Federal Communications Commission
Office of Plans and Policy
1919 M Street, N.W.
Washington, D.C. 20554

A Working Paper on:

3 Speech Bandwidth Reduction

November 1980

By: Philip B. Gieseler
John B. O'Neal, Jr.

The FCC Office of Plans and Policy Working Paper Series presents staff research at various states of development. These papers are intended to stimulate discussion and critical comment both within the FCC as well as outside the agency on issues in contemporary telecommunications policy. Titles in the OPP Working Paper Series will include preliminary "think pieces," interim project progress reports, as well as completed research scheduled for publication on broadcasting, common carrier, cable, and other aspects of telecommunications regulation. The analyses and conclusions presented in the OPP Working Paper Series are those of the authors and do not necessarily reflect the views of other members of the Office of Plans and Policy, other Commission staff, or the Commission itself. Upon request, single copies of working papers will be provided as well as an updated list of current and proposed working paper titles. Given the preliminary character and tentative views contained in some titles in this series, it may be advisable to check with authors before quoting or referencing working paper titles in publications. Periodic public notices will announce additions to the Series.

Copies of the OPP Working Paper Series may be obtained from the Office of Public Affairs, Federal Communications Commission, Room 207, 1919 M Street, N.W., Washington, D.C. 20554, (202) 254-7674. Copies are also available for a fee from the Downtown Copy Center, 1114 21st Street, N.W., Washington, D.C. 20037, (202) 452-1422.

For information concerning the content of the OPP Working Paper Series, contact the Office of Plans and Policy, Federal Communications Commission, Room 838, 1919 M Street, N.W., Washington, D.C. 20554, (202) 653-5940. The Office of Plans and Policy will also have a limited number of copies available for distribution.

SPEECH BANDWIDTH REDUCTION

by

Philip B. Gieseler
Office of Plans and Policy
Federal Communications Commission
Washington, DC   20554

and

John B. O'Neal, Jr.
Electrical Engineering Department
NC State University
Raleigh, NC 27650

November 1979

Authors Note: This report is based on information gathered during 1977 through 1979. The theory and principals described in the report are believed to be accurate, although the specific commercial hardware implementations described may have undergone substantial change. An earlier version of this report was previously published by the National Technical Information Service (PB80-103955; Aug. 79).

SPEECH BANDWIDTH REDUCTION

Executive Overview

by Philip B. Gieseler and Raymond M. Wilmotte*

Voice communication by radio generally involves modulating the speech baseband by AM, single sideband, FM or digital means, and transmitting this information over a communications path. For good intelligibility and speaker recognition it is usually considered that the baseband of voice should include the frequencies from about 300 Hz to 3000 Hz. If good quality can be maintained, any reduction of this baseband will reduce proportionately the radio bandwidth required for its transmission.

The present status of analog reduction techniques indicates the possibility of obtaining reasonable intelligibility in about half the baseband of speech. This gives the potential for doubling the number of available channels, but the slight degradation in speech quality would not necessarily be acceptable in all cases. The status of speech baseband compression is still in the laboratory stage or just emerging. None of it has been applied in any large commercial operation.

Over the past decade the overwhelming attention has been given to digital over analog techniques, largely because the Defense Department and AT&T have special needs that digital can best provide. Digital techniques start with a big handicap as regards bandwidth. Eight thousand samples per second and eight bits per sample make for a total of 64 kilobits per second (kb/s). A commonly accepted goal today is to reduce this bit-rate to 2.4 kb/s. While this bit-rate can be achieved, the cost of equipment is large. Another goal is 9.6 kb/s because such a bit-rate can be transmitted on many twisted pair telephone lines. Generally in radio systems, about 1 Hz bandwidth is required for every digital bit per second in order to obtain reasonable immunity from noise. Therefore, insofar as baseband reduction alone is concerned, digital techniques are presently not as useful as analog techniques. When several communication channels are simultaneously in use, however, the digital form of communication can be very efficient using packet switching, which fills gaps in transmissions with material which has been stored for later transmission.

Among the most interesting developments are analysis-synthesis systems. Rather than transmitting the speech signal itself, these systems transmit a description of the speech, which is used to approximately reproduce the original signal at the output. Analysis-synthesis systems have been under development for some time, and so far have characteristically exhibited a machine-like quality in the speech output. It may someday be possible to design an analyzer-synthesizer system that can transmit good quality speech in digital form in a band of 500 Hz or even less. It would have to meet standards of intelligibility, speaker recognition, robustness and other basic requirements.

* Raymond Wilmotte is with the Private Radio Bureau of the Federal Communications Commission.

Recently, there has been a considerable interest in the development of the speech synthesis part of analyzer-synthesizer systems. Texas Instruments has developed one which is sold for children, called "Speak and Spell." The unit says a word, the child spells it out by pushing appropriate buttons and the unit tells him whether he is right or wrong. There are many variations of this toy. There are also products available in which synthesis is used to produce words in different languages. In view of this interest, continued progress may be expected in the development of synthesizers. Current digital designs of analysis-synthesis systems include two types known as linear predictive coding (LPC) and channel vocoders. Both involve very complex circuitry. Both suffer from a lack of robustness in connection with noise introduced with the voice at the microphone. Intelligibility is significantly lost under noisy conditions where normal circuitry would still be operative. This defect may seriously impair the acceptability of these systems for many land mobile operations.

Little is yet known about cost. The prices that are available are for small quantities. Analog systems are quoted in the range of a few hundred dollars, digital in the range of a few thousand. There are strong arguments in favor of the belief that cost will drop radically, if the demand is large. The electronics revolution has made this a real possibility.

For the typical land mobile type of communication path, the best opportunity for baseband reduction at present appears to be analog, although this situation could change in the future. It is still too early to decide which technique provides the best trade-off between quality and price, and what incentives exist for introducing such spectrum saving into the system.

SPEECH BANDWIDTH REDUCTION

I. Introduction and Summary of Conclusions

There are three ways to reduce the frequency spectrum required for a service like mobile radio. They are:

1. Speech Bandwidth Compression.
   If speech signals which normally require a bandwidth of about 3 KHz can be compressed into a smaller bandwidth, the overall frequency spectrum required for a service like mobile radio can be proportionately reduced.
2. More Efficient Modulation Techniques.
   This refers to the use of low bandwidth modulation systems like single sideband to replace modulation techniques like frequency modulation which require wider bandwidth or wider frequency spacing between adjacent channels.
3. Efficient System Design.
   This refers to more sophisticated uses of the time-frequency matrix to allow more users to share the available spectrum. The new cellular mobile telephone system is an example of this design.

This report is a review of the current state-of-the-art in speech bandwidth compression, the first item listed above. The basic conclusion of this review is that techniques currently available for significantly reducing speech bandwidth produce a concomitant degradation in speech quality. This basic conclusion is amplified, analyzed, and supported in the evidence presented in this report.

Speech bandwidth compression techniques are means of reducing the bandwidth needed to represent the human voice waveform. For bandwidth conservation, the voice bandwidth should be the minimum needed for the degree of quality required. In commercial high fidelity broadcasting, the conventionally accepted input bandwidth is 15 KHz, which is necessary to exactly reproduce musical instruments and voice. For applications where lower quality is sufficient, the entire voice frequency band is not required. In mobile communication, an acceptable voice bandwidth is about 3 KHz. Frequencies above 3 KHz contain little overall voice energy, so that they are little needed for applications other than high fidelity, whereas their elimination results in an appreciable savings in bandwidth. More severe bandlimiting lowers the bandwidth further, but sacrifices quality and intelligibility.

The focus of this study is on minimizing the bandwidth of the voice signal used for mobile and other voice communication systems. If techniques exist for significantly reducing this 3 KHz bandwidth while maintaining sufficient quality, the radio frequency bandwidth could be proportionately reduced. Technology is available for reducing the 3 KHz voice bandwidth by a factor of 10, but at this stage this technology is not useful because a desirable level of performance and cost have not been achieved. The 10 to 1 reduction in bandwidth can be accomplished by a device known as a "vocoder", invented in the 1930s (1) which at that time was relatively complex (therefore expensive) and lacked good voice quality. Vocoders are further described in Chapter III.

Experts in the communications industry do not completely agree whether the most efficient techniques for transmitting speech are analog or digital.

However, most activity today in speech compression is concentrated on digital techniques which provide bit rate reduction rather than bandwidth reduction. There are three reasons for this. First, most researchers are convinced that in the telecommunication world digital transmission and switching will ultimately be more economical than analog transmission and switching. Furthermore, most people doing research on speech compression are employed either directly or indirectly by the telecommunications industry. Secondly, analog speech compression has been studied for over 50 years with little tangible results. Most researchers do not feel it is possible to design speech compression systems which have a bandwidth less than about 3 KHz and which have the requisite speech quality. Finally, it has become easier to do research on digital compression rather than analog compression. Analog compression tends to require sophisticated expensive special purpose hardware whereas research on digital compression is generally done on general purpose computers which can be programmed to perform the compression algorithms.

Digital techniques can have additional advantages such as high security, reliability and ease of multiplexing with other speech or data channels. These advantages appear to make digital well suited for applications such as the military, point-to-point microwave, satellite communication, and short haul telephony. But for a single two-way communications link such as is generally used for mobile radio, the least bandwidth and the lowest cost can presently be obtained with analog techniques. No inherently digital techniques have been found that result in less bandwidth for the same quality as analog although this situation could change in the future. Also important is the spectrum efficiency of digital compared with analog, i.e., users per unit area per MHz. Digital voice can be transmitted in a very small bandwidth when many levels instead of the simple on/off two level system is used, but the ability of this type of signal to withstand noise and interference is comparably reduced, netting little or no improvement in spectrum efficiency. Although overall spectrum efficiency was not analyzed as part of this report, analog appears superior to any known low-cost digital system in bandwidth required and in spectrum efficiency. Again, this situation could change in the future, and applies only to a single two-way communications path. It is likely that there may be other configurations in which digital techniques would be superior.

