

FOR REFERENCE

NOT TO BE TAKEN FROM THIS ROOM

A NONUNIFORM SAMPLING APPROACH TO DATA COMPRESSION OF
SPEECH AND VOICEBAND DATA SIGNALS

by

Ahmet Semih Bingöl

B.S. in E.E., Boğaziçi University, 1982

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of

Master of Science

in

Electrical Engineering

Bogazici University Library



39001100314403

14

Boğaziçi University

1985

To Drs.

Bülent Sankur

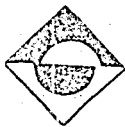
Ömer Cerid

Yusuf P. Tan

Emin Anarım

whom I had the good fortune to work with and learn from

183607



ABSTRACT

Using voiceband data and speech like signals the performance of the Asynchronous Delta Modulator (ASDM) has been investigated and its capabilities as a potential medium-band speech coder have been evaluated.

The asynchronous delta modulator is a versatile source coder in which the sampling rate is determined directly by the activity of the input signal, hence the samples are nonuniform. When the input signal is idling ASDM samples it infrequently. Active segments of the signal are, on the contrary, encoded with many samples. Since the samples are nonuniform in time, the information in the signal is coded into the output bit polarities as well as the time intervals between these bits. Therefore quantization of these inter-bit intervals is necessary as well as a buffer to output the asynchronous input information at a synchronous rate.

This study has mainly concentrated on the quantization and encoding of the inter-bit intervals. For both speech and QPSK data, ASDM sampling statistics have been investigated. Logarithmic, optimum and vector quantization of the inter-bit intervals have been considered. For further compression, entropy coding of the inter-bit interval quantizers have been investigated. Buffer behaviour for different system parameters has been evaluated and the results have been compared with popular speech coders.

ÖZET

Bu çalışmada veri süreçleri ve ses benzeri süreçler kullanılarak Asenkron Delta Kodlayıcısının başarımı incelenmiş ve orta bant ses kodlayıcıları arasındaki yeri araştırılmıştır.

Etkin bir kaynak kodlayıcısı olan asenkron delta kodlayıcısında örnekleme anları giriş iminin kendisi tarafından belirlendiğinden örnekler birbiçimsizdir. Asenkron delta kodlayıcısı giriş iminin durgunumsu olduğu zamanlarda seyrek, hareketli olduğu zamanlarda sık örnek alır. Örnekler zamanda birbiçimsiz olduğundan giriş imindeki bilgi çıkış darbeleri ile bu darbeler arasındaki süreye kodlanmış olur. Alıcının imi yeniden kurabilmesi için darbeler arası süreyi de bilmesi gerektiğinden bu süreler nicelenerek bir tampon belleğe yerleştirilir ve kanala eşzamanlı olarak verilir.

Bu çalışmanın yoğunlaştığı nokta örnekler arası sürenin nicelenmesi ve kodlanmasıdır. Ses ve QPSK veri süreçleri kullanılarak asenkron delta kodlayıcısının örnekleme istatistikleri incelenmiştir. Örnekler arası sürenin logaritmik, (enküçük kareler anlamında) eniyi ve vektör nicelenmesi gerçekleştirilmiş ve daha fazla sıkıştırma elde edebilmek için entropi kodlaması uygulanmıştır. Çeşitli dizge parametreleri için tampon belleğin davranışı da ele alınmış ve sonuçlar yaygın ses kodlayıcılarıyla karşılaştırılmıştır.

TABLE OF CONTENTS

	<u>Page</u>
ABSTRACT	iv
OZET	v
LIST OF FIGURES	viii
LIST OF TABLES	xi
LIST OF SYMBOLS	xii
I. INTRODUCTION	1
II. SPEECH AND DATA SIGNALS	3
2.1 THE SPEECH SIGNAL	3
2.1.1 SPEECH GENERATION	3
2.1.2 CHARACTERISTICS OF SPEECH SIGNALS	6
2.2 DATA SIGNALS	8
2.2.1 CHARACTERISTICS OF DATA SIGNALS	8
2.3 FIDELITY CRITERIA	11
2.3.1 PERFORMANCE MEASURES FOR SPEECH	11
2.3.2 PERFORMANCE MEASURES FOR DATA	12
III. SPEECH CODING	14
3.1 PULSE CODE MODULATION	15
3.1.1 PCM WITH FORWARD ADAPTIVE QUANTIZATION	18
3.2 DIFFERENTIAL PCM	19
3.2.1 ADAPTIVE QUANTIZATION IN DPCM	21
3.2.2 ADAPTIVE PREDICTION IN DPCM	22
3.2.3 ADAPTIVE PREDICTION AND ADAPTIVE QUANTIZATION IN DPCM	24
3.2.4 DELTA MODULATION	25

