

MICROCOPY RESOLUTION TEST CHART
ANSI #2 - 1983

(12)

AD-A134707

RADC-TR-83-89
Phase Report
June 1983



JAM-RESISTANT SPEECH ENCODING

The MITRE Corporation

M. A. Poole and R. Rifkin

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DTIC
SELECTED
NOV 15 1983
A

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, NY 13441

DTIC FILE COPY

83 11 15 010

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-83-89 has been reviewed and is approved for publication.

APPROVED:

Walter R. Richard

WALTER R. RICHARD
Project Engineer

APPROVED:

Bruno BEEK

BRUNO BEEK, Technical Director
Communications Division

FOR THE COMMANDER:

John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (DCCD) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RADC-TR-83-89	2. GOVT ACCESSION NO. AD-A134	3. RECIPIENT'S CATALOG NUMBER 707
4. TITLE (and Subtitle) JAM-RESISTANT SPEECH ENCODING	5. TYPE OF REPORT & PERIOD COVERED Phase Report Oct 81 - Sep 82	
	6. PERFORMING ORG. REPORT NUMBER MTR 8827	
7. AUTHOR(s) M. A. Poole R. Rifkin	8. CONTRACT OR GRANT NUMBER(s) F19628-82-C-0001	
	9. PERFORMING ORGANIZATION NAME AND ADDRESS The MITRE Corporation Dept D-82 Bedford MA 01730	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS MOIE 7210
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (DCCD) Griffiss AFB NY 13441	12. REPORT DATE June 1983	
	13. NUMBER OF PAGES 74	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Same	15. SECURITY CLASS. (of this report) UNCLASSIFIED	
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) Same		
18. SUPPLEMENTARY NOTES RADC Project Engineer: Walter R. Richard (DCCD)		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Vocoder Pulse Code Modulation Entropy Coding Variable Jam Resistance Variable Data Rate Weighted PCM Linear Predictive Coding		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report describes techniques that provide increased jam resistance for digitized speech. Methods for increasing the jam resistance of pulse code modulated data are analyzed and evaluated in listener tests. Special emphasis is placed on new voice encoding approaches that take advantage of a spread spectrum system with a variable (or multiple)-data-rate/variable (or multiple)-AJ capability. Methods for matching a source to a channel in a jamming environment are investigated. Several techniques-		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

that provide about a 4 dB increase in jam resistance have been identified.



Distribution	
Date	
Classification	
Justification	
By	
Distribution	
Availability Codes	
Dist	Availability
Special	
AI	

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

TABLE OF CONTENTS

<u>Section</u>		<u>Page</u>
1.	INTRODUCTION	1
	1.1 OBJECTIVE	1
	1.2 BACKGROUND	1
	1.3 STUDY PROGRAM	2
2.	VARIABLE JAM RESISTANCE FOR PCM BITS	5
	2.1 PCM BIT WEIGHTING	5
	2.1.1 Bit Weighting Analysis	6
	2.1.1.1 Optimum Weighting for Non-Companded PCM	12
	2.1.1.2 Optimum Weighting for Companded PCM	12
	2.1.1.3 Sub-Optimum Weighting	14
	2.1.1.4 Theoretical Performance	14
	2.1.2 Bit Weighting Experimental Results	19
	2.1.3 Practical Considerations	21
	2.2 ERROR-CORRECTION CODING FOR PCM	22
	2.3 CONCLUSIONS	26
3.	ENTROPY CODING FOR PCM	27
	3.1 INFORMATION THEORY AND ENTROPY CODING	27
	3.2 ANALYSIS	28
4.	SPEECH APPLICATIONS OF VARIABLE-DATA-RATE TRANSMISSION	33
	4.1 SOURCE CODERS THAT DO NOT INCREASE JAM RESISTANCE	33
	4.2 SOURCE CODERS THAT DO INCREASE JAM RESISTANCE	35

TABLE OF CONTENTS (concluded)

<u>Section</u>	<u>Page</u>
4.2.1 Externally Contolled Data Rate	35
4.2.2 Data Rate Determined by Data	38
4.2.2.1 Entropy Coding	39
4.2.2.2 Use of Special Bit Sequence	40
4.2.2.3 Use of Different Number of Bits For Voiced and Unvoiced LPC Frames	44
4.2.2.4 Use of Variable-Length Speech Segments for LPC	45
4.3 CONCLUSIONS	46
5. SPEECH APPLICATIONS OF MULTIPLE-DATA-RATE TRANSMISSION	49
5.1 AJ PERFORMANCE OF MUTLIPLIE-DATA-STREAM SPEECH CODERS	49
5.1.1 Residual Excited LPC	51
5.1.2 Baseband Coder	52
5.1.3 Adaptive Transform Coding with Side Information	52
5.1.4 Embedded Transmission	53
5.2 FUTURE INVESTIGATIONS	53
6. SUMMARY AND CONCLUSIONS	55
REFERENCES	57
APPENDIX SPEECH PROCESSING FACILITY	59

LIST OF ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
1	Weighted Voice PCM System	7
2	7-Bit Companded PCM Encoding of Speech	15
3	Companded PCM Encoding of Speech - Optimum Weighting vs. No Weighting	17
4	Companded PCM Encoding of Speech - Sub-Optimum Weighting vs. No Weighting	18
5	Amplitude Distribution of Speech	24
6	Typical Coded & Uncoded Performance	24
7	Variable-Data-Rate Transmission	34
8	Speech Coders That Do Not Increase Jam Resistance	36
9	Speech Coders with Variable Rates Controlled Externally	37
10	Speech Coders with Variable Rates Controlled by the Data	41
11	Multiple-Data-Rate Transmission	50
A-1	Facilities	60
A-2	Speech Processing Facility	62

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1	Comparison of Entropy Coding, Uniform Quantization and μ -law Companding	30

ACKNOWLEDGMENT

The authors gratefully acknowledge the efforts of R. D. Preuss who was the hardware and software architect of the speech processing facility. Design and construction of this facility was a major effort on this project. R. C. Smith and S. J. Washburn provided software support, and R. C. Hooper assembled the system hardware. The authors also wish to thank the reviewers of this report for their helpful comments.

SECTION 1
INTRODUCTION

1.1 OBJECTIVE

The overall objective of Project 7210, Anti-Jam Communication Techniques, is very high jam-resistant tactical communications. The portion of the project described in this report is particularly concerned with the investigation of spread spectrum techniques that take advantage of the special characteristics of digitized voice to improve performance in a jamming environment.*

1.2 BACKGROUND

Spread spectrum systems are used extensively to provide jam-resistant communications. Jam resistance is proportional to processing gain, and spread spectrum processing gain is defined as the ratio of the spread spectrum bandwidth to the data rate, i.e., $PG = W_{ss}/R_b$. Thus, for a fixed spread spectrum bandwidth, jam resistance is inversely proportional to data rate. Therefore, jam resistance can be increased by lowering the data rate.

In an unjammed environment, jam-resistant communication systems provide a great excess of available channel capacity. In the presence of jamming, however, changes in the communication and jammer signal strengths, due to propagation and antenna effects, cause dynamic variations in the available channel capacity. At

*Some of the results of this study are also applicable to video data which has some of the same characteristics.

times the capacity may be sufficient to support the reception of high-fidelity voice. At other times the jammer may defeat the reception of even low-data-rate compressed voice. The ideal voice communication system should use all of the available channel capacity to provide the best performance possible for all jamming levels. Stated another way, this ideal system should match the information source to the channel.

In a previous work program in jam-resistant data link systems, we developed an anti-jam (AJ) spread spectrum system that can directly trade jam resistance for data rate.^[1] This system, referred to as our variable-data-rate/variable-AJ data link, uses various combinations of spread spectrum techniques to cover a range of data rates from 50 bps to 575 kbps. This range is divided into three bands. The signal structure is designed so that data can be transmitted simultaneously in each of these three data-rate bands. Within each band, the data rate can be varied over a 16:1 range.

1.3 STUDY PROGRAM

This study considers many aspects of the variable-data-rate/variable-AJ and multiple-data-rate/multiple-AJ concepts as they relate to increasing jam resistance and matching a voice source to a jammed channel. Specifically, this program investigates the application of variable-data-rate and multiple-data-rate transmission to digitized voice data. Techniques that increase jam resistance by lowering the average data rate are considered. Also studied is the allocation of different amounts of jam resistance to the bits in a digitized data stream because most voice encoders produce bits which require different amounts of jam resistance.

Our analyses are based on minimizing the mean-square error in the reconstructed speech waveform. However, the mean-square error criterion does not always accurately reflect the perceived speech quality. Because speech quality is highly subjective, listening tests are needed. Consequently, a system for voice interface to a computer was assembled. This system provides computer simulation of speech compression algorithms and allows subjective performance evaluation. A detailed system description appears in the appendix.

SECTION 2
VARIABLE JAM RESISTANCE FOR PCM BITS

Pulse code modulation (PCM) is a widely used method for digitizing analog waveforms. The process includes time sampling the waveform and converting each sample to an N-bit digital word by means of an A/D converter. Uniform quantization (equal step sizes) is used for most applications. However, μ -law companding is the standard method for digitizing voice data. By assigning small step sizes at low signal levels and increasingly larger step sizes at higher signal levels, improved performance can be obtained for voice signals in which low amplitudes predominate.

PCM data is usually transmitted by converting each N-bit digital sample to a serial data stream in which all N bits are transmitted with equal energy. Consequently, the probability of error is the same for all N bits. However, an error in the most-significant-bit position creates a larger error in the reconstructed analog signal than an error in the least-significant-bit position. This section considers techniques for improving the quality of PCM encoded voice data by providing increased jam resistance for the most significant bits.

2.1 PCM BIT WEIGHTING

In relatively noise-free environments, i.e., little or no jamming, little advantage is gained by weighting the PCM bits. Because the probability of bit error is so small, speech quality is high. In very heavily jammed environments, the probability of bit error is so great that no communication can occur, whatever the chosen weighting scheme. However, there is a middle region where the probability of error for PCM bits is moderately low. In this



region, intelligible speech may be communicated and bit weighting should be beneficial. This study concentrates on the benefits of bit weighting in that region.

The concept of bit weighting in PCM systems has been analyzed by several investigators. Bedrosian^[2,3] introduced weighted PCM in 1953. He examined the improvement afforded by weighting in the region where the noise produced by incorrect bit decisions equals the quantization noise.* This methodology is not particularly suited to speech communication since intelligible speech can still be communicated at higher levels of bit-error noise. Bedrosian's ideas have been extended in various ways, but very little analysis has been devoted to the specific case of speech communication. This section examines how weighted PCM can be applied to a typical speech communication system.

2.1.1 Bit Weighting Analysis

Figure 1 shows a block diagram of the PCM system under investigation. Time samples of a bandlimited speech waveform are amplitude quantized in an analog-to-digital (A/D) converter according to a particular compression characteristic. In this discussion, the compression function is separated from the A/D conversion process although operationally they are probably performed at the same time. The resulting quantized amplitude samples are then encoded onto a PCM bit stream in which the individual bits are weighted (in energy) according to a formula to

*Quantization noise is the error in the reconstructed signal caused by a finite number of discrete quantization steps.

