discrete clock cycles. When the required gradient is very steep the bit pattern selected maximizes the rate of change of the decoder voltage, otherwise appropriate bits of opposite polarity are inserted to match the amplitude of the decoder voltage with the required voltage at the discrete clock cycles. The performances of gradient encoding and conventional encoding are presented as corresponding signal-to-noise ratios under different inputs and circuit conditions. Further, our preliminary results indicate that gradient encoding can lead to comparable quality of speech at about half the bit rate of the conventional encoding between 32 to 24 kbaud.

I. INTRODUCTION

For stored speech application, one of the ways of generating efficient binary data is tree encoding, which examines and verifies the sequences of a prespecified number of bits by varying the data in every possible combination and selecting the one that yields the best signal-to-noise ratio. This way of exhaustive searching for the best bit pattern demands a large number of computations, and the number of computations expands geometrically as the number of bits in the tree (i.e., the number of sequential bits chosen to explore the range of variation of the decoder voltage) increases. This leads to a further uncertainty about whether the number of bits chosen is satisfactory or not for any given section of speech.

To circumvent this problem we have chosen to seek an alternative algorithm and proceed on a variable length of speech waveform determined by the gradient around the section under investigation. This is accomplished by realizing that all speech wave shapes consist of peaks and valleys, and the duration between these successive extrema should guide the duration of the computation, and that the gradient between them should guide the fragmentation of the compute-and-match procedure attempted during the relaxation* of bits for the intervals between the peaks and valleys or valleys and peaks.

II. THE BASIC APPROACH

When the locations of the peaks and valleys contained in any segment of speech have been determined, the synthesis of the optimal bit sequences may be routinely and systematically determined as follows:

(*i*) Determine the change in amplitude required from the differential decoder and the interval for the change.

(*ii*) Determine the best the decoder can accomplish by forcing a sequence of zeros (for peak to valley fit) or ones (for a valley to peak fit).

(*iii*) If the decoder can exceed the required change, halve the interval for the computation and evaluate the decoder performance by stuffing zeros or ones during half the interval.

(*iv*) Proceed to repeat (*ii*) and (*iii*) until one of the following occurs: (*a*) The decoder performance comes to within a very tight tolerance level of what the original speech wave called for. (*b*) The interval for computation has collapsed to one clock cycle of the decoder and if so choose a (0) or (1) that minimizes the error at the end of that particular clock cycle.

(*v*) When *iv*(*a*) or *iv*(*b*) are completed, update the new peak as the last point processed under *iv*(*a*) or *iv*(*b*) if the search pattern is progressing from a peak to a valley and retain the same valley or update the new valley as the last point processed under *iv*(*a*) or *iv*)*b*) if the search pattern is progressing from valley to peak and retain the same peak.

(*vi*) When the binary bits during the interval have been synthesized, proceed to the next section of the speech wave shape—i.e., to the next peak-valley or valley-peak pair.

III. DIFFERENCE BETWEEN CONVENTIONAL ENCODING AND GRADIENT ENCODING

Conventional encoding ignores the *a priori* information about the location of the next extreme point and can make large errors in achieving the best performance from a decoder. Gradient encoding ignores the

* In this context relaxation implies a systematic iterative selection.

basic premise of conventional encoding by forcing the next bit to be of opposite polarity if the present decoder voltage has exceeded the input signal, and retains a slight error at the present position in an overall attempt to do its best to reach the target extremity of the wave shape. When targets become very far apart, then the intermediate ranges of fit start to shrink and minimize the error at the intermediate points. In an extreme case of absolute silence, gradient encoding and conventional encoding converge in the bit selection of alternate zeros and ones.

The anticipatory characteristics of gradient encoding also prepare the decoder for sudden peaks in wave shapes by sending a series of identical bits before arriving at the peak, so that the extreme point is within a predetermined range of error. This is totally absent in conventional encoding.