	<u>Page</u>
IV. THE ASYNCHRONOUS DELTA MODULATOR	26
4.1 ANALYSIS OF ASYNCHRONOUS DELTA MODULATOR	28
4.1.1 ASDM SAMPLING STATISTICS	29
4.1.2 SIGNAL TO QUANTIZATION NOISE RATIO	30
V. SIMULATION RESULTS	32
5.1 THE ARTIFICIAL SPEECH SIGNAL	33
5.1.1 THE ARTIFICIAL SPEECH GENERATION MODEL	33
5.1.2 THE GLOTTAL EXCITATION	34
5.1.3 GLOTTAL PULSE WHITENING FILTER	35
5.1.4 THE VOCAL TRACT FILTER	37
5.1.5 PSEUDORANDOM VARIATION	38
5.1.6 ENVELOPE MODULATION	40
5.2 THE DATA SIGNAL	42
5.3 SPEECH ENCODER SIMULATIONS	44
5.3.1 SPEECH ENCODING PERFORMANCE OF SIMULATED ENCODERS	45
5.3.2 DATA ENCODING PERFORMANCE OF SIMULATED ENCODERS	47
5.4 SIMULATION OF ASDM	50
5.4.1 AVERAGE SAMPLING RATE	50
5.4.2 SAMPLING STATISTICS	52
5.4.3 SPEECH ENCODING WITH ASDM	54
5.4.4 DATA ENCODING WITH ASDM	60
5.4.5 ENTROPY CODING	66
5.4.6 BUFFER BEHAVIOUR	68
VI. CONCLUSIONS	71
REFERENCES	75

LIST OF FIGURES

		<u>Page</u>
FIGURE 2.1	The conventional speech production model	4
FIGURE 2.2	Waveforms and spectra of voiced "a" and unvoiced "s" sounds	5
FIGURE 2.3	Amplitude PDF of speech along with Gamma and Laplacian PDFs	6
FIGURE 2.4	Long time averaged power spectral density of speech	7
FIGURE 2.5	Long time averaged autocorrelation functions for lowpass and bandpass filtered speech	8
FIGURE 2.6	Power spectral densities of typical data signals and data rates	9
FIGURE 2.7	Power spectral densities of speech and QPSK data signal	9
FIGURE 2.8	Autocorrelation functions of speech and voiceband data	10
FIGURE 2.9	Waveforms of speech and QPSK data	11
FIGURE 3.1	Transmission rates and associated quality of speech coders	14
FIGURE 3.2	PCM encoder	15
FIGURE 3.3	μ -law compressor curves	16
FIGURE 3.4	Forward and backward quantizer adaptation	18
FIGURE 3.5	Differential PCM	19
FIGURE 3.6	Adaptive quantization in DPCM	21
FIGURE 3.7	Adaptive prediction in DPCM	22
FIGURE 3.8	Prediction gain vs. predictor order for fixed and adaptive predictors	23
FIGURE 3.9	DPCM with adaptive prediction and adaptive quantization	24
FIGURE 4.1	ASDM encoder and waveforms	26
FIGURE 4.2	Complete ASDM system	28

FIGURE 4.3	Quantization error spectra for various values of the corridor width	31
FIGURE 5.1	Block diagram of the artificial speech generation model	33
FIGURE 5.2	The glottal pulse waveform	35
FIGURE 5.3	Spectra of the glottal pulse and the whitened glottal pulse	36
FIGURE 5.4	Average long term spectrum and normalized autocorrelation function of telephone speech of ten speakers	37
FIGURE 5.5	Modulus of the transfer function of the 16th order vocal tract filter	38
FIGURE 5.6	Time waveform of the artificial speech signal	39
FIGURE 5.7	Spectrum of the artificial speech signal	39
FIGURE 5.8	Amplitude histogram of the artificial speech signal	40
FIGURE 5.9	Autocorrelation coefficients of the artificial signal and original speech	41
FIGURE 5.10	Simulation block diagram of the voiceband data encoding/decoding process	43
FIGURE 5.11	The sampling phase effect	44
FIGURE 5.12	Performance of simulated speech encoders	46
FIGURE 5.13	Average number of ASDM samples per second vs. corridor width for the artificial speech	51
FIGURE 5.14	Inter-bit interval PDFs for speech	52
FIGURE 5.15	Inter-bit interval PDFs for QPSK data	53
FIGURE 5.16	Speech encoding performance of ASDM with log-quantization of the inter-bit intervals	55
FIGURE 5.17	The micro expansion/compression effect	56
FIGURE 5.18	Speech encoding performance of ASDM with optimum quantization of the inter-bit intervals	57