1A-66,275

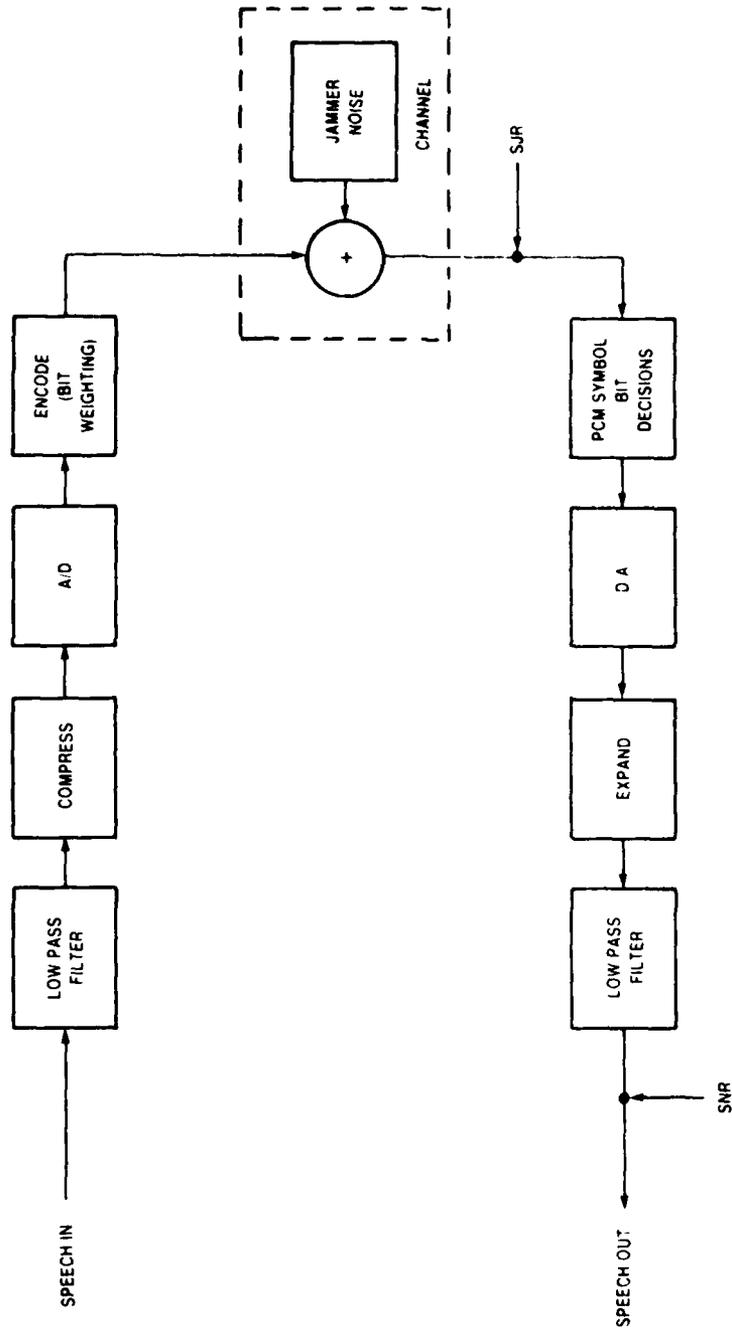


Figure 1 WEIGHTED VOICE PCM SYSTEM

be derived shortly. After transmission over a channel corrupted by additive white Gaussian jamming noise, the received bits are noncoherently detected and decoded. The analog speech signal is then reconstructed by passing the decoded amplitude samples through an expander whose characteristic is inverse to the original compressor characteristic.

The total energy, E , in the N -bit PCM symbol is

$$E = \sum_{i=1}^N E_i \quad (1)$$

where E_i is the energy in the i^{th} bit of the PCM symbol.

The signal-to-jam ratio (SJR) in the channel is defined as the average energy per bit divided by the jammer noise power spectral density, which has been normalized to one. Hence

$$\text{SJR} = E/N \quad (2)$$

For noncoherent detection, the probability of error of the i^{th} bit, p_i , is

$$p_i = \frac{1}{2} \exp(-E_i/2) \quad (3)$$

assuming white jamming noise with unity power spectral density and negligible background noise compared to the jamming noise.

The output signal-to-noise ratio (SNR) of the system is defined as

$$\text{SNR} \equiv P_s / (N_q + N_c + \langle e^2 \rangle) \quad (4)$$

where P_s is the average output signal power, N_q is the quantization noise resulting from amplitude quantization of the input (speech) signal, N_c is the mean-square clipping noise resulting from clipping high-amplitude input signals, and $\langle e^2 \rangle$ is the mean-square error noise resulting from decision errors in the individual PCM symbol bits. Neither N_c nor N_q depends upon the SJR, and PCM symbol bit weighting has no effect upon N_c and N_q .

The average output signal power, P_s , is

$$P_s = 2 \int_0^{V_0} x^2 p(x) dx \quad (5)$$

where V_0 is the maximum value of the input x . Since the information input is a speech waveform, the amplitude probability distribution is assumed to be Laplacian, so that

$$p(x) = (\sqrt{2}/2\sigma) \exp(-\sqrt{2} |x|/\sigma) \quad (6)$$

where σ is the standard deviation. Differences in σ correspond to different volume levels among different speakers. The setting of the clipping voltage, V_0 , relative to σ is somewhat arbitrary. If V_0 is set too low, excessive clipping distortion results. If V_0 is set too high, excessive quantization noise results. The mean-square clipping noise is given by

$$N_c = (\sqrt{2}/\sigma) \int_{V_0}^{\infty} (x-V_0)^2 \exp(-\sqrt{2}x/\sigma) dx \quad (7)$$

If the clipping voltage, V_0 , is set to at least 4σ , the ratio of N_c to P_s is less than -24.6 dB, and clipping distortion is negligible. A value of $V_0 = 10\sigma$ was determined to be reasonable in later experimental tests.

For a companding law described by the function $y=f(x)$ the quantization noise is

$$N_q = (1/12) \int_{-V_o}^{V_o} p(x) (dx/dy)^2 dx \quad (8)$$

The remaining term in equation (4) that must be evaluated to obtain the output SNR is $\langle e^2 \rangle$. The following analysis investigates how $\langle e^2 \rangle$ may be reduced by redistributing the available transmitter energy.

The mean-square error in the reconstructed analog signal is of the form

$$\langle e^2 \rangle = \sum_{i=1}^N p_i B(i) \quad (9)$$

where $B(i)$ is the mean-square voltage error due to a channel-induced error in the i^{th} PCM bit. The exact form of $B(i)$ will be obtained later for both non-companded and companded systems.

The weights, E_i , that minimize $\langle e^2 \rangle$ subject to the constraint given by equation (1) must be determined. This problem may be solved by using Lagrange multipliers, by forming the functional

$$F(E_i, \lambda) = \sum_{i=1}^N B(i) p_i + \lambda(-E + \sum_{i=1}^N E_i) \quad (10)$$

where λ is an undetermined multiplier, and then by solving the set of N equations

$$\frac{\partial F}{\partial E_i} = 0 \quad (11)$$

By using equation (10) in equation (11),

$$E_i = -2 \ln [4\lambda/B(i)] \quad (12)$$

The unknown multiplier λ is eliminated by using the constraint given by equation (1). Then

$$E_i = [E - 2 \sum_{j=1}^N \ln B(j)]/N + 2 \ln B(i) \quad (13)$$

This equation shows that the optimum weights depend on the received SJR. (Recall that the SJR is proportional to E .) Since the transmitter does not generally know the received SJR, this condition can not be satisfied easily in practice. However, in order to determine the best possible performance, assume initially that optimum weighting is possible.

The mean-square error is now obtained by substituting equations (3) and (13) into equation (9).

$$\langle e^2 \rangle_o = \frac{1}{2} N \exp \left[-(E - 2 \sum_{j=1}^N \ln B(j))/2N \right] \quad (14)$$

where subscript o signifies optimum weighting. Equation (14) is a general expression for noncoherent detection. If the bits are not weighted, the mean-square error is again obtained from equation (9) with $E_i = E/N$. This yields

$$\langle e^2 \rangle = \frac{1}{2} \exp (-E/2N) \sum_{i=1}^N B(i) \quad (15)$$

Equations (14) and (15) are now applied to both the non-companded and companded cases.

2.1.1.1 Optimum Weighting for Non-Companded PCM. With no compression, the quantized amplitude samples are represented by a pulse code of the form

$$\sum_{i=1}^N a_i 2^{i-1} \quad (16)$$

where $a_i \in \{0,1\}$. An error in the i^{th} bit results in a reconstructed error of magnitude $D2^{i-1}$, where D is the level spacing in the D/A converter. If the maximum output voltage from this converter is V_o , then $D = 2V_o/2^N$. The mean-square voltage error is given by $B(i) = D^2 4^{i-1}$. The mean-square error does not depend upon the input signal magnitude and therefore does not depend upon its probability distribution. By using this value for $B(i)$ in equation (14),

$$\langle e^2 \rangle_o = 2^{N-2} ND^2 \exp(-E/2N) \quad (17)$$

For the non-weighted case ($E_i = E/N$), equation (15) yields

$$\langle e^2 \rangle = [(4^N - 1)D^2/6] \exp(-E/2N) \quad (18)$$

2.1.1.2 Optimum Weighting for Companded PCM. Assume that the compander follows the standard μ -law compressor characteristic, given by

$$y = \text{sgn}(x) V_o \log(1 + \mu|x|/V_o) / \log(1 + \mu) \quad (19)$$

where x is the input to the compressor, y is the output, and μ is a constant. Typically μ is between 100-255, the standard value being 255. For $|x| = V_0$ the compressor exhibits unity gain. The expander characteristic inverse to equation (19) is given by

$$x' = V_0 \operatorname{sgn}(y') \left(10^{y' \operatorname{sgn}(y') \log(1+\mu)/V_0 - 1} \right) / \mu \quad (20)$$

where primes are used to denote the received samples that have been corrupted by the channel.

A channel error in the i^{th} bit results in an error at the output of the decoder given by

$$y' - y = \pm 2^i V_0 / 2^N \quad (21)$$

where it is assumed that the 2^N quantizer levels are symmetrically divided about zero. From equation (20) the error in the reconstructed signal is

$$x' - x = \operatorname{sgn}(y') C (10^{\operatorname{sgn}(y') P y' - 1}) - \operatorname{sgn}(y) C (10^{\operatorname{sgn}(y) P y - 1}) \quad (22)$$

where $C = V_0 / \mu$ and $P = \log(1+\mu) / V_0$.

The mean-square voltage error, $B(i)$, is obtained by squaring equation (22) and then averaging over the probability distribution of x , such that

$$B(i) = \sum_{\text{levels}} \int_{\text{each level}} (x' - x)^2 p(x) dx \quad (23)$$

where the summation runs over all the quantizer levels, and $p(x)$ is the probability density function. The above equation for $B(i)$ is not a continuous integral over the entire range of x because the

factor $x'-x$ is discontinuous over that range. The value of $B(i)$ depends upon which bit in the PCM word is in error. The mean-square error for optimum weighting is obtained by evaluating equation (23) and substituting it in equation (14). Similarly, the mean-square error for the non-weighted system is obtained by substituting equation (23) in equation (15).

2.1.1.3 Sub-Optimum Weighting. Equation (13) shows that the optimum weights are a function of the SJR, a quantity not usually known at the transmitter. The receiving station could transmit the SJR back to the original transmitting station, but such a transmission could be awkward in practice because it requires two-way transmission capability. Another possibility is to determine the optimum weights for a particular SJR and use these weights at all other SJRs. The weights would be sub-optimum and would increase the mean-square error at all other SJRs. The degradation in output SNR that this entails is determined by selecting a particular total PCM symbol energy, E' , and calculating the optimal weights from equation (13). The mean-square error is then calculated using equation (9) with a probability of error

$$p_i = \frac{1}{2} \exp \left[-N(\text{SJR}) E_i / 2E' \right] \quad (24)$$

When $\text{SJR} = E'/N$, equation (24) reduces to optimum weighting.

