USBM Contract No. H0346045

TECHNICAL SERVICES FOR MINE COMMUNICATIONS RESEARCH

APPLICABILITY OF STATE-OF-THE-ART VOICE BANDWIDTH COMPRESSION TECHNIQUES FOR WIRELESS MINE COMMUNICATION

Richard H. Spencer -- Task Leader
Warren G. Bender

ARTHUR D. LITTLE, INC.

Cambridge, Massachusetts 02140

C-77229

USBM CONTRACT REPORT -- TASK A (Contract No. H0346045)

TASK ORDER NO. 1

July 1975

UNITED STATES

DEPARTMENT OF THE INTERIOR

BUREAU OF MINES

FOREWORD

This report was prepared by Arthur D. Little, Inc., Cambridge, Massachusetts under USBM Contract No. H0346045. The contract was initiated under the Coal Mine Health and Safety Program. It was administered under the technical direction of the Pittsburgh Mining and Safety Research Center with Mr. Howard E. Parkinson acting as the technical project officer. Mr. Michael W. College was the contract administrator for the Bureau of Mines.

This report is a summary of the work recently completed as part of this contract during the period May 1974 to July 1975. This report was submitted by the authors in July 1975.

No inventions or patents were developed and no applications for inventions or patents are pending.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

TABLE OF CONTENTS (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES

I. EXECUTIVE SUMMARY

A. INTRODUCTION

Since the passage of the Coal Mine Health and Safety Act of 1969, the Bureau of Mines has made many advances in mine communications. These efforts have resulted in the following developments and demonstrations:

- Surface-to-surface electromagnetic transmission of baseband voice signals through overburdens as deep as 600 feet. This transmission was achieved with equipment from the Interim Mine Rescue and Survival System Program, and demonstrated that baseband voice communications could be obtained through these overburdens in the absence of mine-generated noise; that is, when the mine power system was shut off. However, during mine operational conditions, the mine-generated noise proved to be so large that effective baseband voice communications to the desired depths could not be effected.

- The development of beacon transmitters for locating trapped miners. This work was again the outgrowth of the Interim Mine Rescue and Survival System. Extensions of this work have led to reliable detection and location of trapped miners to depths of 1000 feet. The system, which makes use of a loop antenna deployed where the trapped miner is located, uses low-duty-cycle, tone-burst transmission. Portable tuned receivers and loop antennas are used on the surface to determine the location of the trapped miner.

  An extension of this work has led to a call-alert paging system concept, wherein a transmitter, similar to that used in the trapped miner system, is used to audibly and/or visually alert key mine personnel carrying pocket receivers that they are wanted at the nearest mine telephone. In addition, a roof-bolt paging system has been developed, wherein commercial paging and mine carrier phone equipment are used in conjunction with a long-wire "roof bolt" antenna to page individuals and relay to them short voice messages.

Despite these advances in communication capabilities within mines, some major unmet needs which require further research and development still remain. One of these major needs is wireless electromagnetic through-the-earth voice communication channels capable of being operated in the presence of both surface and subsurface noise environments. Calculations have been made by both Arthur D. Little, Inc.[(1)] and Collins Radio Company[(2)] relating to the power requirements for such a voice communication system. In both instances the power requirements under operational conditions for depths of 600 to 1000 feet have been quite substantial, in fact, prohibitive, for uplinks and downlinks between the surface and the subsurface. The power requirements range from a few kilowatts up to the 100-kilowatt range. Similar levels were also predicted for side-link communication paths within the mine. These paths may run from a central fixed transmitter to roving miners carrying portable units, or between two portable units. Here the problem is not so much reaching the roving miners from the fixed transmitter location, but the reverse talk-back problem of the roving miner to the central location, or from roving miner to roving miner where transmitter power is severely limited. The Collins Radio Company's calculations for power requirements for the roving miner to roving miner indicate that portable battery-operated equipment appears barely adequate to meet a 600-foot range requirement imposed on such side-link communication systems.*

---

*Recent in-mine EM propagation measurements made by Collins do show an unexpected propagation enhancement for frequencies near 500 kilohertz. It is not known how consistent this enhancement will be over a range of coal mine environments, but should it be found to be consistent, then the problem of achieving adequate side-link communication range will be considerably alleviated. However, it appears that the mechanism that provides this enhancement of side-link propagation will serve to make vertical propagation more difficult and thus the problem of subsurface to surface communication may become worse under such conditions.

One of the means proposed for aiding in overcoming these severe power and range limitations has been the application of voice bandwidth compression techniques. We anticipated that such techniques would permit greater range with the same transmitter power or the same range with less transmitter power.

The purpose of this study was to determine the applicability of state-of-the-art voice bandwidth compression techniques for mine wireless voice communication systems as one means of reducing the power requirements of portable units. One reason why voice bandwidth compression appeared attractive is that at VLF to MF frequencies the noise received by a portable receiver in a mine is largely dominated by external mine-generated noise and not by receiver-generated noise. Therefore, if the received noise power could be lessened by reducing the bandwidth required for the voice signal, without paying a corresponding penalty in processing power and/or required signal-to-noise ratio, the desired range could be either extended with the same transmitter power or obtained with less power. Hence, a 10:1 reduction in receiver-bandwidth could be expected to yield a 10:1 reduction in received noise and hence a 10:1 improvement in signal-to-noise ratio for the same amount of transmitter power. Extending this simplistic notion even further, a 100:1 bandwidth reduction could be expected to yield a 100:1 reduction in transmitter power. These potential values of transmitter power reduction were large enough to undertake this assessment of the potentials of state-of-the-art voice bandwidth compression techniques for achieving the goal of significantly reduced transmitter power consumption for mine wireless communication applications.

B. FINDINGS

1. Real-Time Techniques

Real-time voice bandwidth compression techniques currently do offer some modest savings in transmitter power requirements. The savings, however, are obtained at the cost of increased complexity for transmitters and receivers, and we therefore deem them inapplicable to most mine communication needs. There are several fundamental reasons why this is so:

1. There has been no demonstration as yet of a system which even comes close to coding speech at a 50-bit-per-second (bps) rate, which is the information rate attributed to a voice transmission channel;

2. Although substantial bandwidth reductions are achieved by certain of the voice bandwidth compression schemes, such reductions are achieved at an increase in the required signal-to-noise ratio, thus obviating some of the expected savings of transmitter power that would accrue if the systems were able to operate at the same signal-to-noise ratio as for a conventional voice band;

3. The emphasis of most of the work on voice bandwidth compression has been directed toward means of achieving more voice channels in a given bandwidth. There has been very little concern in this work with the added necessity of increasing power per channel. It has rather been strictly based on getting more channels into a given bandwidth. For For this reason, the results are less applicable to mine communication needs than might otherwise be expected.

Bandwidth compression of modest amounts; that is, more than 4 or 5 to 1 in bandwidth, is generally achieved with analysis-synthesis systems. In this type of system, the input voice signal is first analyzed. Then voice parameters extracted by the analysis, parameters which will allow the voice signal to be reconstructed, are transmitted in a narrow band to the receiver. At this point these extracted parameters are used to reconstruct the original voice message. Research in this area in the 1930's, 1940's, and 1950's was directed toward using analog circuits to implement these systems. Work centered on digital systems in the 1960's and 1970's, and progress has been significant.

To digitize conventional telephone speech requires approximately 50,000 bps. This assumes a bandwidth of about 3000 Hz and a sampling rate of about 10,000 Hz with quantizing to around 8 bits. Modern digital systems can encode the analyzed version of the speech into 1200 bps and, in some extreme cases, into as little as 700 bps. This indicates a compression of 40:1 is obtainable by these systems. However, the more advanced systems of this type require considerable computer power, placing them considerably beyond their application to mine communication requirements.

2. Non-Real-Time Techniques

In the area of non-real-time voice bandwidth compression techniques, we feel there may be some merit to the stretched-out voice system wherein an original message T seconds long is in effect stretched to NT seconds long for transmission. All the quality of the original speech may be preserved at the same signal-to-noise ratio as would be required for full bandwidth transmission, but with a bandwidth that is shrunk in direct proportion to the stretch-out factor. Thus, true transmitter power requirement savings can be made, but only at the expense of a delay between the time the message is spoken and the time it is available to be heard at the receiver.

C. CONCLUSIONS

From our study we conclude that real-time state-of-the-art voice bandwidth compression techniques are not suitable as a practical means for improving mine wireless communications performance at present. However, non-real-time voice bandwidth compression techniques and real-time, non-voice systems do offer means for meeting the needs of through-the-earth emergency subsurface to surface communications. We conclude that the advantage of limiting the voice signal waveform to improve received signal-to-noise ratio should be applied to mine wireless communication systems.

D. RECOMMENDATIONS

Based on the findings and conclusions of our study, we recommend that the stretched-out voice systems be further explored as a means of meeting certain of the mine communication needs, particularly for links through the earth for emergency purposes, and most important, the link from underground to the surface. It is particularly true that power savings here can be important because of the limited amount of energy available to a miner trapped underground and the requirements for intrinsic safety for electrical communication equipment. We feel that in an emergency the delay time imposed by a stretched-out speech system becomes a much less important objection than it is for use in the day-to-day operation of the mine; and that a suitable system could be made to operate at a substantially reduced power using this means.

We also recommend that non-voice-type systems where truly narrow-band operation can be achieved be studied further. One such system might be a very low-bandwidth teletype-quality system wherein the operator uses an alphabet keyboard to spell out his message and a LED display to present both incoming and outgoing messages.

State-of-the-art electronic components are available today which make both of these alternatives to a real-time voice system extremely attractive for emergency mine communication needs, and capable of being demonstrated in the form of prototype hardware with only a modest development effort.

II. THE NATURE OF VOICE SIGNALS

Speech is produced by the excitation of a complicated acoustic filter by raw acoustic energy. The filter consists of the cavities of the throat, nose, and mouth. The energy sources are either produced by air vibrating the vocal cords (voiced sounds) or by turbulent air flow between the teeth or between the tongue and the teeth or the palate (unvoiced sounds) or by some such combination.

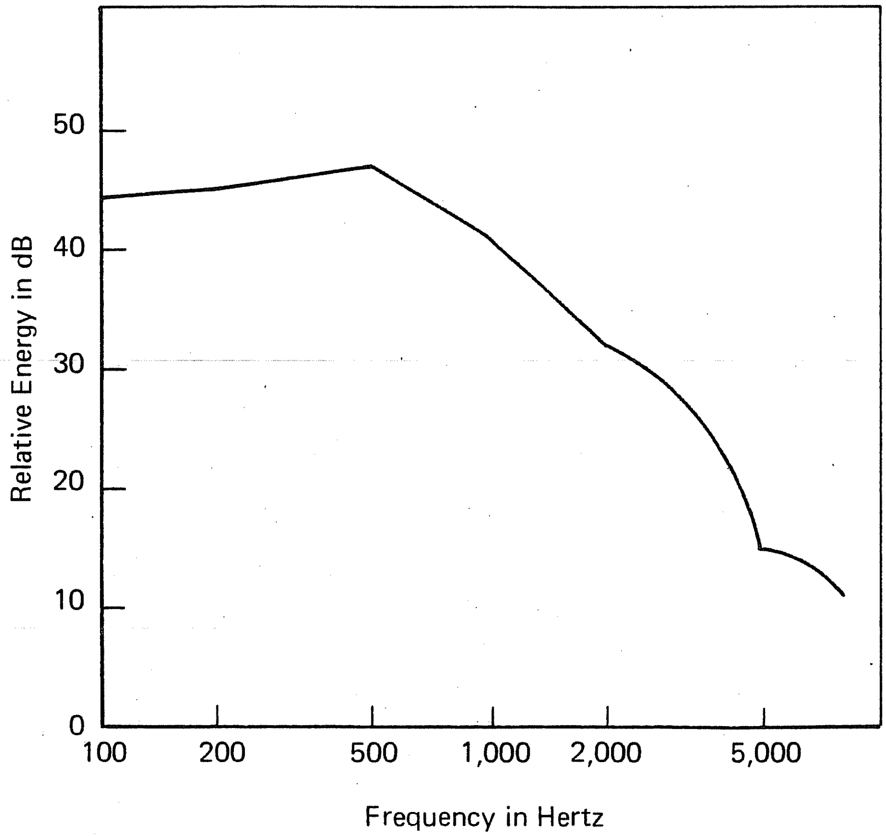
Typical vocal tract filter characteristics consist of a series of resonances and anti-resonances with as much as 10 to 20 dB between peaks and valleys. At the peaks of these resonances some of the harmonics in the exciting forces are reinforced relative to the others. These frequencies where reinforcement takes place are referred to as formants, of which there may be as many as five.

For unvoiced sounds the spectrum of the exciting signal is like noise spread out over the audio range. If they go through the filter, they exhibit formant frequencies. Hissing sounds that do not go through the filter (ssss and sh) are noise-like, with a very wide frequency band.

Voiced sounds look like a train of impulses with a fundamental frequency between 60 and 240 Hz (for men, but higher for women and children) with a rich harmonic structure that falls off with frequency up to perhaps 4000 Hz. These harmonics are modified by the filter and the formants form an envelope on the harmonics.

Formants differ from individual to individual, making the voices of different individuals distinctive. However, the combined effects of harmonics and filter always produce a rapidly falling spectrum. Figure 1 shows the envelope of spectra for male speakers and illustrates this feature.

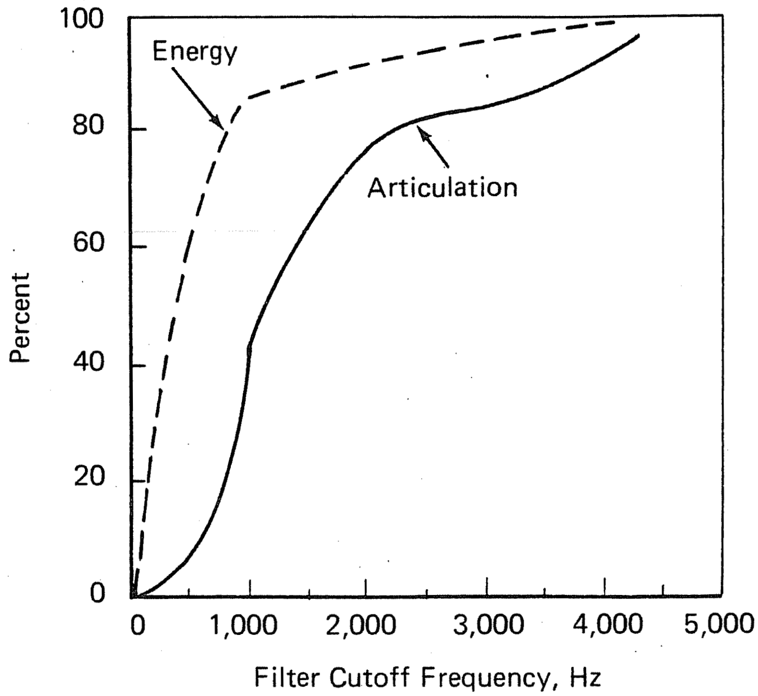
One of the most straightforward ways of decreasing the frequency required for voice transmission is simply to chop some of it off. Extensive tests have been conducted on the intelligibility of telephone speech as a function of bandwidths.



Source: Ref. 3

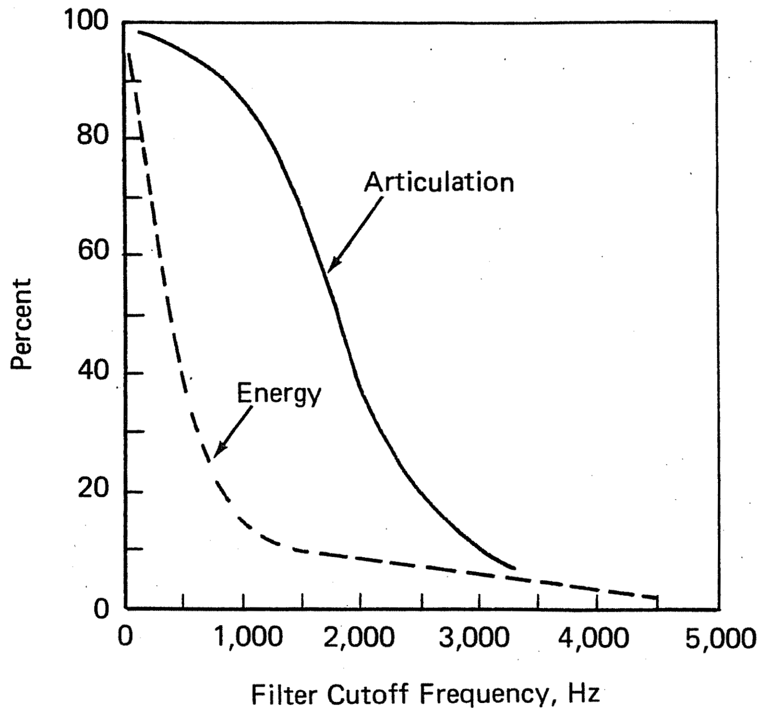
FIGURE 1 ENERGY DENSITY SPECTRUM FOR MALE SPEECH

Figure 2 shows the effect of changing the upper end of the speech band with a variable cutoff low-pass filter. If the band is restricted to 1000 Hz, intelligibility is decreased to 40%, but 83% of the energy is still present. This curve shows that low frequencies contribute to energy much more than to articulation. Figure 3 shows the effect of eliminating low frequencies with a variable cutoff high-pass filter. This shows that elimination of frequencies below 900 Hz leaves intelligibility above 90%, but decreases energy to 17% of the total. While intelligibility is preserved, speaker recognizability deteriorates severely.



Source: Ref. 4

FIGURE 2 ENERGY AND ARTICULATION FOR SPEECH SIGNALS WITH LOW-PASS FILTER



Source: Ref. 4

FIGURE 3 ENERGY AND ARTICULATION FOR SPEECH SIGNALS WITH HIGH-PASS FILTER

(This page intentionally left blank.)

III. METHODS OF VOICE BANDWIDTH COMPRESSION

Systems designed to compress the bandwidth of speech fall into two categories:

1. Those which act or modify the speech signal directly, and
2. Those which analyze the signal and transmit facts about this analysis to the far end where they can be used to reconstruct (synthesize) the original speech.

In the following sections we treat methods of bandwidth compression of these two classes.

A. MODIFICATION OF SPEECH SIGNALS

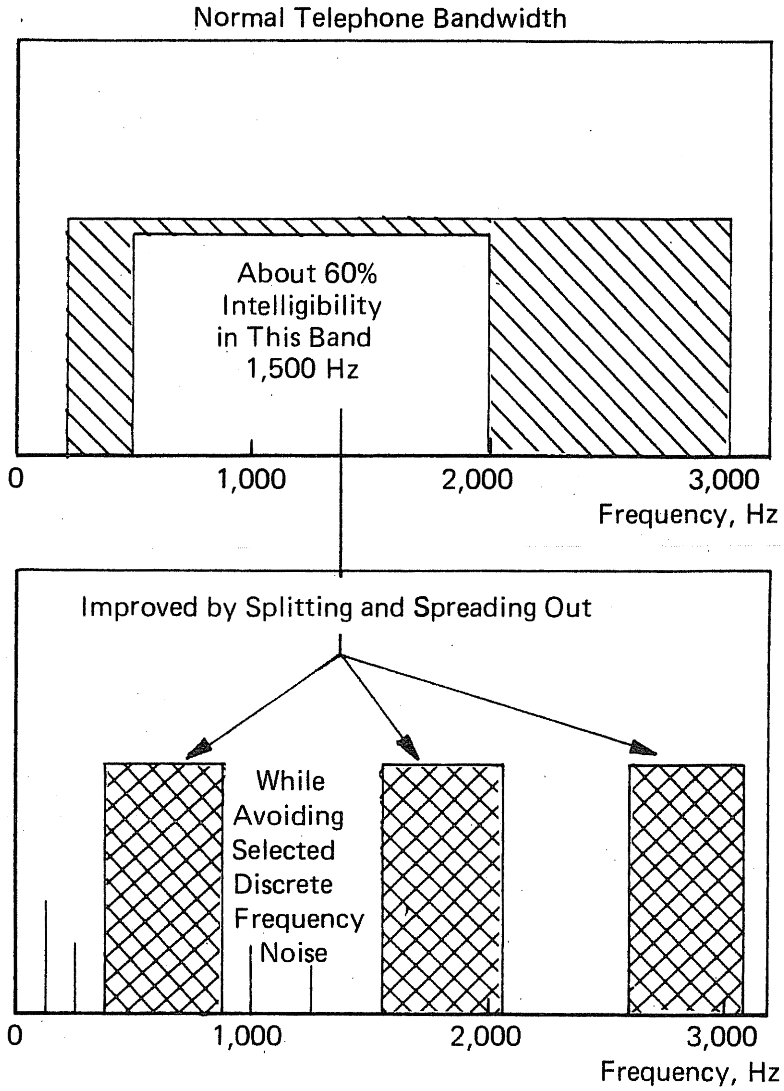
1. Band Limiting

An obvious way of decreasing bandwidth is to cut some of it off with a filter or filters. We have observed in Figures 2 and 3 that very little articulation is lost by cutting off a substantial portion of the lower frequencies or a moderate amount of the upper end of the band. Restricting the bandwidth decreases the noise at the receiver. Cutting off lower frequencies is especially attractive in a noise environment with large discrete frequency noise contributions at the power frequency and its first several harmonics.