Page

FIGURE 5.19	The reconstruction levels of a 2-bit optimum speech IBI quantizer for various values of w	58
FIGURE 5.20	Speech encoding performance of ASDM with vector quantization of the inter-bit intervals	59
FIGURE 5.21	The reconstruction levels of a 2-bit optimum data IBI quantizer for various values of w	62
FIGURE 5.22	Buffer behaviour for various values of the corridor width w	69

LIST OF TABLES

		<u>Page</u>
TABLE 5.1	Results of the data encoding performance of simulated speech coders	48
TABLE 5.2	Average number of ASDM samples per second for the artificial speech and QPSK data	50
TABLE 5.3	Results of the data encoding performance of ASDM with no IBI quantization	61
TABLE 5.4	Results of the data encoding performance of ASDM with 2 to 5 bit log-quantization of the IBI sequence	63
TABLE 5.5	Results of the data encoding performance of ASDM with 2 to 5 bit optimum quantization of the IBI sequence	64
TABLE 5.6	Results of the data encoding performance of ASDM with vector quantization of the IBI sequence	65
TABLE 5.7	Results of Huffman coding of the optimum IBI quantizer outputs for speech	66
TABLE 5.8	Results of Huffman coding of the vector quantizer outputs for speech	67
TABLE 5.9	Results of Huffman coding of the optimum IBI quantizer outputs for data	68

LIST OF SYMBOLS

A	Data signal amplitude
\mathbf{a}	Vector of predictor coefficients
$b(t_i)$	Transmitted bit sequence
$E(t)$	Data envelope
$e(n)$	Error signal
$\hat{e}(n)$	Quantized error signal
\bar{f}	Average number of samples per second of the ASDM encoder
G_p	Prediction gain
H	Source entropy
\bar{N}	Average code rate in bits/symbol
PDF	Probability density function
PSK	Phase shift keying
p	Predictor order
$p(n)$	Envelope of artificial speech
QPSK	Quadrature phase shift keying
R	Matrix of autocorrelation coefficients
\mathbf{r}	Vector of autocorrelation coefficients
rms	Root mean square
$r(n)$	n 'th autocorrelation coefficient
T	Symbol time in data transmission
$x(n)$	Input signal
$\hat{x}(n)$	Predicted or quantized or reconstructed signal
$\hat{x}(nT)$	In phase component in demodulated data
$\hat{y}(nT)$	Quadrature component of demodulated data
α	A coefficient to be used to optimize performance
Δ_n	Sampling phase

$\delta, \delta(nT)$	Phase error
$\phi, \phi_n, \phi(nT)$	Transmitted phase
$\hat{\phi}, \hat{\phi}_n, \hat{\phi}(nT)$	Received phase
σ_e^2	Error signal variance
σ_n^2	Variance of noise
σ_x^2	Variance of input signal
$\sigma_{\hat{\phi}}^2$	Phase error variance
ω_c	Carrier frequency

I. INTRODUCTION

With the evolving digital IC technology, there has been growing interest in finding methods for efficient digitization of speech. Apart from other speech processing applications such as recognition and enhancement, digitization forms the front-end of two important processes; speech transmission and storage. Efficient digitization, which translates into low bit rate coding, is necessary to conserve bandwidth in transmission and memory in storage systems.

The benchmark for all speech coders is log-PCM which is widely used in the telephone network. Since PCM necessitates a high bit rate, many alternative lower bit rate coders have been developed but these have found limited commercial use. If one reason for this is the vast amount of investment on log-PCM by the telephone companies, the other reason is the rather stringent requirements that a successful speech coder has to satisfy.

A "telephone quality" coder must provide a certain signal to noise ratio (SNR) over a large input signal range; 40 dB dynamic range is typical for speech. Furthermore, in the present half analog network a signal may undergo a number of analog to digital conversions in tandem on its route. A commercially acceptable speech coder must be sufficiently robust to such tandem connections.

Many low bit rate speech coders are matched to the statistics of the speech signal. However, these statistics are not constant and have inter-speaker as well as intra-speaker variability. Adapting the coder to changing input signal statistics may provide a solution but adaptive coders are confronted with the obvious problems of complexity and cost.

The problem is further complicated by the fact that in the switched telecommunications network the coder must be able to handle not only speech, but various kinds of voiceband data signals. If the network were all analog there wouldn't be any need to code the voiceband data signal for digital transmission. On the other hand, if the network were all digital there wouldn't be any voiceband data at all. However, as long as the present hybrid network exists, a speech coder must successfully code voiceband data signals without demodulating them. Since the statistics of voiceband data signals differ significantly from that of speech the performances of many low bit rate coders, which are matched to speech statistics, are seriously degraded in the presence of data.