2.1.1.4 Theoretical Performance. The results of the bit weighting analyses have been evaluated under many conditions. Because 7-bit companded PCM is the standard for voice encoding, major emphasis is placed on the performance improvement that can be obtained using bit weighting for this type of encoding. Figure 2 shows the theoretical results for unweighted, optimally weighted and sub-optimally weighted 7-bit companded PCM encoders. The optimum curve was formed

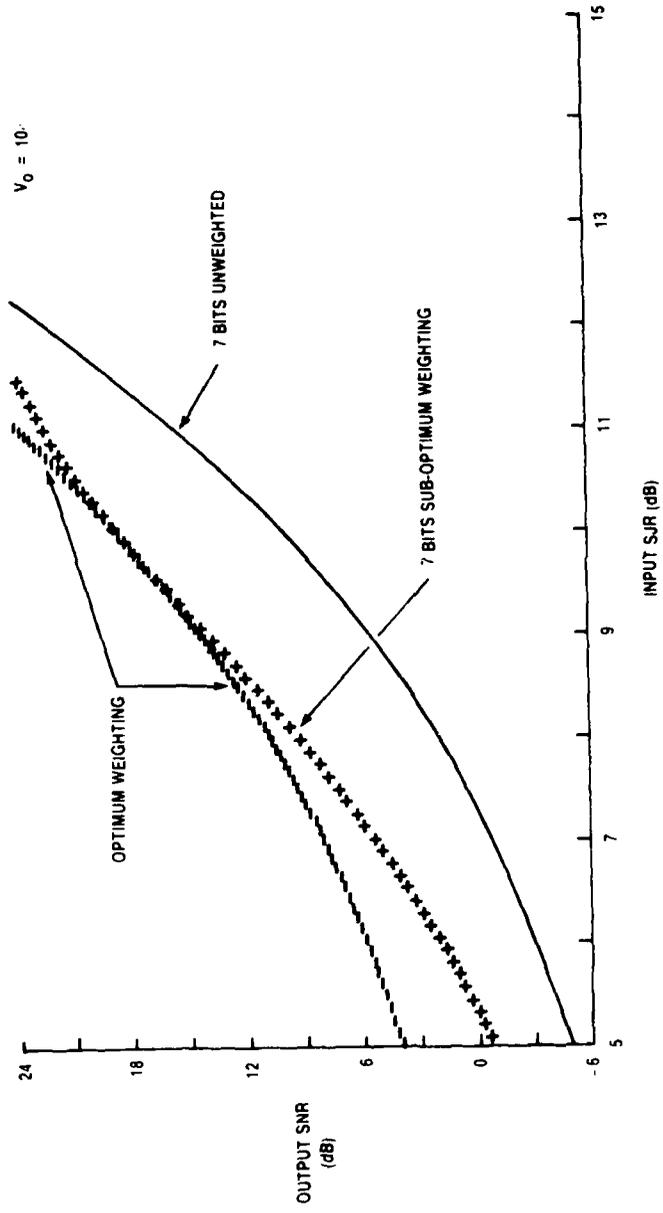


Figure 2 7-BIT COMPANDED PCM ENCODING OF SPEECH

by allowing the number of bits to vary. Optimum weighting provides approximately 2 dB improvement in signal-to-jam ratio over an unweighted system. The distinction between improved jam resistance and improved output SNR should be emphasized. From the figure, an improvement of 2 dB in jam resistance corresponds to an improvement in output SNR of approximately 9 dB. Although the improvement in speech quality at a given jamming level is of importance, the major emphasis of this study is on improvement in jam resistance which is 2 dB. Figure 2 also shows that sub-optimum weighting, using a fixed set of weights for all SJRs, provides almost as much improvement as optimum weighting except at high SJRs. Since speech quality at high SJRs is already very good, sub-optimum weighting is recommended as a practical alternative to optimum weighting.

The effect of reducing the number of bits has also been investigated. Figure 3 shows the performance of 3-, 5-, and 7-bit encoding using optimum weighting and equal weighting. This comparison is made using the same total PCM symbol energy, E , regardless of the number of bits. As the number of bits is reduced, the energy per bit and jam resistance per bit increase. This improvement in performance at low SJRs for fewer bits is shown in the figure. However, at high SJRs as the number of bits is reduced, quantization noise limits the highest output SNR that can be obtained. Comparison of the weighted and unweighted curves shows that optimum weighting is effective only when the number of bits is high.

Figure 4 compares 7-bit sub-optimum weighting with unweighted 3-, 5- and 7-bit performance. Seven-bit suboptimum weighting provides the best performance at medium and high SJRs. Only at low SJRs does 3-bit unweighted coding do better. However, the use of 3-bit unweighted coding severely limits the highest output SNR that

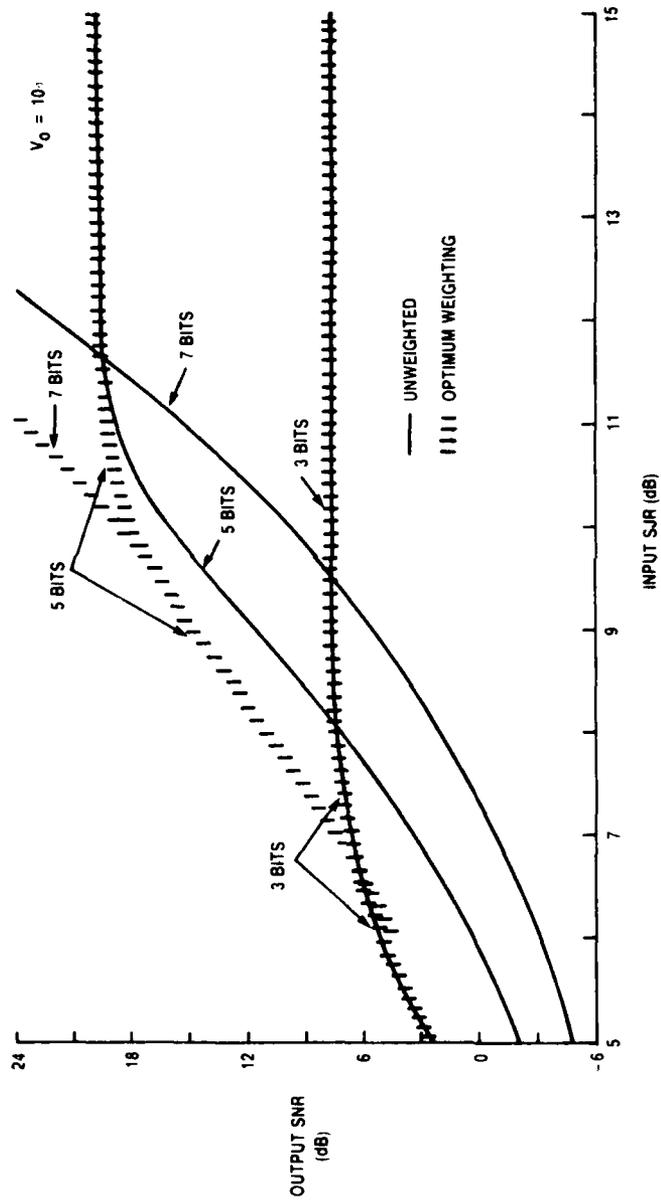


Figure 3 COMPANDED PCM ENCODING OF SPEECH — OPTIMUM WEIGHTING VS NO WEIGHTING

IA-66-270

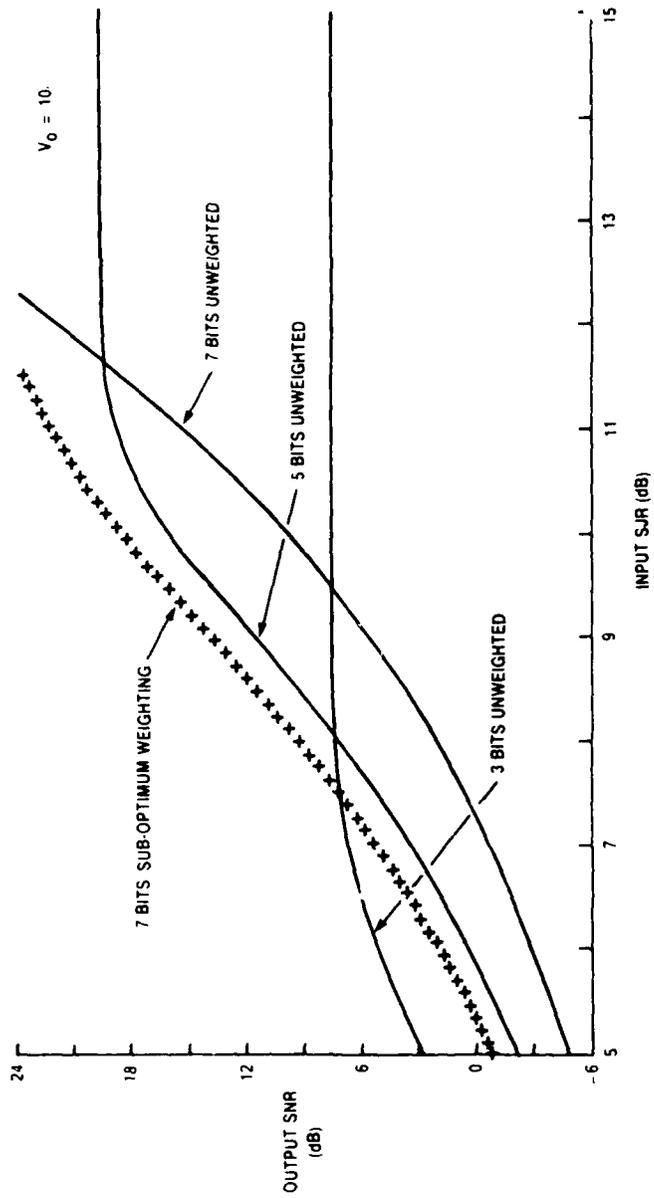


Figure 4 COMPANDED PCM ENCODING OF SPEECH - SUB-OPTIMUM WEIGHTING VS NO WEIGHTING

can be obtained due to the quantization noise. Seven-bit sub-optimally weighted coding offers the best performance over the widest range of SJRs.

2.1.2 Bit Weighting Experimental Results

The previous analyses used mean-square error and output SNR as the figures of merit for evaluating speech quality. It is well known that no single figure of merit adequately predicts speech quality. Thus the above results, based upon a simple mathematical model, are only guides in predicting the relative performance of several approaches. Simulations using real speech are required to determine exactly how these improvements are perceived.

Facilities for computer simulation of voice coding algorithms have been constructed as part of this program. The Audio Signal Conversion Laboratory was assembled to provide the I/O interface to a VAX 11/780 computer. These facilities, which are described in more detail in the appendix, have been used to evaluate the performance of PCM bit weighting. The following paragraphs describe the experimental methods and the results of the PCM bit weighting listening tests.

The reference for all experiments is a 7-bit μ -law companded speech signal. This reference is obtained by bandlimiting the original voice signal to 4 kHz and then sampling it at twice that frequency. These time samples are uniformly quantized to sixteen bits and then μ -law quantized to seven bits using one's-complement coding. By expanding this compressed digitized signal back to analog form, a reference is established for 7-bit μ -law companded speech passed through a perfect channel, i.e., no channel errors.

Speech signals for evaluation in a jamming environment are generated like the reference, and then bit errors in the channel are simulated as follows: At a particular input SJR, the total symbol energy $E = (SJR)N$. This total energy is apportioned according to the appropriate weighting scheme, and the probability of error for the bit is then obtained from equation (3). A pseudo-random number sequence over the range (0,1) is used to determine which bits are in error. For each bit in the PCM symbol, a random number is generated. If that number is less than the calculated probability of error for that bit, the bit is flagged as an error, i.e., the bit is complemented. If the random number is greater than the probability of error for that bit, the bit is flagged as correctly received. This process is repeated for all of the bits in the PCM symbol and for all of the symbols within the message. The corrupted symbol bits are then expanded and converted back to an analog signal.