During World War II the Bell System installed "emergency banks" that provided two voice channels in a standard 4000 Hz slot intended for one voice channel. This method was accepted as a wartime measure and provided at least marginally useful channels.

Experiments have shown that restricting the band to 500 to 2000 Hz results in about 60% intelligibility for isolated words, but provides adequate performance for connected speech with suitable repetitions. Kryter[(5)] has found that intelligibility with a 1500 Hz bandwidth can be improved by dividing the band into three separate 500 Hz bands centered within the range 500- to 3000-range (see Figure 4).

Pursuit of this simple approach may be an attractive possibility for mine communication, since the bands (not necessarily limited to



Source: Arthur D. Little, Inc.

FIGURE 4 BAND LIMITING SPEECH SIGNAL

three, nor to a total of 1500 Hz) may be strategically located to avoid dominant noise sources.

2. Time Assignment Speech Interpolation

Time assignment speech interpolation (TASI) makes use of silent periods in speech in a given direction -- either while the source is listening or in silent intervals between words or sentences -- by interleaving words from different speakers. This has been effective on transatlantic cables where the medium is costly enough to warrant the expensive terminls, but it does not seem applicable to mine communications.

3. Temporally Interrupted Speech

Channel utilization can be improved by continuously interrupting the speech signal at a fairly high frequency. If a 50% duty cycle is used, another channel may be interleaved in the same band. The resultant intelligibility is a little better than that obtained by halving the frequency bandwidth. If the samples are run together, they can be transmitted at half speed and put back together at the receiving end, halving the bandwidth required in the channel.

4. Clipping

Where the signal-to-noise ratio is low, some improvement in intelligibility results from simple clipping. This results from increasing the energy in lower amplitudes at the expense of distortion on the peaks. Simple clipping differs from companding in that no compensating expansion takes place at the receiver.

Clipping may be carried to the limit, giving a constant-amplitude "infinitely clipped" rectangular waveform. The only information preserved is the zero crossings of the original speech. Surprisingly, little intelligibility is lost, but speaker recognizability is almost obliterated and the quality is unpleasant.

An improvement in effective signal power of as much as 12 dB can be obtained from clipping. This represents a large number of watts at the power levels we are considering.

5. Frequency Division (Stretched-out Voice)

One of the most effective ways of decreasing bandwidth is to divide all frequencies by a constant before transmission and multiply them back to their original values at the receiver. If all cycles of all frequencies are transmitted, this implies a slowing down of transmission by the constant used. An elementary way to achieve this is to record speech on a tape recorder and play it over the line at a slower speed with a corresponding decrease in bandwidths. At the receiver the speech is again recorded and played back sufficiently speeded up to reconstitute the original.

In principle, there is no limit to the savings in bandwidth achievable in this way. From an information theory viewpoint there is absolutely no saving in channel capacity since a greater time is used. Time and bandwidths are exchanged. Further processing may reduce the actual channel capacity needed. Interrupted samples may be joined and sent as discussed in Section A-3.

Real-time compressed speech may be sent by dividing in real time with, for example, phase-locked loops. The divided frequency is sent for the amount of <u>time</u> -- not cycles -- that it occurs and is multiplexed back up for that time at the receiver. This is actually a form of "synthesis-analysis" compression where a frequency is encoded into 1/n times itself and decoded by multiplying by n at the receiver. This type of processing goes beyond modification of the voice signal and will be discussed further in Section B-2, Harmonic Compression. This form of frequency division is limited to about 3 to 1.

B. ANALYSIS-SYNTHESIS TECHNIQUES - VOCODERS

The compression methods discussed so far try to preserve, however crudely, the original voice waveform. Another approach to bandwidth compression is to transmit only the important <u>properties</u> of the speech signal with the intention of synthesizing at the receiver a signal that sounds like the original. Systems that work in this way are called "vocoders," a term coined in 1928 by Homer Dudley of Bell Telephone Laboratories.

Three important properties of human hearing are fundamental in the analysis-synthesis process in vocoders:

1. The ear is a spectrum analyzer with a relatively short integrating period;
2. For monaural sounds, the ear is virtually insensitive to phase; and
3. Hearing is extremely sensitive to pitch variations.

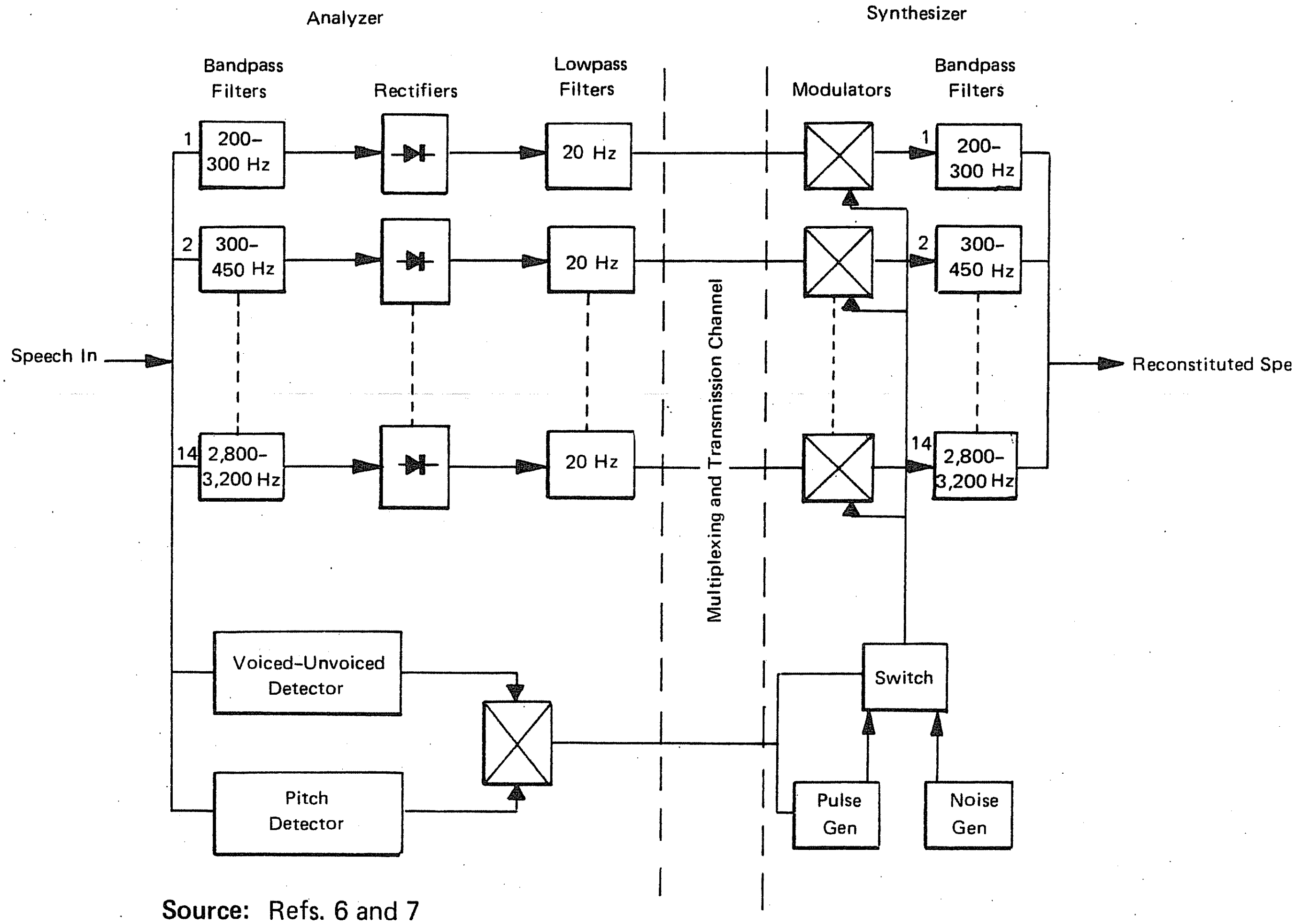
For a vocoder to be successful then, the short-term power spectrum has to be preserved without worrying about the phase relationships. The generic system for doing this is the spectrum channel vocoder.

1. Spectrum Channel Vocoder

Figure 5 is the block diagram of a spectrum channel vocoder. A set of bandpass filters divides the voice signal into frequency channels. To determine the power in each band, the signals are rectified and passed through low-pass filters. The outputs of these filters are continuous estimates of the speech power spectrum in each channel. Working on the whole input signal, a pitch extractor determines the pitch and a detector determines whether the speech sound is voiced or unvoiced.

The separate channels are transmitted by any of a number of means, either analog or digital, to the receiver synthesizer which reconstructs the signals using the spectrum energy estimates to modulate the reconstructed pitch and voicing signal. During voiced segments a pulse generator puts out short pulses at the pitch rate. During unvoiced sounds, a noise generator output is fed to the filter bank modulators.

Vocoder research was active in the 1950's and concentrated on the number of channels and optimization of channel filter design. With increasing emphasis on digital transmission and availability of powerful computers in small spaces, recent channel vocoder research has been directed to digital implementation.



Source: Refs. 6 and 7

FIGURE 5 SPECTRUM CHANNEL VOCODER

2. Voice-Excited Vocoders*

Naturalness in channel vocoders can be increased by eliminating the pitch detector and voicing detectors and substituting a low-frequency portion of the baseband signal (hopefully containing the fundamental). The excitation function is generated at the synthesizer by rectifying, filtering, and clipping the baseband signals, thus creating a spectrally flat signal with energy at pitch harmonics for voiced sounds, noise for unvoiced sounds, and silence for silence. The set of filters does not need to include the frequencies covered by the low-bandpass filter since these signals can be directly added in at the output (see Figure 6).

3. Cepstrum Vocoder

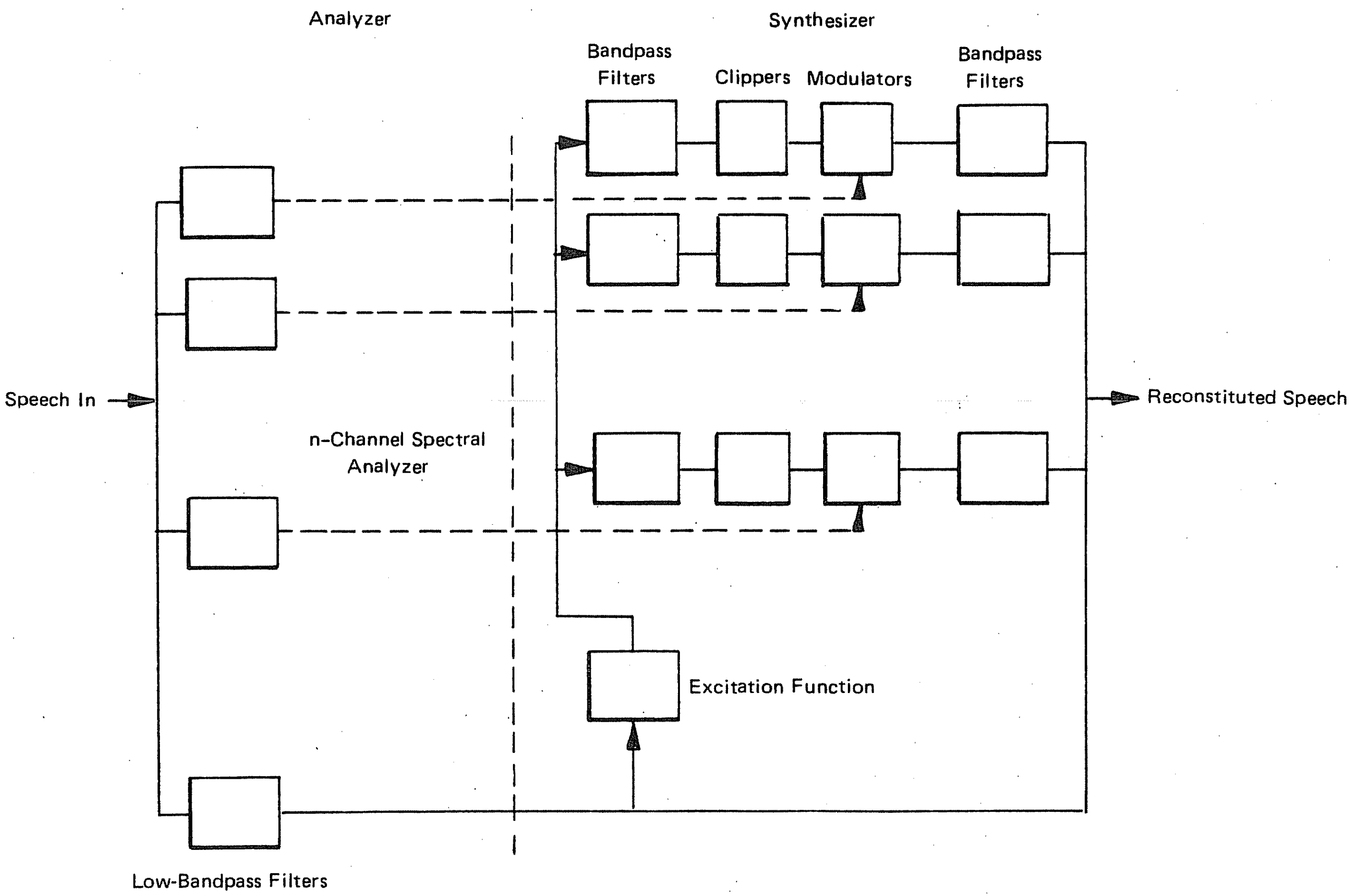
The channel vocoder essentially transmits the Fourier transform of the speech signal. Because of the nature of typical voiced speech, significant changes are more readily detectable in the Fourier transform of the logarithm of the power spectrum which has been named the Cepstrum. Cepstrum information can be more easily used by the synthesizer to separate the envelope of the spectrum (slowly changing vocal tract configuration) and excitation function.

4. Harmonic Compression

As described in Section A-2, frequency division without compression simply slows transmission. A harmonic compressor, as shown in Figure 7,

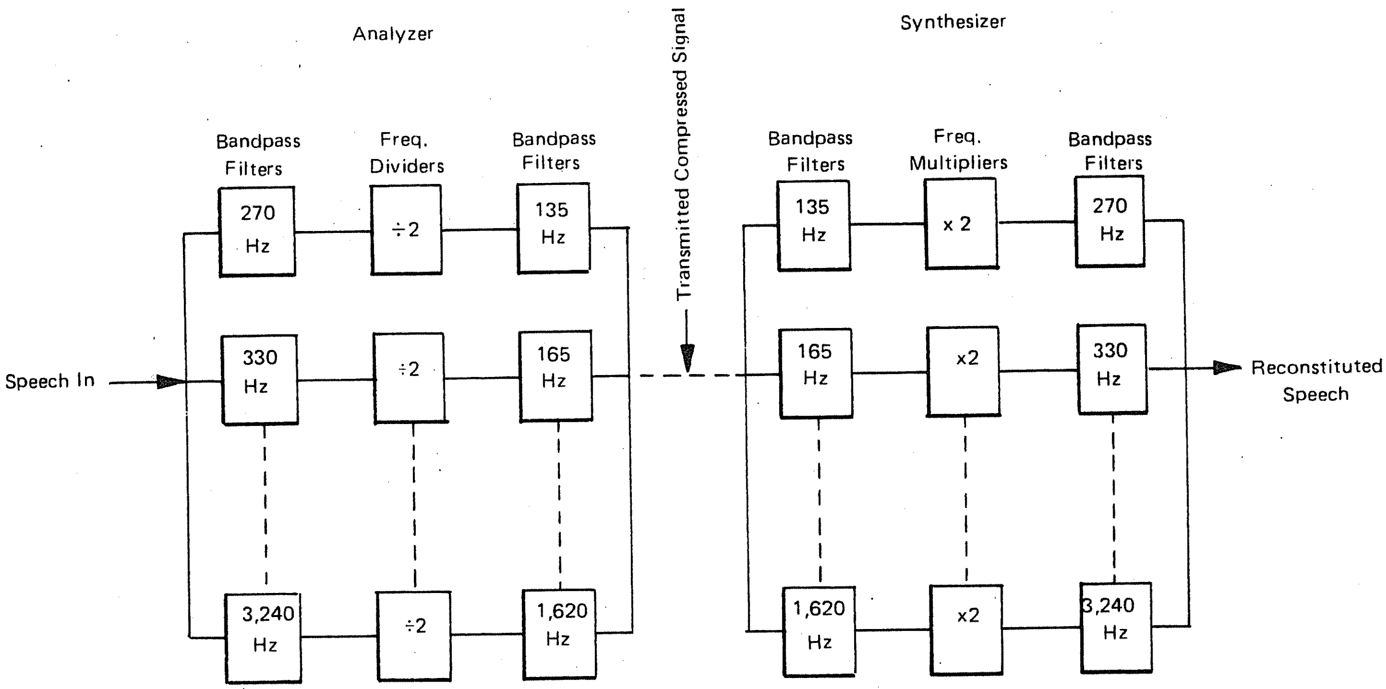
---

*The voice-excited vocoder is an excellent example of the simplification possible for mine communications compared with telephone applications. For an acceptable telephone voice-excited vocoder, the low band must include the fundamental frequencies for voices of men, women, and children. The wider this band, the more it eats into possible bandwidth compression. By limiting the band to the range of fundamentals of men's voices only and optimizing the design around accompanying expected formants, much more encouraging results may be obtained than telephone research would lead one to expect.



Source: Ref. 8

FIGURE 6 VOICE-EXCITED VOCODER



Source: Refs. 9 and 10

FIGURE 7 HARMONIC COMPRESSOR

halves the instantaneous frequency of each band, transmits the compressed signal, and resynthesizes the input by doubling in each of the (halved) bands.

A harmonic compressor can be applied to the baseband signal of a voice-excited vocoder (Section B-2) to decrease the bandwidth required by that type of vocoder.