IV. PERFORMANCE OF GRADIENT ENCODING—ALGEBRAIC WAVES

4.1 Summed sine waves

Figure 1a indicates the performance of the conventional ADM encoding technique when a sampling frequency of 12 kHz has been used to excite the encoder which is following an input wave generated as the sum of two sine waves, one at 400 Hz with an amplitude of 80 mV, and the other at 1200 Hz with an amplitude of 100 mV. In contrast, Fig. 1b indicates the performance of the gradient encoder technique under the same conditions. In Fig. 1a it can be seen that the point 3 on the dotted line being very close to 3 on the full line can materially change the next bit polarity and thus change the ensuing bit pattern; whereas in Fig. 1b the gradient encoding tolerates errors at 3, 4, 5 and 6 in order to match the segment 2–6 on the decoder curve (dotted line) as closely with the section 2–6 of the input, (solid line) and concentrates the transition of 010 near the peak where it should logically be placed. Other such variations are also noticeable by comparing 1a and 1b.

4.2 Interrupted sine wave

The anticipatory character of gradient encoding is evident by comparing Figs. 2a and 2b. The input to the two types of encoders is a sine wave at 1200 Hz at 100 mV interrupted at a frequency of 400 Hz. In Fig 2a it can be noted that the encoder starts to respond by a series of fixed values only after the input wave has actually been presented at the encoder, whereas the gradient encoder in Fig 2b starts to process the bits prior to actual impact of the wave, and adjust its bits accordingly.

V. SIGNAL-TO-NOISE RATIO ANALYSIS

5.1 Program description

The performances of conventional and gradient encoders have been modeled on a Nova 800 minicomputer by a sequence of machine language

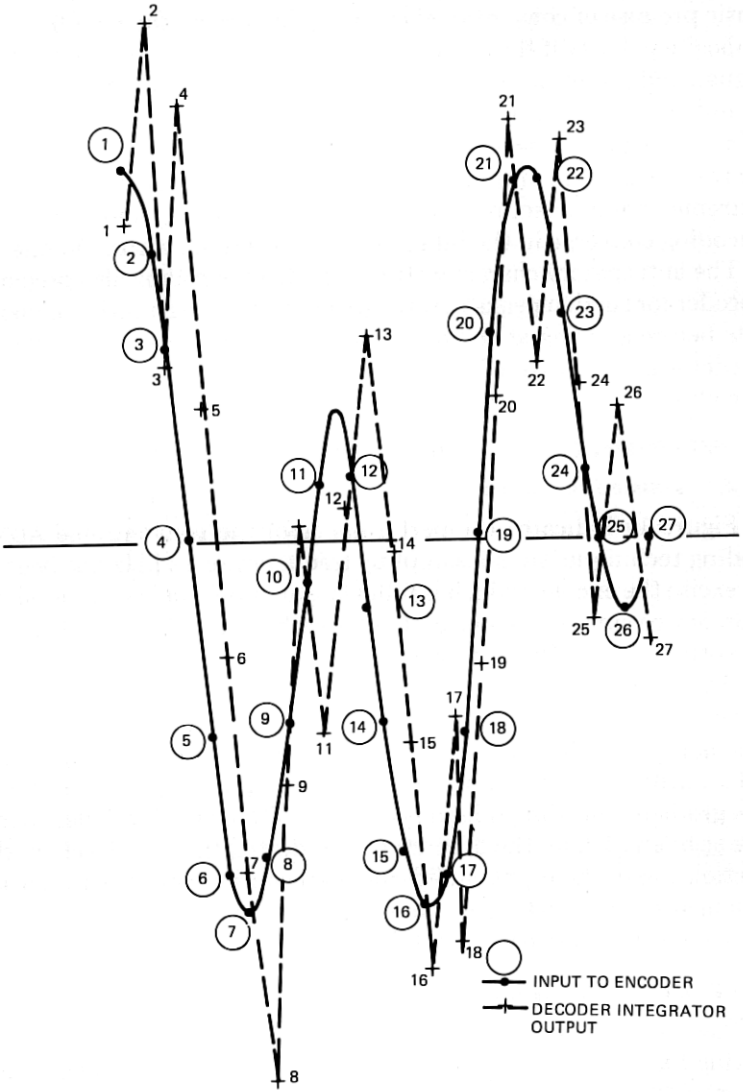

Fig. 1—(a) Conventional encoding.

and Fortran programs. The numerical computations are confined to a block of replaceable Fortran programs and the data handling from disc is performed by a set of machine language subroutines. Both communicate with the operator in a conversational mode and the circuit parameters are input controlled. It is thus possible to compare the relative performance of the two types of coding schemes under any set of conditions. The results are presented in the following sections.