In the listening tests, the best coding method is difficult to choose at SJRs above 11 dB because there are so few bit errors. With equal bit weighting, the errors manifest themselves as occasional background clicks which are easily integrated out by the brain. The sub-optimum system reduces the frequency of these clicks at the expense of a slight muddiness in the speech. Both of the critical listeners preferred the uniformly weighted system. In figure 2, the output SNR is lower for the uniformly weighted system than for the sub-optimally weighted system. Thus for SJRs greater than about 11 dB, mean-square error is not too helpful in predicting the perceived quality of the speech. At SJRs below 11 dB the advantages of the sub-optimum system become apparent in the listening tests. In this SJR range, the bit errors manifest themselves as significant background clicking noise. The sub-optimally weighted system reduces the level of this noise and

thereby improves the speech quality. This improvement is again obtained at the expense of some muddiness in the speech, but the significant reduction in background noise more than offsets the increased muddiness. At an SJR of about 5 dB, the sub-optimally weighted system achieves approximately 2 dB additional jam resistance over the equally weighted system. This improvement is in reasonable agreement with the results obtained using the mean-square error criterion and plotted in figure 2. Other listening tests with fewer PCM bits also generally agree with the curves in figures 3 and 4 at low and medium SJRs.

2.1.3 Practical Considerations

All of the weighting schemes require redistribution of the available transmitter energy. This may be accomplished either by varying the transmitter power for each bit and keeping the bit duration constant or by keeping the transmitted power for each bit constant and varying the bit duration. Either variation would, of course, be proportional to the desired weight for each bit. Since the transmitter is probably peak-power limited, weighting the bits in power reduces the average transmitted power. This method therefore does not use all of the power capabilities of the transmitter efficiently and is probably only useful in low-probability-of-intercept applications. Using bit duration as the weighting technique is more reasonable because the transmitter can always be set at its maximum power level. Bit-duration weighting is especially easy to implement if the desired weights are related by powers of two or at least by integer ratios.

The best weighting method depends upon whether or not the transmitter knows the SJR at the receiver. If there is a feedback path from the receiver to the transmitter, the SJR at the receiver

can be used by the transmitter to adjust weights depending on the jamming level. This method can be used only if the jamming signal is continuous rather than pulsed. The effect of various strategies with receiver-to-transmitter feedback is shown in figure 3. Consider two strategies: optimum weighting and equal weighting with fewer bits. If bits are dropped while maintaining the same total energy per PCM symbol with equal energy in the remaining bits, the performance is very close to that for optimum bit weighting. Considering the complexity of implementing optimum bit weighting, the simpler strategy (dropping bits and redistributing the total symbol energy equally among the remaining bits) is the recommended choice.

If there is no feedback path from the receiver to the transmitter, weights must be preselected and cannot be changed. Figure 4 depicts four possible weighting possibilities. The 7-bit sub-optimum weighting curve provides the best performance over the widest range. However, in applications where high jamming levels are expected, 3 bits unweighted provide intelligible speech with better quality. High quantization noise does occur, and this degrades performance at high SJRs. Although other speech encoding techniques were not compared in this study, continuously variable slope delta modulation (CVSD) encoding at 16 or 32 kbps may provide better performance at high jamming levels than does 3-bit PCM encoding at 24 kbps.

2.2 ERROR-CORRECTION CODING FOR PCM

The jam resistance of the most significant bits can be increased by using a special error-correction coding technique called mean-square error coding. This method minimizes the mean-square error in the reconstructed waveform by providing more

protection for some bits than for others. Mean-square error coding has been discussed by several authors over the last ten years, most recently by G. R. Redinbo.^[4,5] Each PCM-encoded N-bit word represents a number (signal amplitude) of the form

$$s = \sum_{i=0}^{N-1} s_i 2^i \quad (25)$$

where all $S = \{s_0, s_1, \dots, s_{N-1}\}$ are equally likely. For a given BCH (n,N) code and a specified channel error rate, ϵ , the mean-square error algorithm determines the optimum generator and parity check matrices. When the number of channel errors exceeds the error-correction capabilities of the code, the errors in the decoded word are confined to the least-significant-bit positions so that the mean-square error $E[(s-\hat{s})^2]$ is minimized. This method results in the mean-square error at low SJRs being lower than for standard coding methods. At high SJRs, the mean-square error is low regardless of the coding method. The reduction in mean-square error at low SJRs corresponds to an improvement in performance of 1-2 dB over standard coding methods. This improvement is obtained by using a very complex algorithm.

Several problems arise when mean-square error coding is applied to PCM-encoded voice data. The assumption that all PCM-encoded samples are equally likely is not true. The probability distribution of speech is shown in figure 5. If a uniform quantizer is used, most samples correspond to low amplitude signals. On the other hand, if companding (logarithmic quantization) is used, the probability distribution of the samples is more nearly uniform. In equation (25) does not apply. The non-uniform distribution of samples can be handled by modifying the algorithm, thereby increasing its complexity even more. The matter of logarithmic quantization requires further study.

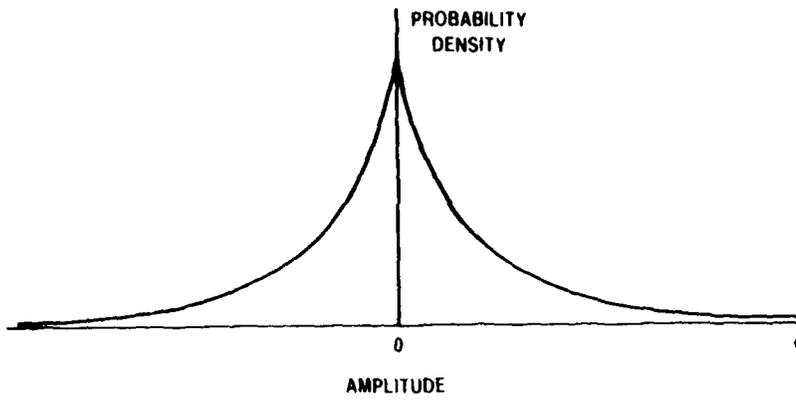


Figure 5 AMPLITUDE DISTRIBUTION OF SPEECH

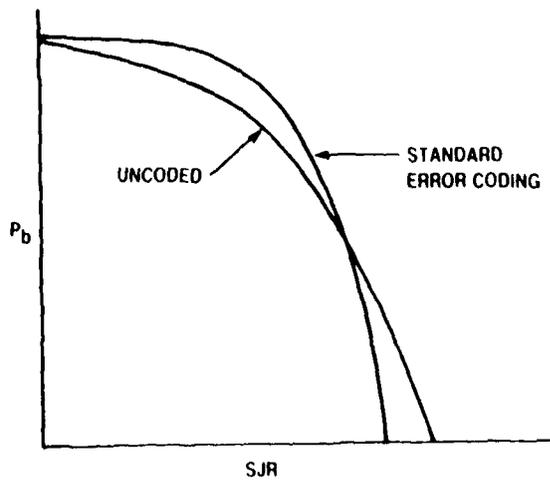


Figure 6 TYPICAL CODED & UNCODED PERFORMANCE

IA-66 27

Application of mean-square error coding in a jamming environment poses another problem. The channel error rate, ϵ , is not known a priori but is determined by the jammer. Although ϵ is not very critical in the algorithm at high SJRs, it becomes increasingly important at low SJRs. Thus, in order to achieve the 1-2 dB improvement in performance at low SJRs, both terminals must use the correct value in the algorithm. Thus, the channel error rate must be measured by the receiver and sent back to the transmitter. The need for a return transmission path limits the usefulness of this algorithm in a jamming environment. In addition, there is a problem in handling a pulsed jammer that can constantly change ϵ .

The utility of adding coding must also be considered. Typical performance curves for standard coding methods are shown in figure 6. The use of coding improves performance when the bit error rate is low but degrades performance when the bit error rate is high. If PCM-encoded voice data is being transmitted, it is impossible to distinguish between the two curves at high SJRs by listening. However, at low SJRs, better performance is obtained without coding than with coding, and this is the region where the difference is very noticeable. Mean-square error coding improves performance over standard error coding at low SJRs and therefore probably sounds much like uncoded data. Thus, the performance does not justify the complexity required to implement mean-square error coding.

All of these arguments suggest that mean-square error coding may actually degrade PCM-encoded voice data. On the other hand, mean-square error coding may be useful for coding the coefficients in linear predictive coding (LPC). Some type of coding may be advisable for these coefficients, and mean-square error coding should be considered while keeping in mind the special requirements for implementing it as discussed here.

2.3 CONCLUSIONS

Optimum bit weighting increases jam resistance by approximately 2 dB for 7-bit companded PCM encoding. However, calculation of the optimum bit weights depends upon the SJR at the receiver. Assuming that the transmitter knows the SJR and considering the difficulty of implementation, the best strategy is to adjust the number of bits in the PCM symbol and to transmit all of those bits with equal energy. If the SJR is unknown, 7-bit companded PCM with sub-optimum weighting, i.e., assignment of fixed weights, provides the best performance over the widest range.

Mean-square error coding is not recommended for PCM encoding of voice data.

SECTION 3

ENTROPY CODING FOR PCM

The jam resistance of a voice communication system can be increased by reducing the number of bits required to transmit a given amount of speech information in a given time. The reduction in data rate results in more energy and higher jam resistance for each of the transmitted bits.

Speech contains much redundancy. The correlation of speech in time is one form of redundancy. Many voice-compression techniques reduce this time correlation by using a predictor. A second form of redundancy is caused by non-uniform amplitude distribution of the speech waveform. Since low-amplitude samples occur more often, they convey less information than high-amplitude samples. Entropy coding can reduce this type of redundancy.

3.1 INFORMATION THEORY AND ENTROPY CODING

When the output of an information source is a sequence of equally likely symbols, the sequence carries the maximum amount of information. If the symbols are not equally likely, entropy coding, e.g., Huffman coding, may be used to convert the sequence into a new sequence that carries the maximum amount of information per encoded bit. Entropy coding assigns shorter code words to the more likely symbols and longer code words to the less likely symbols. The average information carried by each encoded bit is then a maximum, and the average encoded bit rate in the channel is reduced. Another characteristic of the data stream produced by the entropy coder is that each encoded bit is equally likely to be a one or a zero. Such an encoded bit carries the maximum amount of information, i.e., one bit of information.

3.2 ANALYSIS

In this section, the entropies of both uniformly quantized and μ -law quantized speech waveforms are compared to the entropy of an ideal entropy coder.

Successive PCM-encoded speech samples may be thought of as a discrete information source. The entropy, H , of that source is calculated from

$$H = - \sum_{i=1}^N h(x_i) \log_2 h(x_i) \quad (26)$$

where N is the number of quantization levels, x_i is the i^{th} value of the amplitude x , and $h(x_i)$ is the probability of the i^{th} level being occupied. The entropy calculated from equation (26) gives the expected value of the information carried per speech sample.

To a good approximation, the amplitude probability distribution of the speech waveform, $p(x)$, follows a Laplacian distribution

$$p(x) = (\sqrt{2}/2\sigma) \exp(-\sqrt{2}|x|/\sigma) \quad (27)$$

where σ is the standard deviation.

To obtain the entropy of uniformly quantized speech samples, $p(x)$ is set equal to $h(x)$ in equation (26). For non-uniform quantization, however, the amplitude probability distribution of the speech samples after companding must be determined first. This problem may be analyzed as a random variable X with distribution

function $p(x)$ being redistributed according to a function $g(x)$. The new random variable $Y = g(X)$ then has a distribution given by

$$h(y) = \begin{cases} p(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & a < y < b \\ 0 & \text{otherwise} \end{cases} \quad (28)$$

where $a = \min [g(-\infty), g(+\infty)]$, $b = \max [g(-\infty), g(+\infty)]$.

The function $g(x)$ is the μ -law compressor characteristic, given by

$$g(x) = \text{sgn}(x) V \log(1+\mu|x|/V) / \log(1+\mu) \quad (29)$$

where x is the input to the compressor, and V and μ are constants. Equations (27) and (29) are substituted into equation (28) to obtain the new amplitude distribution of speech samples,

$$h(y) = \sqrt{2} \ln(10) \log(1+\mu) A \exp(-\sqrt{2} V(A-1)/\sigma\mu) / 2\sigma\mu \quad -\infty < y < \infty \quad (30)$$

where $A = 10^{|y| \log(1+\mu)/V}$

The post-compression speech distribution, $h(y)$, may now be inserted into equation (26) to determine the entropy of the speech samples.