5. Autocorrelation Vocoder

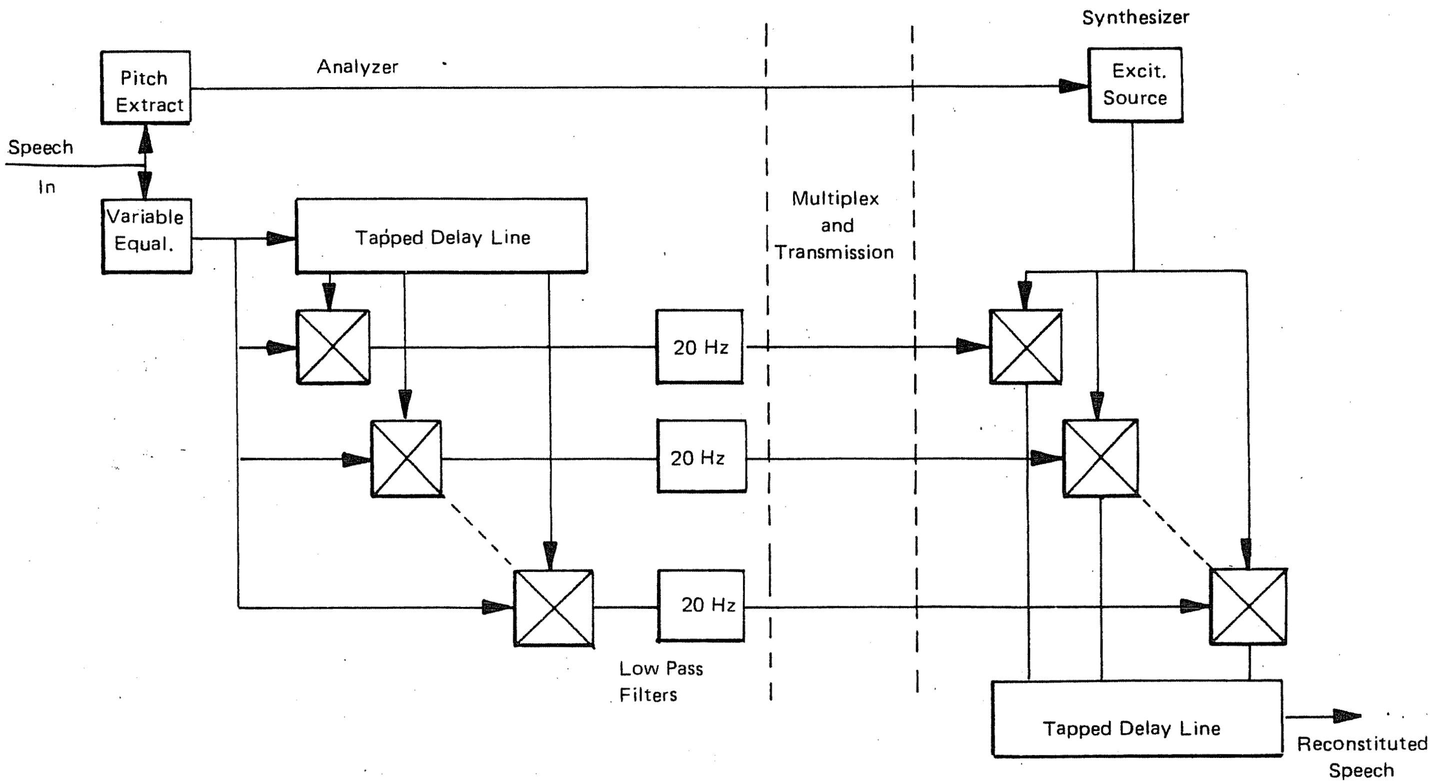
A vocoder that operates on the autocorrelation function of speech is shown in Figure 8. In the analyzer the short time autocorrelation function is derived for a number of discrete delays. The autocorrelation function is band-limited to the same band as the input signal and is therefore completely specified by the instantaneous values for discrete delays whose period is no more than the reciprocal of the Nyquist interval. A delay step of 0.1 msec corresponds to a 5000-Hz band-limited signal. Speech quality is influenced by the maximum delay (number of steps). For a 2.5-msec delay 26 delay channels are available. Each represents a time-varying sample of the short-time autocorrelation function. If each channel is limited to 20 Hz, a total gross transmission bandwidth without multiplexing or excitation of 520 Hz is required.

At the synthesizer the original signal is regained by a process which is essentially a time-varying transverse filter multiplying the excitation spectrum by the power spectrum of the original speech.

The autocorrelation vocoder is not susceptible to degradation resulting from arbitrary choice of bands of frequency that may not correspond with the harmonic structure of voiced speech sounds. However, the analogous problem crops up in the choice of maximum delay relative to the fundamental frequency being encoded.

6. Phoneme Vocoder

Continuous speech may be broken into perhaps 40 distinct elements, called "phonemes." Typical phonemes are all vowel and consonant sounds in a language. A system that would recognize these phonemes and send a simple code to the far end where the phoneme would be generated again



Source: Ref. 7, 10

FIGURE 8 AUTOCORRELATION VOCODER

probably offers the greatest band saving possible. Vocoders of this type have not been very successful, although the synthesis process alone is useful for speaking machines operated by a keyboard or a computer. Since pre-arranged phonemes are used, speaker recognizability is lost in a phoneme vocoder.

7. Adaptive Predictive Coding

A great deal of current research is based on adaptive predictive coding. This embodies a combination of differential quantizing and vocoder techniques, and depends on sophisticated computational ability in real time. In this approach both the transmitter and the receiver, using the same algorithms, predict what the current value of a signal should be based on its history. The difference between this value and the actual value is transmitted to the receiver where the difference is added to the predicted value to produce the next sample. The predictor at each end is adaptive -- that is, it changes its ground rules periodically to minimize the error.

Work in this field is taking place using simulation on large digital computers rather than hardware implementation of systems. Workers in the field estimate that high quality speech may be realizable with rates as low as 1000 bps.

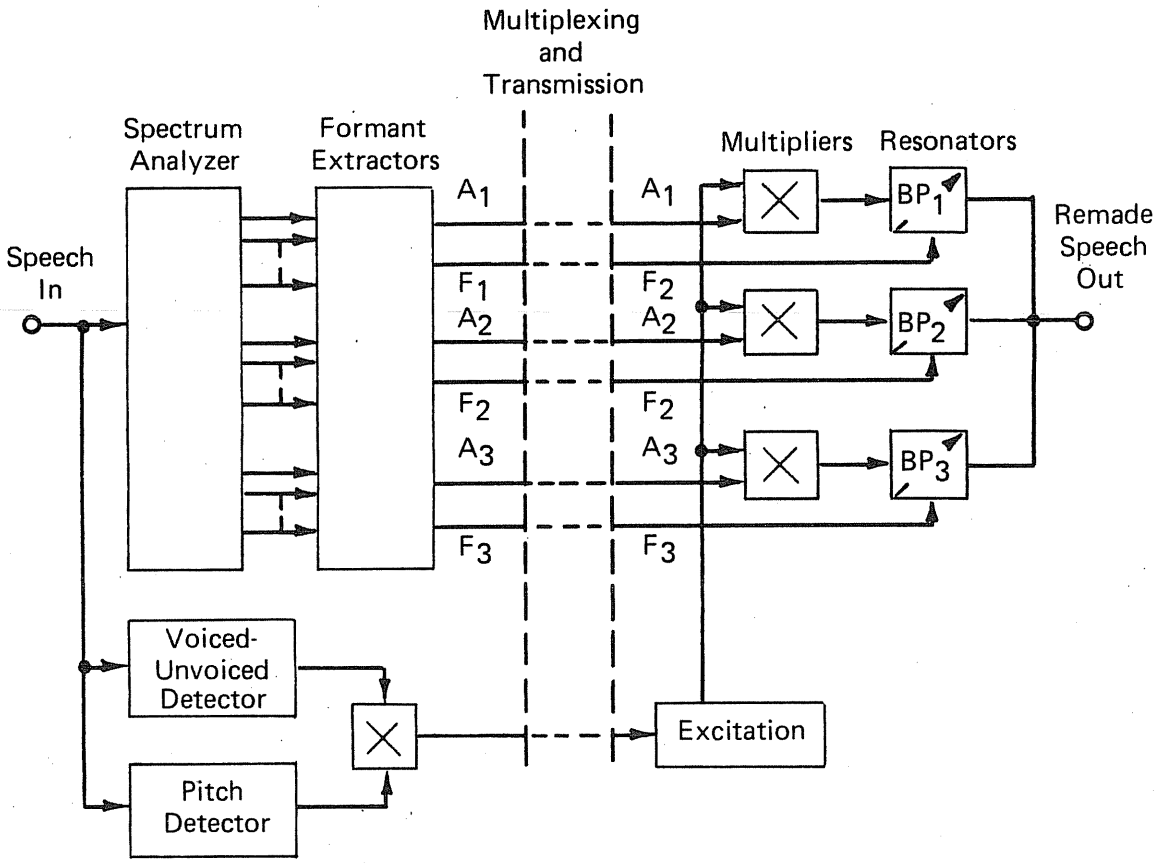
These systems tend to be vastly confused by signals other than single talkers. Noise and double talking can make them useless. This is a serious drawback for use in mines.

8. Pattern Matching Coding

Caldwell Smith[11] has developed a speech coding system which uses a set of patterns as the base for coding speech. A channel vocoder operates on the speech, and the time history of the output responses from each channel forms a pattern over a preselected period of about 20 msec. For each pattern thus obtained, a best match is found to one of 1024 stored patterns, for example. The basic information transmitted to the receiving location is a series of pattern numbers, thus developed. The speech is reconstructed by a generator which produces a series of sounds corresponding to the sequence of patterns. Data rates for such systems are from 500-1000 bps, depending primarily on the number of stored patterns used. Intelligibility ranged from 83 - 92% in tests of such systems.

9. Formant Vocoder

The spectral envelopes of speech sounds are characterized by several prominent maxima. These represent the resonances of the vocal tract called the formants. Below 3000 Hz three formants are typically found in adult speech. In a formant vocoder, the analyzer attempts to determine the frequency locations of the major formants. Signals corresponding to the formant locations and strengths are transmitted and utilized to control the resonances of a formant synthesizer consisting of three or more single-tuned resonant circuits. The block diagram of a formant vocoder is shown in Figure 9.



Source: Ref. 7

FIGURE 9 FORMANT VOCODER

IV. FUNDAMENTAL LIMITS TO BANDWIDTH COMPRESSION

A number of investigators have treated the problem of the information rate associated with speech communication. They tend to agree on a figure of about 50 bps as being the upper limit of the information rate communicable to a person listening to speech. This rate falls far short of the nominally accepted 50,000-bps rate necessary to code a digital representation of speech directly. This value is arrived at by assuming a sampling rate of 8000 bps and a 7-bit amplitude resolution (corresponding to 1 part in 128). It is this savings of 1000:1 that is the lure for the development of compressed speech for use in communication systems such as might be employed in the mine environment. No one has yet developed a real-time voice communication system that achieves this ultimate in bit rate. However, if such an achievement were made, it should be possible to transmit the 50-bps rate in a communication system whose bandwidth in Hz is approximately this value. Indeed, by using modern digital data transmission techniques, it would be possible to confine the bandwidth to even less than this ultimate 50 Hz. However, in general, the reduction of bandwidth beyond that represented by the basic bit rate would be bought at the expense of added transmitter power as can be seen from an examination of Appendix IV, Section E.

There is another point of comparison that can be used in arriving at one of the limits to the compression of voice in real time, and that is to accept the number of bits per second required in a phonemic-type voice reproducer which responds to phonemic commands and produces an understandable voice output from these phonemic commands. Such a device is the Votrax manufactured by the Vocal Interface Division of Federal Screw Works[25]. If pitch and inflection commands are deleted from those of the Votrax, then a six-bit characterization results for each phoneme. If one assumes a nominal speaking rate of 4 or 5 phoneme words per second, a usable output results at a 90-bps rate. These two limits, then, seem to be the boundaries against which voice bandwidth compression systems can work. It is instructive to note that a high-speed electric typewriter which types 1 character at a time can achieve a typing rate of about 3 words per second, each word comprising, say, 6 characters. Thus, 5 bits per character results in a 90-bps rate of transmission.

V. DISCUSSION OF BANDWIDTH COMPRESSION TECHNIQUES

Table 1 shows the classes of bandwidth compression schemes discussed in Section III and compares their respective parameters. The characteristics compared include: bandwidth or bit rate, signal-to-noise ratio required, band reduction ratio, estimate of sentence intelligibility, estimate of computational requirements for analysis and synthesis, analog signal processing requirements, acoustic background influences, the deterioration of performance with decreasing signal-to-noise ratio, pitch preservation, speaker identification, real time, delay time if it is not a real-time system, the state of development, cost, size, weight, power consumption, and an estimate of voice quality. In a number of instances the bases for comparison were relative, not absolute, with the data presented meant to provide a sense of the relative behaviors of the various systems. In addition to various types of bandwidth compression techniques presented in the matrix, we also presented three others, which are not directly bandwidth compression schemes, but represent other means of preserving or saving signal-to-noise ratio or decreasing the required transmitter power, such as method of clipping and companding.

The basis for comparing the compressed systems is a direct baseband voice system. The prominent features of each of the voice bandwidth compression techniques tabulated in this comparison are discussed below.

A. ADAPTIVE PREDICTIVE CODING

Adaptive predictive coding is an area of current ongoing research and development. The emphasis in this area is on producing a telephone-quality voice with a small bps rate. The impetus for this work is to get more voice channels per digital channel in communication systems. The general system requires computer operation to develop the codes to represent the speech from the sending station. However, the mine environment is a drawback to the use of such a technique. There is heavy reliance on very sophisticated and complex digital data processing at both ends of the system, and the computer power required is quite significant. Another problem is that the adaptive predictive coding system is quite

COMPARISON MATRIX

	Adaptive Predictive Coding	Spectrum Channel Vocoder	Formant Vocoder	AM Voice	Band Limiting	Temporally Interrupted Speech	Frequency Division (Stretched Voice)	Harmonic Compression	Clipping	Phoneme Vocoder	Pattern Coding	Companding
Bandwidth	Not fixed	300 - 500 Hz	300 - 500 Hz	500-3000Hz	1400 Hz	1400 Hz	Any	1400 Hz	500-3000 Hz	Promises to be small	Not fixed	500-3000Hz
Bit Rate	700 - 1500	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	300 - 500	N/A
S/N	15 - 20 dB	15 dB	15 dB	6 dB	10 dB	10 dB	6 dB	10 dB	6 dB	?	15 dB	6 dB
Band Reduction Ratio	N/A	10/1	10/1	1/1	2/1	2/1	Any	2/1	2/1	Promises to be high	10/1	1/1
Sentence Intelligibility	>90%	>90%	>90%	>90%	>90%	>90%	>90%	>90%	>90%	90% Achievable	~90%	>90%
Computational Requirements	Large	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Medium	Medium	N/A
Analog Signal Processing	Negligible	Modest	Modest	None	Modest	Modest	Medium	Medium	Little	Medium	Medium	Little
Acoustic Background Influence	Extreme	Quite Tolerant	Quite Tolerant	Unchanged	Unchanged	Unchanged	Unchanged	Unchanged	Troublesome during pauses	Tolerant	Tolerant	Troublesome during pauses
Deterioration with Decreasing S/N	Sudden	Yes	Yes	Direct with S/N	Direct with S/N	Direct with S/N	Direct with S/N	Direct with S/N	Direct with S/N	Sudden	Sudden	Direct with S/N
Pitch Preservation	Yes	Yes or No	Yes or No	Yes	Yes	Yes	Yes	Yes	Yes	Yes or No	Yes or No	Yes
Speaker Identification	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not likely	No	Yes
Real Time	Essentially	Essentially	Essentially	Yes	Yes	Yes	No	Yes	Yes	Essentially	Essentially	Yes
Delay Time	25 ms						Large Depends on Ratio			25 - 50 ms	25 - 50 ms	
State of Development	On-going	Complete	Complete	Complete	Some Experiments Needed	Complete	Hardware Design Required	Complete	Experiments Could Be Useful	Not Developed	Complete	Complete
Cost	High	Medium	Medium	Low	Medium	Medium	Medium	Medium	Low	Medium	Medium to High	Low
Size	Large	Medium	Medium	Small	Medium	Medium	Medium	Medium	Small	Medium	Medium	Small
Weight	Heavy	Medium	Medium	Low	Medium	Medium	Medium	Medium	Low	Medium	Medium	Low
Power Consumption	High	Low	Low	Low	Low	Low	Low	Low	Low	Low	Medium	Low
Subjective Voice Quality	Excellent	Good	Good	Good	Good	Good	Good	Good	Fair	Fair	Poor	Good

sensitive to background acoustic noise that mixes with the original voice message. The system attempts to translate the total input signal into a representation as if all signals were of the voice message. Hence, the performance in a noisy environment could become quite poor. The ongoing research indicates that, for telephone-type use, excellent system performance will be achieved at a level of about 700 or 800 bps within the next few years. The two features of extreme computer power required and the susceptibility to background noise make this particular system quite unattractive for mine communication systems.

B. SPECTRUM CHANNEL VOCODER

Spectrum channel vocoding, which was highly developed 20 years ago, is an old art compared to adaptive predictive coding. Relatively unsophisticated circuitry would suffice to provide systems of this type. The subjective voice quality from a system of this type is classed as good. There will be machine-like characteristics to the reproduced voice, but this should offer little problem in a mine environment. An interesting advantage of this type of vocoder is its relative immunity to background-interfering acoustic noise at the sending station. This is so because the device uses a bank of tuned filters and tends to emphasize the resonant formants of the speaker's voice to the exclusion of background noises.

C. FORMANT VOCODERS

The developments in formant vocoders follow the lines of the spectrum channel vocoders; the characteristics are generally the same. Some feel that the perceived subjective voice quality of the reconstructed voice is somewhat more machine-like than that of the spectrum channel vocoder. A common problem to both the spectrum channel vocoder and the formant vocoder is the problem of extracting accurate pitch information for reconstruction of the voice. For mine-type applications it may be possible to drop the requirement for pitch entirely. This will let intelligible messages be transmitted, but the received voice would have a monotone characteristic to it, or the received voice could be made to sound like a whisper and it would then be entirely of an unvoiced character (see Appendix III).

D. BASEBAND VOICE

Baseband is included as a basis of comparison for the other systems for mine communication uses. One of the characteristics of the direct baseband voice -- or a band allocated to voice -- is that its deterioration with deteriorating signal-to-noise ratio is gradual and well understood. By reference to Figure 11*, it can be seen that there is no sudden deterioration of the ability to discern the message. It deteriorates gradually with decreasing signal-to-noise ratio, whereas most of the processed speech systems fall apart suddenly and completely destroy the ability to communicate.

E. BAND LIMITING

Band-limiting systems provide an opportunity to make relative simple, very modest ratio reductions in bandwidth required for voice communication. Such a system simply blocks out certain bands and transmits the remaining part. Performance from such band-limiting systems is presented in Appendix I. The conclusion reached is that no great benefit can be obtained with such systems.

F. TEMPORALLY INTERRUPTED SPEECH

Many experiments on temporally interrupted speech have been conducted. They were directed toward speeding up or slowing down recorded speech for either rapid comprehension of voiced messages, or during a learning process, to present the voiced message at a slow rate for greater comprehension of difficult subject matter. It is completely evident from this work that one can drop one-half of the voice message without disturbing the ability to understand the spoken messages. In this way one can, by various techniques, cut the bandwidth required for transmission to one-half of the original band. We do not expect that there will be any savings in transmission power for such a system, however, because it represents a removal of redundancy in speech, rather than a repackaging of the form of the speech. Again, simple systems can be made and they do work; the applicability to mine communication needs is not evident, however.

---

*See page 41.

G. FREQUENCY DIVISION (STRETCHED-OUT VOICE)

It is obvious that an appreciable reduction of required bandwidth for transmission of voiced signals can be obtained simply by recording the original voiced message and then playing it back at a much reduced rate or stretched out in time. The bandwidth associated with it will shrink directly in proportion to the stretch-out factor one uses. In this event, one can claim a direct transmitter power savings proportional to the stretched-out ratio. The major problem associated with such stretched-out voice systems is that the users of the system must adapt to the enforced added time delay between the sender speaking and the receiver hearing the spoken message.

While an emergency voice transmission system might be able to accommodate quite large delay times imposed by this restriction, it is unlikely that day-to-day communications would tolerate a stretch-out factor of more than 10 to 1 at the most. A 10-to-1 ratio would correspond to a 10-to-1 ratio in transmitter power savings. Applicable experience with systems of this type comes from the U.S. Navy's underwater telephone systems. In seawater the velocity of propagation of sound is approximately 5000 ft./sec., and it is well known that transmissions over 5 nautical miles are obtainable with such underwater telephone systems. Hence, one-way time delays of 6 seconds are experienced. The telephone operators know that time delay is inherent in their communication system and thus are able to adapt to it without severe difficult. For an untrained person, however, such a time delay could be annoying if he were not experienced with it nor expecting it. Certainly for the day-to-day communication needs of mines, delays of this type could be considered an undesirable characteristic of the system. For a true emergency voice communication system, we do believe that delays of this length, or even longer, would not pose a severe problem. The application of these systems can provide a reduction of power capacity required for voice at the expense of the imposed time delay. Such a system would best operate with very short voice messages, because it is the length of message <u>and</u> the reduction ratio that determine the delay time between the transmission and reception of the signal. The possibilities for using this system are discussed in Section VII in more detail.

H. HARMONIC COMPRESSION

Harmonic compression schemes of many kinds have been developed. In principle, these systems take the original voice spectrum and by real-time signal processing reduce the bandwidth by a factor of, say, 2 to 1. Thus, an original voice band between, say, 200 and 3000 Hz would be reduced to 100 to 1500 Hz, cutting the bandwidth in half. Upon reception, the signal frequencies in the received waveform would be analyzed, multiplied by 2, and presented to the listener.

Band reduction ratios of 2 to 1 are readily obtainable in this way. Again, a somewhat higher signal-to-noise ratio than for a baseband signal would be required, and there appears to be almost no advantage in terms of power requirements for a voice link obtained by this means.