In this study the performance of the Asynchronous Delta Modulator (ASDM) has been investigated. Chapter II presents the characteristics of speech and data signals and establishes the performance metrics. Simulated speech coders are described in Chapter III. Chapter IV introduces the Asynchronous Delta Modulator. Simulation results are given in Chapter V and conclusions and suggestions for further research in Chapter VI.

II. SPEECH AND DATA SIGNALS

This chapter gives a brief outline of the characteristics of speech and data signals. In particular, properties that are exploited in one way or another by current speech coders are described rather than properties of academic interest.

2.1. The Speech Signal

Speech is the most natural form of human communication. This wonderful signal conveys much more information to a listener than the mere textual content of the message. Because of this reason, speech is probably the most "compressible" information source ever known to man. It can be compressed by more than a factor of 100 and still retain intelligibility.

2.1.1. Speech Generation

Analysis of a long record of the speech waveform reveals two basic structures :

- a) High energy, quasi-periodic segments
- b) Low energy, noise-like segments.

The high energy segments, called "voiced" sounds, are generated by the periodic vibration of the vocal cords at the so called pitch frequency.

Pitch frequency ranges from about 50 to 200 Hz for men and 200 to 400 Hz for women and children. The low energy segments, termed as "unvoiced" sounds, exhibit no periodicity. The vocal cords do not vibrate and the sound is generated by forcing the air through a constriction somewhere in the vocal tract, thus creating a noise-like turbulent air flow.

When the vocal tract is excited by one of the two excitation types described above, it behaves like a nonuniform acoustical tube and shapes the output signal spectrum. Just like an organ pipe, the vocal tract has resonance frequencies called "formants". Each different shape of the vocal tract corresponds to a different set of formants. The bandwidths and frequencies of the formants as well as their time course are extremely important for the intelligibility of speech.

The earliest but still commonly used model of the speech generation process is shown in Fig.2.1. The proper type of excitation

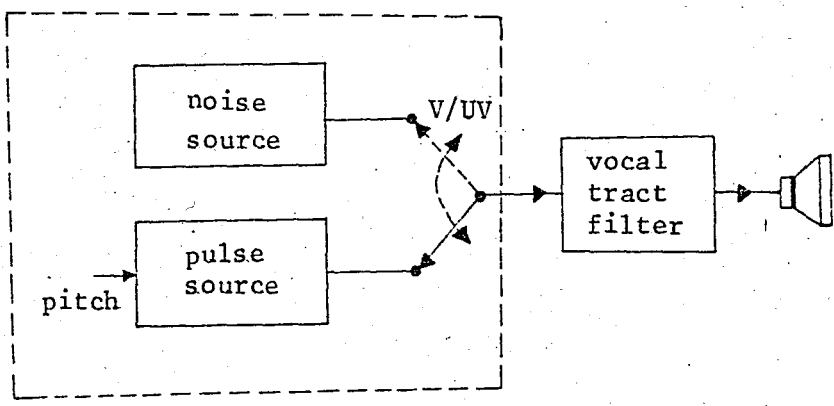


Figure 2.1. The conventional speech production model

is selected by the voiced/unvoiced (V/UV) switch. The vocal tract is represented by the filter F . For computational simplicity, it is customary to represent the vocal tract as an all-pole filter of order 10-16. Although this simple voiced/unvoiced designation is not adequate for many sounds (e.g. voiced fricatives like "v" and "z"), the model has been successfully used up to now [1].

Obviously the vocal tract is time varying, but for short time analysis (typically 10 to 40 ms) it can be assumed as a time invariant system. Short time waveform segments and spectra of voiced "a" and unvoiced "s" are shown in Fig.2.2. Note that "s" has significant high frequency content while the opposite is true for "a". For both sounds the general shape of the spectrum is determined by the vocal tract while for "a", the rather regular fine structure is due to the pitch frequency.