Following is a summary of the principle results of this study:

1. At this time analog transmission uses less bandwidth and seems more appropriate for mobile radio than digital transmission, even though the cost of digital devices and equipment is falling precipitously.

2. A compression system which significantly reduces the bandwidth of a speech signal below about 3 KHz without degrading the quality or required robustness of the signal has not been demonstrated.

3. Speech compression systems have been demonstrated which reduce the required bandwidth by about one half. Usable speech with good intelligibility is produced under ideal conditions. But the current state of research does not indicate whether quality speech transmission below about 3 KHz will ever be a reality for practical speech transmission channels.

4. There is an unavoidable tendency for a compressed speech signal to be more susceptable to noise and distortion than an uncompressed signal. Since speech transmission systems must always operate in a noisy environment reducing the bandwidth of the speech signal is not necessarily advantageous.

5. Currently most research on speech compression is concentrated on digital transmission rather than on analog transmission. Researchers involved in digital speech compression attempt to lower the bit rate needed to transmit speech over digital transmission lines. The communications industry obviously feels that there is some payoff for research on digital bit rate compression but very little for research on analog bandwidth compression.

6. For digital transmission linear predictive coding and vocoder techniques can be used to produce intelligible but synthetic speech at bit rates as low as 1 kb/s. Such systems operating at 2.4 and 4.8 kb/s are now in use for special military applications.

7. Speech bandwidth and bit rate compression systems are commercially available now. Table 1 of Appendix A compares several commercially available speech compression systems. This table gives the bandwidth or bit rate of each system and evaluations of intelligibility and cost.

II. Speech Quality

The bandwidth or bit rate needed to transmit a speech signal is directly proportional to speech quality. High quality speech generally requires high bandwidth or bit rate. Table 1 lists some categories often used to describe the quality of speech communication systems (19). At the top of the list is high fidelity, the quality we normally associate with modern stereo equipment. The bandwidth of 20-40 KHz assumes one (20 KHz) or two (40 KHz) channels. Program quality is the quality of the signal delivered to radio stations for broadcast. A good AM radio will often produce this quality in your home. FM radio is capable of producing speech quality somewhere between program quality and high fidelity. Toll quality is that quality provided by modern telephone carrier systems under favorable conditions. The majority of long distance telephone calls probably do not achieve toll quality but the communications carriers are aiming for toll quality on all such calls. Communications quality is that quality achieved by mobile and CB radio channels. Synthetic quality describes unnatural speech sounds characteristic of computers which have been programmed to generate speech.

The quality designations in Table 1 are not meant to be exact specifications of quality. Speech quality (like almost any quality measure) is a continuum from synthetic and below to high fidelity and beyond. The signal-to-noise ratios and bandwidths shown are approximate attributes of each quality designation. These attributes have been omitted from synthetic quality because they have little meaning here.

Table 1. Descriptions of Speech Quality

Quality	Approximate Speech Bandwidth	Approximate S/N ratio	Example
High Fidelity	20 - 40 kHz	60 dB	high quality stereo equipment
Program	5 kHz	50 dB	AM radio
Toll	3 kHz	30 dB	good long distance telephone call
Communications	3 kHz	20 dB	CB radio
Synthetic	--	--	computer generated speech

The quality of a speech communication system is determined by many attributes. The most important of these are listed as follows:

1. Intelligibility. Intelligibility reflects the ability of a system to transmit understandable words and sentences. There are several standard ways to measure intelligibility. For example, the diagnostic rhyme test detects the capacity of a communication system to transmit sounds which allow a listener to distinguish between words that sound very nearly alike. In most applications high intelligibility is a necessary but not sufficient condition for a usable speech communication system.

2. Speaker recognition. This is the ability of a listener to recognize the voice of someone he knows. This is a very important attribute of military communication systems. All speech qualities in Table 1 except the synthetic quality provide very good speaker recognition.

3. Naturalness. This is closely related to speaker recognition. It is the attribute of sounding natural - like a human utterance, not machine-made. Synthetic quality does not preserve naturalness but all other quality designations in the table do.

4. Lack of noise and distortion. Background noise and distortion of the speech signal can degrade the speech quality of any system. Noise is usually perceived as a constant hissing sound while distortion may produce a rasp-like sound or echos during speech but disappears when the speech is not present.

5. Frequency content. This is another way of saying high voice bandwidth but in this context frequency content means the bandwidth of the speech signal that the listener hears. This may be quite different than the transmission bandwidth required by the communication system. Speech contains frequencies as high as 8 - 10 KHz and music contains frequencies considerably higher. The frequency content of a speech signal may be lowered by removing selected frequency bands within the speech bandwidth as well as by lowering the maximum frequencies transmitted.

6. Robustness. This is that property of a transmission method which makes the resulting signal impervious to various forms of interference and distortion which occur in the transmission medium. The opposite of a robust or forgiving signal is a fragile one. Generally, any technique which reduces the bandwidth or bit rate of a signal tends to make that signal less robust.

7. Echoes. Echoes are the presence of audible sounds which echo the original sound. They are a form of distortion usually caused by impedance mismatching in telephone systems. Echoes can also give rise to a hollow sound in speech often described as "speaking in a barrel".

8. Interference. This degradation is caused by unwanted signals getting into the wrong frequency band or channel. The sources of interference are usually man-made. It may be perceived as a tone, noise, buzz or clicks.

9. Crosstalk. This is speech, often intelligible, from another talker whose signal gets into the wrong channel or frequency band.

Obviously, the intended use of a communication system determines the speech quality that should be provided. Also important is the speech quality that a user has come to expect. Good intelligibility while it is an obvious requirement for any speech communication system is never enough. This is one reason why vocoders, while useful for some purpose, have never been widely used. People want higher quality - they don't necessarily need higher quality for all applications.

III. Speech Bandwidth Reduction

Although a large number of bandwidth reduction methods for speech have been patented, none has ever been extensively used and few seem to have any promise at all. With the "digital revolution", serious research on analog speech transmission systems has taken a back seat to digital processing of speech signals. Just as the analog computer has given way to the digital computer, analog speech processing has given way to digital speech processing. In this section we describe three of the potentially promising methods of bandwidth compression and then discuss why it is so difficult to compare the quality of such systems.

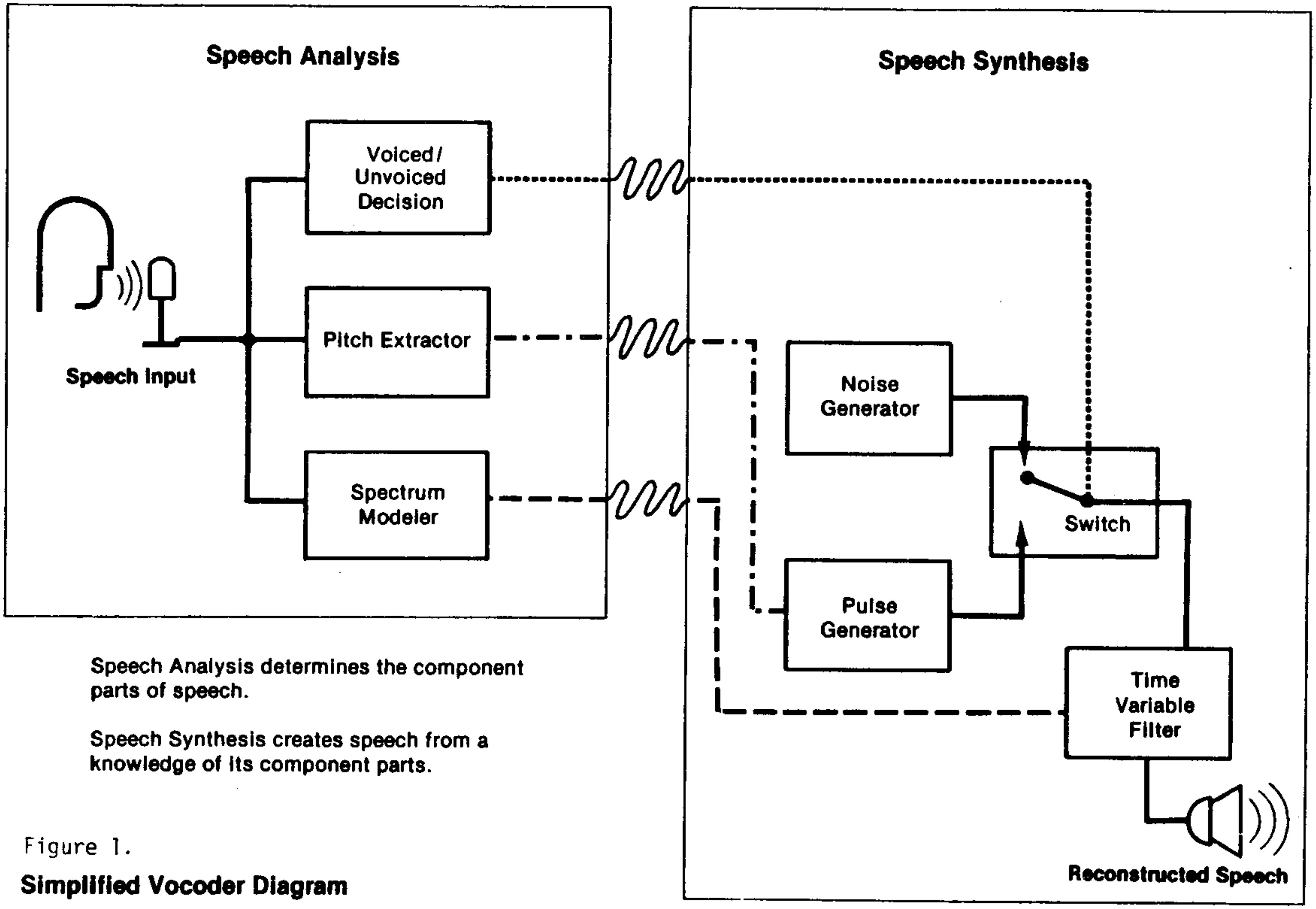
A. Catalog of Speech Bandwidth Reduction Methods

Bandwidth reduction methods introduce approximations and eliminate unneeded information in the speech signal. Bandlimiting the voice signal to eliminate frequencies above 3 KHz is a simple bandwidth reduction method. The possibility exists for other techniques to even further reduce the necessary voice bandwidth. A catalog of some of these techniques follows.