I. CLIPPING

Clipping of transmitted waveforms has long been used, particularly in instances where the limitation on transmitter power is the peak power rating of the transmitter. The ability to communicate depends on the received signal-to-noise ratio which is proportional to the average power of the transmitter. The average power can be much higher when clipping is practiced, thus causing the transmitter to operate at maximum power all of the time.

However, there are some problems with clipping systems. In particular, there is a harshness to the quality of the voice as received. If a high degree of clipping is used, the quiet passages where the voice is not present tend to fill up and respond to noise. This creates an irritating property to the output speech presented. There are modifications of this technique, such as Lincompex discussed in Appendix V of this report. In Lincompex, in addition to hard limiting or clipping, a narrow-band modulating signal is transmitted which is used to control the level at the receiver. This operation makes the voice natural and also avoids the problem of the quiet periods being filled up with noise.

Clipping can be of considerable use in a voice communication link if the transmitter power is peak-limited. The degree of improvement can be as high as 12 dB (more than 10 to 1 in power). Thus, it would appear to be a desirable feature to include by itself in a mine voice communication system, either with or without the Lincompex type of control signal.

J. PHONEME CODER

There has been much talk and little resulting hardware related to a phonemic-type coding system. One can make up a perfectly adequate English language representation with as few as 32 phonemes. There is hardware and equipment that accepts phonemic commands. From these commands it can generate understandable English or an artificial speech. The Votrax is a device of this class.

There does not presently exist the capability of accepting spoken English messages and from these messages obtaining the extraction of the phonemic structure to a degree that would be adequate for a voice communication system. It also appears that a device capable of extracting the phonemic structure of general spoken English would be a computer-based, complicated, non-mineworthy piece of equipment.

It should be pointed out that present voice communication research is directed toward the telephone-quality voice systems, such as those represented by the adaptive predicted coding already discussed, and that little work is being done on the phonemic coder. It may be appropriate to keep track of the progress being made in the area of voice-controlled computers (a part of continuing voice research), because many of the voice input systems, as part of their basic structure, need decoders similar to those required in a true phonemic coder.

K. PATTERN CODING

The pattern vocoding systems developed by Caldwell Smith are an interesting approach to producing voice communication systems. The basic Caldwell Smith approach is to look at the patterns of spectral energy distribution with respect to time produced by the various sections of spoken words and, by examining these patterns over periods of

25 to 50 milliseconds, try to find a family of patterns from which the spoken messages can be coded, then transmit by codes the sequence of patterns found in the spoken stream being analyzed.

Experiments by Caldwell Smith show that bit rates of 300 to 500 bits/second can provide intelligible message transmission. For mine communication needs the amount of storage of information and computer processing required to generate the proper interpretation of the patterns is quite large. Caldwell Smith's system used a magnetic disk storage unit, entirely inappropriate for use in the mine communication systems presently contemplated.

L. COMPANDING

Although not a voice bandwidth compression device, companding is frequently used in communication systems to permit a relatively narrow dynamic range system to accommodate a substantially wider dynamic range of message input. It automatically compresses the loud passages of spoken messages and automatically increases the level of quiet passages for transmitting purposes. Upon reception, the opposite is done to restore the correct level to the passages of the spoken message. In some ways it suffers from the same drawback as the limiting or clipping of speech in that during quiet passages background noises will be amplified and transmitted.

VI. COMPARISON OF SYSTEMS

A. METHODS OF COMPARISON

There are a number of ways in which various voice bandwidth compression systems can be compared. These include degree of circuit complexity, weight, size, adaptability for mine environment, power consumption of processing circuit, quality of voice, processing power consumption, inherent delays in the system, and so forth. Bandwidth reduction ratio, or the factor by which the voice band can be compressed into a transmission band, is a parameter which is also central to all voice-compression techniques. However, it should be borne in mind that the application to mine wireless communication systems is one in which the prime measure of the effectiveness of candidate voice bandwidth compression techniques is not the bandwidth reduction ratio per se, but how much power reduction can be made while maintaining an adequate signal-to-noise ratio or comprehension level for the reconstructed voice message at the receiver. This consideration is therefore quite different from the considerations related to much of the past and present work on voice bandwidth compression.

In the area of past and current progress in voice bandwidth compression, emphasis has been placed, not on saving transmitter power, but rather strictly on the conservation of bandwidth. Much attention has been directed toward adding capacity in terms of voice channels to already existing bandwidth-limited communication paths, which focusses the conservation efforts at bandwidth and not at power. Thus, the objectives sought in much of the voice bandwidth compression work are not necessarily compatible with the objectives of the Bureau of Mines for voice bandwidth compression techniques. For this reason we believe that the method of comparison between various candidate systems should be based strictly on the amount of transmitter power required compared to a standard level required for full-voice bandwidth transmission. Such a posture will clarify the real power demands placed on the transmitter of such a system. Then the importance of bandwidth compression ratio vanishes from the comparison criteria and the power requirements are given prime importance, together with the complexity required to achieve the reported performance.

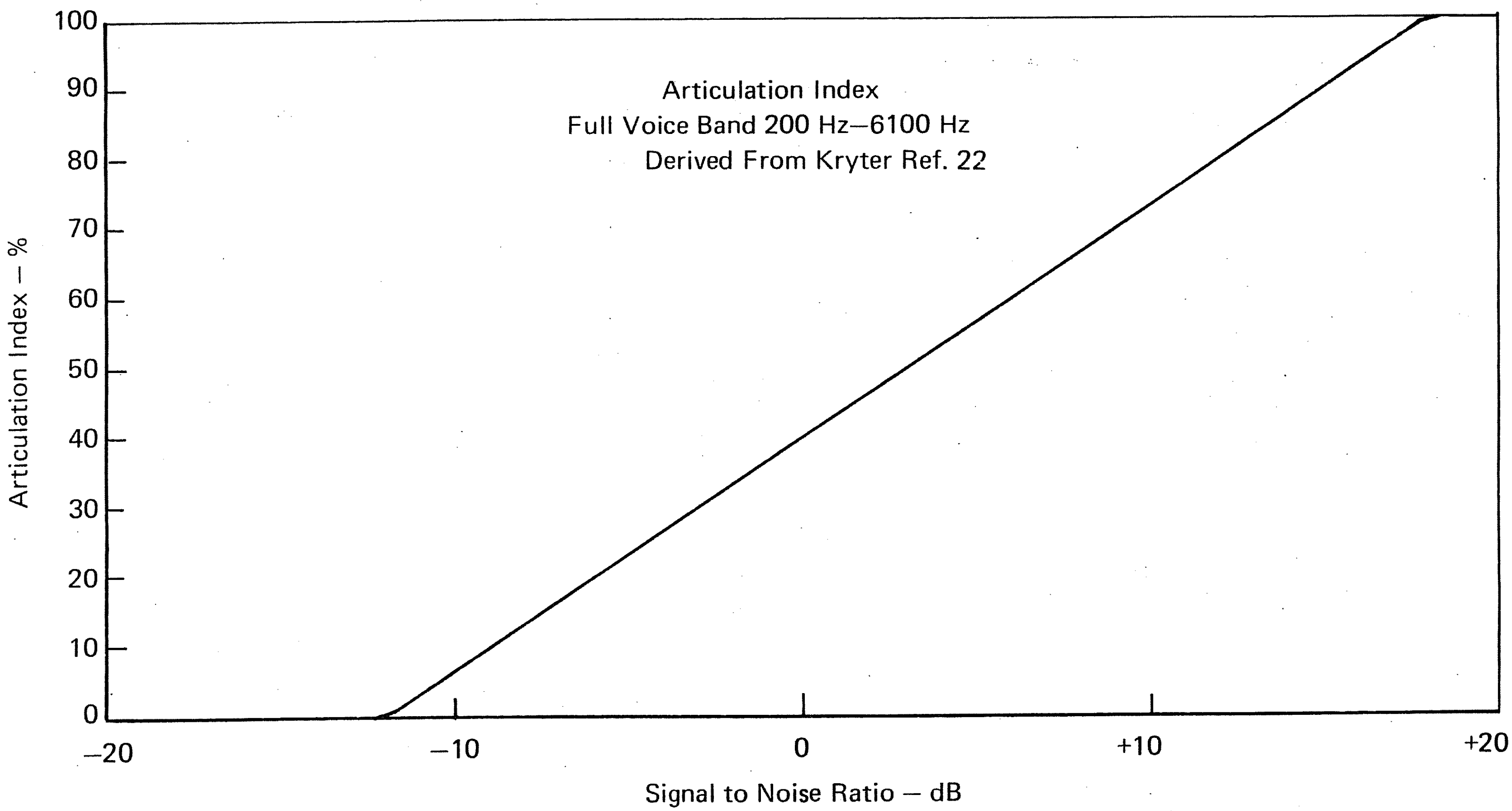
There is a well established body of data from which the various measures of channel quality can be assessed for full-bandwidth baseband voice communication. In particular, we used the articulation index as a measure of the capability of a voice channel. Figure 10 shows how this quality varies as a function of signal-to-noise ratio in the channel. There are several conditions associated with the performance indicated. The first assumption is that the background noise has a flat spectrum over the voice band, and the second assumption is that the signal-to-noise ratio is constant over the band. This assumption means that modest pre-emphasis of the voice signal has been used and a method of restoring the spectral shape of the original voice is used at the output end of the system. Under these conditions the plot illustrates the articulation index as a function of signal-to-noise ratio. The signal power for these plots is the long-term average signal power at the output end of the system. Similarly, noise power is the long-term average noise power.

The articulation index (A.I.) is a measure of a listener's correct interpretation of voice sounds. There are several features of the Figure 10 plot that are worthy of note:

1. At a signal-to-noise ratio of +18 dB, articulation index is approaching the asymptotic value, and any further increase in the signal-to-noise ratio would be judged solely as an improvement in quality of the channel, not in intelligibility of the messages received through the channel.

2. At a -12 dB signal-to-noise ratio, essentially all articulation index is lost and communication becomes almost impossible.

This form of articulation index plot can be converted to one in which the abscissa is signal power rather than signal-to-noise ratio merely by relating the total average signal power to the average power found in a 1-Hz bandwidth due to the received noise. If this is done, various systems of bandwidth compression can be compared strictly on the basis of power necessary on reception to achieve various levels of articulation index. What this new plot will show are the relative power requirements



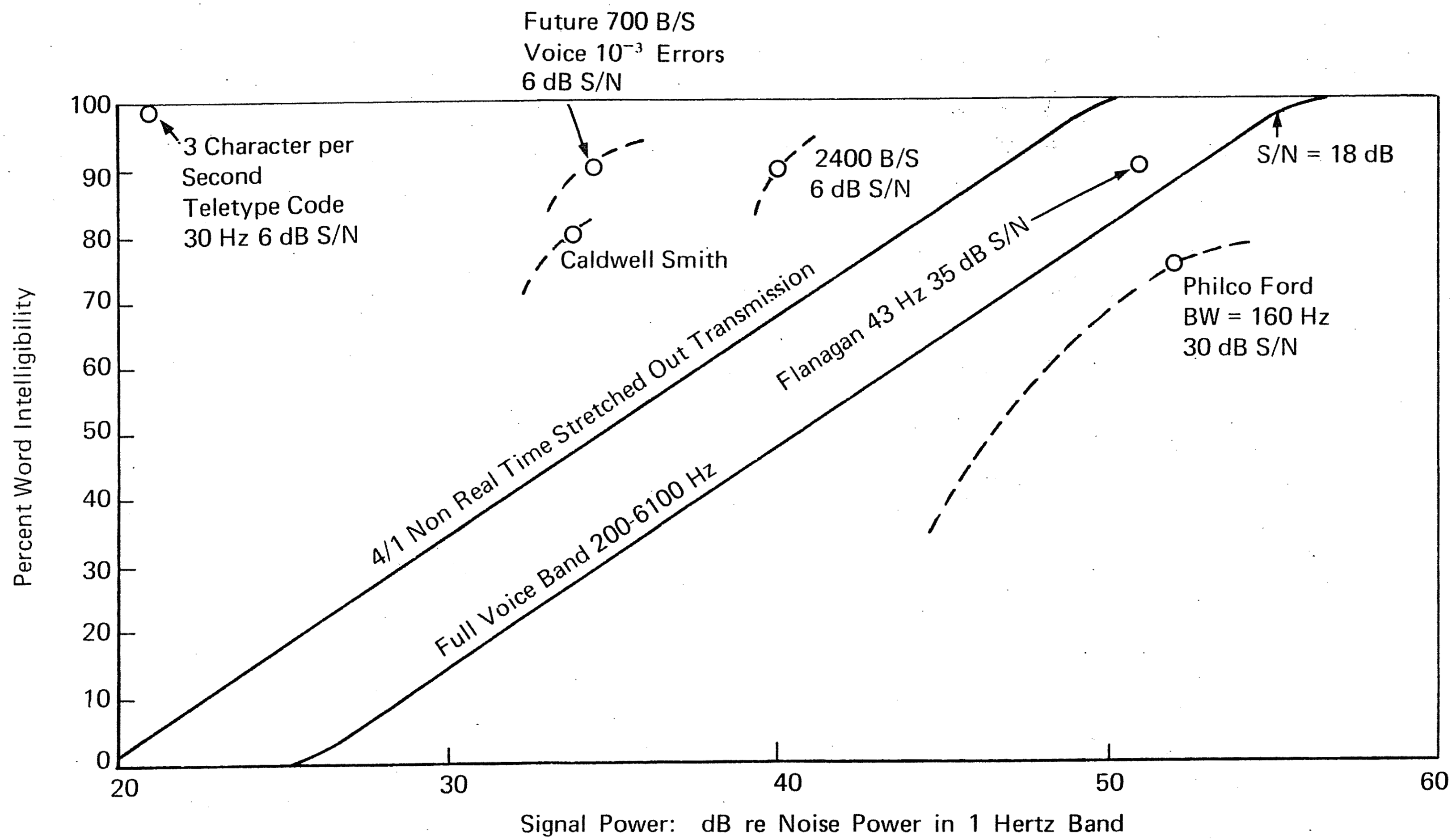
Source: Arthur D. Little, Inc.

FIGURE 10 ARTICULATION INDEX AS A FUNCTION OF SIGNAL TO NOISE RATIO for the various bandwidth compression techniques, together with a measure of what intelligibility or articulation index can be expected for the various systems. Although many of the bandwidth compression techniques do not have fully identified articulation indexes as the signal-to noise ratio deteriorates, they do identify minimum required values of the signal-to-noise ratio in specific bands, and sometimes give a measure of articulation index, on which we can base a comparison with other systems.

B. DISCUSSION OF COMPARISON PLOT

The first item of concern in the modified articulation index comparison plot of Figure 11 is the way that articulation index varies as a function of signal power. For the full voice band (200 to 6100 Hz), it can be seen that once the total signal power exceeds approximately 56 dB relative to the noise power in a 1-Hz band, there is no further improvement in the ability to interpret voice sounds spoken through a communications system. A drop of 15 dB in the signal power leads one to a point where the articulation index is around 50%. As one decreases further in signal power, a corresponding loss of articulation index results. Generally, a communication network that works with articulation index below 30% becomes difficult to use, although not impossible. Consequently, only systems that achieve something of the order of 30% articulation index should receive serious consideration.

The plot of Figure 11 also shows points or regions where some of the voice bandwidth compression techniques fall. The ideal region for techniques for mine wireless communications is the upper left corner; they are also not complex. The most inappropriate system shown on this plot is the two-formant frequency vocoder developed by Philco Ford[23]. For this vocoder, it is better to use the full-voice bandwidth instead of the 160-Hz wide channel of the vocoder because the vocoder requires a 30-dB signal-to-noise ratio, thus leading to a higher power requirement at a 75% articulation index than for the full-voice band. In fact, at the same power level an articulation index of about 90% results for the full-voice band as compared to a 75% level for the Philco Ford vocoder. Indeed, because of the nature of that vocoder, even as signal-to-noise



Source: Arthur D. Little, Inc.

FIGURE 11 COMPARISON OF COMPRESSION TECHNIQUES

ratio improves, it is unlikely that the articulation index would improve at all. Therefore, this system is inappropriate for mine wireless communication applications.

The next system of interest is one identified with Flanagan[12] in which a 43-Hz wide bandwidth is required for vocoded speech, with a signal-to-noise ratio of 35 dB yielding only a 1-dB improvement over the full-voice bandwidth. Again, the system is inappropriate for application to mine communication needs.

Another system of interest is the 2400-bps voice-compression system. Such a system, typical of the digitally processed adaptive predictive systems with a signal-to-noise ratio of 6 dB, would result in an improvement of some 8 dB over the full-voice band. This improvement would be at the cost of complicated digital processing equipment, and therefore we do not expect that it would be economical or mineworthy at the present time.

The Caldwell Smith pattern-recognizing vocoder seems to be occupying a favorable position. The use of this technique is dependent on equipment that is quite complex. Caldwell Smith used a learning process as well as a disk storage system to perform the analysis and synthesis required for his processing.

A 4 to 1 non-real time, stretched-out speech system is also shown on the plot of Figure 11. This type of processing preserves the characteristic of the full-voice bandwidth system as a function of signal-to-noise ratio. Therefore, this relatively simple system, stretched-out voice, will have a 50% word intelligibility at the same power level that Caldwell Smith's pattern-recognition system would show an 80% word intelligibility. Thus, in terms of the communication systems, it is entirely reasonable to expect acceptable performance from the stretched-out voice system.

As a final point of comparison, power requirements for a three character-per-second teletype code system operating over a 30-Hz bandwidth with a 6-dB signal-to-noise ratio are shown. This system appears to be extremely favorable from the point of view of power consumption, and does provide a data rate that would be useful for some purposes. This should be borne in mind when considering alternatives to voice bandwidth compression systems (discussed later).

The general overall conclusion obvious from the plot of Figure 12 is that available state-of-the-art voice bandwidth compression techniques do not offer sufficient improvements in power requirements to make their use universally applicable to mine wireless communication systems.

(This page intentionally left blank.)

VII. TIME-DELAYED SPEECH

As stated in Chapter VI, present real-time voice bandwidth compression systems have an extremely limited capability for decreasing the power requirements of voice communication systems. The non-real-time systems, on the other hand, do offer some improvements in this respect. However, this improvement is bought at the expense of time delay in the voice channel. It is well recognized that modest time delays in the voice channels offer no serious problems. For example, earth-to-moon voice communications have been demonstrably easy to conduct. In this instance, one-way delay times of about 1-1/2 seconds and round-trip delayed times of about 3 seconds are experienced.

Underwater acoustic voice channels suffer a delay of 1.2 seconds per nautical mile and are known to be useful for ranges beyond 5 nautical miles. At 5 nautical miles the one-way transmission delay time is 6 seconds and the round-trip time is 12 seconds. Operators who use these underwater systems have no trouble using the system, but they do know that delay is to be expected and they so adapt.

It therefore seems reasonable to expect that a mine wireless communication system having delays of 12 seconds could prove to be a usable system. Indeed, an underground-to-surface system between fixed or semi-portable stations could well employ such a system. It noted that for such a system it may be appropriate to use different stretch-out factors for the two directions of transmission. A large stretch-out factor would be used for the in-mine transmitter and a small stretch-out factor would be used for the surface transmitter. This difference in stretch-out factor is possible because the power of surface transmitter is not as limited as that for the in-mine transmitter, both from the safety aspect and the availability of raw power.