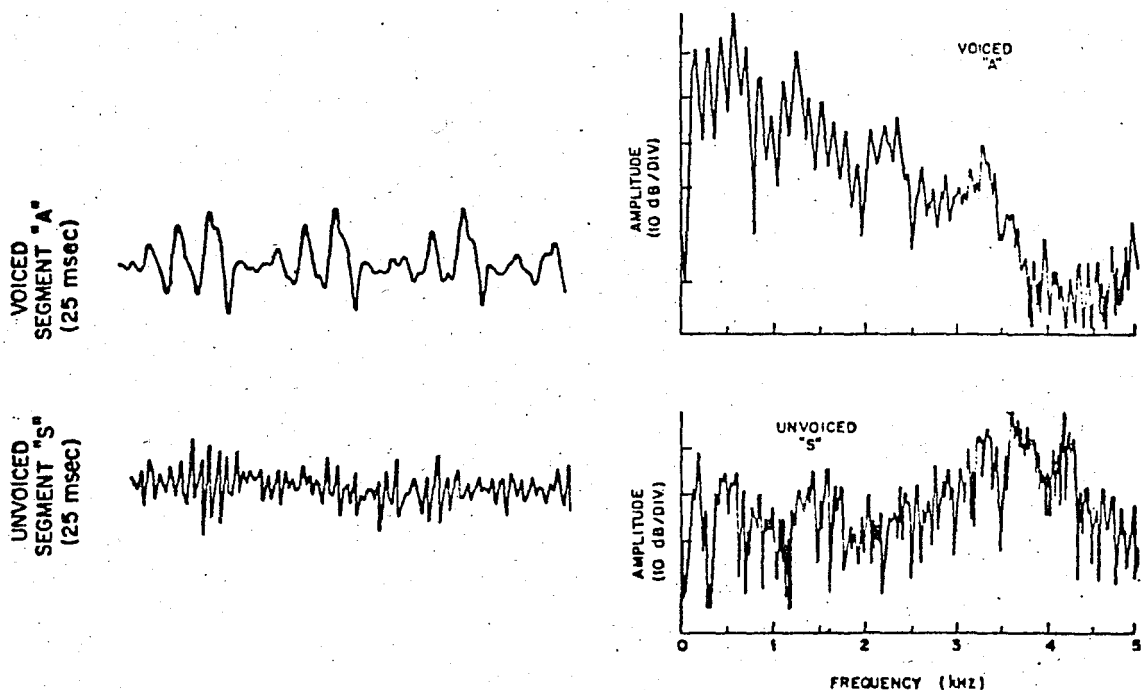


Figure 2.2. Waveforms and spectra of voiced "a" and unvoiced "s" sounds (After Flanagan et.al. [2]).

2.1.2. Characteristics of Speech Signals

Perhaps the most striking feature of the speech signal is its nonstationary nature. As mentioned before, for short time analysis the vocal tract filter can be assumed to have a fixed configuration and the principles of stationarity can be invoked. Another important property is the silent intervals which amount to more than fifty percent in ordinary conversational speech. There are methods like DSI (Digital Speech Interpolation) which exploit this property to reduce the bit rate. Other properties which are of interest to us are described below.

Amplitude PDF

Observation of a long record of the speech waveform reveals a very high probability of near zero amplitudes due mainly to unvoiced segments and silent periods. The speech PDF therefore has a peak at zero and decreases monotonically with increasing amplitude. This long time PDF is best approximated by a Laplacian or Gamma PDF. The speech PDF is shown in Fig.2.3 along with Laplacian and Gamma PDFs.

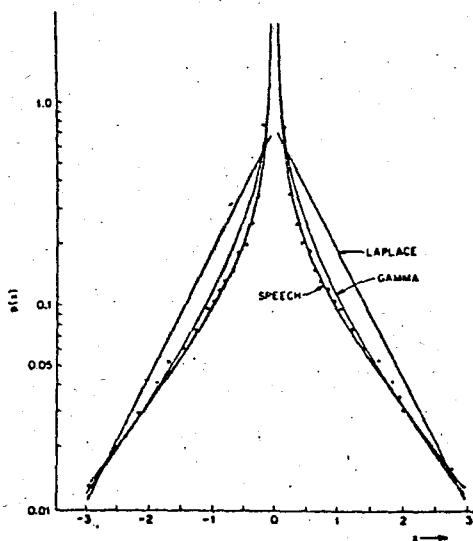


Figure 2.3. Amplitude PDF of speech along with Laplacian and Gamma PDFs (After Paez and Glisson [3]).

Power Spectrum

Speech waveforms are inherently bandlimited because of the speech production process. Another bandlimiting factor is the filtering of speech prior to encoding. Conventional telephone circuits have a bandwidth that extends from 200 to 3200 Hz. Long time averaged power spectral density of speech is shown in Fig.2.4. Although speech is globally a low-pass signal, short time speech segments can have high-pass spectra as exemplified by Fig.2.2. These high-pass segments must be preserved by the coder because they contribute a lot to overall intelligibility.