In some cases, methods can be combined and used together. General techniques will be described in this section; some specific commercial implementations will be described in Appendix A.

1. Speech analysis and synthesis (vocoding)

The vocoder is probably the best known analysis synthesis (2) device. Most vocoders now use digital transmission but analog transmission as originally used by Dudley is quite feasible. A vocoder communication system consists of an analyzer at the input, a low bandwidth transmission path, and a synthesizer at the output. This is shown in Figure 1. The analyzer section determines if



Speech Analysis determines the component parts of speech.

Speech Synthesis creates speech from a knowledge of its component parts.

Figure 1.

Simplified Vocoder Diagram

the instantaneous speech input is voiced or unvoiced,* and determines the shape of the speech spectrum. This information is transmitted to the synthesizer section in analog or digital form. The synthesizer consists of a pulse generator to represent the voiced speech, and a noise generator to represent the unvoiced speech. Proper pitch is maintained by the pulse generator. The noise and pulse generator outputs are fed through a variable filter whose parameters are set to duplicate the original speech envelope. In combination, the noise generator, pulse generator and variable filter attempt to produce a duplicate of the original speech. Several different types of vocoders have been developed, some of which are described in Appendix B. Their differences usually involve the methods employed in modeling the vocal tract.

A large potential exists for voice bandwidth reduction using analysis-synthesis techniques, since information about voice characteristics can take up much less spectrum space than the voice itself. The disadvantages of vocoders are that they are quite expensive, produce synthetic machine-like speech and are sensitive to noise in the transmission medium. Although vocoders have been under virtually continuous experimentation for over 35 years they have never been extensively used except for certain military applications. Recently, a number of commercial products have appeared that use voice synthesis. These include educational toys, language translators and home computers that speak to the user. Thus a portion of the analysis-synthesis function is available to the mass market, which provides potential for large price decreases. Vocoders are further described in Appendices A and B and in (2).

2. Frequency Band Elimination

In simple frequency band elimination some of the frequency spectrum in the middle of the 3 KHz band is simply thrown away. The frequency components retained can be transferred into the frequency gaps created and the bandwidth of the speech signal thereby compressed. For example, one system called "frequency companding" by its inventor (14) begins with a 3100 Hz speech signal. The energy in the bands from 350 to 600 Hz and from 1500 to 2500 Hz is retained while other frequencies are discarded. The energy from 1500 to 2500 Hz is then moved down by appropriate modulation so that it occupies the frequencies from 600 to 1600 Hz. The resulting speech signal is now contained in the frequencies from 350 to 1600 Hz. The bandwidth of the original speech has therefore been approximately halved. Some guard bands are required between the low and high bands. At the receiver, the energy in the 600 - 1600 Hz band is translated back up to the 1500 - 2500 Hz region. The resulting signal contains energy in the bands 350 - 600 Hz and 1500 - 2500 Hz. Other frequencies have been eliminated. The missing frequencies (600 - 1500 Hz in this example) cause the speech to sound somewhat hollow - often described as the "speaking in a barrel" sound. The speech is, however, quite intelligible. This type of frequency elimination has been used for many years in the telecommunications industry to create fre-

* Speech is composed of voiced sounds and unvoiced sounds. Voiced sounds occur when the vocal chords are vibrating and correspond roughly to vowels and nasal sounds. In unvoiced sounds the vocal chords are not vibrating - typical unvoiced sounds are the fricatives "sss's", "sh..." and "f", which occur in words like "six", "shoot", and "fish", respectively.

quency space in the speech band which could be used for data transmission or signalling (23).

Another idea for frequency band elimination takes advantage of the fact that the voiced and unvoiced components tend to be separated in time and frequency (21). Voiced sounds have most of their energy in the spectrum between 1,00 and 1,000 Hz. Unvoiced sounds have much of their energy in the spectrum between 1,000 and 3,000 Hz. A voiced and an unvoiced component will generally not occur at the same time (just as vowels and consonants occur at different parts of a word). Thus, little relevant information is contained in the spectrum 100 to 1,000 Hz when an unvoiced sound occurs; and little relevant information is contained in the spectrum from 1,000 to 3,000 Hz when a voiced sound occurs. Techniques that take advantage of this time and frequency separation can potentially reduce the voice bandwidth.

Variable sub-band coding is the name given to a dynamic frequency band elimination scheme which is currently in the experimental stage (20). Although this technique is proposed for use primarily with digital coding of speech, it is basically an analog technique. In variable band coding the speech energy in several frequency bands is analyzed continuously. Only energy in the bands which contain the most significant energy is transmitted. Means must be provided to inform the receiver which frequency bands are being transmitted. This technique is similar to sub-band coding and to adaptive transform coding, schemes proposed for efficient digital encoding discussed in the next chapter.

3. Frequency compression-expansion

The previous technique is linear in that, strictly speaking, frequencies are not compressed, but some frequencies are simply eliminated from the output. Frequency compression-expansion is a non-linear technique which compresses all frequencies and eliminates none. This technique has a counterpart in the time domain called an amplitude compandor. (Frequency compression-expansion is not called frequency companding here because some of the recent literature uses this term to describe a frequency band elimination technique described in the previous section.)

Amplitude compandors compress amplitude levels before transmission and expand the amplitudes at reception. Frequency compression-expansion similarly compresses the frequencies at the transmitter and expands them at the receiver. For example, the normal voice bandwidth for radio transmission is about 200 Hz to 3,000 Hz, or roughly 3 KHz wide. If frequencies could be easily divided and multiplied by 2, bandwidth could be halved. The 200 to 3,000 Hz voice spectrum could be converted to 100 to 1,500 Hz, transmitted, received, and reconverted to 200 to 3,000 Hz. The voice bandwidth during transmission would only be 1,500 Hz wide.

One technique called rooting operates by taking the square root of frequencies at the transmitter and then squaring the frequencies at the receiver. Although proposed in many forms, this type of frequency compression-expansion has never proved to be a useful speech bandwidth compression technique.

Another type of frequency compression-expansion, which we shall call speech gapping, operates in the time domain. Observation of sustained vowel sounds reveal quasi-periodic signals - signals which repeat themselves in almost identical cycles. This is shown in Figure 2 for the phrase "we pledge". The "e" sound is made up of about 20 nearly identical pitch periods. The "g" sound is similar to random noise of no particular pattern. If these signals are chopped into properly-sized intervals, half of the intervals discarded and the remainder transmitted, a receiver can reconstruct usable speech by simply repeating the intervals that are there to fill in the absent gaps. This technique is often explained by numbering the speech signal intervals 1, 2, 3, 4, 5, ... The transmitted speech is the sequence 1, 3, 5, 7, ... The reconstructed speech at the receiver is 1,1,3,3,5,5 .... Basically this technique creates time gaps - the transmitted signal is absent half of the time. These time gaps can be translated into a bandwidth savings by stretching out the signal in each of the transmitted intervals so that it covers the gaps created. A savings of one-half in the time domain can be translated into a bandwidth savings of one-half. Alternatively a second channel can be time division multiplexed into the gaps so created allowing two speech signals to occupy the bandwidth normally required for one signal.

The speech resulting from this technique is intelligible but contains noticeable artifacts resulting from the gapping.

B. Comparison of Speech Bandwidth Reduction Methods

The bandwidth reduction methods described in the previous section are very difficult to compare. Since bandwidth reduction has not been considered to be feasible by the large communication companies, they have not made meaningful comparisons between such systems. Interest in these bandwidth reduction techniques has been sustained by a few small companies who are oriented toward producing a product rather than providing a service and who have not carefully compared the quality of various forms of these bandwidth reduction methods under practical operating conditions.

Comparison of these systems has meant listening to audio tapes or perhaps actual transmissions set up and controlled by the proponents of each system, or perhaps comparing the results of intelligibility tests. A quality rating for these bandwidth reduction methods is inextricably tied to their implementation, the noise in the system and the particular speakers used in testing. Since it is so difficult to ascertain the potential of these methods, we do not speculate on their quality here. Instead we give some subjective and tentative evaluation of some specific commercially available systems in Appendix A of this report.