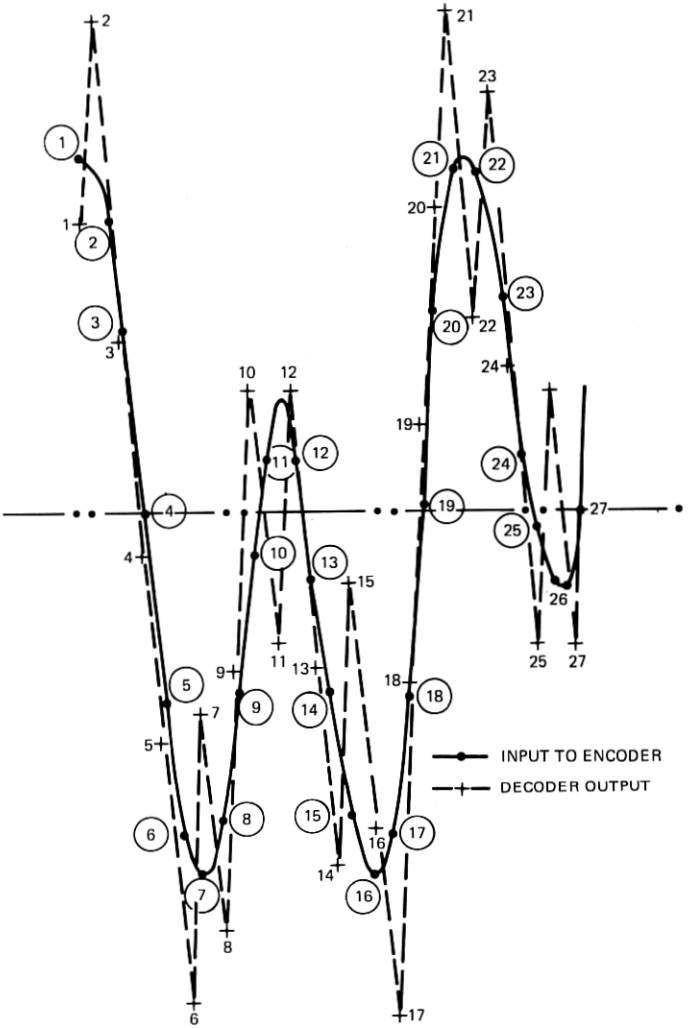

Fig. 1—(b) Gradient encoding.

5.2 Cross comparison of performance at the same clock rate

Four clock rates (32, 24, 16, and 12 kHz) are chosen to compare the performance with the ADM codec described in Ref. 1. The charging time constants of the step size capacitor and the levels have been optimized to achieve the best S/N ratios with different number of bits for companding (see Section II, Ref. 1). However, in the case of the 4 bit companding the charging time constant has been retained as 3 msecs as it