There are a number of ways that such stretched-out voice messages could be handled. Figure 12 illustrates some of the possibilities. At the top of this figure is shown an initial voice message of 2-seconds duration. The second line illustrates the stretched-out version of this

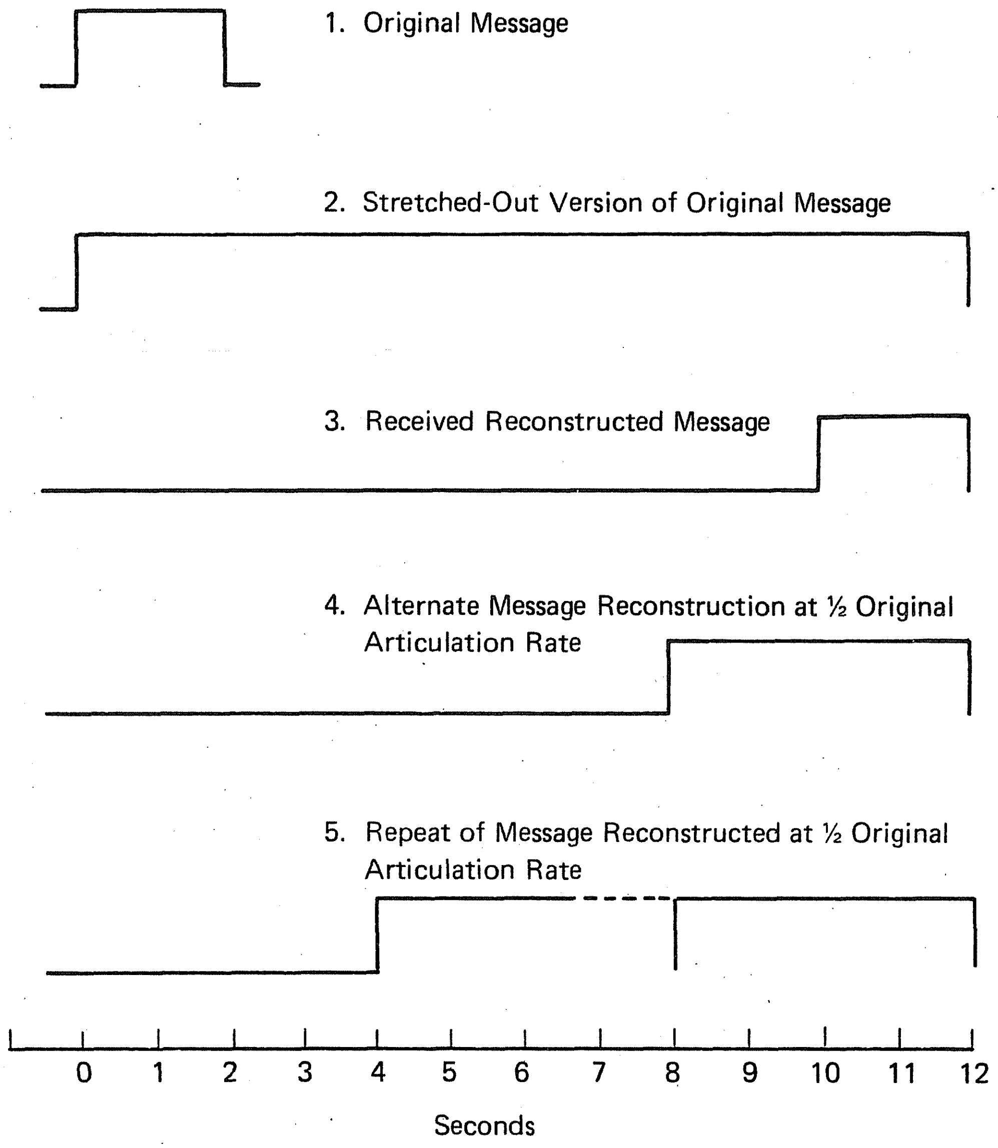


FIGURE 12 FORMATS FOR RECONSTRUCTION OF STRETCHED-OUT MESSAGES AT 6 TO 1 STRETCH-OUT RATIO

message, the time stretch-out factor being 6 to 1. The generation of this stretched-out version can start immediately; it does not require that the full message be completed before transmission of the stretched-out version starts.

There are a number of alternative ways that the receiver end can be handled. Perhaps the most obvious is compressing the received message by 6 to 1, such that the last part of this reconstruction corresponds to the last part of the stretched-out version. This mode is shown on line 3. A second interesting alternative is to compress the received version by only 3 to 1, but keeping the pitch to a natural one (this is a familiar technique for the slowed-down speech systems promoted for learning foreign languages from tape recordings). This compression would be done by repeating on a subphonemic level short passages of the received and compressed waveform. The output presented at the receiver would occupy 4 seconds and end as the stretched-out version ends. This mode is shown on line 4. A still further possibility is to use 3 to 1 compression and repeat it twice, as illustrated on line 5. It is interesting to note that this mode results in received messages starting to reach the operator only 4 seconds after the start of transmission.

Stretched-out transmission systems seem to offer one of the few viable means of reducing transmitter power requirements for wireless through-the-earth voice communication links. As shown above, there also exists considerable flexibility in the way such systems can be organized.

(This page intentionally left blank.)

VIII. REFERENCES

1. Arthur D. Little, Inc., USBM Final Report, Contract No. H0122026, "Survey of Electromagnetic and Seismic Noise Related to Mine Rescue Communications," January 1974.

2. Collins Radio Group, Rockwell International, Preliminary Mine Wireless Capability Investigation for Contract No. H0346067, August 1974.

3. H. Schlinder, "Delta Modulation," IEEE Spectrum Vol., pp 69-78, 1970 (includes bibliography with 36 entries).

4. A. Albert, "Electrical Communication," 3rd Edition, New York, Wiley, 1950. (Chapter 2 has 47 references on fundamentals of speech and acoustics.)

5. K. D. Kryter, "Speech Bandwidth Compression through Spectrum Selection," J. Acoust. Soc. Am., Vol. 32, pp 547-556, May 1960.

6. J. Williams, "Some Bandwidth Compression Systems for Speech Transmission," Radio & Electronic Engr., Vol. 27, pp 459-472, June 1964.

7. M. Schroeder, "Vocoder: Analysis and Synthesis of Speech," Proc. IEEE, Vol. 54, pp 720-734, May 1966.

8. J. Bayless, S. Campanella and A. Goldberg, "Voice Signals: Bit-by-bit," IEEE Spectrum, Vol. 10, pp 28-34, October 1973.

9. J. Savidge, "Ease Radio-Spectrum Congestion by Bandwidth Compression and Expansion," EDN, pp 74-77, May 20, 1973.

10. M. Schroeder, B. Logan and A. Prestigiacoma, "New Methods for Speech Analysis-Synthesis and Bandwidth Compression," 4th Int. Congress on Acoustics, Copenhagen, 21-28 August 1962.

11. Caldwell Smith, "Perception of Vocoder Speech Processed by Pattern Matching," J. Acoust. Soc. Am., Vol. 40 No. 6, pp 1562-1571, December 1969.

12. J. L. Flanagan, "Bandwidth and Channel Capacity Necessary to Transmit the Formant Information of Speech," J. Acoust. Soc. Am., <u>28</u>:4, July 1956.

13. K. D. Kryter, "Validation of the Articulation Index," J. Acoust. Soc. Am., <u>34</u>:11, November 1962.

14. K. W. Cattermole, "Principles of Pulse Code Modulation," London, Iliffe Books, Ltd., 1969.

15. B. Smith, "Instantaneous Companding of Quantized Signals," Bell System Technical Journal, Vol. 36, pp 653-709, May 1957.

16. G. Raisbeck, "Information Theory," MIT Press (1964), p. 37.

17. W. R. Bennet and J. R. Davey, "Data Transmission," New York, McMillian, 1965.

18. J. Martin, "Telecommunications and the Computer," Englewood Cliffs, New Jersey, Prentice-Hall, 1969.

19. J. Salz, "Communications Efficiency of Certain Digital Modulation Systems," IEEE Trans. On Comm-Technology, Vol. COM-16, pp 149-156, February 1968.

20. A. Tomozawa and H. Kaneko, "Companded Delta Modulation for Speech Transmission," IEEE Trans. Comm-Technology, Vol. COM-16, pp 149-156, February 1968.

21. J. Greefkes and K. Riemens, "Code Modulation with Digitally Controlled Companding for Speech Transmission," Philips Tech. Review, Vol. 31, pp 115-133, 1970.

22. K. D. Kryter, "Methods for the Calculation and Use of the Articulation Index," J. Acoust. Soc. Am., 34:11, November 1972.

23. Philco Ford, "Research and Development of an Optimized Laboratory Speech-Compression System for Spacecraft Application," Final Report, Contract NAS9-4523, Prepared for Manned Spaceflight Center, NASA, Houston, Texas, May 1966.

24. F. L. van den Berg, "The Lincompex Radio-Telephone Terminal Equipment of the RY741 Series," Philips Telecommunication Review, 30:2, April 1972.

25. Vocal Interface Equipment Group, Federal Screw Works, 3401 Martin Ave., Detroit, Mich. 48210.

26. J. Greefkes and F. DeJaeger, "Continuous Delta Modulation," Philips Research Report No. 22, pp 233-246, 1968.

APPENDIX I

NARROW BAND VOICE USING FREQUENCY SLICING

It is a well known fact that the full-voice band from 200 to 6100 Hz is not necessary to provide completely acceptable voice channel communications. Indeed, a telephone channel whose transmission band is confined to 300 to 3000 Hz is generally considered a good voice channel, and, at adequate signal-to-noise ratio, would far exceed the requirements for voice communications within coal mines. One of the means of achieving narrow-band voice communications is to restrict the transmission band even further than is done for telephone service.

A convenient way to access the performance to be expected is by using the articulation index developed for voice transmission systems. The articulation index (AI) ranges from 0 to 100%. If a system has an AI of 100%, it indicates that all isolated words spoken into the system would be correctly recognized by a listener. (A moderate degree of skill is expected for both speaker and listener.) Correspondingly, an AI of 50% would mean correct identification of spoken words 50% of the time.

Kryter[13,22] and others have developed means for predicting the AI for voice channels. The method is based on bands of equal contribution to articulation index and the signal-to-noise ratio in each band. Twenty bands are used, each contributing a maximum of 5% to the AI. If the S/N ratio (long-term average signal power to average noise power) is 18 dB or greater, the maximum contribution of 5% per band is achieved. If this S/N ratio is -12 dB or less, no contribution is made to the AI. Between these two values, AI contribution depends linearly on the S/N ratio in dB. Thus, if the S/N in a band is +3 dB, that band contributes 2.5% to the AI. By using this means, it is possible to estimate the performance of any proposed frequency-restrictive voice channel.

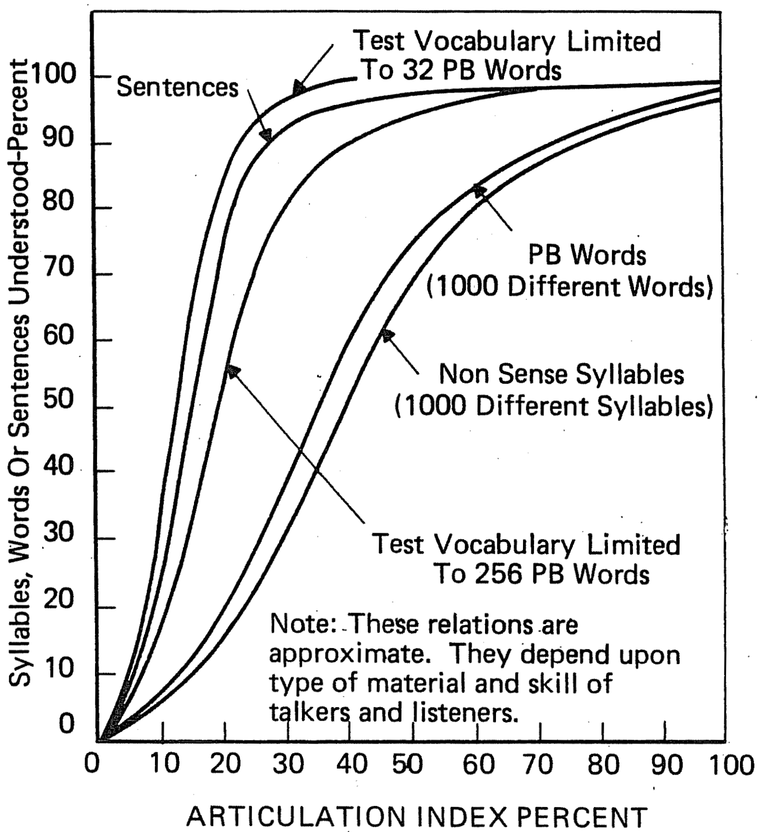
The bands of equal contribution to articulation index are tabulated in Table I.

TABLE I-1

TWENTY FREQUENCY BANDS OF EQUAL CONTRIBUTION
TO SPEECH INTELLIGIBILITY

Band No.	Limits	Mid-Frequency	Band No.	Limits	Mid-Frequency
1	200 to 330 cps	270 cps	11	1660 to 1830 cps	1740 cps
2	330 to 430	380	12	1830 to 2020	1920
3	430 to 560	490	13	2020 to 2240	2130
4	560 to 700	630	14	2240 to 2500	2370
5	700 to 840	770	15	2500 to 2820	2660
6	840 to 1000	920	16	2820 to 3200	3000
7	1000 to 1150	1070	17	3200 to 3650	3400
8	1150 to 1310	1230	18	3650 to 4250	3950
9	1310 to 1480	1400	19	4250 to 5050	4650
10	1480 to 1660	1570	20	5050 to 6100	5600

From this table we can easily determine that a telephone channel from 330 to 3200 Hz, for example, shows an articulation index of 75% at best -- that is, if the S/N ratio in all bands is greater than 18 dB. If the S/N in all bands decreases to 8 dB (1/3 of the way to -12 dB), then the AI falls to 50%. For mine communication needs, this value of AI is probably more than adequate. The reason this is so can be seen from Figure I-1, which shows that for an AI of 50%, a sentence intelligibility near 95% can be expected.



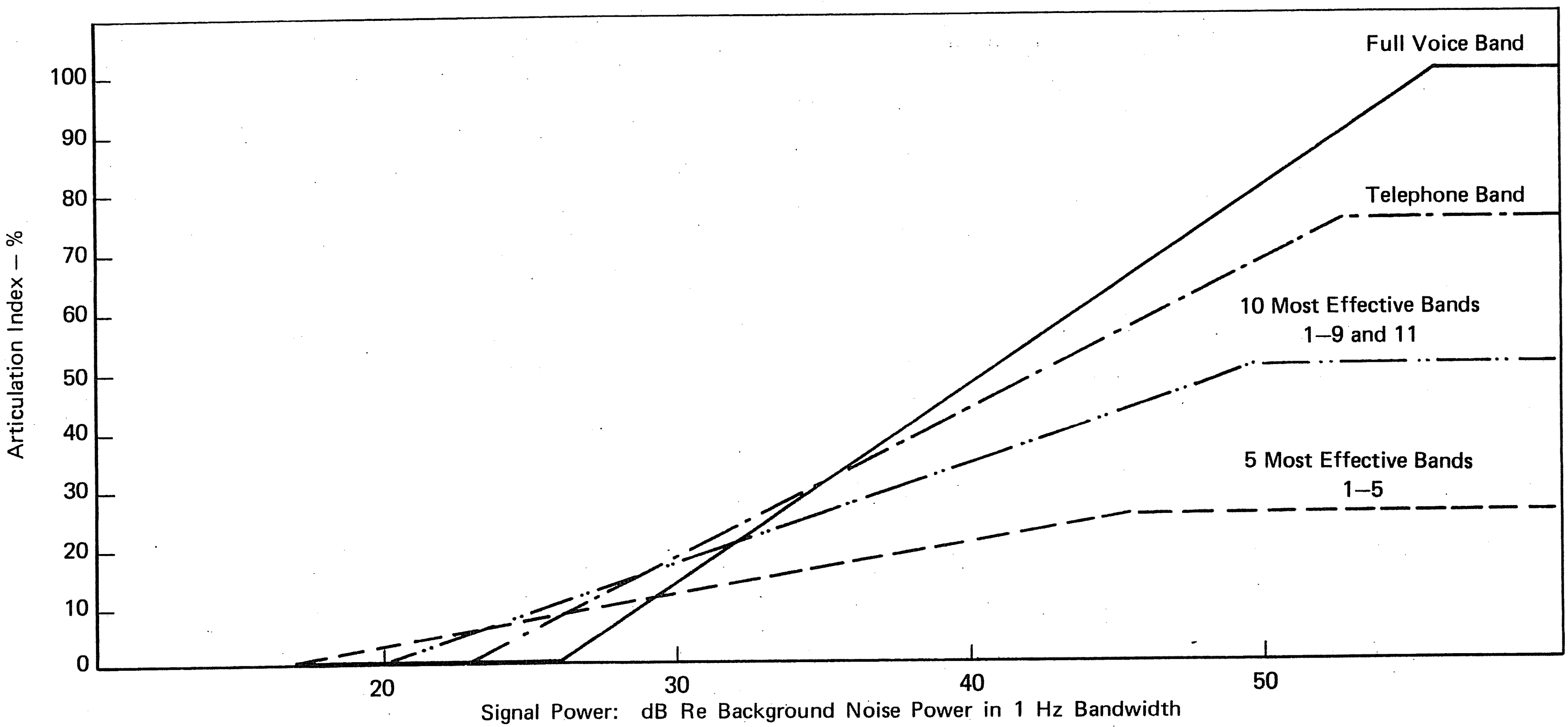
Source: Ref. 22

FIGURE I-1 RELATION BETWEEN AI AND VARIOUS MEASURES OF SPEECH INTELLIGIBILITY

It is instructive to examine how power requirements can be expected to vary for a frequency-sliced voice communication system. We start with the assumption that received noise is flat across the voice band. We also assume that the signal-to-noise ratio of proposed systems is equalized across the band of interest: 200 to 6100 Hz. In an actual system, this would probably be done by pre-emphasis of the voice signals. Because the noise is flat across the band, we can equate the total noise power directly to the bandwidth of a hypothetical system, and we can plot AI as a function of signal power level for such hypothetical systems.

The results of this exercise are shown in Figure I-2 for four conditions: full-voice band of 200 to 6100 Hz; a telephone band of 330 to 3200 Hz; bands 1 through 9 and 11, the 10 bands of largest contribution to AI per Hertz of bandwidth; and bands 1 through 5, the five bands of largest contribution to AI per Hertz of bandwidth.

It is obvious from these plots that if AI's larger than 30% are desired, there is no merit whatever in terms of signal power requirements to the use of frequency slicing. Indeed, the merit is found in the use of the widest possible bandwidth.



Source: Arthur D. Little, Inc.

FIGURE I-2 ARTICULATION INDEX FOR VARIOUS FREQUENCY SLICED SYSTEMS AS FUNCTION OF SIGNAL POWER

(This page intentionally left blank.)

APPENDIX II

ALTERNATIVES TO VOICE BANDWIDTH COMPRESSION SYSTEMS

Although we were not asked in the contract to explore alternatives to voice communication channels, the subject of narrow-band, coded systems of a non-voiced type surfaced at almost every interview and discussion, both amongst our own staff and those of other organizations with whom we discussed the voice bandwidth compression techniques. We tabulate these topics and ideas here for the sake of completeness and because of the fact that most voice bandwidth compression techniques do not appear to have as much merit as was orginally expected of them.

A. LIMITED SYMBOL NARROW-BAND CODE SYSTEM

This system would use an approach very similar to that of Westinghouse in their Interim Coal Mine Rescue and Survival System wherein a limited set of symbols is transmitted to provide a communication channel. In that instance, pushbuttons representing four or five messages were used to send very narrow-band, coded signals between the subsurface and the surface and the surface and the subsurface to provide a limited message-capability communication system. This approach was mentioned quite frequently by people with whom we discussed the general mine communication problem.

B. MORSE CODE-TYPE SYSTEMS

It is certainly possible to communicate over a very narrow bandwidth system with a Morse Code-type of communication. The problem with using Morse Code is that miners are likely to be unfamiliar with the Morse Code, and the difficulty of reading from a list of alphabetical characters and their Morse Code equivalents could pose a problem, particularly during emergency or dark conditions.