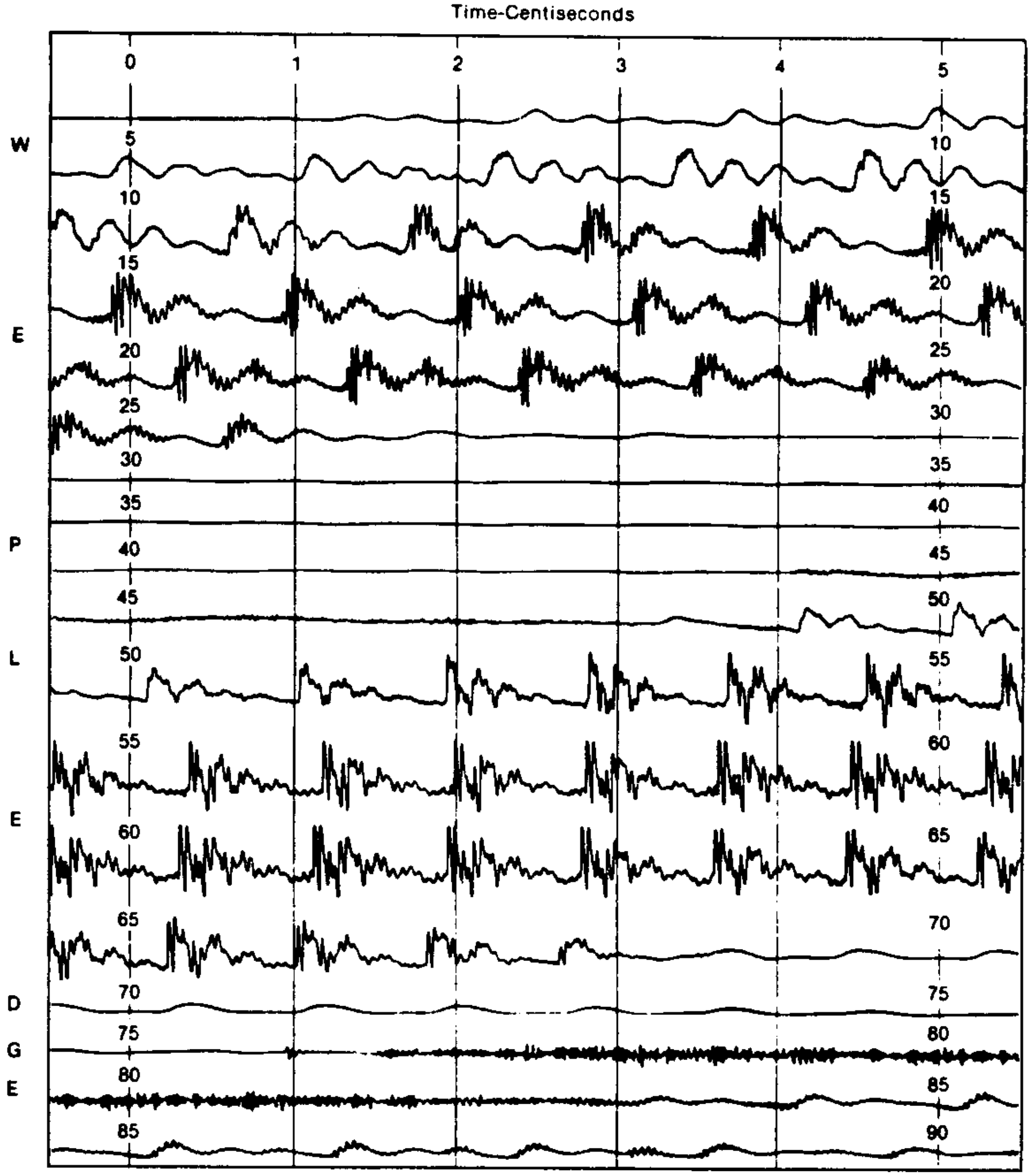


Figure 2.

Speech waveform for the phrase "We pledge...." From Ref. (22).

IV. Speech Bit Rate Reduction

Since T1 carrier, the workhorse short haul PCM carrier system, was introduced in the early 1960s there has been a continuously increasing interest in bit rate reduction methods for speech signals. The T1 carrier terminals encode 24 speech signals into a 1.544 mb/s bit stream and the T1 repeatered lines carry these digital signals over ordinary (but usually treated) twisted pairs of telephone cables for distances of usually less than 50 miles. If the bit rate required for each speech signal could be reduced, these T1 carrier lines could carry more than 24 channels and considerable economy would result.

Digital transmission has not proved to be profitable over long haul telecommunications facilities on the surface of the earth or in undersea cable. It does not, at this time, seem to be applicable where bandwidth is a serious constraint even though satellite communications systems often use digital transmission. Digital transmission does not seem to be a serious contender for mobile radio or standard broadcast services.

The advantages of digital transmission are several. Digital communication may be retransmitted thousands of times with insignificant loss of quality. Some digital methods are extremely tolerant to noise and interference, which means that digital can use less power than analog for the same operating range. Security and full privacy are easy to maintain. Digital voice can be handled on a routine basis over existing data lines, just like any other digital information.

The bandwidth of digital voice systems is influenced by two opposing factors. First, digital inherently increases the bandwidth of voice signals. Digital obtains its tolerance to noise and interference by utilizing a large bandwidth. Second, digital processing is available for reducing speech bandwidth whereas comparable techniques are unavailable or cumbersome for analog processing. At present, analog techniques require less bandwidth than digital techniques for the same quality. This situation is unlikely to change in the immediate future.

The most straightforward digital technique is pulse code modulation (PCM). The standard PCM system used by the telephone companies requires 8,000 samples per second. Each sample needs about 8 digital bits to accurately represent it. Thus PCM uses 64,000 bits per second (3). Since the transmitting system usually puts roughly one digital bit into every Hertz of bandwidth, PCM requires 64 KHz for one voice channel. This is a tremendous expansion compared to FM at 16 KHz, AM at 6 KHz and single sideband at 3 KHz. The bandwidth required by PCM and other digital techniques can be reduced by using a transmitting scheme able to pack more bits into each Hertz of bandwidth, but this results in a signal more susceptible to noise and interference (24).

PCM is an example of the first point above, the expansion inherent in digital. Many digital encoding techniques illustrate that digital voice may be easily processed to lower the bandwidth requirements (4). Delta modulation (DM) compares the amplitudes of each sample with the amplitude of the previous sample. The difference between the two is coded. This requires a lower bit rate than encoding each sample independently as in PCM. An extension of delta modulation, continuously variable slope delta modulation (CVSD) additionally processes the rate of change of amplitudes. CVSD operates with good intelligibility at a bit rate of 16 kilobits per second. This is a bit rate reduction of 4 to 1 over PCM, but, at one bit per Hertz, is just beginning to approach the bandwidths often used for FM. Other digital techniques that use lower bit rates are available, some of which are detailed in what follows.

Robustness is an important factor for digital systems. PCM, delta modulation and CVSD are robust techniques in that they can operate in relatively noisy environments. Techniques that operate at the lower bit rates, 4.8 kilobits per second and less, typically are less robust.

A. Catalog of speech bit rate reduction methods

During the past 30 years many systems have been proposed for reducing the bit rate required to encode speech signals. This activity has been especially active in the past 10 years. The precipitous drop in the cost of processing signals digitally has fueled an intensive search to find ways of compressing the bit rate required to transmit high quality speech signals.

The primary method used to encode speech signals today is Pulse Code Modulation (PCM). Standards in the telecommunication industry call for a 64 kb/s bit rate. PCM is not a bit rate compression system but its standard bit rate and the resulting high quality produced have become the standard of compression for bit rate reduction systems. PCM was invented by A.H. Reeves of AT&T in 1938 (5), was illucidated by Shannon et.al. in 1948 (6), and was put into service by the telecommunications industry in the early 1960s.

In what follows some of the more important digital encoding systems are outlined. Most systems described operate on a single speech signal. However, two systems are cataloged--digital TASI and DSI--which reduce the bit rate by operating on many speech channels and take advantage of the fact that, even when in use, a speech channel is inactive more than half of the time. Strictly speaking, these two techniques belong to the catagory of "Efficient System Design" mentioned in the introduction.

1. Pulse Code Modulation (PCM)

PCM is the most common type of digital encoding system for speech. It is widely used by the telephone companies in the U.S. and abroad. In the U.S. more speech circuits use PCM than analog carrier systems. With present standards the speech signal is sampled at 8 KHz and each sample value is encoded into an 8-bit companded word resulting in a bit stream of 64 kb/s. This is not a bit rate compression system. Rather it is the system to which all bit rate compression systems are compared. The S/N ratio provided by a single PCM encoding is about 38 dB. The noise is therefore imperceptible. PCM systems in which the step size adapts or adjusts itself to the strength of the speech signal are called Adaptive PCM (APCM).

2. Differential Pulse Code Modulation (DPCM) (4)

DPCM takes advantage of the fact that when sampled at 8 KHz, each sample value can be accurately predicted by the previous sample value. In simple DPCM only the difference between adjacent sample values is encoded and transmitted. In more sophisticated DPCM systems, a sequence of past sample values is used to form the estimate of the current sample value and the difference between this estimate and the actual sample value is encoded and transmitted. Practical DPCM systems are Adaptive DPCM or ADPCM systems in which the quantization levels or step sizes adjust or adapt to the input signal level. ADPCM systems are always better than PCM for speech but, unfortunately, not for voice band data and other non-speech-like signals which are present on the telephone network. When only speech is to be transmitted ADPCM systems operating at 32 - 40 kb/s give speech quality which most observers find identical to 64 kb/s PCM - the quantizing noise cannot be detected by the normal user.

3. Delta Modulation (DM) (7)

DM by far is the simplest digital encoding system. It is a form of DPCM in which the difference between adjacent sample values is encoded into a one-bit word. The sampling rate is equal to the bit rate. Practical systems have an adaptive step size similar to ADPCM and are designated as ADM. ADM systems generally operate at bit rates from 16 to 32 kb/s. At 32 kb/s the quantizing noise is virtually inaudible. AT 16 kb/s the noise is disturbing but does not impair intelligibility.

4. Digital Vocoders (4)

Although some present day vocoders are analog, most are designed to transmit digital signals. Vocoders are the most senior of the bandwidth conservation techniques. They were invented in the 1930s. A vocoder analyzes the speech signal and extracts the information necessary for its synthetic reconstruction at a distant location. Therein lies their disadvantage--vocoders produce speech with a computer-like synthetic quality. In analyzing the speech signal the vocoder is attempting to determine an electrical analog for the geometrical configuration of the human speech production mechanism and of the vocal vibrations which excite it. This requires the use of models of the speech production mechanism. These models are inevitably imprecise and the resulting speech never sounds exactly like the person talking. There are many types of vocoders--channel, autocorrelation, formant, cepstral or homomorphic, voice excited, linear predictive and phase vocoders to name a few. Vocoders can operate at the very low bit rate of 1 kb/s and perhaps even lower. Intelligibility is good but speaker recognition may be impaired especially for the lower bit rate systems.