Table 1 shows the results of the calculations outlined above for $N=7$, $\mu=255$, and various values of V relative to σ . (Other values of N give comparable results.) PCM encoding with or without companding produces a data stream in which each channel bit (symbol) conveys less than one bit of information. Therefore, more channel bits must be transmitted for a fixed information source. The higher channel data rate results in a loss in jam resistance as shown in

TABLE 1

Comparison of Entropy Coding, Uniform Quantization
and μ -Law Companding

V	Information bits/symbol			Loss in Jam Resistance	
	No Companding	μ -law Companding	Ideal Entropy Coding	No Companding	μ -law Companding
5 σ	5.6	6.6	7.0	1.0 dB	0.3 dB
10 σ	4.6	6.5	7.0	1.8 dB	0.3 dB
20 σ	3.6	6.4	7.0	2.9 dB	0.4 dB
40 σ	2.6	6.1	7.0	4.3 dB	0.6 dB

Table 1. This table shows that entropy coding provides a significant improvement in jam resistance over uniformly quantized PCM samples. This improvement is due to the Laplacian distribution of uniformly quantized samples. The table also shows that the companded system comes very close to the maximum of seven bits per PCM sample. Therefore ideal entropy coding can provide an increase in jam resistance of only 0.3 to 0.6 dB over that of companded PCM encoding. This small improvement is explained by the fact that companding produces samples that are almost equally likely. Thus, the μ -law compander approximates an ideal entropy coder.

SECTION 4
SPEECH APPLICATIONS OF VARIABLE-DATA-RATE TRANSMISSION

For a spread spectrum system with a fixed bandwidth, jam resistance is inversely proportional to data rate. Therefore, any technique that lowers the data rate required to transmit voice data can also provide improved AJ performance in a well designed system. Thus, the continuing emphasis on low-data-rate speech encoding for transmission over narrowband channels has also resulted in increased jam resistance for digitized speech. However, lowering the data rate degrades speech quality. The best performance is obtained by adjusting the data rate according to the jamming level or, equivalently, by matching the source to the channel.

Previous work in developing our variable-data-rate/variable-AJ data link has demonstrated the practicality of directly trading data rate for jam resistance.^[1] Applications of this variable-data-rate concept to speech coding are investigated in this section.

Variable-data-rate source coders can be divided into categories as shown in figure 7. This section gives examples of source coders in each of these categories and discusses their significant characteristics. Emphasis is placed on the jam resistance provided by these source coders and on inherent problems with implementation. Variable-data-rate source coders that produce a single output data stream rather than multiple data streams are considered here.

4.1 SOURCE CODERS THAT DO NOT INCREASE JAM RESISTANCE

Two examples of source coding that do not increase jam resistance are packet transmission and time assignment speech interpolation (TASI). Block diagrams and timing for these systems



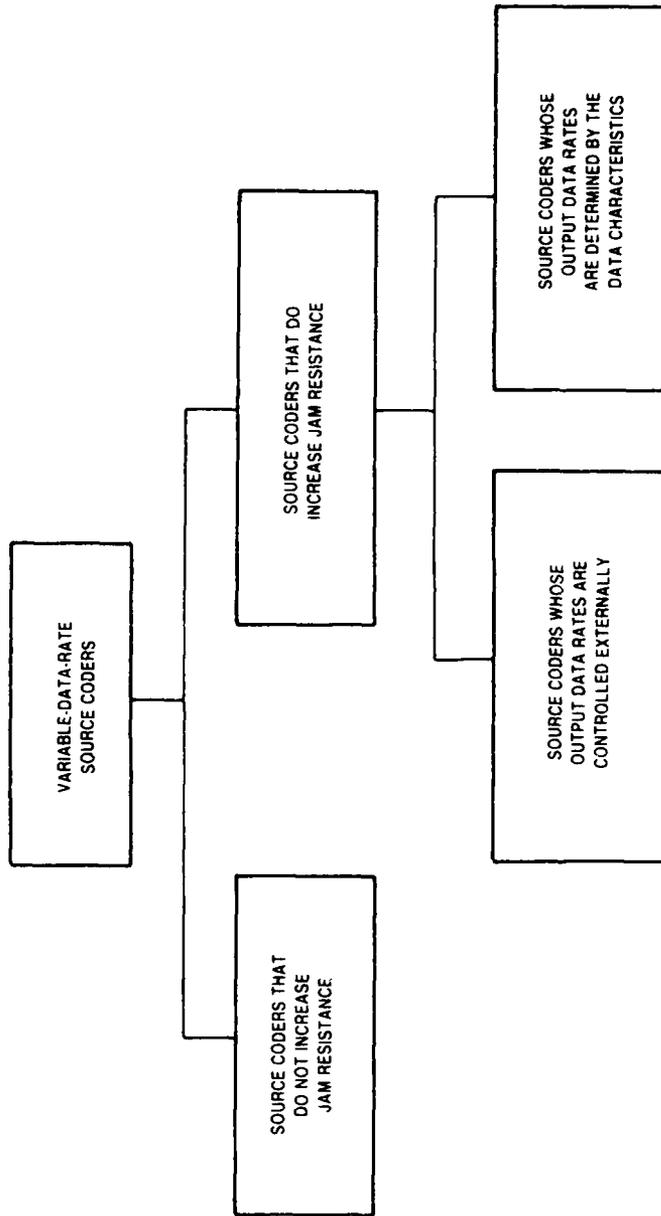


Figure 7 VARIABLE-DATA-RATE TRANSMISSION

are shown in figure 8. In both examples, gaps in time are left in the signal. These gaps allow several signals to be multiplexed to make more effective use of the channel capacity. However, no attempt is made to take advantage of the low average data rate to increase jam resistance. The use of a high-peak-power transmitter with a low duty cycle to obtain high jam resistance is not possible because of the long duration of the data bursts. Source coders in this category are not considered any further because they do not increase jam resistance.

4.2 SOURCE CODERS THAT DO INCREASE JAM RESISTANCE

In contrast to the source coders described previously, the source coders discussed in this section transmit continuously. Improved AJ performance is obtained by lowering the channel data rate.

4.2.1 Externally Controlled Data Rate

The data rate of a variable-data-rate/variable-AJ coder can be controlled externally, specifically by a feedback path from the receiver to the transmitter. The necessity for this feedback path limits the usefulness of this method. Another problem is that the feedback path must be error-free in order to maintain synchronization between the transmitter and receiver. Thus, applications for this type of source coder are limited. Nevertheless, these methods do match the information source to the channel.

Three examples of this type of variable-rate coder are shown in figure 9. In each example, the receiver monitors performance and returns a control signal to the transmitter. The signal lowers the

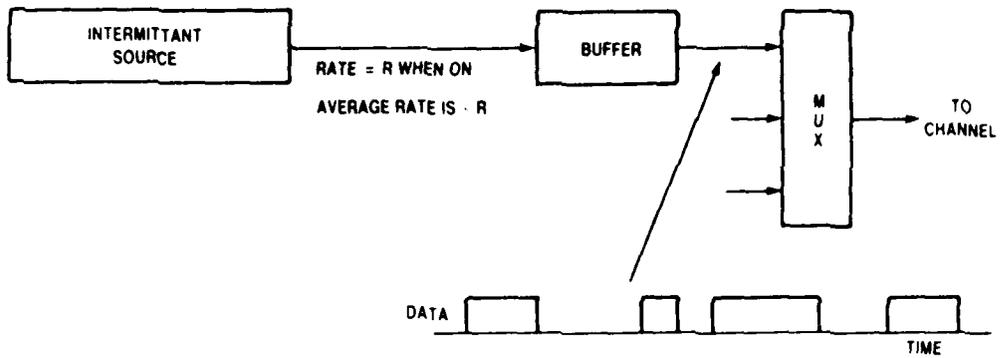


Figure 8a PACKET TRANSMISSION

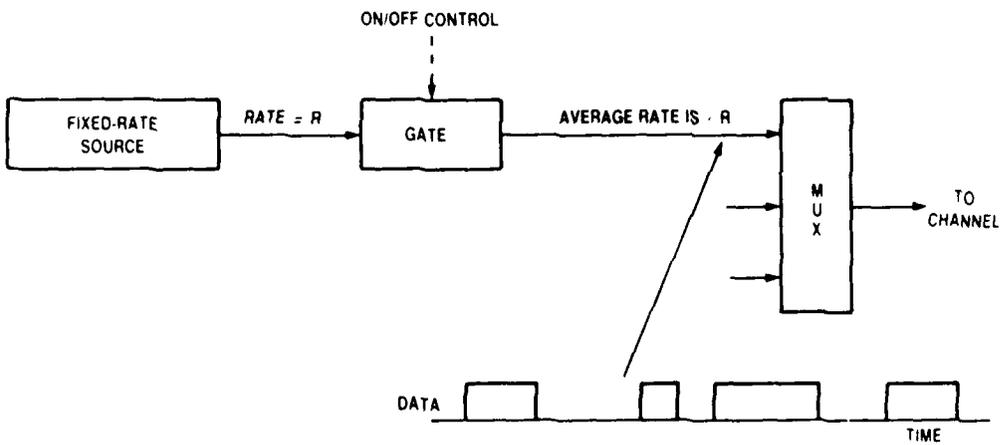


Figure 8b TASI

JA-66-273

Figure 8. SPEECH CODERS THAT DO NOT INCREASE JAM RESISTANCE

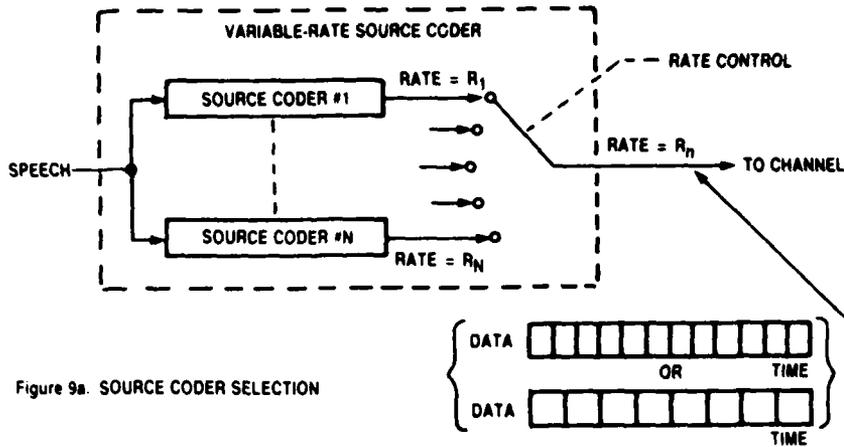


Figure 9a. SOURCE CODER SELECTION

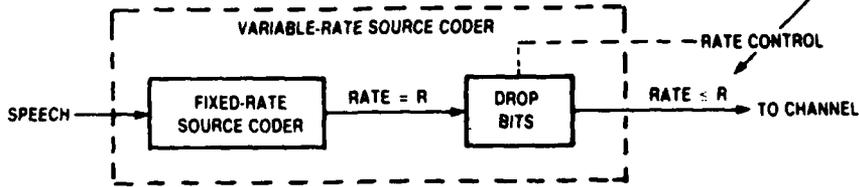


Figure 9b. SOURCE BIT SELECTION

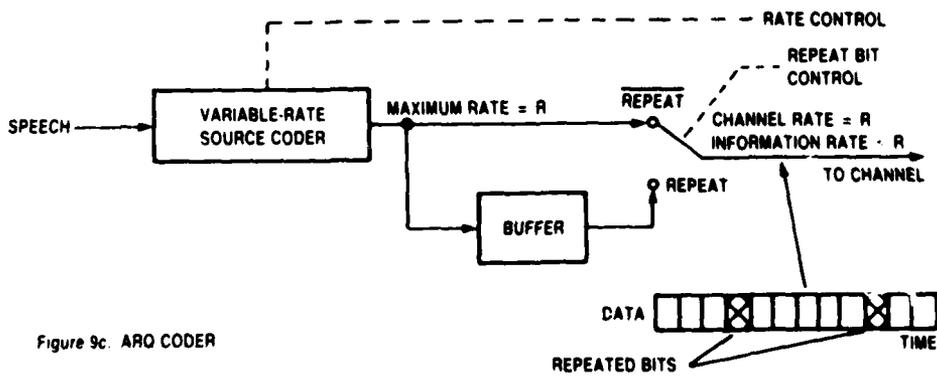


Figure 9c. ARQ CODER

IA-66.274

Figure 9. SPEECH CODERS WITH VARIABLE RATES CONTROLLED EXTERNALLY

data rate when the jamming level increases. Variable-rate encoding is accomplished in figure 9a by selecting one of the N encoders. For speech signals, the choices could be companded PCM at 56 kbps, CVSD at 16 kbps, and LPC at 2.4 kbps. The variable-rate coder in figure 9b is especially well suited to PCM transmission. In a high-level jamming environment, the least significant bit(s) can be eliminated, thereby lowering the data rate and increasing the jam resistance of the bits transmitted (see figures 3 and 4). Both of these methods automatically provide the best performance that can be obtained with a given jamming level. These methods produce data streams at specific discrete data rates.