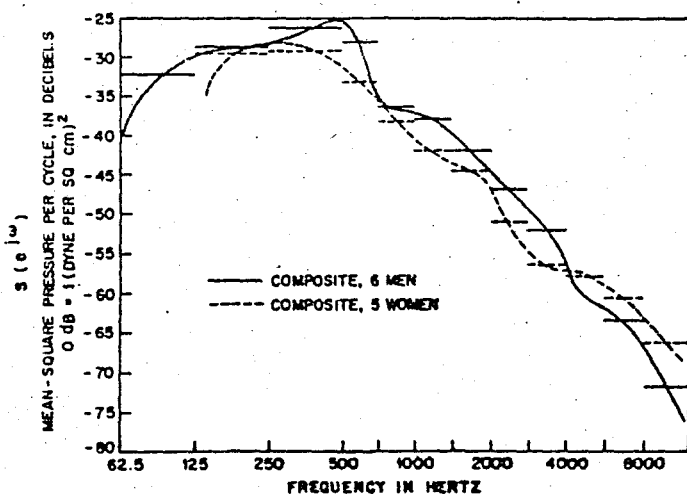


Figure 2.4. Long time averaged power spectral density of speech (After Flanagan et.al. [2]).

Autocorrelation Function

Telephone bandwidth speech is usually sampled at 8 kHz. The normalized autocorrelation functions of lowpass (0 to 3400 Hz) and bandpass (200 to 3400 Hz) filtered and 8 kHz sampled speech are shown in Fig.2.5. As the high value of the first autocorrelation function indicates, there is significant correlation between successive samples in 8 kHz sampled speech. One intuitively expects this correlation to

increase as the sampling rate is increased. It is precisely this principle that is utilized in Delta Modulation (DM).

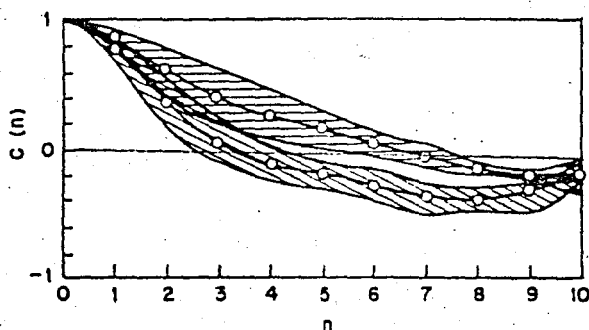


Figure 2.5. Long time averaged autocorrelation functions for lowpass (upper) and bandpass filtered speech (After Noll [4]).

2.2. Data Signals

In contrast to speech, relatively little work has been done in digital encoding of voiceband data signals. Voiceband data signals are analog signals created from digital data to be transmitted over telephone channels. Although data rates up to 14.4 kbits/s have been successfully transmitted over voice grade (3.5 kHz) telephone lines using sophisticated trellis encoding techniques, we shall be concerned with lower data rates.

2.2.1. Characteristics of Voiceband Data Signals

Power Spectrum

Power spectral densities for four different voiceband data signals are shown in Fig.2.6. In the rest of this thesis we shall deal with the 1200 baud QPSK signal (2400 bps) whose power spectrum is shown in Fig.2.6.(b).

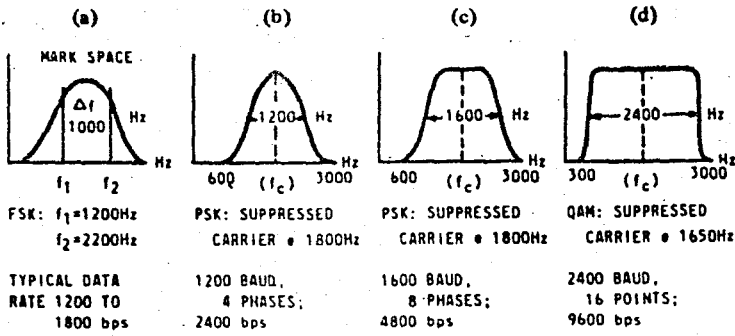


Figure 2.6. Power spectral densities of typical voiceband data signals and data rates (After Kretzmer [5]).

The power spectra of this QPSK data signal and speech are displayed together in Fig 2.7. Observe that while speech has most of its energy concentrated at frequencies below 800 Hz, the energy of the data signal is more evenly spread over the available bandwidth and centered at about 1800 Hz (carrier frequency).