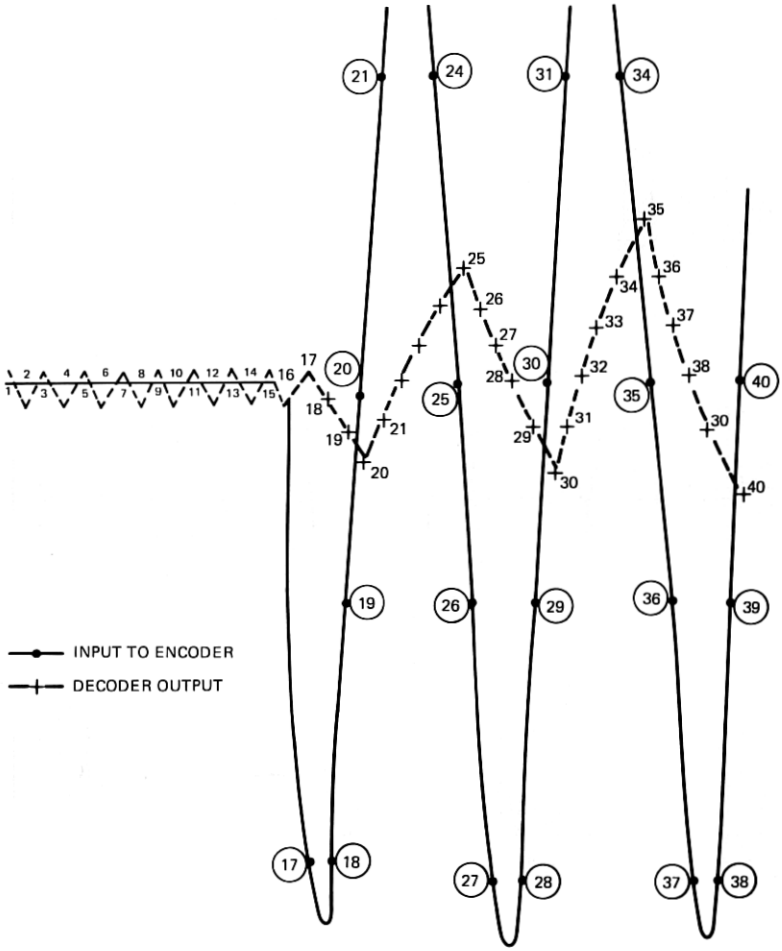

Fig. 2—(a) Conventional encoder.

currently exists. Table I contains the computed S/N ratios for 32 and 24 kbit per second clock rates and Table II contains the result from 16 and 12 kbits per second simulations.

5.3 Half-rate gradient encoding performance

Tables III and IV compare the performances of the 16 and 12 kbit per second gradient encoding against 32 and 24 kbits per second conventional encoding respectively. The time constants and levels have again been optimized to yield the best performance from the codec.

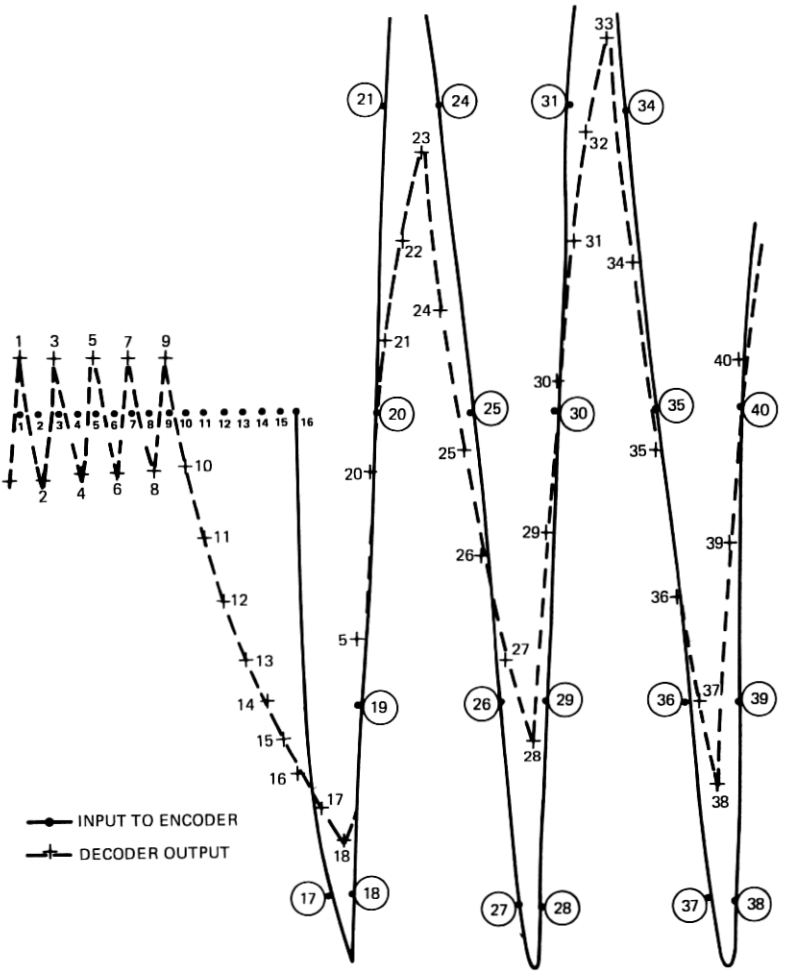

Fig. 2—(b) Gradient encoder.