The automatic repeat request (ARQ) coder in figure 9c differs from the other two in that it generates a data stream whose rate is a continuous function. Whenever the receiver detects a bit with a low signal-to-jam ratio, the receiver requests the transmitter to repeat the bit. As the jamming level increases, the information data rate decreases automatically. There are several problems associated with this type of system other than the requirement for an error-free feedback path. Bit repetition causes delays, especially when a bit must be repeated several times. Also, the jam resistance of each transmitted bit does not change with jamming level because the signal level and bit duration are constant. Therefore, at very high jamming levels many bits must be repeated. The resulting lowered overall information data rate is not utilized to reduce the probability of error of each transmitted bit.

4.2.2 Data Rate Determined by Data

A second type of variable-data-rate/variable-AJ coder has its data rate controlled by the data. All methods of this type provide increased jam resistance by lowering the average data rate.

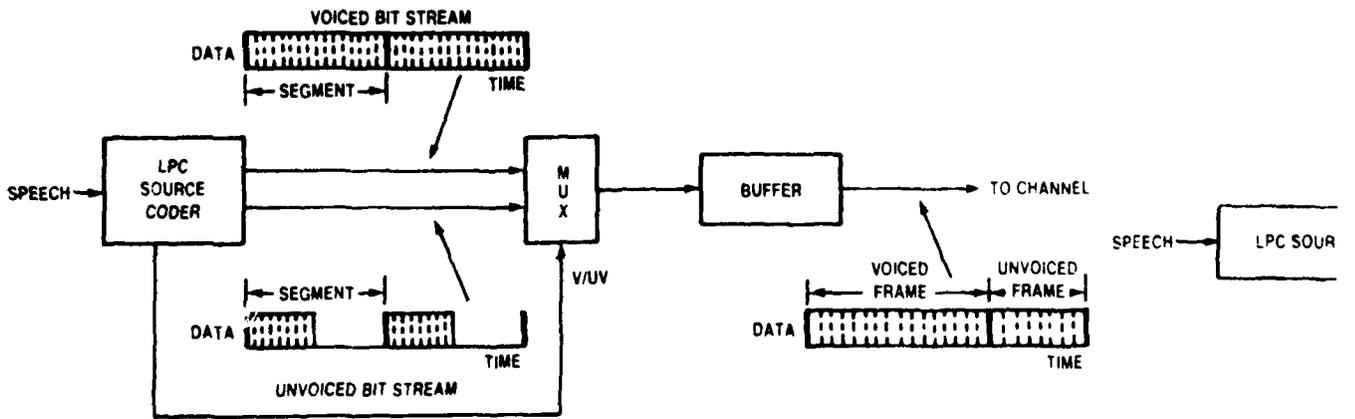
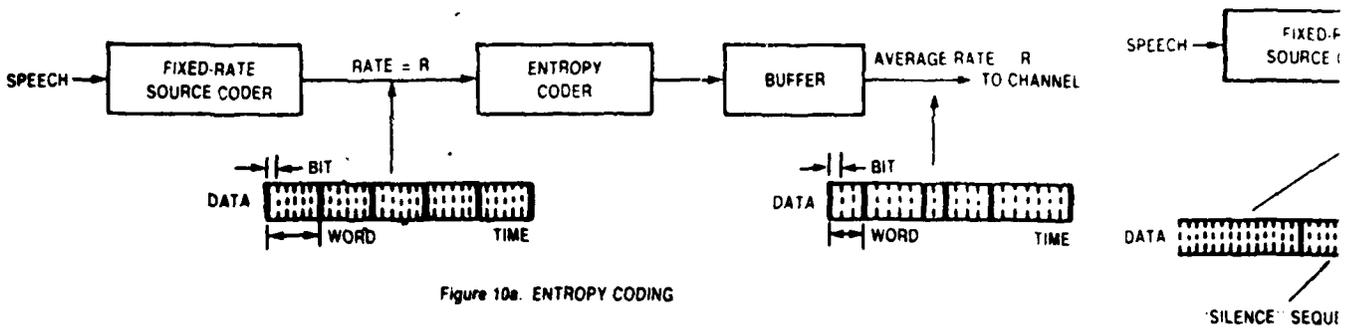
However, the instantaneous data rate varies. The encoding methods discussed in this section vary the number of transmitted bits per PCM sample or per LPC speech segment. The data can be transmitted by varying the duration of the bits and holding the PCM word (or LPC frame) duration constant. The other choice is to hold the bit duration constant and vary the duration of the PCM word (or LPC frame). The choice determines whether the jam resistance of the bit or of the PCM word (or LPC frame) is varied. Most of the methods discussed in this section require the same jam resistance for all bits. It is not practical to design a bit detector for bits with variable duration unless the detector knows in advance what the bit duration is. A transmission signal that appears to the receiver to have random bit durations is easy to jam because the receiver does not know when the bit begins or ends. Therefore, the encoding methods discussed in this section assume a constant bit duration (and constant bit rate) in the channel. This fixed channel bit rate must be designed into the system. However, the required channel bit rate varies with the speech signal and may not have the same average rate that the system is designed to handle. As a result, buffers must be added to all of these variable-data-rate encoding schemes. These buffers smooth out short-term fluctuations in data rate but do not solve the problem of long-term differences. The only solution to this problem is buffer control to delete or add bits when the buffer overflows or underflows. The buffers should be large enough to minimize the probability of overflow or underflow, but they must be small enough to minimize the delay between the input and output speech signals.

4.2.2.1 Entropy Coding. Entropy coding such as Huffman coding was discussed in section 3 for PCM transmission. Although entropy coding did not prove to be advantageous for PCM-encoded speech, it may be useful in other speech applications.

Figure 10a depicts entropy coding using PCM for illustrative purposes. Each word at the output of the PCM source coder consists of N bits. The entropy coder produces variable-length words whose average number of bits is less than N . Consequently, the bit duration at the output of the buffer is longer than at the input to the entropy coder. Longer bit duration means higher bit energy and higher jam resistance. The buffer shown in the figure will overflow or underflow occasionally. Because of the robustness of speech, bit insertion or deletion does not seriously degrade performance. In addition, with the proper entropy code, the entropy decoder in the receiver recovers quickly from inserted or deleted bits and from bit errors introduced in the channel.

For companded PCM encoding, entropy coding of the N -bit samples can reduce the data rate to provide approximately a 0.3 to 0.6 dB improvement in jam resistance. This improvement is not worth the additional hardware complexity required. Entropy coding of a CVSD output data stream is capable of providing a 10-25% reduction in bit rate (0.5-1.2 dB improvement in jam resistance).^[6] Entropy coding in this type of system serves to reduce the time correlation of the data stream. However, the improvement in jam resistance is also small. Another possible application of entropy coding is the transmission of LPC coefficients. Although the data rate reduction using entropy coding has not been evaluated for LPC, experimental results with vector quantization (a related technique that considers coefficient correlations and probabilities) have produced data-rate reductions by factors as large as two (3 dB).

4.2.2.2 Use of Special Bit Sequence. Speech typically has a preponderance of silent periods. If a special sequence of only a few bits can be used to indicate silence, the average bit rate can



(1)

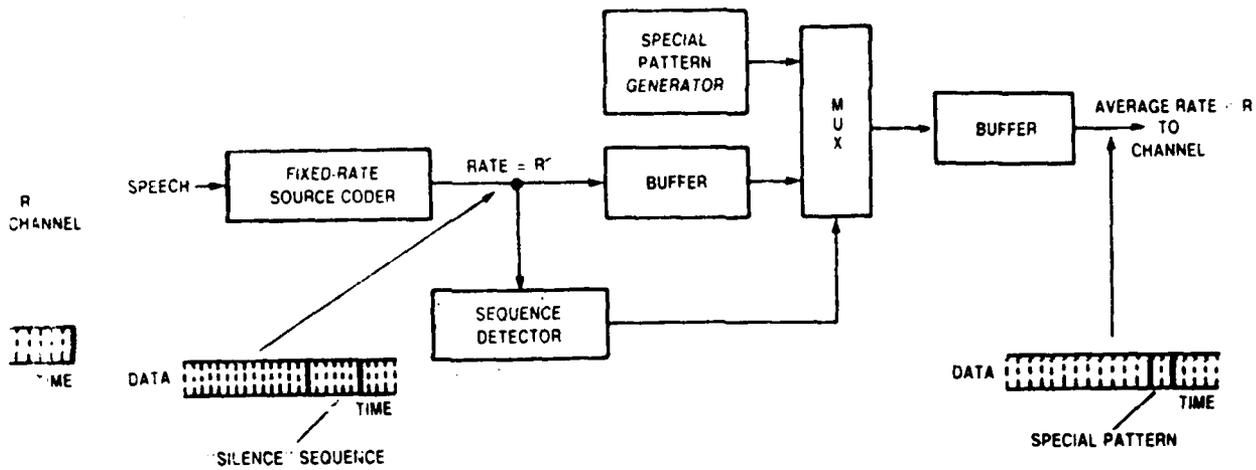


Figure 10b. USE OF SPECIAL BIT SEQUENCE

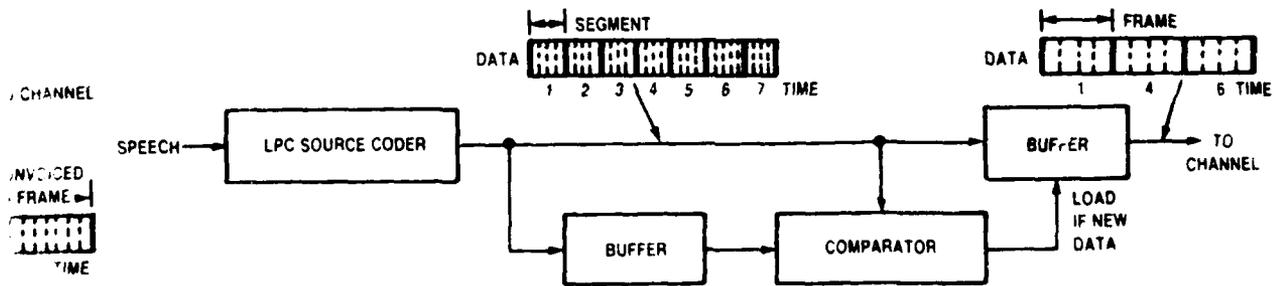


Figure 10d. NON-TRANSMISSION OF REPEATED LPC SEGMENTS

2

Figure 10. SPEECH CODERS WITH VARIABLE RATES CONTROLLED BY THE DA

be reduced. This concept is a modified form of entropy coding in which the number of bits required for silence (which occurs frequently) is reduced. In order to make full use of this technique, there should be a means for indicating the length of the silent period within the special silence code. Although extra bits are needed to indicate length, the reduction in bits required for long silent periods should more than compensate for these extra bits. For a CVSD encoder, the 1010 idle sequence indicates silence. For a PCM encoder, the zero codeword means silence. For both encoders, the output pattern must be distinguishable from data. The output PCM silence pattern should also be an integer number of words long so that word synchronization can be maintained.