5. Linear Predictive Coders (LPCs) (4)

LPC is a type of digital vocoder implementation that has been gaining popularity in the last several years. These coders are complex versions of ADPCM in which the predictive algorithm as well as the step size adapts to the speech signal. In LPC the vocal tract model is based on the principle that speech can be reasonably predicted by weighing the sum of previous speech samples. This involves solving a set of linear equations to obtain "predictor coefficients". LPC codes are generally expensive and complex and generally operate at the bit rates from one to eight kb/s. Like other vocoders they may have a synthetic quality about them although intelligibility and speaker recognition are generally preserved.

6. Tree Encoding (8)

This is the most complex encoding system in this catalog. Also called delayed encoding, this method contains considerable encoding delay. Coding involves following complex tree structures to obtain the best quantization levels. Theoretically tree encoding along with ADPCM or transform coding can be shown to be an optimum encoding method for certain well-behaved classes of signals. It is doubtful whether optimum tree encoders will ever be practical for speech. However, simple suboptimum application of the principles may prove useful in conjunction with other coding techniques.

7. Sub-band Coding (SBC) (9)

In SBC the speech frequency band is divided into four to eight sub-bands by a bank of bandpass filters. Each sub-band is low-pass translated to zero frequency by a modulation process equivalent to single-sideband modulation. Each band is then sampled at its Nyquist rate (twice the width of the band) and digitally encoded with an adaptive-step-size PCM (ADPCM) encoder. In this process, each sub-band can be encoded according to perceptual criteria that are specific to that band. On reconstruction, the sub-band signals are decoded and modulated back to their original locations. They are then summed to give a close replica of the original speech signal.

The SBC techniques are relatively complex to implement. They seem to provide superior coding quality in the region from about 16 to 24 kb/s.

8. Transform Coding (10)

Transform coding, like sub-band coding, is a frequency domain technique. A mathematical transformation, like a Fourier, cosine or Hadamard transform, is first applied to segments of the speech signal. Then the transformed signal is quantized and transmitted. The primary advantage of this technique is that the system can be adaptive so that transform coefficients which are unimportant can be omitted and not transmitted at all. This has the effect of transmitting only those frequencies which are present in the speech signal. Eliminating these unnecessary frequencies is a dimensionality reduction procedure which is impossible in DPCM techniques as they are presently understood. In speech coding the Discrete Cosine Transform is usually used and the system has an adaptive bit assignment procedure. It's called Adaptive Transform Coding (ATC).

9. Nearly Instantaneous Companding (NIC) (12)

This is a form of adaptive PCM in which the step size is transmitted roughly every 12 samples. This technique can be used to reduce the bit rate by a factor of about 1/4 while retaining 8-bit companded PCM quality.

10. Digital Time Assignment Speech Interpolation (TASI) (13)

This is a technique used to reduce the bit rate on multi-channel communication links. TASI has an analog version and a digital version. Both operate on the same principle. In a two-way telecommunications channel each user is speaking only about 40 percent of the time and therefore requires a transmission channel only during this period. During the remaining 60 percent of the time he is pausing or listening. If the presence of speech can be reliably detected then a large number of users can share a fewer number of channels. In digital TASI the available bit rate is shared among those who need it at any given moment.

11. Digital Speech Interpolation (DSI) (11)

This is a specific application of the TASI technique to multi-channel speech systems using digital transmission. A voice switch is used to determine which of the users is speaking. Only those users who are actually speaking share the available digital transmission capacity of the line. Their speech is efficiently encoded and transmitted. Those users who are listening or pausing between words are not using any of the line's capacity. A bit rate reduction of typically 2:1 can be achieved with DSI systems. There are many techniques used to accomplish the DSI function.

12. Hybrid Techniques

We describe as hybrid techniques, those encoding systems which use two or more of the above coding algorithms. For example, ADPCM/TASI techniques are the subject of ongoing experiments. DSI systems are often forms of PCM/TASI. DPCM/transform coding techniques have been used for image signals and may be applicable to speech.

B. Comparison of speech bit rate reduction methods

Comparisons of the bit rates of various coders are shown in Figure 3. At the high quality end of the scale PCM is currently universally used. The strength of PCM is that it furnishes high quality for any signal--not just speech. Telecommunications channels must carry a wide variety of voice band data, facsimile and telemetry signals which are not speech-like. PCM encodes all of these signals equally well. The other coders listed in the figure are tailored for speech signals and do not operate well with other types of signals.

The ADPCM encoders operate well in the region from above 32 kb/s and at about 40 kb/s furnish toll quality. So far ADPCM coders have not been widely used. The ADM coders operate from about 16 kb/s where they are quite noisy to 40 kb/s where the noise is imperceptible. ADM has found use in multi-channel

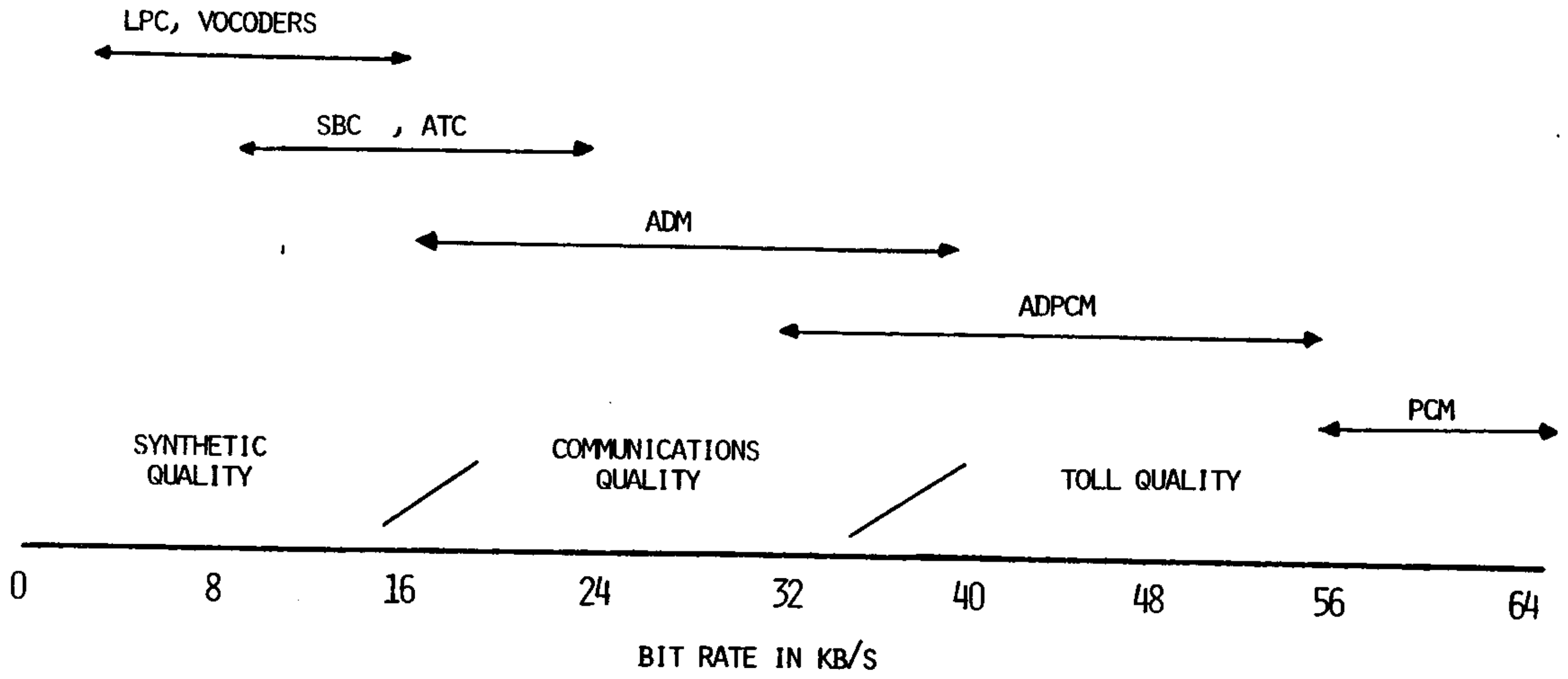


FIG. 3 BIT RATES OF SPEECH ENCODERS

Abbreviations

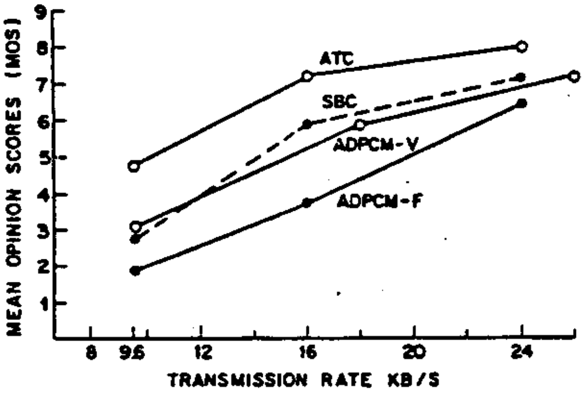
LPC	- Linear Predictive Coders
SBC	- Sub Band Coders
ATC	- Adaptive Transform Coders
ADM	- Adaptive Delta Modulation
ADPCM	- Adaptive Differential Pulse Code Modulation
PCM	- Pulse Code Modulation

rural carrier systems in telephony and in certain satellite communication systems. The SBC and ATC operate best in the region from about 12 kb/s to 24 kb/s. AT 12 kb/s they tend to be noisy and somewhat synthetic. About 24 kb/s their quality is quite good but simpler coders such as ADM and ADPCM can furnish roughly the same quality and are cheaper. Neither SBC and ATCs have been used for anything but experimental purposes. Below about 12 kb/s it has been impossible to achieve anything but a very noisy signal (ADM) or the synthetic quality furnished by LPCs or vocoders. These have had some application to military communication systems where encription and digital transmission over modems operating at 2.4 - 9.6 kb/s is used. The current use of LPCs and vocoders is probably dependent upon this need for secrecy. These coders furnish satisfactory but synthetic speech.