VI. PERFORMANCE WITH SPEECH

Gradient encoding outperforms conventional encoding at the same bit rate. As the bit rate is reduced for gradient encoding a region of indifference is encountered between 50 to 60 percent of the rate of conventional encoding. The attack time constant (i.e., the product of the resistance for charging the step size capacitor and its value) starts to influence the higher frequency response but enhances the signal to noise ratio at the lower end of the audio frequency response and vice versa. A compromise is necessary to achieve the best response over the range

Table I — Computed *S/N* ratios at 32 and 24 kHz

Clock:		32 KHz		24 KHz	
Coding:		Conventional	Gradient	Conventional	Gradient
Audio frequency, Hz (sine waves)	No. of bits	*S/N*, dB	*S/N*, dB	*S/N*, dB	*S/N*, dB
5700	3	—*	10.2	—	
	4	—	0.8	—	
4700	3	—	12.8	—	
	4	—	13.5	—	
3950	3	9.0	8.0	—	13.8
	4	7.2	10.3	—	2.5
3150	3	12.0	14.7	—	12.3
	4	1.4	13.3	—	12.4
2900	3	9.7	15.8	7.9	10.0
	4	6.4	15.6	5.7	9.2
2490	3	11.9	20.3	3.7	16.6
	4	12.3	15.6	2.1	12.0
1660	3	19.6	24.5	16.0	19.3
	4	16.0	21.4	13.0	17.0
830	3	29.7	35.2	25.1	30.0
	4	26.0	32.3	23.2	26.6
415	3	38.5	40.7	31.6	33.0
	4	34.0	39.1	31.0	34.1

* — indicates near-zero values

of importance for telephone conversation. However, since the quantization noise in gradient encoding is scarcely present due to the optimization of the selected bit pattern, the region of indifference between gradient encoding and conventional encoding tends to be biased in favor of the former at a slight expense of the higher audio frequency response. Informal subjective testing has indicated that the 12 kbit per sec, 2 bit companded, gradient-encoded speech is comparable with the 24 kbit per sec, 4 bit companded, conventionally encoded speech.* However, the 24 kbit per sec sequentially companded speech[1] shows a favorable margin of performance over the 12 kbit per sec gradient-encoded speech.

The computation time depends on the bit rate and concentration of peaks and valleys. Lower frequency wave forms demand more computations in order to perform intermediate compute-and-match attempts. Higher frequencies on the other hand are adequately fitted by fewer overall gradient matching trials. When long telephone announcements

* This improvement has been made possible because gradient encoding does not attempt to maximize the signal to noise ratio but instead, matches the extremities of the wave shape. When the incoming wave shape offers a large cyclic change in the transition (due to change in pitch) of the gradient between and peak-valley or valley-peak pair together with a steep gradient between the points, gradient encoding ignores the cyclic variation in the gradient whereas an attempt to maximize the S/N (as it is done in tree encoding) tries to accommodate the cyclic change and can lead to a perceptually poorer quality of speech. To this extent gradient encoding outperforms tree encoding.

Table II — Computed *S/N* ratios at 16 kbits and 12 kbits/sec

Bit rate:		16 kbits/sec		12 kbits/sec	
Coding:		Conventional	Gradient	Conventional	Gradient
Audio frequency, Hz	Companding No. of bits	*S/N*, dB	*S/N*, dB	*S/N*, dB	*S/N*, dB
3150	2	—*	7.2	—	3.4
	3	—	2.8	—	—
2900	2	—	9.2	—	5.3
	3	—	4.0	—	0.3
2490	2	—	13.9	—	5.9
	3	—	7.3	—	1.0
1660	2	2.1	18.0	1.6	11.7
	3	0.92	19.0	—	7.5
830	2	19.5	28.5	18.4	26.0
	3	21.2	28.4	17.8	27.3
415	2	30.4	30.6	19.9	30.0
	3	32.2	30.2	27.5	29.3

*— indicates near-zero values

Table III — Comparison of 12 kbits/sec gradient and 24 kbits/sec conventional coding

Bit rate:	12 kbits/sec		24 kbits/sec	
Coding:	Gradient		Conventional	
Audio frequency, Hz	Companding No. of bits	*S/N*, dB†	Companding No. of bits*	*S/N*, dB
3150	2	3.4	3	—
	3	—‡	4	—
2900	2	5.3	3	7.9
	3	.32	4	5.7
2490	2	5.9	3	3.7
	3	1.0	4	2.1
1660	2	11.7	3	16.0
	3	7.5	4	13.0
830	2	26.0	3	25.1
	3	27.6	4	23.2
415	2	30.0	3	31.6
	3	29.3	4	31.0