Figure 10b shows this type of system. Buffer overflow and underflow can best be handled by shortening or lengthening the silent periods. Buffer control is necessary to anticipate overflow and underflow and to decide whether to add or delete CVSD bits or PCM words. Channel errors in the silence pattern can change the length of the silent period or cause the code to be interpreted as data. Errors elsewhere in the data stream can cause data to be interpreted as silence. None of these errors causes serious problems or drastic degradation of speech quality when the error rate is not too high.

In telephone conversations between two people, non-silent speech transmission in one direction occurs 1/3 to 1/2 of the time. With sufficient added buffering, the improvement in jam resistance could be as much as 3 to 4.7 dB. However, the delays associated with this much buffering would be unacceptable. Assuming that one person is talking more or less continuously, silent periods occur less than 20% of the time so that the potential improvement in jam resistance is less than 1 dB. Thus, the improvement in jam

resistance that can be obtained using this method is highly dependent upon the speech source. Therefore, this technique is recommended only for applications in which short segments of speech are interspersed with periods of silence. The theoretical AJ performance improvement figures stated previously will be reduced by the overhead involved in transmitting the silence pattern.

A special pattern in LPC encoding can be used to indicate that the previous LPC frame should be repeated N times. This special pattern should be the standard LPC frame length and should be distinguishable from frames with speech parameters. For normal LPC encoding and for an average speech segment of 80 ms and a 22.5 ms frame duration, an average of 3.6 frames per speech segment is sent. Using a repeat pattern, the number of frames transmitted for each speech segment is two or less. Therefore, the expected improvement in jam resistance is ≈ 2.5 dB.

4.2.2.3 Use of Different Number of Bits for Voiced and Unvoiced LPC Frames. The number of LPC bits needed to represent a speech segment depends upon whether the segment is voiced or unvoiced. Voiced segments require 54 bits whereas unvoiced segments require only 27 bits. LPC data is normally structured to use 54 bits for all segments.

Figure 10c shows an encoder that uses different numbers of bits for voiced and unvoiced segments in order to reduce the average data rate. The voiced/unvoiced bit specifies the number of bits in the segment. An error in deciding whether a segment is voiced or unvoiced can cause the rest of the message to be garbled. Therefore, special precautions must be taken to ensure that the transmitter and receiver do not lose synchronization. The jam resistance provided for the voiced/unvoiced bit must be very high.

Increased jam resistance can be obtained by adding redundancy and possibly error coding. The result is that the theoretical improvement in jam resistance is partially offset by the need for added redundancy in the voiced/unvoiced bit. Assuming that 30% of all speech segments are unvoiced, the theoretical increase in jam resistance using different length frames is 0.7 dB. The net improvement in jam resistance due to lowering the average data rate is reduced to about 0.5 dB by the redundancy in the V/UV bit. Therefore, this method is not recommended.

4.2.2.4 Use of Variable-Length Speech Segments for LPC^[7,8].

Voice data is normally divided into 22.5 ms segments for LPC encoding. Actual speech segments average about 80 ms; they can be as short as 4 ms or as long as 360 ms. The length of 22.5 ms is short enough to minimize the loss of short segments but long enough to allow good estimates of speech parameters. If segment lengths are varied, the potential for increased jam resistance is approximately $80/22.5 = 5.5$ dB, but suitable means must be found for implementing a system in which data is transmitted only when the speech signal changes significantly.

One implementation approach involves devising an algorithm for segmenting the voice waveform. This is a much more complicated problem when handled by a computer than when done manually using a plot of the waveform in time. The number of bits used to encode each speech segment should also be varied according to the length of the segment. Because the accuracy of the encoded speech parameters for short speech segments is not as good as that for long speech segments, fewer bits should be used for short segments. In order to transmit this type of speech data, a header specifying segment length and bit allocation (frame length) must be appended to each frame. This variable frame length causes problems in maintaining

synchronization in the presence of errors caused by jamming. Redundancy must be added in the header to improve its jam resistance. Assuming that the header information increases the number of transmitted bits by 40%, the increase in jam resistance using this method should be about 4 dB.

A somewhat different implementation approach is depicted in figure 10d. The speech waveform is divided into the usual 22.5 ms segments, but a frame of encoded data is transmitted only when there is a significant change in speech parameters. This approach avoids the difficult problem of devising an algorithm for segmenting the speech waveform. In this case a header must also be added to specify the number of 22.5 ms speech segments represented by each frame of transmitted data. Assuming the same 40% increase in transmitted data due to the header, the improvement in jam resistance should also be about 4 dB. This approach is similar to the use of a repeat-frame code as discussed in section 4.2.2.2. The difference in jam resistance is due to the use of a header of 40% of a frame length rather than a whole extra frame. Therefore, the method described here is preferred. The improvement in jam resistance may justify the additional complexity in implementing this method.

4.3 CONCLUSIONS

In two-way communications with an error-free feedback path, several techniques exist for matching a speech source to the channel. Two techniques, source coder selection and source bit selection, adjust the data rate in discrete steps according to the jamming level so that performance is optimized at all jamming levels. This ideal type of system has limited usefulness because of the need for an error-free feedback path. A pulsed jammer also poses a problem.

When the jamming level is unknown at the transmitter, the communication system cannot match the source to the channel. For these applications, data-rate-reduction techniques provide increased jam resistance. The techniques described require buffers to compensate for the difference between the variable instantaneous and average data rates. These buffers add delay and require some means for handling overflow and underflow. Of the various techniques studied, only two offer a significant increase (≈ 4 dB) in jam resistance due to a reduction in the average data rate. These two are: non-transmission of repeated LPC segments and use of a special bit sequence for silent periods. The latter technique is restricted to applications where short periods of speech are interspersed with longer periods of silence.

SECTION 5

SPEECH APPLICATIONS OF MULTIPLE-DATA-RATE TRANSMISSION

Our previously designed variable-data-rate/variable-AJ system can simultaneously transmit three data streams at three different data rates with three different amounts of jam resistance. This capability is achieved by a signaling structure in which various spread spectrum techniques are combined in different ways for each of the data streams. The application of this multiple-data-rate/multiple-AJ concept to speech encoding is discussed in this section.

5.1 AJ PERFORMANCE OF MULTIPLE-DATA-STREAM SPEECH CODERS

In addition to the obvious method of simultaneously transmitting the outputs of several speech coders such as PCM, CVSD and LPC, several speech encoding methods that produce two data streams have been described in the literature. The source coders for these methods are depicted in figure 11. In all of these methods, the lower-rate data stream requires more jamming protection than the higher-rate data stream. Hence, all of these methods are possible source coding candidates for use with a multiple-data-rate/multiple-AJ spread spectrum system in a jamming environment. In the presence of jamming, errors occur first in the high-data-rate bit stream and cause some degradation of voice quality. As the jamming level increases further, the number of errors in the high-data-rate bit stream increases until that data stream conveys no information. Speech quality is then determined by the information conveyed by the low-data-rate bit stream.

The multiple-data-stream encoders being considered transmit the same data regardless of the jamming level. Therefore, the source is not matched to the channel. Some inherent loss in AJ performance



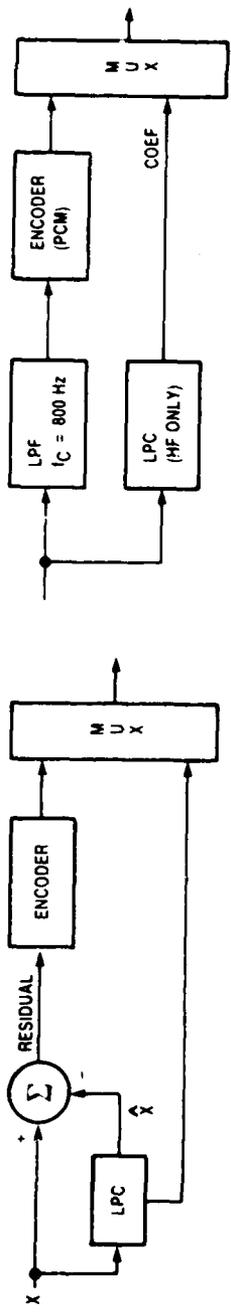


Figure 11a. RESIDUAL EXCITED LPC

Figure 11b. BASEBAND CODER

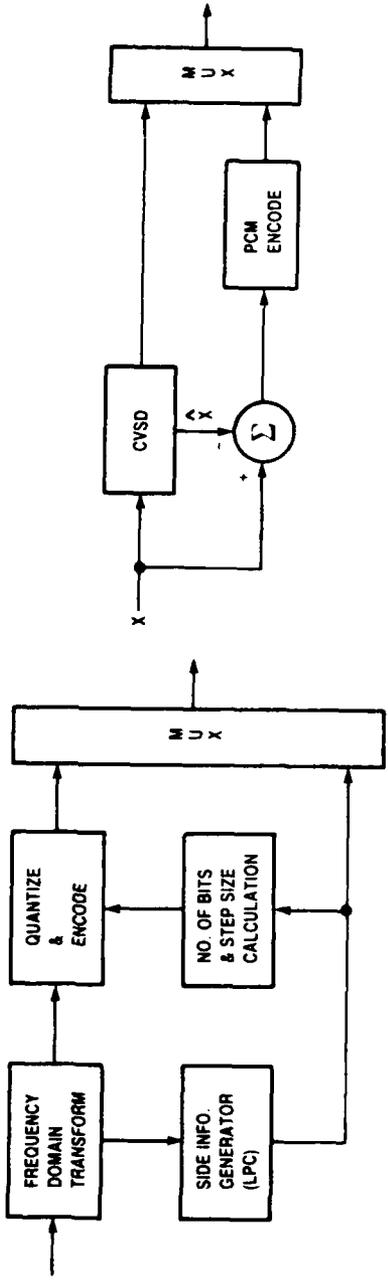


Figure 11c. ADAPTIVE TRANSFORM CODING WITH SIDE INFORMATION

Figure 11d. EMBEDDED TRANSMISSION (CVSD, PCM, DPCM)

Figure 11. MULTIPLE-RATE TRANSMISSION

occurs at high jamming levels because some of the transmitter power is wasted on the high-data-rate stream which contains so many errors that it conveys no information. The advantage of multiple-data-stream encoding is the large range over which speech quality varies with jamming level without the need for a feedback path.*

The multiple-data-stream speech coders depicted in figure 11 are described in the following paragraphs.

5.1.1 Residual Excited LPC

LPC encoding imparts a mechanical quality to the reconstructed speech signal because LPC does not attempt to reconstruct the speech waveform. Although the predictor section of the algorithm tries to duplicate the waveform, the predictor in the receiver is excited artificially by either white noise or pulses at the pitch period. The result is an unnatural sound in the voice output.

Residual excited LPC provides improved quality over standard LPC encoding. As indicated in figure 11a, the difference between the actual waveform and the predicted waveform is encoded separately, perhaps using PCM, and multiplexed with the predictor coefficients for transmission. In the receiver, the high-data-rate residual data is used to excite the predictor specified by the low-data-rate predictor coefficients. Consequently, a better reconstruction of the waveform is obtained, and the output speech sounds more natural.