A comparison of some low bit rate wave form coders is shown in Figure 4,
taken from reference (13). This figure shows a comparison of an Adaptive
Transform, sub-band coder SBC and two types of adaptive differential pulse
code modulation systems--ADPCM-V with a variable predictor and ADPCM-F with
a fixed predictor. The figure plots mean opinion scores of these coders as
rated by 20 untrained listeners. The listeners were asked to rate the quality
on a scale from 1 (poor) to 9 (excellent). The averaged scores show that the
ATC system is consistently judged to be superior in quality to the other coding
systems for every bit rate shown. As the bit rate and quality increases, how-
ever, the perceived quality difference between the coders decreases. The super-
iority of the ATC system is in its ability to distribute the quantizing noise
judiciously over the frequency band and to dynamically and selectively elimi-
nate frequency bands which are not important in the speech signal. This
ability becomes less of an advantage as the bit rate increases.

The next figure, Figure 5, is an extension of Figure 4. It represents
a speculation of the subjective quality of some of the more important types
of coders over a wider range of bit rates than are shown in Figure 4.
coders over a wider range of bit rates than are shown in Figure 4. Each
shown is optimum for some set of bit rates. The precise crossover points are
a matter for speculation.

Figure 6 shows how the coders rank in cost. Only relative costs are shown.
The ADM system is by far the cheapest. The cost of the PCM and the ADPCM sys-
tems is dependent of whether or not the coders are used as single channel or
multi-channel coders. PCM especially can be quite inexpensive in multi-channel
applications such as T1 carrier because one coder can be shared over a large
number of channels.



Mean opinion listening scores
of four speech encoders

Fig. 4

Taken from Ref. [13]

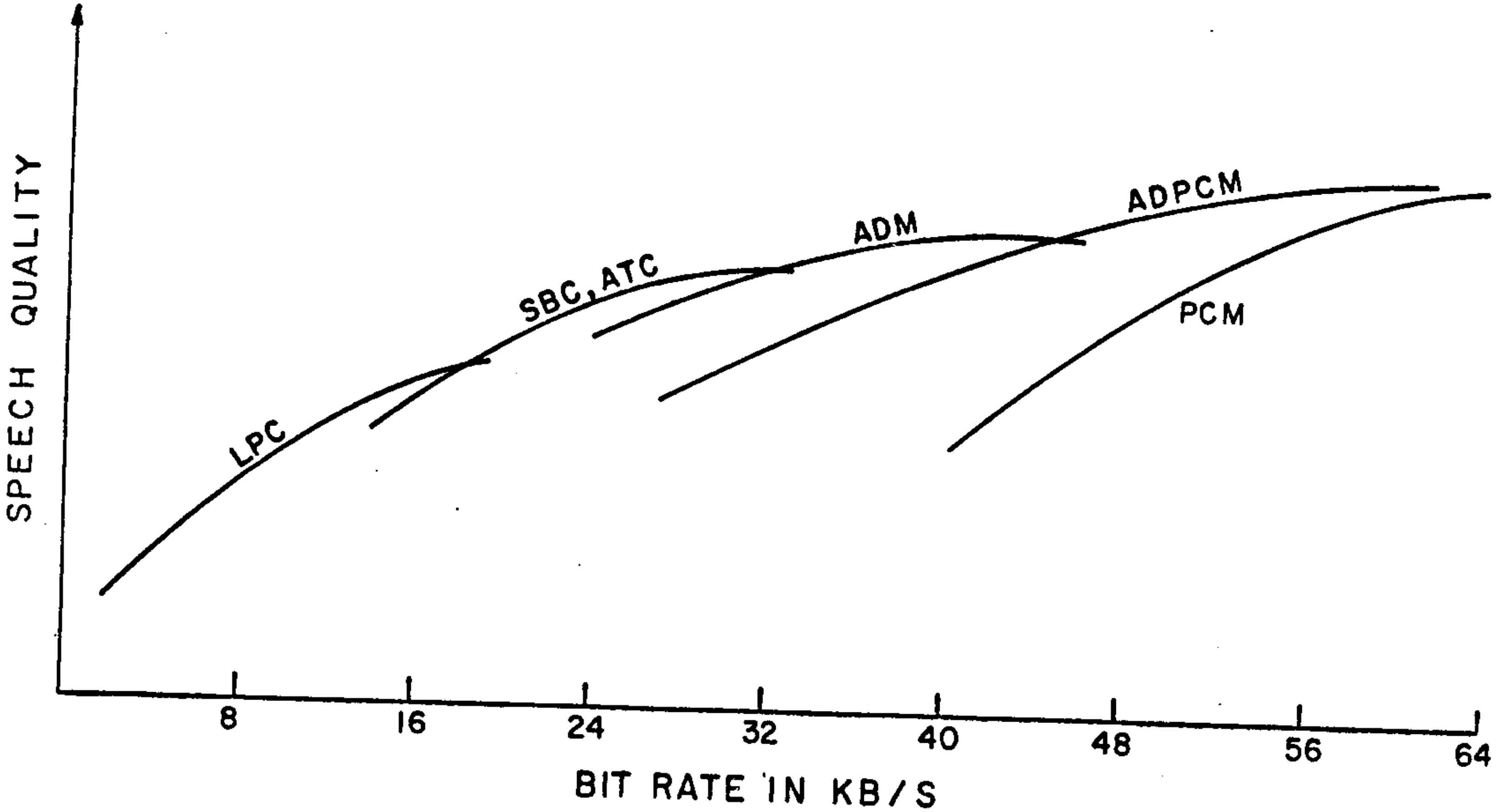


Figure 5.

APPROXIMATE QUALITY VS RATE FOR SPEECH ENCODERS

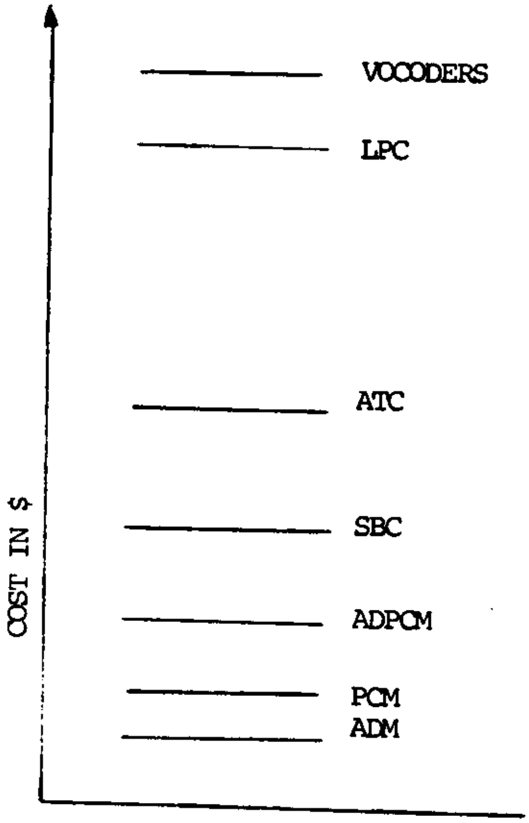


Figure 6. Relative cost of speech encoders

APPENDIX A

Notes on Commercially Available Reduced Bandwidth and Bit Rate Systems

This section contains a description of some commercially available bandwidth and bit rate reduction systems. Some information in these notes has been obtained from equipment manufacturers who may have unwarranted high opinions of their own products. No attempt has been made to verify information regarding cost and performance of this equipment. Four analog systems with bandwidths from 1000 to 2100 Hz and two digital systems with bit rates of 2400 to 4800 bits/sec are covered here. These systems should be considered to be representative of those available rather than a complete listing. Although the ensemble considered here contains more analog than digital systems, there are actually more digital than analog systems commercially available.

Table A-1 contains a list of the systems considered, their bandwidth or bit rate, estimates of their intelligibility and their approximate cost. Information in this table was furnished by the manufacturers or was approximated using the best available information. Cost figures are merely estimates of the cost to manufacturers and should not be treated as precise. The user cost is judged to be about three times the cost to a manufacturer, to allow for overhead, assembly and profit.

At the present time none of the analog systems have been used extensively by those organization furnishing communications services or equipment. The digital systems have been used by the military in situations where digital encryption and subsequent digital transmission over a modem is required.

It is difficult to speculate on the future of the analog systems. It is not known whether users will accept the performance provided by them and it is not known how these systems will operate over real transmission channels. The speech quality of these reduced bandwidth systems is inferior to the performance of other full bandwidth speech channels.

It is fairly well recognized that the digital systems produce speech that is usable but inferior to full bit rate (64 kb/s PCM) digital channels.

A. The Harris System

VBC, Inc. was formed in 1977 for the purpose of developing and marketing an invention for reducing voice bandwidth. Their implementation, the Harris system, uses both frequency and amplitude companding (14). The frequency companding technique compresses bandwidth by electronically folding high frequencies (which contain unvoiced sounds). At reception the spectrum is "unfolded" to reproduce the original waveform. By proper choice of parameters, the distortion introduced by the folding and unfolding procedures can be minimized. Amplitude companding, used in other communications systems for many years, is not used in the Harris system for speech bandwidth compression. Instead it is used to reduce interference and noise on weak signals. The amplitude and frequency companding are apparently independent of each other and can be used separately.

Table A-1. SUMMARY OF SOME COMMERCIALLY AVAILABLE VOICE SYSTEMS

Manufacturer and Model	Technique	Bandwidth (Hertz) or Bit Rate (bits per second)	Intelligibility*	Approximate Current Cost** (low demand)	Speculation on Cost in a Mass Market**
A. VRC, Inc. Stockton, California	frequency band elimination				
Harris 1600		1600 Hz		$300	$30 or less
Harris 2100		2100 Hz	90%***	$300	$30 or less
B. Numa Corporaion Orlando, Florida	frequency compression-expansion (speech gapping)				
Waveform Iteration 2:1		1500 Hz	90%	$400	$50 or less
Waveform Iteration 3:1		1000 Hz	89%	$400	$50 or less
C. E-Systems, Inc. Garland Division Dallas, Texas	channel vocoder				
Vadac 5		2400-4800 bps	83-85%	$10,000	less than $1,000
D. Time & Space Processing Cupertino, California	linear predictive coder				
TSP-100		2400 bps	92%	$10,000	less than $1,000
E. --------	unprocessed voice (for comparison)	3000 Hz	97%	$0	$0

* Intelligibility scores are from the Diagnostic Rhyme Test of Dynastat, Inc. Scores were provided by the manufacturers and have not been verified. Scores may not necessarily be directly comparable as different speakers may have been used in each case.