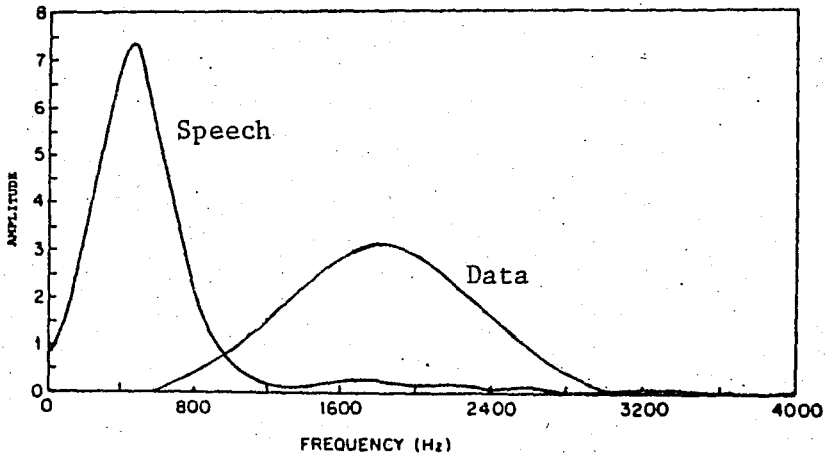


Figure 2.7. Power spectral densities of speech and the QPSK data signal.

Autocorrelation function

Autocorrelation functions of speech and data are shown in Fig.2.8. Note that 8 kHz sampled speech samples have higher correlation

than 8 kHz sampled data which, of course, is in agreement with the power spectrum plots of Fig.2.7.

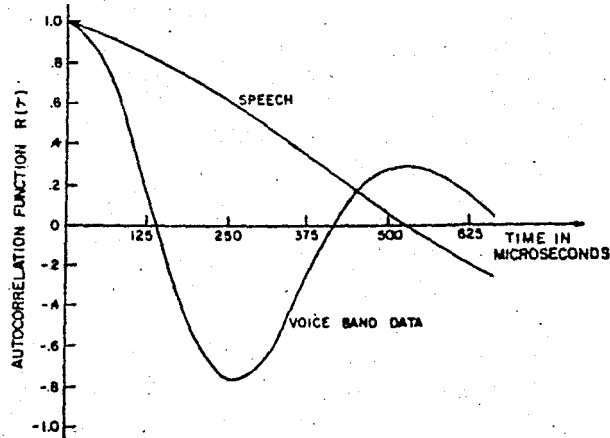


Figure 2.8. Autocorrelation functions of speech and voiceband data (After O'Neal [6]).

This lower correlation is the major reason that coders which exploit inter-sample redundancy (like DM and DPCM) perform poorly with data signals. Another important reason is the predictor mismatch in fixed predictor coders.

Time Waveform

Fig.2.9 displays the time waveforms of speech and QPSK data signals. Apparently, speech and data waveforms have drastically different characteristics. Speech contains bursts of high energy with long silent and low energy segments in between. The data signal on the other hand, has a smooth flow of energy.

The crest factor (ratio of peak value to rms value) is about 100 for speech while it is between 1 and 2 for voiceband data (approximately 1.6 for QPSK).

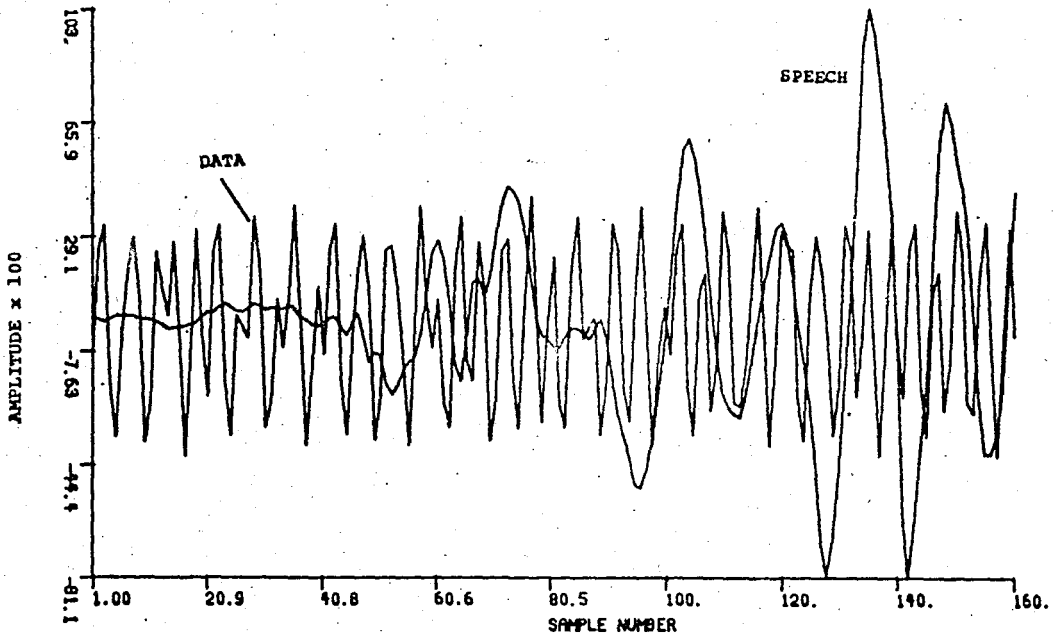


Figure 2.9. Waveforms of speech and QPSK data (After O'Neal [6]).