*Some of the coding methods depicted in figure 11 could make use of jamming-level feedback to match the source to the channel and provide even better performance.

Performance in a jamming environment depends upon the method of recombining the two data streams in the receiver. When the high-data-rate bit stream is heavily jammed, it is desirable to detect this condition and then use only the low-data-rate LPC data to generate the output signal. With this type of receiver, excellent speech quality can be obtained when the jamming level is low, and fair to good speech quality can be obtained using standard LPC when the jamming level is high.

5.1.2 Baseband Coder

The baseband coder (voice-excited coder) shown in figure 11b also provides improved quality over LPC-encoded speech. High- and low-frequency components are encoded separately. The low-frequency components are encoded using a waveform coding technique such as PCM, thereby generating a high-data-rate bit stream. The high-frequency components are encoded using LPC, thereby producing a low-data-rate bit stream. The combined data rate is typically about 9.6 kbps.

In the presence of jamming, the method of generating the output signal in the receiver should be modified according to the jamming level. The basic method described for the residual excited LPC receiver also applies to the baseband coder receiver. Such a receiver provides graceful degradation of speech quality with increased jamming.

5.1.3 Adaptive Transform Coding With Side Information

As seen in figure 11c, the first step is to take the discrete cosine transform of a segment of data. Side information, which is generated using LPC, is transmitted at a low data rate and is also

used to determine the method of quantizing and encoding the transform coefficients which are transmitted at a high data rate. This method provides very good speech quality at about 9.6 kbps.

Each type of data transmitted requires a different amount of jam resistance. Existing algorithms use error-correction coding to provide increased jamming protection for some data. The use of multiple data streams allows the jam resistance of each type of data to be adjusted as required. Multiple-data-stream transmission may provide better performance over a wide range of jamming levels.

5.1.4 Embedded Transmission

An example of embedded transmission is shown in figure 11d. The input signal is encoded using a CVSD algorithm at some rate, perhaps 16 kbps. The difference between the input and the CVSD reconstructed signal is encoded using PCM. The output bit rate of the PCM encoder is higher and depends upon the number of PCM bits per sample.

Without jamming, speech quality is equivalent to that of PCM encoding. As the jamming level increases, the errors in the PCM bit stream should be detected, and the method of combining the bit streams should be modified. At high jamming levels, the PCM bit stream should be disregarded so that performance is determined by the CVSD encoder.

5.2 FUTURE INVESTIGATIONS

For all of the methods described, the manner of recombining the two data streams in the receiver needs to be optimized to obtain the

best possible speech quality at all jamming levels. Optimization of the speech coder and signal reconstruction methods in the receiver will require considerable effort. This optimization must be based on subjective evaluation of speech quality. MITRE's Audio Signal Conversion Laboratory and Corporate Research Computer Facility provide the means for computer simulation of the coding methods shown in figure 11. After optimization of speech quality using these facilities, breadboards and/or our variable-data-rate/variable-AJ data link system will be used to demonstrate performance in a jamming environment.

SECTION 6
SUMMARY AND CONCLUSIONS

Bit weighting and mean-square error coding have been investigated for their ability to provide different amounts of jam resistance for the bits in a PCM-encoded data stream. Bit weighting provides about a 2 dB increase in jam resistance for 7-bit companded PCM. When the bit weighting method is compared with the simpler method of dropping bits, the AJ improvement is not as great. The major advantage of bit weighting is near-optimum performance over a wide range of jamming levels. Error-correction coding with a mean-square error criterion does not improve performance.

Several anti-jam techniques that exploit the special characteristics of speech to reduce the average channel data rate have been studied. All of these techniques require buffers to handle the variations in data rates and special provisions for handling overflow and underflow. Entropy coding increases the jam resistance of companded PCM by less than 1 dB. Two techniques that may increase jam resistance by approximately 4 dB are 1) non-transmission of repeated LPC segments, and 2) use of a special bit sequence for silent periods in applications where short periods of speech are interspersed with longer periods of silence.

Matching a speech source to a jamming channel provides the best possible performance, but it requires an error-free feedback path from the receiver to the transmitter to optimize the source coder to the jamming level. The need for this feedback path and problems caused by a pulsed jammer limit the usefulness of this concept. Nevertheless, several source-coding methods for obtaining this ideal performance have been described, assuming that suitable applications exist.

In the absence of a feedback path, multiple-data-stream speech encoding offers the possibility of good performance over a wide range of jamming levels. Several such encoders have been identified for investigation next year. Optimization of the encoder and determination of the best way to recombine the data streams in the receiver require considerable study.

Facilities for computer simulation of speech algorithms have been constructed and utilized as part of this program. These facilities will be used much more extensively in next year's investigation of multiple-data-stream speech encoding methods.

REFERENCES

1. B. D. Campbell, E. A. Palo, and M. A. Poole, "A Data Link Design with a Variable-Data-Rate and Variable-Jam-Resistant Capability", MTR-3850, Sept. 1979, (also published as RADC-TR-80-214).
2. E. Bedrosian, "A Weighted Pulse Code Modulation Communications System", PhD Thesis Northwestern University (1953).
3. E. Bedrosian, "Weighted PCM", IRE Trans. Info. Theory, Vol. IT-4, March 1958, pp. 45-49.
4. G. R. Redinbo, "Optimum Symbol-by-Symbol Mean-Square Error Channel Coding", IEEE Trans. Info. Theory, Vol. IT-25, July 1979, pp. 387-405.
5. G. R. Redinbo, "The Optimum Mean-Square Decoding of General Block Codes", Information & Control, Vol. 31, 1976, pp. 341-363.
6. J. L. LoCicero and P. M. Bocci, "Entropy Coding of ADM Speech Signals", IEEE Int. Conf. Acoustics, Speech & Signal Processing, May 1980, pp. 42.7.1-42.7.5.
7. R. D. Preuss, "Practical Considerations in the Development of a High Performance Low Data Rate Variable Frame Length Digital Voice Communications System", MTR-8322, 15 July 1981.
8. R. A. J. Clapp, "Acoustic Segments in Natural Speech: Analysis and Statistics", MTR-8405, 20 August 1981.

APPENDIX
SPEECH PROCESSING FACILITY

The purpose of this facility is to provide a simple and flexible environment for the development, test, and evaluation of digital signal processing algorithms for audio band signals (below 20 kHz). In particular, the laboratory was developed primarily for processing voice band signals (below 5 kHz). With very few exceptions, processing functions are currently limited to monaural half-duplex operation.

The Audio Signal Conversion Laboratory (ASCL) is an extension to the Corporate Research Computer Facility (CRCF). The primary processing system for the CRCF is a Digital Equipment Corporation VAX 11/780 which hosts the Digital Systems Corporation DSC200 signal conversion system in the ASCL. The Digital Systems Corporation DMA11 provides the physical interface between the VAX/UNIBUS and the DSC200/XBUS. This system is pictured in figure A-1.

Equipment within the ASCL currently consists of

- Infinity RS_e two-way speaker system
- Klark-Teknik graphic equalizers
- Frequency Devices 901F lowpass filters, (4, 2 with batteries)
- TEAC A-2340SX four-channel tape recorder
- Technics RS-M270X cassette recorder
- Marantz PMD-340 portable cassette recorder
- Jenson JE-11S-L balancing transformers
- DSC240 monaural amplifier and power supply





Figure A-1a. Audio Signal Conversion Laboratory



Figure 1-Ab. Corporate Research Computer Facility

Figure A-1. Facilities

- DSC200 signal conversion system
 - mono A/D board
 - stereo D/A board
 - stereo data buffer module
 - enclosure and power supply
- DSC DMA11 interface
- Sennheiser HMD224X headset
- Realistic NOVA40 headphones
- Tektronix type 545A oscilloscope
- General Radio random noise generator
- General Radio 1310-B oscillator
- Digital Electronics Corporation VT-100 terminal
- Digital Engineering VT-640 retro-graphics terminal

In addition to the equipment located in the ASCL, a Digital Equipment Corporation RMO3 disk drive was purchased for use with this system. Magnetic tape drives with the VAX computer are also used in system operation. Figure A-2 shows the hardware configuration.

Software support for the system consists of the UNIX operating system for the VAX computer, the Interactive Laboratory System (ILS) software package for signal processing and speech analysis, and the ASCL software library of special purpose programs. Many programs have already been written or obtained from other sources for the ASCL library.

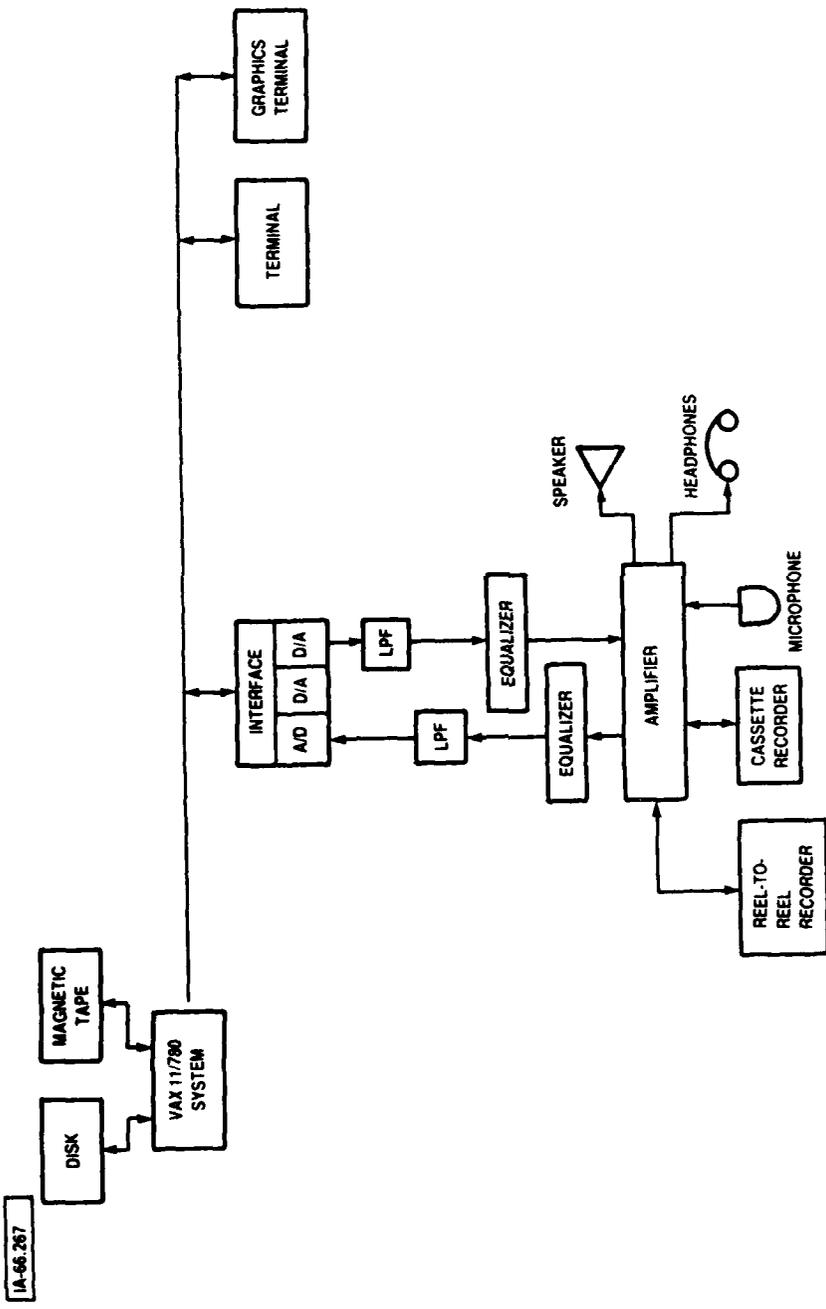


Figure A-2. SPEECH PROCESSING FACILITY



MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C³I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.

ATE
LME