C. AN ALPHABET SYSTEM WITH KEYBOARD

Communication and data-link equipment manufacturers have brought about the development of compact, hand-held keyboard and display units, such as that illustrated in Figure II-1. A 5-bit code is entirely

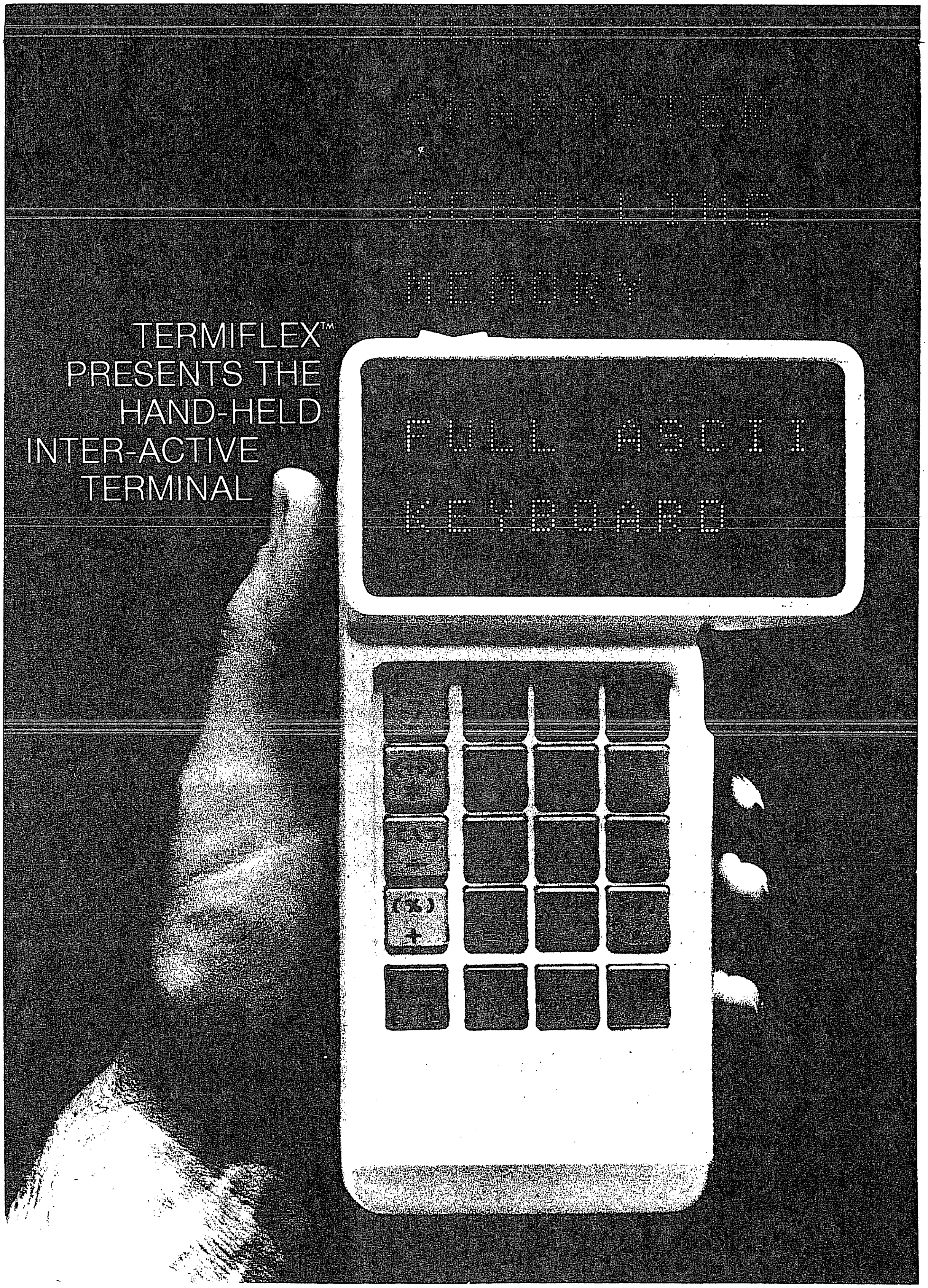


FIGURE. II-1 HANDHELD KEYBOARD AND DISPLAY adequate to represent a full alphabet of characters and messages would simply be spelled out in sequence by the operator and subsequently encoded and transmitted by the unit in a narrow-band transmission scheme. The keyboard on this portable unit is intended for one-finger operation; thus, the character rate is probably limited to three or four per second and the bit rate is approximately 20 per second, a number easily handled by a narrow-band communication system. Received messages could be presented on the optical display part of the hand-held unit such that received messages are spelled out in front of the operator. Such a low bit-rate system would appear to be compatible with performance achieved by the trapped miner detection and location equipment presently undergoing tests by the Bureau, and units suitable for use in mines should not be difficult to develop.

D. FLAG SIGNAL SYSTEMS

On two occasions when we were talking with people about communication requirements in coal mines, they immediately suggested the use of flag codes as used at sea for developing a message structure capable of transmitting a wide variety of messages with a rather small number of symbols. Naturally, the flag would be replaced by coded signals, but the technique of communicating could be made the same. Unfortunately, in this instance a look-up book would be required at both transmitter and receiver to enable the operator to put together messages and to interpret received messages.

E. PANTOGRAPH SYSTEM

In conversations with other people concerned with the narrow-band communication systems, we found the consensus to be that the Pantograph System used in some restaurants to communicate a written order from the waitress to the kitchen might be feasible for mine use. In such a system, two analog signals can characterize the position on a piece of paper of a pen from which the original written message can be reconstructed. We have not given much consideration to this as a viable means, but it is an interesting possibility for communication.

F. THE ELECTRONIC PAD

Some of the people with whom we talked suggested that use could be made of an electronic pad which, again, is another way of transmitting a written or printed message from the sender to the receiver. In this instance, a stylus would be run across an electronic pad, writing characters, symbols, pictures, or words. These would be electronically encoded and transmitted for reconstruction at the surface. Again, it could be expected that the system could operate effectively in a fairly narrow band. Little further attention was given to this as an alternate possiblity.

APPENDIX III
WHISPERED SPEECH

The analysis/synthesis type systems which form the major bulk of the voice bandwidth compression systems require that pitch be measured and a decision be made as to whether the transmitted part is voiced or unvoiced. Data from Flanagan[12] show that the channel capacity required for the pitch and the voiced/unvoiced decision is approximately 125 bps. Therefore, if it were possible to drop these two measurements from the analyzing equipment, a corresponding reduction in required channel capacity would occur. How much of a savings this reduction in required channel capacity achieves has to be referred to the bits per second required for the full capability of the system. These numbers range from 1000 to 30,000 bps, so that at best a 12% savings of channel capacity, and hence, power, could be achieved by the adoption of this scheme.

One might well inquire as to the quality or ease of dealing with conversations held in whispered voice -- the result of dropping both the voice/unvoiced and the pitch information in an analysis/synthesis system. The reproductions of speech from the control signals would be produced as whispered speech. While one might believe it to be difficult to carry on a conversation in whispered speech, this perception is probably based on the difficulty of articulating with whispered speech, but in the systems we are contemplating the input would be in normal voice and the analysis/synthesis equipment would be used to create the whispered speech. Bernald Gold of M.I.T.'s Lincoln Lab, in a private communication, told us that whispered voice is perfectly understandable. We have two other points of comparison for this conclusion. During our visit to Caldwell Smith at Air Force Cambridge Research Laboratories, we witnessed demonstrations of his equipment. One of the demonstrations was to have the synthesis equipment operate in a whispered fashion: the whispered voice resulting was perfectly intelligible. The second occasion was during a Votrax demonstration. By operating it in a whispered mode too, the output was found to be perfectly intelligible.

Thus, it can be said with a fair degree of confidence that an analysis/synthesis system can be made which does not use pitch or voiced/unvoiced information, and communicates in a satisfactory manner to the listeners in a whispered fashion. Therefore, it is a perfectly acceptable system to consider. At the present time the shortcoming of this system would be that its application has only a marginal effect on the channel capacity required for communication purposes, and thus would bring no very large benefits to a system. It could be noted, however, that by dropping the above two requirements from the system, the equipment becomes simpler to fabricate. It has generally been recognized that the pitch extraction requirement for analysis/synthesis systems always pose problems of one sort of another. The dropping of this requirement would therefore simplify the equipment moderately.

APPENDIX IV

THE NATURE OF TRANSMISSION CHANNELS

A. TRANSMISSION OF ANALOG SIGNALS ON DIGITAL CHANNELS - PCM

In the 1950's, long before the explosive growth of computer communications, Bell Telephone Laboratories developed a pulse code modulation (PCM) transmission system to handle voice signals over digital lines -- the T1 carrier system. In the original T1 carrier system (whose terminals are now designated as D1 channel banks) each voice channel used a 64,000-bps pulse stream. Despite the apparent squandering of bandwidth, T1 was economically competitive with contemporary analog carrier systems. This was a classical tradeoff of bandwidth for other advantages -- principally the ability to regenerate signals as they went along so that degradations did not accumulate with distance.

It is interesting to observe what D1 channel banks do with their 64,000-bps per channel: the voice bandwidth is limited to well below 4000 Hz and then sampled at the Nyquist rate, 8000 times per second. (Note that the sampling rate could be lower. With expensive filters the telphone band could be sharply limited to 3000 Hz and the sampling done at a little more than 6000 per second. Since filters represent a per-channel cost rather than a common equipment-shared cost, the decision was made to use a sampling rate of 8000 per second and avoid the use of expensive filters.)

In D1 each sample is encoded into a 7-bit binary word. An eighth bit is added for signalling and supervision. Thus, each channel uses:

$$(7 + 1)\ 8000 = 64{,}000 \text{ bps.}$$

The noise in a PCM system is mainly due to quantizing distortion which is caused by the approximation in assigning a discrete binary word to the value assumed by a continuous speech voltage waveform. A well-known theoretical analysis which assumes a uniform probability density of the quantizing error over $\pm 1/2$ the size of the least significant bit has the result that for an n bit binary word the ratio of the rms quantizing error to the full range of values is $1/2^n\sqrt{12}$. Thus, for the 7 bits of D1 the theoretical ratio of the rms quantizing noise is (in dB):

$$20 \log [(0.707) (\sqrt{12}) (2^7)/2] = 44 \text{ dB}$$

which is surprisingly good. If it were not for the restricted bandwidth, it might qualify as high fidelity. A linear PCM system, however, is very poorly matched to the amplitude statistics of a speech signal. D1 consequently was modified to provide optimum performance for speech signals by compressing the speech signal before encoding and expanding it after decoding. This compressing plus expanding (called companding) follows an experimental law such that for the compressor alone:

$$E_{out} = \frac{\ell n \ (1 + \mu E_{in})}{\ell n \ (1 + \mu)}$$

where μ = 100 for D1. The effect of this logarithmic companding is to smear the original signal-to-noise ratio out over a wider dynamic range. With linear encoding, 7 bit theoretical S/N is 44 dB at full load and drops 1 dB per dB as the level decreases (reaching 0 dB at -44 dB signal level. This behavior is shown by the dotted line in Figure IV-1. With μ = 100 the signal-to-noise ratio remins at 30 to 35 dB from 0 to -30 dB signal level (see Figure IV-1). Logarithmic companding is therefore a tradeoff, since it provides more noise distortion for large amplitudes and less for small aplitudes, but that is just what voice signals need.

If D1 channels had been meant for some other signal, such as facsimile or fsk data, the companding (if any) would certainly have been completely different.

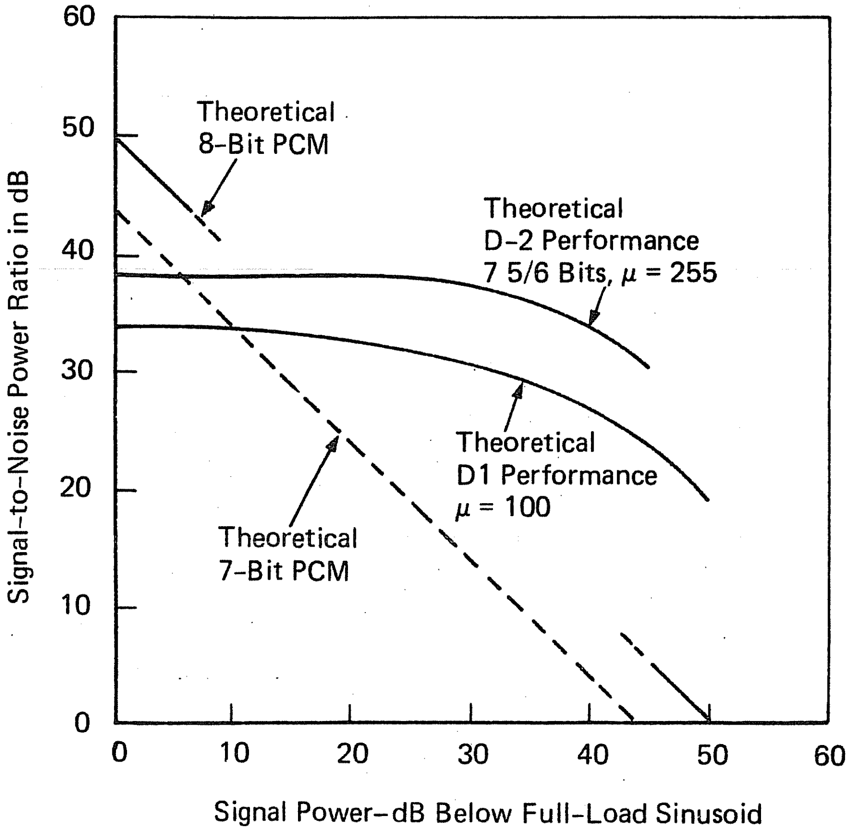
A value of 100 for μ was chosen for T1 because of the technology of the late 1950's. The compressor and expander nonlinear characteristics were determined by diode networks, and the higher the μ, the more difficult was the problem of the two characteristics tracking to cancel each other. It should be noted that the logarithmic characteristic used in T1 is by no means universal and other countries use different companding laws.

D1 channel banks were designed for exchange trunk transmission and did not meet the requirements for longer haul trunk carriers. The D2 channel bank was designed about a decade after the D1 to make use of more modern technology to provide toll transmission without changing the basic T1 line format of 1,544,000 bps. It was clear that using a whole bit per sample (8000 per second) to transmit switchhook and dialing information was wasteful, so D2 uses this bit for voice signal transmission 5 times out of 6. (A slight change is required to keep track of channel signal framing as well as channel number framing.) In D2, μ was changed to 255. No longer are coder and compander separate -- the two functions are combined in a single, nonlinear coder performing both functions. These two changes result in a theoretical S/N ratio better than 35 dB over about a 39-dB dynamic range which meets toll requirements in the direct dial hierarchy (see Figure IV-1).

It should be realized that logarithmic companding in itself is a powerful <u>bandwidth compression scheme</u>. By capitalizing on the instantaneous amplitude distribution of speech, it offers a great economy in number of bits needed. For example, to pick a range on the D1 curve, if we want to maintain a 17.5 dB S/N ratio at a signal power 40 dB below full load with linear PCM, 11 digits would be needed per sample instead of 7. Thus $\mu = 100$ companding reduces the line bit rate by 36%.

The curves in Figure IV-1 are ideal curves. Actual PCM systems may be 1 to 3 dB poorer for any of a number of reasons such as small errors in sampling time, finite rather than instantaneous sampling, compressor-expander characteristic mistracking, noise in decision circuits, and general system noise.

The analog telephone channels derived on T systems are used interchangeably with all other analog channels in the telephone plant. Thus, it is not unusual for a 1200- or 2400-bps modem signal to be riding on a T carrier channel using 64,000 bps. This dual standard poses tariff inconsistencies for telephone companies. For example, to provide a 50-kbs data channel, 12 voice channels must be removed from an analog carrier but only one from a digital. Tariffs are not geared to the particular facility providing service. Is the fair price for a 50-kbs channel



Source: Ref. 14, 15

FIGURE IV-1 PCM QUANTIZING NOISE

roughly 1 or 12 times that for a voice channel? This inconsistency is one of the pressures that is leading to all digital systems such as Datran and Bell's Dataphone Digital Service (DDS).

B. TRANSMISSION OF DIGITAL SIGNALS ON ANALOG CHANNELS - MODEMS

At the same time that Bell was finding it economically attractive to transmit analog signals on digital chanels, the inverse problem arose and grew rapidly. Telephone plant had evolved over the years to handle voice efficiently and was not well suited for raw digital data. For example, phone channels generally do not transmit DC or indeed any frequency below a few hundred Hertz, so they are not suitable for baseband DC pulses. To transform the data information into a format compatible with the telephone channel, it is necessary to modulate it in some way. To reconstitute it, it must be demodulated at the other end. This combination of a modulator-demodulator is called a modem or a data set.

By the mid 1960's data sets for 1200 to 2400 bps using frequency shift keying (FSK) or phase modulation were readily available. It is instructive to consider how much better transmission might be. A well known result from information theory is Shannon's expression[16] for the information capacity of a band-limited channel in the presence of gaussian noise:

$$C = W\log_2 \left(1 + \frac{s}{n}\right)$$

where C is the channel capacity in bps, W is the bandwidth in Hertz, s is the signal power, and n the noise power in the band W. Shannon's equation is analogous to the Carnot cycle in thermodynamics; it sets an upper bound on theoretically achievable performance, but does not tell <u>how</u> to achieve or even approach this bound.

An interesting consequence of Shannon's equation is that channel capacity can be increased without bound for any W or S/N ratio provided there is no limit on the other. Another limit to transmission rate is the Nyquist rate 2W -- the upper bound on which <u>symbols</u> can be sent through a band W. This is not a limit on channel capacity since there is no limit on the number of bits that can be encoded in each symbol; indeed it is intuitively clear that the larger S/N ratio is, the more bits per symbol a decoding arrangement can distinguish. Section E shows some of the properties of Shannon's equation.

What does Shannon tell us about a telephone channel? For an untreated channel the band may be 400 to 3000 Hz for a W = 2600. If the S/N ratio is 30 dB, we get:

$$C = 2600 \log_2 \ (1 + 1000) = 25{,}900 \text{ bps.}$$

Using complex combinations of phase and multilevel modulation, modems for 9600 bps are now available, considerably short of the 25,900 bits theory says are possible. If we take the true capacity of the channel as 25,900 bps, we can assess the present state of the art, in terms of efficiency:

$$\text{For digital transmission on analog channel: } \frac{9{,}600}{25{,}900} = 37\%$$

$$\text{For analog transmission on digital channel: } \frac{25{,}900}{56{,}000} = 46\%$$

This is the money changer situation -- you lose both ways. In the next two sections we examine what hope there is for improving these two efficiencies.

C. BPS PER HERTZ IN DIGITAL TRANSMISSION ON AN ANALOG CHANNEL

While it is a side issue, it may be worth noting that 9600 bps on a telephone channel, even though it represents only 37% efficiency, may be close to the best that will be accomplished for a long time, maybe ever. Shannon's equation speaks only to the problem of gaussian noise and there are many other degradations in phone channels including phase distortion, impulse noise, crosstalk, phase jitter, frequency offsets, and nonlinear distortions. It has only been the development of means of coping with the first of these, phase distortion, with automatic transversal equalizers that speeds have risen from the 2400-bps range to 9600 bps. Cure of some of the remaining ills is beyond the present state-of-the-art.

It is useful to examine the theoretical efficiencies of various modulation techniques. Details of the techniques themselves are beyond the present scope of the report; see for example, Bennet & Davey[17] or Martin[18]. For band-limited gaussian noise, the efficiencies are plotted in Figure IV-2 (from Salz - Ref. 19). These curves assume that transmitters and receiver filters are optimized to discriminate against noise. Smooth curves are drawn connecting the efficiencies of 2-level, 4-level, 8-level, etc., systems.