* 2-bit companding at 24 kbits/sec leads to extremely noisy silence periods.
† Changing the time constants of the attack circuit (see Section II, Ref. 1) changes the distribution of *S/N* ratios between the low and high audio frequencies. For instance with a 66 percent time constant the values of the *S/N* ratios are 4.8, 9.4, 9.6, 13.4, 21.6, 23.5 dB from 3150 to 415 Hz respectively with 2-bit companding.
‡ — indicates near-zero values

are synthesized we have noticed a one-third second (corresponding to 256 sixteen-bit words at 12 kbaud) of speech occasionally demanding as long as 20 minutes of Nova-800 minicomputer CPU time. This particular machine has an 800-nsec cycle time and hardware floating point multiply-divide facility. Conversely other one-third second speech

Table IV — Comparison of 16 kbits/sec gradient and 32 kbits/sec conventional coding

Bit rate:	16 kbits/sec		32 kbits/sec	
Coding:	Gradient		Conventional	
Audio frequency, Hz	Companding No. of bits	S/N, db	Companding No. of bits	S/N, dB
3950	2	6.1	3	9.0
	3	1.0	4	7.2
3150	2	4.7	3	12.0
	3	2.8	4	1.4
2900	2	6.5	3	9.7
	3	4.0	4	6.4
2490	2	8.5	3	11.9
	3	7.3	4	12.3
1660	2	15.3	3	19.6
	3	19.0	4	16.0
830	2	27.7	3	29.7
	3	28.4	4	26.0
415	2	36.8	3	38.5
	3	30.2	4	34.0

segments are synthesized in as little as four minutes. Averaged over three and a half minutes of speech synthesis, the computational time is roughly half an hour per second of real time speech signifying two thirds billion arithmetic operations† for every second of message. Stated alternatively one may expect one third million numerical functions between a typical peak and valley of the speech waveshape.

The computations during the silence periods are not trivial since gradient encoding is always alert to the incidence of the next peak (or valley). During the interval the dispersion of zeros and ones alternately is limited to that period which is too long to prepare the decoder for the next peak (or valley).

VII. REAL TIME IMPLEMENTATION

The real time implementation of gradient encoding is feasible in two distinct ways: (*i*) by a multiplicity of decoder circuits with feedback paths, each one being excited by a bit pattern of zeros or ones over a finite intervals and then selecting the pattern of the decoder which yields the waveform closest* to the waveform of the original speech or (*ii*) by one decoder circuit whose internal timing has been hastened dramatically by decreasing all the time constants in the circuit accordingly, and then

† This includes the modeling of all nonlinearities as they exist in the codec, the algebraic representation of most circuit elements, all the address computations, changing the synchronization rates between the scanning A/D converter and the codec clock rate, etc.

* Such as to maximize the S/N.

choosing that bit pattern which yields the waveform closest to that of the incoming speech. The former technique proposes that the data from the encoder transmitted to the final real time decoder is selected as the data of that particular real time decoder whose output came closest to that of the incoming speech. The latter technique proposes that the data of the encoder transmitted to the final real time decoder is selected as that binary combination of bits which had brought the waveform of the faster non-real-time decoder closest to that of the original speech.

Both of these techniques explore every branch of tree encoding to determine which one of the binary sequences yields the best performance. Our preliminary estimations show that eight decoder circuit for implementing technique (i) and an accelerated clock rate at about 100 kHz for a 12 kbaud data rate would be a reasonable compromise between complexity of the encoder design and optimality of bit configuration from the encoder design and optimality of bit configuration from the encoder. Such an arrangement is expected to enhance overall signal to noise ratio by 2 to 4 dB during the transmission of speech and the accelerated decoder circuits are completely capable of responding at about 100 kHz. Further tree encoding with 3-bit look-ahead option achieves most of the advantages obtained by these encoding schemes.

VIII. CONCLUSION

The success of the gradient encoding lies in the complete knowledge of the incoming speech waveshape. This prior information has been employed to optimize the bit pattern and thus reduce the bit rate. Attempts to reduce the storage requirement by one-half appear to have achieved an encouraging degree of success. The combined effects of sequential companding and gradient encoding are being investigated for bit rate reductions ranging between 60 and 66 percent for the same quality of speech.

REFERENCE

1. S. V. Ahamed, "Sequentially Companded ADM for Low Clock Rate Speech Coder Applications," B.S.T.J., this issue.