** Cost figures represent cost to a radio manufacturer. A factor of three can be used as a rule of thumb to convert from manufacturer's cost to user's cost.

*** This intelligibility score is for a unit with 1700 Hz bandwidth.

An implementation of the Harris system using a 1700 Hz bandwidth scored 90 percent on the Diagnostic Rhyme Test (DRT), a widely used intelligibility measure. By way of comparison, the telephone has a DRT intelligibility of about 94 percent, and an 85 percent DRT is sometimes used as the boundary between acceptable and unacceptable performance in military applications. The hardware for the Harris system is relatively straightforward, and the circuitry can be placed on one or two integrated circuit chips. In a mass market, the potential is good for a cost of $30 or less (manufacturers' cost).

The utility of the Harris system is currently the subject of considerable controversy. This controversy is summarized by P.S. Henry writing in Communications Magazine (15). Further details can be found in (16, 17, 18).

Since the intelligibility tests and publication of articles mentioned above, improvements have been made in this system. Field tests in a real-life radio environment are to begin shortly.

B. Waveform Iteration

Several implementations for reducing the bandwidth for voice have been developed by the Numa Corp., which was formed in 1977. Their waveform iteration technique takes advantage of the redundant nature of speech by isolating sections that are repetitive and disposing of them.

It operates by placing a section of the input speech waveform into temporary storage. This section corresponds to one pitch period in the case of voiced speech. For unvoiced speech, which does not exhibit the same periodicities, sections approximately 8 milliseconds long are used. The stored waveform can be read out at 1/2 or 1/3 of the rate at which it was read in, converting all frequencies to 1/2 or 1/3 of their original value. To allow for this frequency conversion on a real-time basis, the next waveform section or the next two waveform sections are ignored, depending on whether the compression ratio is 2 to 1 or 3 to 1. This elimination is possible due to the similarities of adjacent sections. The waveform may thus be transmitted with 1/2 or 1/3 of its original bandwidth, and, upon reception, the procedure is essentailly reversed. The low-frequency waveform section is placed into temporary storage and repeatedly read out at twice or three times the speed, so as to regenerate an approximation of the original speech.

For the frequency conversion to properly operate, a synchronization process is necessary, to enable the compressor and expander use the same method of determining the pitch period and waveform sections. Sync is maintained by a knowledge of when pitch periods begin.

The limitation of waveform iteration is in the transitions from one speech phoneme to another, such as the transition in Figure 2 from the "e" sound to the "d" sound in "pledge". At this transition, adjacent waveform sections are not particularly similar, resulting in distortion. But at a compression ratio of 2 to 1, an intelligibility of 90 percent is achieved.

At compression ratios greater than 3 to 1, a hoarseness is introduced and the speech becomes less intelligible. In a mass market, manufacturer's cost is estimated to be $50 or less. Numa also has developed other bandwidth reduction devices. A "spectral shaper" has a compression ratio of 2:1, and is less complex than waveform iteration. An "interferometric processor" can have compression ratios of 2:1, 3:1, 4:1 and 6:1, and is more complex than waveform iteration.

C. Vadac 5 Channel Vocoder

The E-Systems Vadac Model 5 Speech Processor is a channel vocoder with 16 channels and can operate at 2400 or 4800 bits per second. It scored 83 percent on the Diagnostic Rhyme Test at 2400 bits per second; intelligibility is estimated to be about 85 percent at 4800 bits per second. The maximum achievable intelligibility with further development is stated by the manufacturer as being about 90 percent. Cost of individual units is greater than $10,000.

In general, vocoder speech is not as robust as speech that has not undergone extreme bandwidth compression. By this it is meant that the presence of noise and interference will degrade vocoder speech before it will degrade uncoded speech. For digital transmission, vocoder speech can tolerate a bit error rate of 10^{-2} or 10^{-3} while more robust systems like PCM or CVSD can tolerate an error rate of up to 10^{-1} or more. Comparable performance would require a less error-prone transmission system, increased power or a redundant coding scheme, each of which has some tradeoffs in overall spectrum efficiency.

Vocoders offer a large bandwidth reduction but factors that limit their widespread use are high cost, lack of naturalness and susceptibility to noise. The Vadac unit is designed for limited-purpose military and high security applications.

D. TSP-100 Linear Predictive Coder

The Time and Space Processing TSP-100 is an example of a linear predictive vocoder. This unit scored a DRT intelligibility of 92 percent, highest of the units analyzed. The TSP unit approached telephone quality in intelligibility, but a mechanical character limits naturalness somewhat. It costs greater than $10,000 per unit.

Though this linear predictive implementation is significantly more intelligible than the Vadac 5 channel vocoder, one cannot necessarily conclude that LPC inherently outperforms the channel vocoder. At the present level of research, neither vocoder type has demonstrated clear superiority. LPC is preferred by some because it may offer the possibility for a simpler implementation than the channel vocoder.

The TSP coder employs a relatively slow-speed, of-the-shelf microprocessor rather than a high-speed custom processor. The manufacturer claims that this is possible due to the efficient algorithm employed.

It can be reasonably assumed that vocoder quality and price will improve in the future, but the magnitude of improvement is speculative. This is true for linear predictive, channel and other vocoder types. Assuming the current level of research continues, a safe prediction for vocoder costs in a mass market is less than $1,000.

APPENDIX B

Further Description of Vocoders

The vocoder (an acronym for "voice coder") uses analysis and synthesis for reducing voice bandwidth. The input waveform is analyzed based on parameters that have been determined to be important for human speech, hearing and language. Parameter values are transmitted rather than the voice waveform itself. At the receiver, the parameters are used to set up electronic circuits for constructing a replica of the original speech.

There are several different types of vocoders, but most current applications employ digital techniques. This emphasis is due to the military's interest in encrypted speech, to digital's attractiveness to the telephone industry in high-speed switching, and to the ease of analyzing and simulating digital techniques by computer. At this stage analog vocoders appear to offer the potential for a larger bandwidth reduction than digital vocoders, roughly 500 Hz versus 2400 bits per second.

Vocoders are among the most complex and expensive bandwidth reduction techniques, though a great deal of work is being done to reduce the cost. Three of the many vocoder types are described below. Reference (2) is recommended for a further understanding of this field.

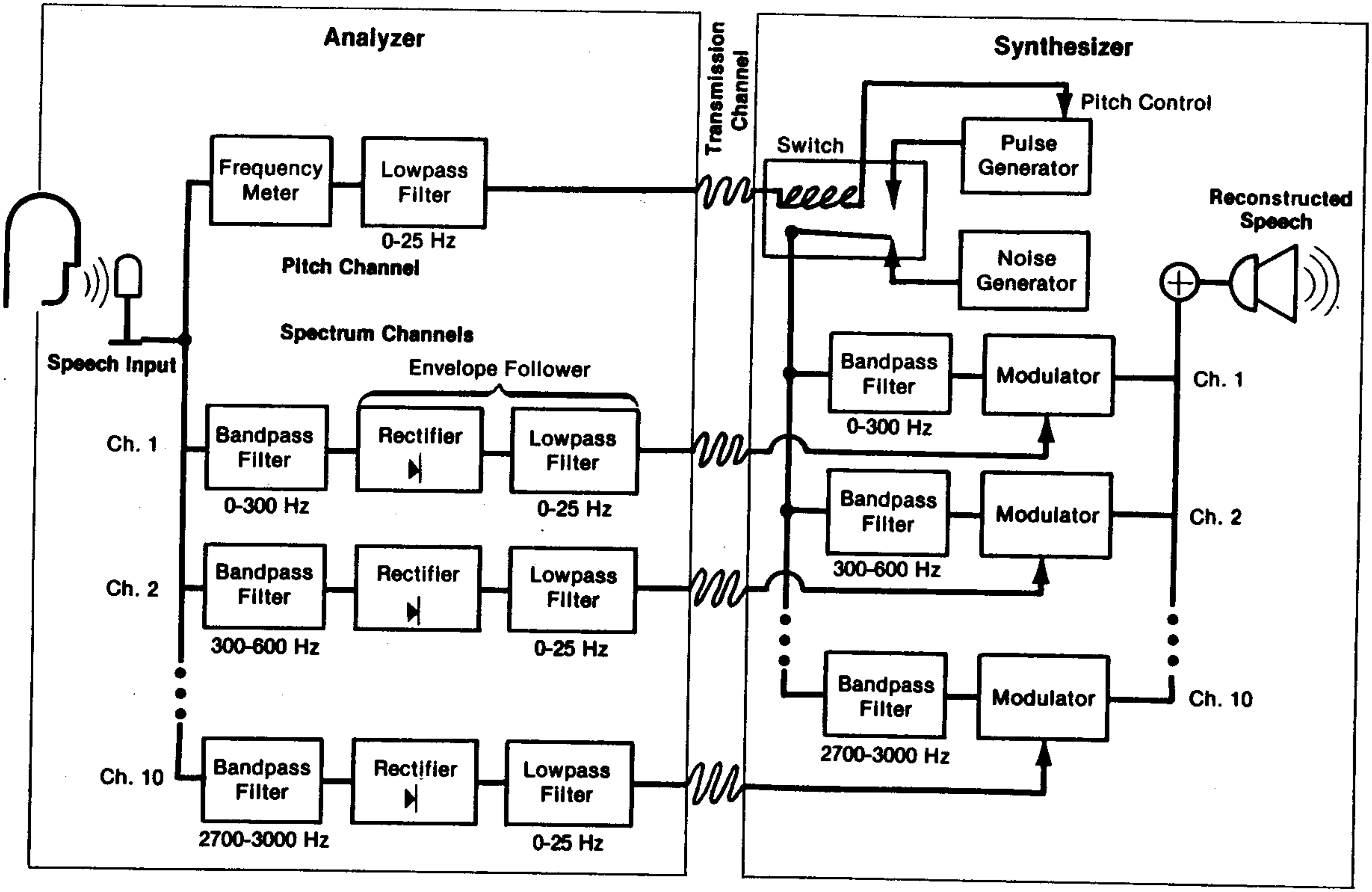
1. Channel Vocoder

Homer Dudley of Bell Laboratories invented the vocoder in the 1930s, and his device is now known as the channel vocoder (1). Figure B-1 is a simplified diagram of Dudley's scheme, which used ten channels. The number of channels can be varied, and have range between 8 and 100.