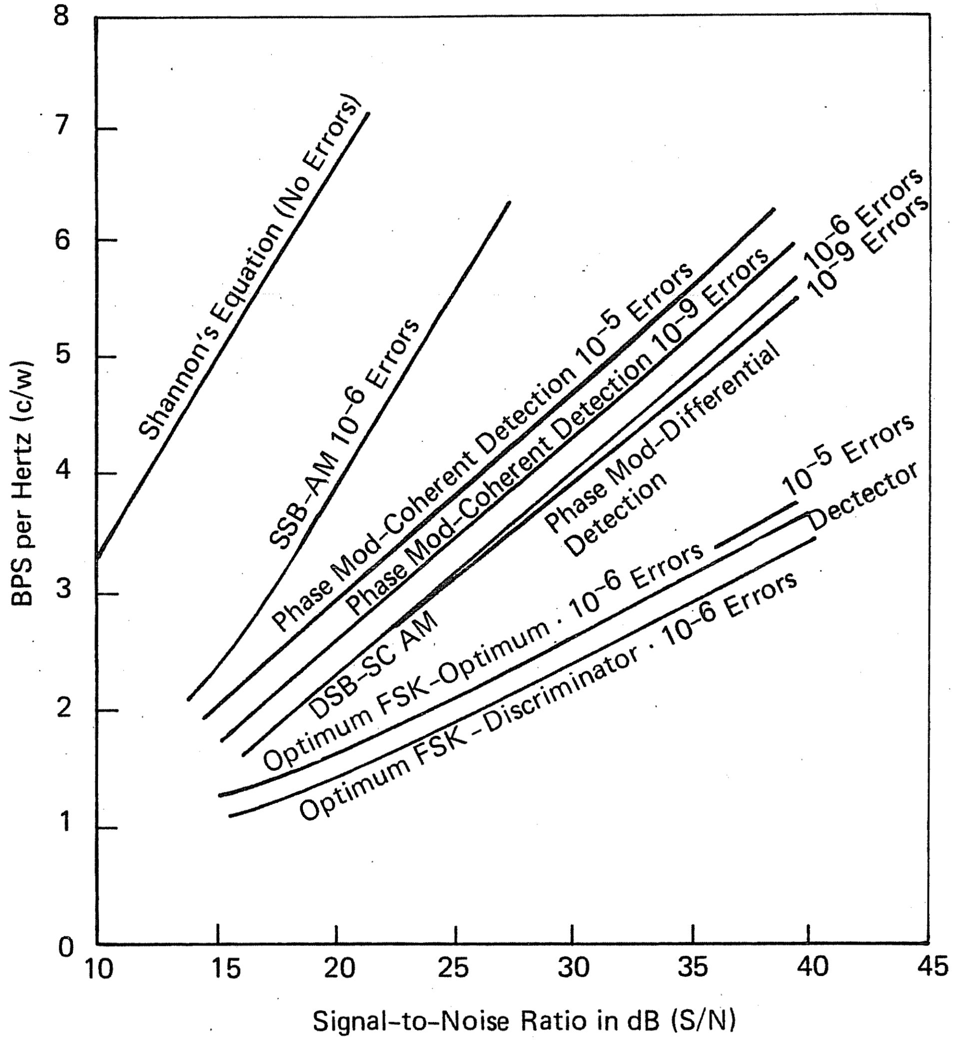
Single sideband is the most efficient system although it falls far short of Shannon's equation. FSK is least efficient. Some of the current research in modems use novel modulation not shown on Figure IV-2, for example, simultaneous combination of amplitude modulation and phase modulation.

The problem of mine communication with bandwidth compression techniques is the joint problem of processing and transmission through a very hostile medium. Efficient interaction of the two is of utmost importance. An experimental program in bandwidth compression must take into account the overall system. For example, Figure IV-2 hints at the rapid changes in error rate with small changes in S/N ratio. Two equally efficient processing schemes may react quite differently to a distribution of digit error limits. It should also be noted that Figure IV-2 deals only with gaussian noise. The peculiar noise environment in a mine may give very different results.

D. HERTZ PER BPS IN ANALOG TRANSMISSION ON DIGITAL CHANNELS

We have seen that companding reduces the number of bits for D1 channel performance from 11 to 7, but that the resultant efficiency still leaves room for improvement. A great deal of telephone-oriented research has been taking place in Delta modulation and related differential quantizing techniques. The prime motivation of the research has not been bandwidth saving per se, but rather coder-decoder simplification.

A nonlinear PCM encoder-decoder is quite expensive. This is not serious in T carrier, since it is shared by 24 channels so the per channel



Source: Ref. 19

FIGURE IV-2 BITS/SECOND/HERTZ VS SIGNAL/NOISE

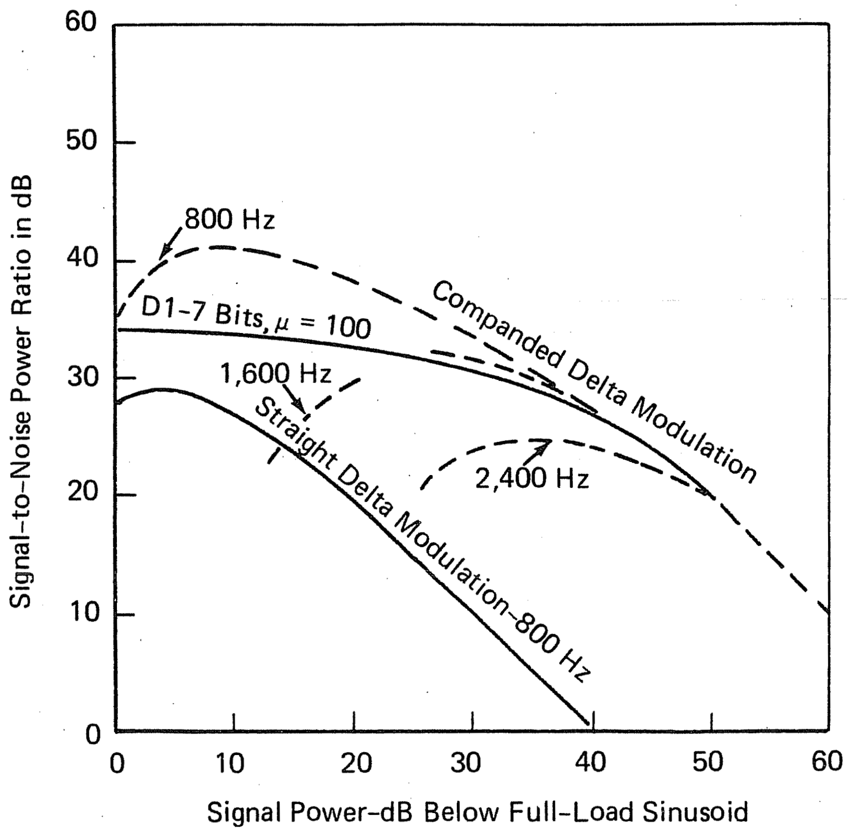
cost is not excessive. There are applications in subscriber carriers for much smaller bundles than 24, and here the cost of the PCM encoder-decoder becomes a problem. A Delta modulation coder-decoder is very simple compared with its PCM counterpart. If Delta modulation can be made as good as PCM for a given bit rate, the economics of small digital systems become attractive.

The basic idea of all differential quantizing schemes is to quantize and encode changes in the signal rather than instantaneous samples of the signal itself. The simplest type is Delta modulation which can be thought of as 1-bit differential PCM.

A Delta coder consists of a comparator in which an analog signal derived from the coder output is compared with the input signal. At intervals controlled by a clock, an output pulse is sent if the input is larger than the output and no pulse is sent if it is smaller. Decoding consists of a simple integration process. The more often pulses are sent (higher bit rate) the better the approximation. Delta modulation is subject to a unique type of distortion known as slope-overload distortion. Here the rate of change of the input signal becomes too large for a series of ones to follow. As a result, curves of S/N ratio vs level are different for different frequencies. Fortunately, the level for a given distortion falls off at almost exactly the same rate as a normal speech energy spectrum. As a consequence, if an adequate S/N ratio exists at 800 Hz, speech will tend to sound natural.

Figure IV-3 compares differential quantizing schemes with D1 theoretical performance. For normalization, a voice signal rate the same as D1 (i.e., 56 kbps) is assumed. It will be noted that straight Delta modulation is quite inferior. Delta modulation is appreciably improved by companding. Various companding schemes are detailed in the references. One of the most promising (Tomozawa and Kaneki[20]) is shown in Figure IV-3.

More sophisticated schemes have been proposed. One of these (Greefkes and Riemens[21]) is for digitally controlled companded Delta modulation (DCDM). Here the quantization strip size is varied at the transmitter



Source: Ref. 20

FIGURE IV-3 DIFFERENTIAL QUANTIZING AT 56KBPS

and receiver simultaneously under the control of the digital signal. To indicate the potential usefulness of this for mine communications, the 800-Hz performance of a 16-kbps DCDM channel is compared with 5 kbps D1 PCM for reference and shown in Figure IV-4.

E. CHANNEL CAPACITY RELATIONSHIPS

In the investigation of the applicability of voice bandwidth compression techniques to mine communication systems, it is important to know the general way in which channel capacity varies as a function of channel bandwidth, noise, and signal power. The classical relationship is contained in Section B and has been recast below:

$$C = W \log_2 \left(1 + \frac{P_o}{WN_o}\right)$$

where

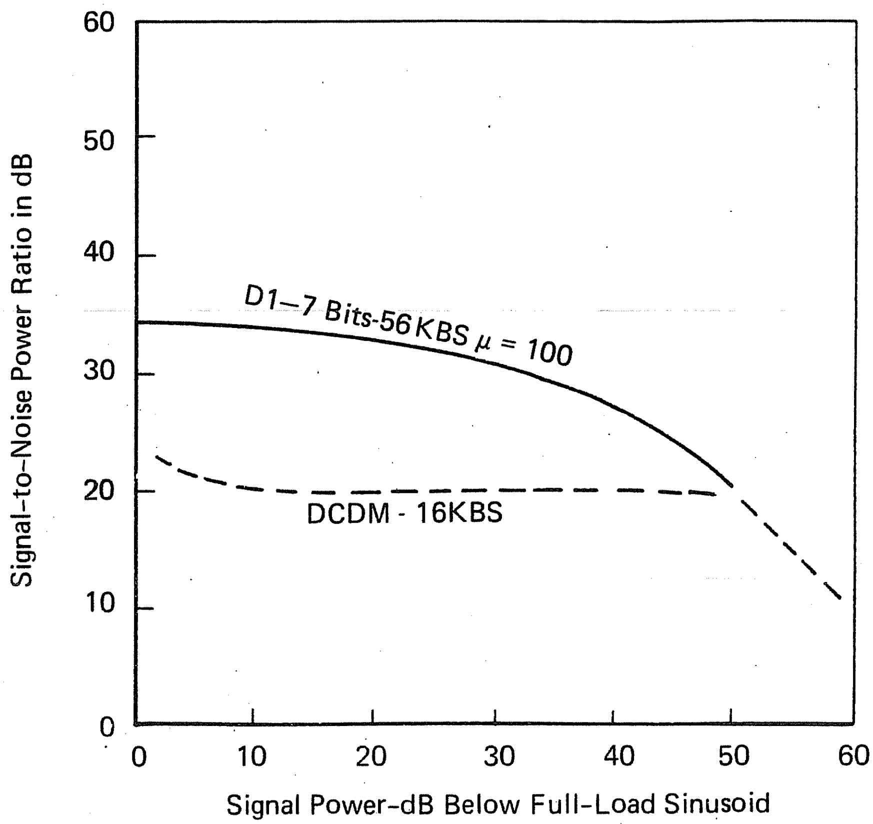
C = channel capacity in bps,

W = bandwidth in Hertz,

P_o = received signal power,

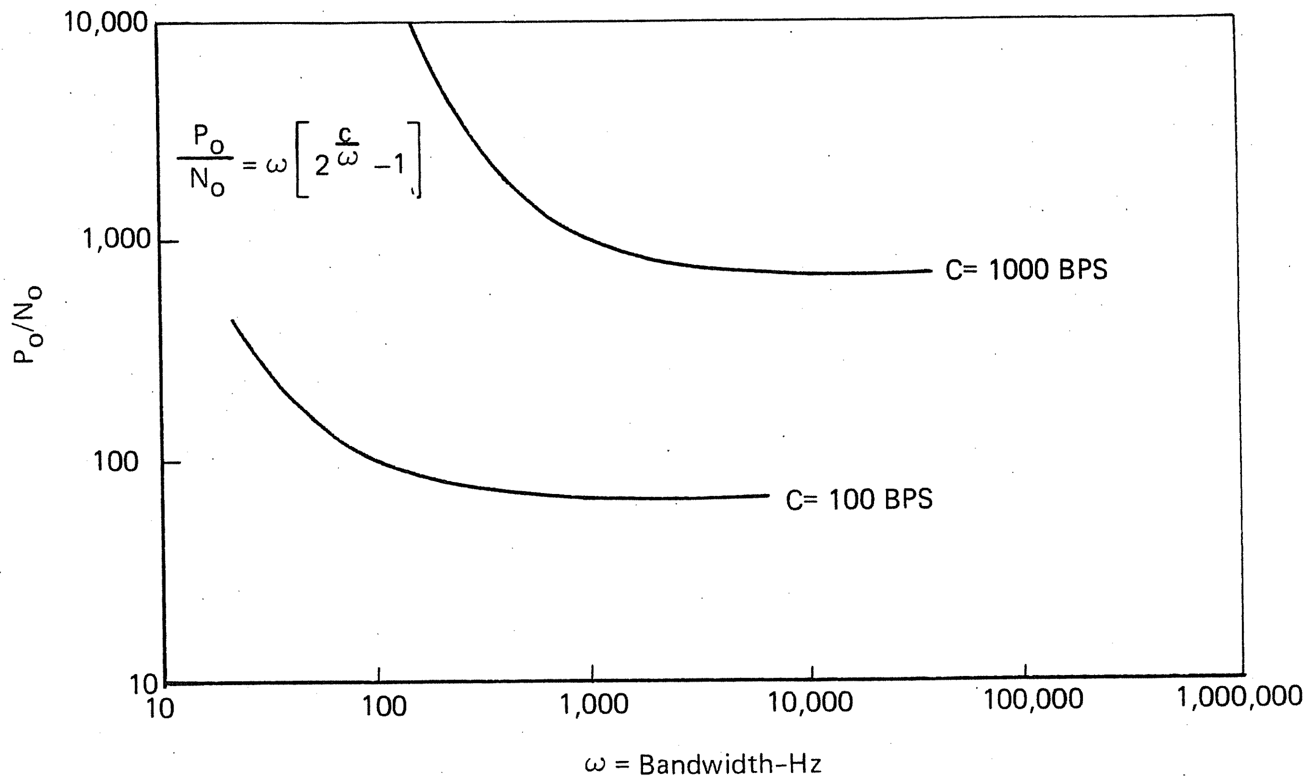
N_o = received noise spectral density (flat noise assumed).

Two curves (Figures IV-5 and IV-6) derived from this relationship are plotted. Figure IV-5 seeks to show how transmitter power (and thus received power) depends on bandwidth for a fixed-channel capacity requirement. This plot illustrates the penalty in power that is exacted if one insists on a communication system that asks for even a few bps per Hertz of bandwidth. Figure IV-6 shows how channel capacity varies for a fixed power level and variable bandwidth.



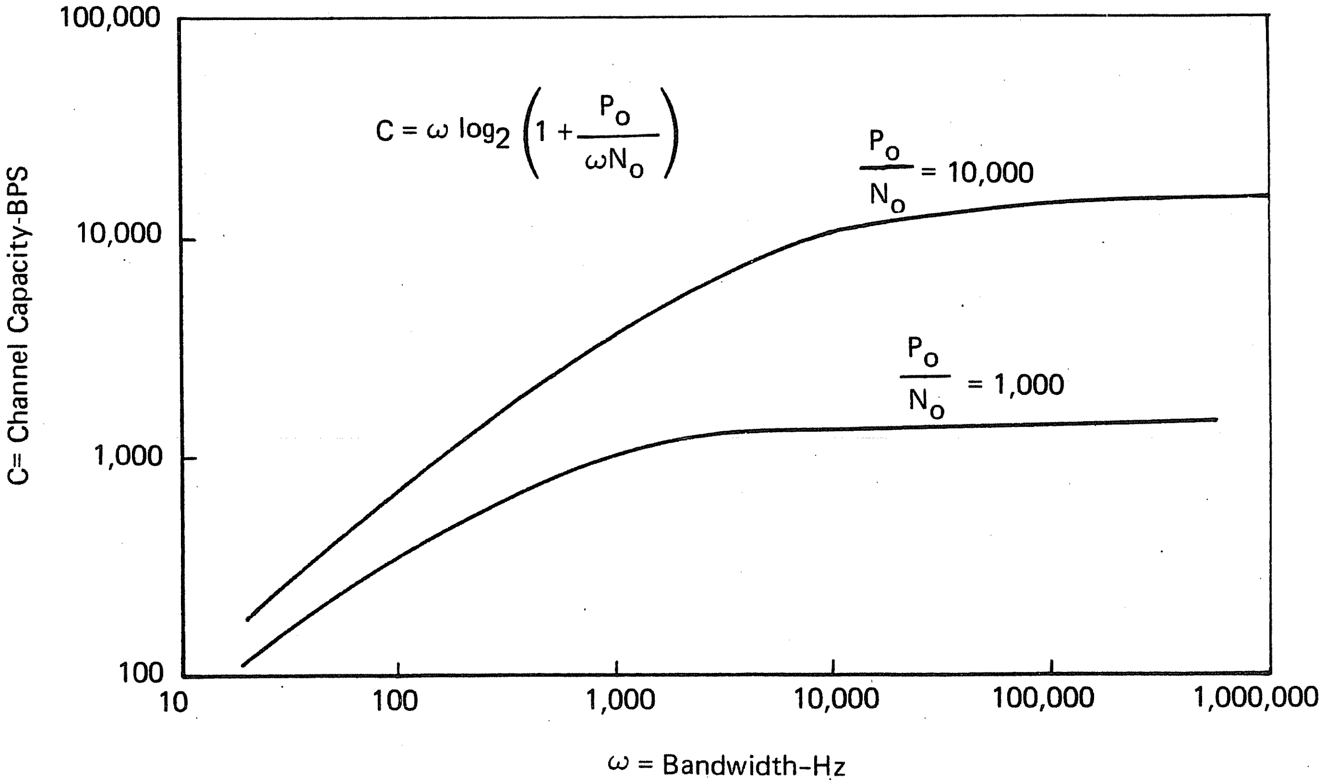
Source: Ref. 21

FIGURE IV-4 16KBPS DCDM PERFORMANCE



Source: Arthur D. Little, Inc.

FIGURE IV-5 RELATIVE POWER REQUIRED AS FUNCTION OF BANDWIDTH FOR FIXED CHANNEL CAPACITY IN PRESENCE OF FLAT NOISE



Source: Arthur D. Little, Inc.

FIGURE IV-6 CHANNEL CAPACITY AS FUNCTION OF BANDWIDTH FOR FIXED SIGNAL POWER IN PRESENCE OF FLAT NOISE

APPENDIX V

LINCOMPEX

Perhaps the most interesting method of improving signal-to-noise ratio in voiced AM radio communication developed in the past decade is the Lincompex[24] (for Linked Compressor and Expander) technique. The basic idea is to transmit the speech signal at the maximum level capability of the channel. Amplitude variations in the speech signal are almost entirely removed by a compressor. The amplitude information is transmitted as a narrow-band frequency modulated signal. In the receiver, the amplitude signal is used to control an expander in order to restore the normal amplitude variations in the speech signal.

A simplified block diagram of the Lincompex system is shown in Figure V-1. The input speech is supplied to an amplitude measurement circuit as well as a direct speech channel. The delay in the speech channel is necessary to synchronize the amplitude measurement, which is delayed by a low-pass smoothing filter, with the envelope of the speech signal. The amplitude signal controls the compressor, a variable gain amplifier, so as to reduce the gain at high amplitude and increase the gain at low amplitude, resulting in a compressed speech signal with constant amplitude. The speech signal is then filtered to remove any components above 2700 Hz.

The amplitude signal is also applied to a voltage-controlled oscillator to obtain a narrow-band frequency-modulated signal varying from 2840 to 2960 Hz, which corresponds to an amplitude range of 60 dB (2 Hz per dB). This FM amplitude signal is now added to the filtered compressed speech signal to modulate the SSB transmitter.