The channel vocoder recognizes two important constraints on speech. First, proper perception depends on preserving the shape of the amplitude spectrum. Second, speech is a combination of voiced and unvoiced sounds, which can be approximated by a pulse generator and a noise generator.

In the top branch of Figure B-1 the fundamental pitch of voice-like sounds is determined. Unvoiced sounds usually have insufficient energy to operate the frequency meter, so this branch also serves to indicate the voiced/unvoiced decision. In the lower branch, ten channels measure the amplitude spectrum of the speech at ten frequencies. An envelope follower* circuit converts each 300 Hz wide section to an amplitude spectrum only 25 Hz wide. At the

* An envelope follower is a detector with a low pass filter. The voice signal coming in has long-term (low frequency or envelope) and short-term (high frequency) variations. The envelope follower eliminates the short-term variations in favor of the long-term variations, thus lowering the frequency and producing a picture of the peak amplitudes of the voice spectrum.



Dudley's Invention of the Channel Vocoder

Figure B-1.

synthesizer, the speech is reconstructed. If an unvoiced sound occurs the output of a broadband noise generator is applied to a set of bandpass filters identical to those at the analyzer. If a voiced sound occurs, a switch applies the output of the pulse generator to the filter bank. The pulse generator is set to duplicate the proper fundamental pitch frequency. The filters receive the pulse generator or noise generator signal, which is shaped by the amplitude spectrum information into an approximation of the original waveform. It is the amplitude spectrum information that establishes the difference between, for example, the voiced "e" and the voiced "a", or the unvoiced "p" and unvoiced "t".

Dudley's 10-channel vocoder reproduced speech using less than 300 Hz bandwidth. Present configurations usually use 15 to 20 channels and hence need slightly larger bandwidth. With proper design, channel vocoders of 500 Hz bandwidth and less can be made quite intelligible, though the actual quality or "acceptability" of the speech may be poor. Vocoders are sensitive to speaker differences. If optimized for male speakers, higher-pitched female speakers are not processed as well. In addition, though intelligibility may be acceptable, a machine-like quality is characteristic of the output, which some users find objectionable. Vocoders are not entirely machine speech in that they provide some speaker recognition capability. This is important for military applications where the listener might need to recognize the voice of the person giving an important order. Indications are that research will continue to increase the quality of vocoder speech.

The digital implementation of channel vocoders converts the speech waveform into bits which are then processed to obtain the vocoder channels, pitch, and voiced/unvoiced information. Digital channel vocoders require a larger bandwidth than analog, so the advantages of digital channel vocoders do not appear particularly suited for the commercial radio market. However, analog vocoders have received less recent research attention in spite of their bandwidth reduction capability.

2. Linear predictive vocoder

A technique that is more inherently digital that has developed popularity in the last five years is the linear predictive vocoder. The linear predictive vocoder filter is based on vocal tract modeling in the time domain as described in chapter IV. In linear predictive coding the speech is dynamically analyzed to detect correlations in the time domain. Knowledge of these correlations are used to solve a set of linear equations, and parameters called predictor coefficients are obtained. The predictor coefficients may be transmitted and used to define the shape of the synthesizer filter response. Actually, most methods do not transmit the predictor coefficients directly, but use them to derive other parameters which may be more accurately digitized and transmitted.

3. Voice-excited vocoder

The voice excited vocoder offers the capability for increased naturalness and voice quality at the expense of bandwidth. These devices require a bandwidth of about 1,000 Hz, while analog channel vocoders can operate with 500 Hz and less. Still the voice-excited vocoder offers a bandwidth compression of three to one over 3 KHz voice.

Channel vocoders and linear predictive vocoders are limited by difficulties in obtaining accurate pitch measurement and determining whether a sound is voiced or unvoiced. Speech naturalness depends on these qualities. The voice-excited vocoder avoids these problems by transmitting a low-frequency sub-band of the original speech, thus preserving pitch and voiced/unvoiced information. With the voice-excited vocoder, the noise generator and pulse generator indicated in Figure 1 of Chapter III are not used. Rather the low-frequency sub-band is artificially flattened and broadened to cover the frequency range up to 3 KHz or so. This distorted waveform is then used to excite conventional vocoder channels above the sub-band frequencies. Use of the sub-band preserves the specific structure of speech, as well as the pitch and voiced/unvoiced information, resulting in increased naturalness. Thus the voice-excited vocoder can offer higher quality speech than the channel (or linear predictive) vocoder.

One implementation of a voice-excited vocoder is described in reference (2). In this device the frequency range from 250 to 940 Hz is transmitted unprocessed; the range from 940 to 3650 Hz is covered by 17 vocoder channels. The overall bandwidth is about 1200 Hz. The following test results are quoted from (2), page 257; "Overall speech quality of the voice-excited vocoder was assessed, along with that for three other transmission methods, by presenting listeners with sentences in isolation. The subjects were asked to rate each sentence "as good as normal telephone" or worse than "normal telephone". In 72 percent of the cases, the voice-excited vocoder was rated as good as normal telephone. In the same test, for comparison, a long-distance carrier telephone circuit rated 82 percent, an 1800 cps low-pass circuit rated 35 percent, and a regular 18-channel vocoder rated 17 percent. The results show the voice-excited system to be clearly superior to the spectrum channel vocoder and to approach the quality of conventional voice circuits."

This description applies to a laboratory implementation of the voice-excited vocoder.

V. References

(1) H. Dudley, "Remaking Speech", *J. of Acoust. Soc. of America*, Vol. II, 1939.

(2) J.L. Flanagan, *Speech Analysis, Synthesis and Perception*, Second Edition, Springer-Verlag, NY, 1972.

(3) *Transmission Systems for Communications*, Bell Telephone Laboratories Technical Staff, published by Western Electric Co., 1971.

(4) L.R. Rabiner and R.W. Shafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.

(5) A.H. Reeves, French patent no. 852,183, Oct. 1938.

(6) B.N. Oliver, J.R. Pierce and C.E. Shannon, "The Philosophy of PCM", *Proc. IRE*, 36, Nov. 1948, pp. 1324-1331.

(7) R. Steele, *Delta Modulation Systems*, John Wiley & Sons, NY, 1975.

(8) F. Jelinek, "Tree Encoding of Speech", *IEEE Trans. Inf. Theory*, November, 1976.

(9) R.E. Crochiere, "On the Design of Sub-band Coders for Low Bit Rate Speech Communication", *BSTJ*, Vol. 56, p. 747-770, 1977.

(10) A. Jain, "Transform Coding for Data Compression", report submitted to Office of Plans and Policy, FCC.

(11) "Digital Speech Interpolation", Session 14, 1978 National Telecommunications Conference Proceedings, December, 1978.

(12) D.L. Deittweiler and D.G. Messerschmitt, "Nearly Instantaneous Companding as Applied to Mobile Radio Transmission", Conf. Record of 1975 ICC, June, 1975.

(13) J.M. Tribolet, P. Noll, B.J. McDermott, R.E. Crochiere, "A Study of Complexity and Quality of Speech Waveform Coders", IEEE Int'l. Conf. on SAASP, Tulsa, 1978.

(14) R.W. Harris and J.F. Cleveland, "A Baseband Communications System", Part 1, *QST*, November, 1978, pp. 14-21.

(15) P.S. Henry, "Single Sideband Mobile Radio, a Review and Update", IEEE *Communications Magazine*, March, 1979.

(16) R.W. Wilmotte and B. Lusignan, "Spectrum-Efficient Technology for Voice Communications", Federal Communications Commission UHF Task Force Report, February, 1978.

(17) B. Lusignan, "Single Sideband Transmission for Land Mobile Radio", IEEE *Spectrum*, July, 1978, pp. 33-37.

(18) Comments on the Federal Communications Commission UHF Task Force Report entitled "Spectrum-Efficient Technology for Voice Communications", a report of the EIA RT-8 Ad Hoc Committee, R.C. Buetow, Chairman. The report is on file at the Land Mobile Communications Council, 1150 17th Street, NW, Suite 1000, Washington, DC 20036.

(19) J.L. Flanagan, et. al., "Speech Coding", IEEE Trans. on Communications, Vol. COM-27, No. 4, April, 1979, pp. 710-737.

(20) R.E. Crochiere and M.R. Sambur, "A Variable Sub-Band Coding Scheme for Speech Encoding at 4.8 kb/s", BSTJ, Vol. 56, pp. 771-779, 1977.

(21) R.W. Harris, J.F. Cleveland and H.M. Howland, "A Unique Narrowband Voice Modulation System", paper presented at the IEEE Acoustics, Speech and Signal Processing International Conference, May 10, 1977.

(22) B. Gold, "Digital Speech Networks", Proceedings of the IEEE, Vol. 65, No. 12, p. 1638, December, 1977.

(23) Reference Data for Radio Engineers, fifth edition, Howard W. Sams and Co., Inc., New York, 1968.

(24) R.G. DeWitt, "Digital Microwave Radio", Telecommunications, p. 24, April, 1975.