The speech and amplitude components of the received Lincompex signal are separated by filters, as shown in Figure V-1. The FM amplitude signal is detected by a discriminator, and this output is used to control the expander. The speech signal is also applied to the expander. The expander is a variable gain amplifier operated in the reverse manner from the compressor. Now the gain is increased by high amplitude signals

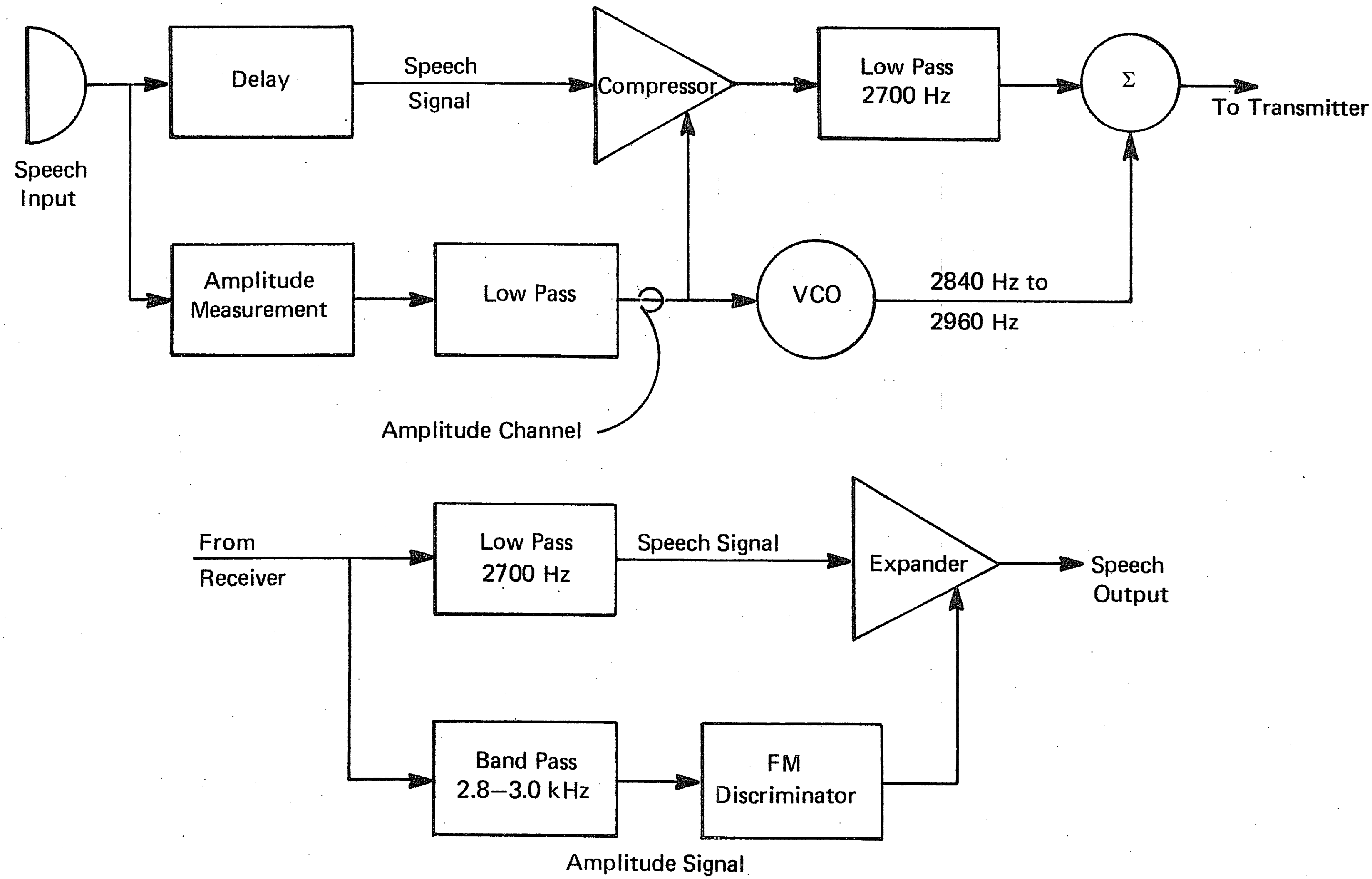


FIGURE V-1 SIMPLIFIED BLOCK DIAGRAM OF LINCOMPEX SYSTEM and reduced by low amplitude signals from the discriminator, with the result that the speech output of the expander has the amplitude variations of the original speech signal. Obviously, it is very important that the speech and amplitude signals be kept in synchronism if the system is to work properly. This necessitates several other delays not shown in the simplified block diagram. A high degree of frequency stability is also required because of the narrow-band FM used to transmit the amplitude information. The overall system is complex and expensive.

While there is general agreement that Lincompex does provide significantly improved quality conversation under certain poor signal-to-noise conditions, there is a difference of opinion on the desirability of Lincompex for certain applications. The negative view contends that, in general, the improvement is not worth the expense, and that most of the time, transmission conditions are sufficiently good so that the improvement due to Lincompex is marginal.

(This page intentionally left blank.)

BIBLIOGRAPHY

PCM and Companding

K. W. Cattermole, "Principles of Pulse Code Modulation," London, Iliffe Books, Ltd., 1969.

H. Mann, H. M. Straube and C. P. Villars, "A Companded Coder for an Experimental PCM Terminal," Bell System Technical Journal, Vol. 56, pp 173-226, January 1962.

B. M. Oliver, J. R. Pierce, C. E. Shannon, "The Philosophy of PCM," Proc. I.R.E., Vol. 36, pp 1324-1331, November 1948.

B. Smith, "Instantaneous Companding of Quantized Signals," Bell System Technical Journal, Vol. 36, pp 653-709, May 1957.

Digital Transmission on Analog Channels

W. R. Bennet and J. R. Davey, "Data Transmission," New York, McMillian, 1965.

J. Martin, "Telecommunications and the Computer," Englewood Cliffs, New Jersey, Prentice-Hall, 1969.

J. Salz, "Communications Efficiency of Certain Digital Modulation Systems," IEEE Trans. on Comm-Technology, Vol. COM-18, pp 97-102, April 1970.

C. E. Shannon, "Communication in the Presence of Noise," Proc. I.R.E., Vol. 37, pp 10-21, January 1949.

Delta Modulation, Differential PCM, etc.

S. Brolin and J. Brown, "Companded Delta Modulation for Telephony," IEEE Trans. Comm-Technology, Vol. COM-16, pp 157-162, February 1968.

P. Cumminsky, N. Jayant and J. Flanagan, "Adaptive Quantizing in Differential PCM Coding of Speech," Bell System Technical Journal, Vol. 52, pp 1105-1118, September 1973.

J. Greefkes and F. DeJaeger, "Continuous Delta Modulation," Philips Research Report No. 22, pp 233-246, 1968.

J. Greefkes and K. Riemens, "Code Modulation with Digitally Controlled Companding for Speech Transmission," Philips Tech. Review, Vol. 31, pp 115-133, 1970.

N. Jayant, "Adaptive Quantization with a One-Word Memory," Bell System Technical Journal, Vol. 52, pp 1119-1144, September 1973.

F. Johnson, "Calculating Delta Modulator Performance," IEEE Trans. Audio and Electroacoustic Vol., pp 121-129, March 1968.

E. Protonotarios, "Slope Overload Noise in Differential Pulse Code Modulation Systems," Bell System Technical Journal, Vol. 46, pp 2119-2161, 1967.

H. Schlindler, "Delta Modulation," IEEE Spectrum Vol., pp 69-78, 1970 (includes bibliography with 36 entries).

A. Tomozawa and H. Kaneko, "Companded Delta Modulation for Speech Transmission," IEEE Trans. Comm-Technology, Vol. COM-16, pp 149-156, February 1968.

Characteristics of Speech

A. Albert, "Electrical Communication," 3rd Edition, New York, Wiley, 1950 (Chapter 2 has 47 references on fundamentals of speech and acoustics.)

C. Cherry, "On Human Communication," 2nd Edition, New York, Dover, 1961.

K. D. Kryter, "Methods for the Calculation and Use of the Articulation Index," J. Acoust. Soc. Am., 34:11 (November 1962).

Speech Bandwidth Compression Systems

B. Atal and M. Schroeder, "Adaptive Predictive Coding of Speech Signals," Bell System Technical Journal, Vol. 49, pp 1973-1986, October 1970.

B. H. Batson, "An Analysis of the Relative Merits of Various PCM Code Formats, National Aeronautics and Space Administration, Internal Note No. MSC-EB-R-68-5 (November 1, 1968).

B. H. Batson and R. W. Moorehead, "A Digital Communications System for Manned Spaceflight Applications", National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, Houston.

J. Bayless, S. Campanella and A. Goldberg, "Voice Signals: Bit-by-Bit," IEEE Spectrum, Vol. 10, pp 28-34, October 1973.

B. Bogert, "The Vobanc - A Two-to-One Speed Bandwidth Reduction System," Jour. Acoust. Soc. Am., Vol. 28, pp 399-404, May 1956.

S. J. Campanella, D. C. Coulter, L. A. Barnes, R. E. Irons and J. M. Pickett, "Formant Tracking Vocoder System, Melpar Inc. a Subs. of Westinghouse Air Drake Co., 3000 Arlington Blvd., Falls Church, Virginia (September 1, 1963 to November 30, 1963), Contract No. DA36-039-AMC-0006(e), DA Project No. 1G640309D491; DA Task No. 1G640309D49105.

S. J. Campanella, D. C. Coulter and R. E. Irons, "A 1000 Bit Per Second Speech Compression System," Melpar, Falls Church, Virginia, a report of work on Contract AF33(600)-39962.

J. Daguet, "Summary Codimex, Speech Compression System," Paper G 2 in Proc. of the Speech Communication Seminar, Vol. II (Stockholm 1963).

H. Dudley and T. Tarnoczy, "The Speaking Machine of Wolfgang von Kempelen," Jour. Acoust. Soc. Am., Vol. 22, pp 151-166, March 1950 (description of a successful pneumatic-mechanical phoneme vocoder built in 1780).

J. L. Flanagan, "Development and Rating of a Formant-Coding Speech Compression System," Jour. Acoust. Soc. Am., 28:6, November 1956.

J. L. Flanagan, "Speech Analysis, Synthesis and Perception," Academic Press, New York, 1965.

J. L. Flanagan, "Bandwidth and Channel Capacity Necessary to Transmit the Formant Information of Speech," Jour. Acoust. Soc. Am., Vol. 28, pp 592-596, July 1956.

J. E. Flood and D. I. Urquhart-Pullen, "Time-Assignment Speech Interpolation in Time-Compression-Multiplex-Transmission, Proc. IEEE 111, (4): (April 1964).

G. Font and K. N. Stevens, "Systems for Speech Compression," Fortschr. Hochfrequenztechn., 5 (1960).

A. J. Goldberg, "Vocoded Speech in the Absence of the Laryngeal Frequency," B. Gold, Editor Group 62, Technical Note 1967-9, 3 April 1967-9 [AD651 253].

A. J. Goldberg, M. J. Rose and A. Areese, "A Predictive Voice Processor for Narrowband Communication Channels," IEEE, ICC, June 1974.

H. Gruen and B. Olevsky, "Increasing Information Transfer," Advanced Communications Laboratory, Philco Corp.

J. Haskew, J. Kelly, R. Kelly, Jr. and T. McKinney, "Results of a Study of the Linear Prediction Vocoder," IEEE trans. on Comm., Vol. COM-21, pp 1008-1115.

H. E. Jones, "Bandwidth Economy for Multiplexed Digital Signals," Westinghouse Electric Corporation Defense and Electronic Systems Center, Baltimore, Maryland.

J. W. Jones and W. R. Roberts, "Performance Standards and Test Specifications for Voice Message Amplitude and Bandwidth Compression Equipment," Final Report, Contract AF30(635)-12857.

K. D. Kryter, "Speech Bandwidth Compression Through Spectrum Selection," Jour. Acoust. Soc. Am., Vol. 32, pp 547-556, May 1960.

K. D. Kryter, "Validation of the Articulation Index," Jour. Acoust. Soc. Am., 34:11, November 1962.

D. T. Magill, "Adaptive Speech Compression for Packet Communication System," Stanford Research Institute, Menlo Park, California 94025.

C. H. Manney, "Applicability of Speech Bandwidth Compression Techniques in Mine Electromagnetic Communications," National Bureau of Standards Report 10728, Sponsored by Bureau of Mines Agreement H)111019.

J. Markel and A. Gray, Jr., "A Linear Prediction Vocoder Simulation Based Upon the Autocorrelation Method," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-22, pp 124-134, April 1974 (includes 15-item bibliography).

W. F. Meeker and A. L. Nelson, "Vocoder Evaluation Research," National Technical Inf. Service, U.S. Dept. of Commerce, Springfield, Va. 22151 (February 29, 1964), Contract No. AF19(604)8507, Project No. 4610, Task No. 4610002, AD-602955 (prepared for Air Force Cambridge Research Laboratories Office of Aerospace Res. U.S. Air Force, Bedford, Mass. AFCRL-64-453; CR-64-558-1).

H. L. Morgan, "Bandwidth Compression and Expansion by Frequency Division and Multiplication," U.S. Patent 3,349,184, October 24, 1967.

A. Noll, "Cepstrum Pitch Determination," Jour. Acoust. Soc. Am., Vol. 41, pp 293-309, February 1967.

J. B. O'Neal, "Efficient Digital Encoding Methods for Improved Military Communications Systems," NTIS National Technical Inf. Service, U.S. Dept. of Commerce, Air Force Office of Scientific Res., Contract No. F44620-69-C-0033 (April 1, 1972).

G. Power and C. Speaks, "Intelligibility of Temporally Interrupted Speech," Jour. Acoust. Soc. Am., Vol. 54, pp 661-667, 1973.

B. O. Pyron and F. R. Williamson, Jr., "Signal Processing By Infinite Clipping and Related Techniques," National Technical Inf. Service, U.S. Dept. of Commerce, Spring, Va. 22151 (August 16 to March 31, 1964), Project No. A-727, Contract No. DA 49-092-ARO-21 AD600 863 (prepared for Army Research Office, Dept. of the Army, Washington, D.C.

L. R. Rabiner, L. B. Jackson, R. W. Shafer and C. H. Coker, "A Hardware Technology, Vol. COM-19, No. 6, December 1971.

J. Savidge, "Ease Radio-Spectrum Congestion by Bandwidth Compression and Expansion," EDN, pp 74-77, May 20, 1973.

M. Schroeder, "Vocoder: Analysis and Synthesis of Speech," Proc. IEEE, Vol. 54, pp 720-734, May 1966.

M. Schroeder, B. Logan and A. Prestigiacoma, "New Methods for Speech Analysis-Synthesis and Bandwidth Compression," 4th Int. Congress on Acoustics, Copenhagen, 21-28 August 1962.

G. S. Sebestyen and D. Van Meter, "Investigation of Automation of Speech Processing for Voice Communication," Litton Systems, Inc. Data Systems Division (May 28, 1962), Contract No. AF19(604)-8828 AD 287 887 (reproduced by the Armed Services Technical Information Agency, Arlington Hall Station, Arlington, Virginia).

H. Seo, "Speech Compression," University Microfilms Limited, High Wycomb, England (submitted in partial fulfillment in the Department of Elec. Eng., University of South Carolina) 1967.

C. Smith, "Perception of Vocoder Speech Processed by Pattern Matching," Jour. Acoust. Soc. Am., Vol. 40 No. 6, pp 1562-1571, December 1969.

C. P. Smith, "A Phoneme Detector," Jour. Acoust. Soc. Am., 23:4, July 1951.

C. P. Smith, "Speech Compression Studies," Armed Services Technical Information Agency, Arlington Hall Station, Arlington, Virginia (September 1962) AD 291 740 [presented at the 1962 Northeast Electronics Research and Engineering Meeting (NEREM), Boston, Mass.]

J. L. Steward, "Two New Speech Comm-nication Systems," Technical Documentary Report No. MRL-TDR-62-10, March 1962, Biomedical Laboratory, 6570th Aerospace Medical Res. Labs, Aerospace Medical Div., Air Force Systems Command, Wright-Patterson Air Force Base, Ohio; Project No. 7233, Task No. 723305 (prepared under Contract No. AF33-(616)-7800 AD 278 711.

H. A. Straight and O. C. King, "Speech Processing Studies (Voice Data Processing System), Melpar Inc., 3000 Arlington Blvd., Falls Church, Virginia, Contract No. AF19(628)214, Project No. 4610, Task No. 461002, Final Report, October 1963 [AD4304611] (prepared for Cambridge Research Laboratories, Office of Aerospace Research, U.S. Air Force, Bedford, Mass.

W. Wathen-Dunn and D. Lipke, "On the Power Gained by Clipping Speech in the Audio Band," Jour. Acoust. Soc. Am., Vol. 30, pp 36-45, January 1968.

R. M. Wilkinson, "An Adaptive Pulse Code Modulator for Speech Signals," Signals Research and Development Establishment (January 1972), Report No. 72001, AD 750-167, Procurement Executive Ministry of Defense S.R.D.E., Christchurch, Hampshire, England.

J. Williams, "Some Bandwidth Compression Systems for Speech Transmission," Radio and Electronic Engr., Vol. 27, pp 459-472, June 1964.

R. Viswanathan and J. Makoal, "Current Issues in Linear Predictive Speech Compression," presented at EASCON, October 1974.

Philco Ford, "Research and Development of an Optimized Laboratory Speech-Compression System for Spacecraft Application," Final Report, Contract NAS9-4523 (prepared for Manned Spaceflight Center, NASA, Houston, Texas, May 1966.

"Speech Analysis/Synthesis Survey," 1967, sponsored by Office of Aerospace Research, USAF.

"Speech Analysis Techniques," Contract No. DA-91-591-EUC-2626 (July 1, 1962 - June 30, 1963), Final Technical Report No. 4 AD 417356, catalogued by DDC Defense Documentation Center for Scientific and Technical Information, Cameron Station, Alexandria, Virginia.

"1972 Conference on Speech Communication and Processing, Institute of Electrical and Electronics Engineers, Inc., Air Force Cambridge Research Laboratories, Cat. No. 72 CHO 596-7 AE (April 24-]6, 1972) AD 742236.

"An Evaluation of Speech Compression Techniques," Bolt, Beranek and Newman, Inc., Technical Report No. RADC-TDR-63-90 (February 23, 1963), (prepared for Rome Air Development Center), Research and Technical Division, Air Force Systems Command, U.S. Air Force, Friffiss Air Force Base, N.Y.; Project No. 4519, Task No. 45350, Contract No. AF 30(602)-2235) AD 402 604.

Recent Conferences

IEEE Transactions on Audio and Acoustic, Vol. AU-21, No. 3, June 1973 - Special Issue on 1972 International Conference on Speech Communication and Processing, Newton, Mass., 24-]6 April 1972.

National Telecommunication Conference, 26-28 November 1973, Atlanta, Ga. - Session 29, "Voice Processing for Narrowband Channels," Vol. 2 of IEEE, Cat. No. 73 CHO 86 2NTC.

Other

Arthur D. Little, Inc., USBM Final Report, Contract No. H0122026, January 1974, "Survey of Electromagnetic and Seismic Noise Related to Mine Rescue Communications."

Collins Radio Group, Rockwell International, "Preliminary Mine Wireless Capability Investigation" for Contract No. H0346067, August 1967.

J. L. Flanagan, C. H. Coker, L. R. Rabiner, R. W. Schafer and N. Umeda, "Synthetic Voices for Computers," IEEE Spectrum, October 1970.

G. Raisbeck, "Information Theory," MIT Press (1964).

F. L. van den Berg, "The Lincompex Radio-Telephone Terminal Equipment of the RY741 Series," Philips Telecommunication Review, 30:2, April 1972.

Vocal Interface Equipment Group, Federal Screw Works, 3401 Martin Ave., Detroit, Michigan 48210.

BIBLIOGRAPHIC DATA SHEET	1. Report No. ADL-77229-Task A	2.	3. Recipient's Accession No.

4. Title and Subtitle Technical Services for Mine Communications Research Task A (T.O.1) Applicability of State-of-the Art Voice Bandwidth Compression Techniques for Wireless Mine Communication	5. Report Date Preparation July 1975
	6.
7. Author(s) Richard H. Spencer and Warren G. Bender	8. Performing Organization Rept. No.
9. Performing Organization Name and Address Little (Arthur D.), Inc. 25 Acorn Park Cambridge, Mass. 02140	10. Project/Task/Work Unit No.
	11. Contract/Grant No. USBM-H0346045 Task Order No. 1
12. Sponsoring Organization Name and Address U.S. Department of the Interior, Bureau of Mines Pittsburgh Mining and Safety Research Center 4800 Frobes Avenue Pittsburgh, Pennsylvania 15213	13. Type of Report & Period Covered Final (Task A) May 1974 - June 1975
	14.

15. Supplementary Notes

Initiated under the Coal Mine Health and Safety Program

16. Abstracts The applicability of voice bandwidth compression techniques to mine wireless communication systems is examined. Promise had been attached to such techniques because the lessened bandwidth gave hope for lessened received noise and hence increased operating range or reduced transmitter power requirements. Analog vocoders and digital adaptive predictive coders are examples of real-time techniques examined. For these techniques, as the bandwidth gets narrower, the signal-to-noise ratio required rises, so that expected benefits of reduced transmitter power are not fully achieved. Benefits are achieved only through the use of sophisticated and costly equipment deemed impractical for most mine applications. However, non-real-time bandwidth compression achieved by stretching out the transmission time for a voice signal to several times its original duration can provide a savings in transmitter power but with an attendant time delay.

17. Key Words and Document Analysis. 17a. Descriptors

Bandwidth Compression
Mines-Communications
Thru-the-earth Communications-Electromagnetic
Coal Mines
Vocoders
Mine Wireless Communication
Mines-Coal

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 96
	20. Security Class (This Page) UNCLASSIFIED	22. Price