2.3. Fidelity Criteria

In order to compare various encoders we need some fidelity criteria. In the following sections we give a concise discussion of various performance measures for speech and data signals.

2.3.1. Performance Measures for Speech

The most common objective criterion for evaluating coder performance is the signal to noise ratio (SNR) defined as

$$\text{SNR}(\text{dB}) = 10 \log_{10} \frac{\sum_n x^2(n)}{\sum_n e^2(n)} \quad (2.1)$$

where $x(n)$ denotes the original sample and $e(n)$ is the error between the original and reconstructed samples.

Since the ultimate consumer of speech is the human ear, subjective measures based on listening tests are more meaningful. Furthermore SNR, as defined above, does not correlate well with subjective evaluations and various refinements have been proposed. One of them is segmental SNR which is the arithmetic mean of SNRs computed over 10-20 ms intervals [7]. This method prevents the masking of coder noise in low energy segments by high signal energy segments thereby avoiding an artificially high overall SNR, and agrees better with subjective preferences.

Still the conventional SNR given in (2.1) is widely used as an objective measure of coder performance because of its mathematical tractability. We have used the SNR as the performance metric throughout this thesis bearing in mind that it does not reflect subjective criteria well enough.

2.3.2. Performance Measures for Data

Clearly, the only performance criterion for any data transmission system is the probability of error (P_e). With additive white Gaussian noise error probabilities can be directly derived from the SNR. However, the quantization noise generated by most coders is neither Gaussian nor additive [8].

For the QPSK data signal that we shall exclusively be concerned with, better performance measures are the mean, variance and maximum value of the phase error where the phase error δ is a random variable defined as

$$\delta = \phi - \hat{\phi} . \quad (2.2)$$

In (2.2) ϕ and $\hat{\phi}$ denote the transmitted and received phases, respectively.

The phase error variance can be used to derive an equivalent SNR given by

$$\text{SNR}_e = 10 \log_{10} \frac{1}{2 \sigma_\delta^2} \quad (2.3)$$

which is the SNR that would be obtained if the coder noise were Gaussian [9]. However, since the coder noise is not Gaussian SNR_e cannot be used to derive error probabilities.

III. SPEECH CODING

Speech coding has been an active research area in the last two decades and more than a dozen speech coders and refinements and variations thereof have emerged. Speech coders can be broadly classified into two categories: Waveform coders and parametric coders (vocoders). Waveform coders, as the name implies, code the actual speech waveform while parametric coders first extract and then code the parameters of an assumed model of the speech production process. As expected, parametric coders require more complicated hardware than waveform coders. Operating bit rates, attainable signal quality and relative complexity of speech coders are summarized in Fig.3.1.

Coder	Waveform			Parametric
Complexity	Low	Low Medium	High	Very high
Quality	Broadcast (Commentary)	Toll 200-3200 Hz SNR > 30 dB dist. < 2-3%	Communication	Synthetic
Bit rate (kbps)	200	64	16	7.2 1

Figure 3.1. Transmission rates and associated quality of speech coders (After Flanagan et.al. [2]).

The quality of waveform coders increases with increasing bit rate. For parametric coders on the other hand, increasing the bit rate does not bring much improvement because the quality can at most be as good as the assumed speech production model.

An interesting region in Fig.3.1 is the vicinity of 9.6 kbits/s where a low complexity toll quality coder (as defined in Fig.3.1) could represent considerable economic potential for the telecommunications companies.

It is beyond the scope of this thesis to present even a general outline of existing speech coders. We shall instead investigate a few coders against which ASDM is compared in Chapter V. The reader is reminded that there are coders with superior performance than those that will be discussed and is referred to [10] for a full treatment of waveform coders. General aspects of parametric coding can be found in [11].

3.1. Pulse Code Modulation (PCM)

PCM is the most commonly used speech coding system due mainly to its simplicity and, probably, chronological precedence. The block diagram of a PCM system is shown in Fig.3.2. It simply consists of a sampler (not shown) and a "quantizer" (denoted by Q in Fig.3.2) which, in practice, is an A/D converter.

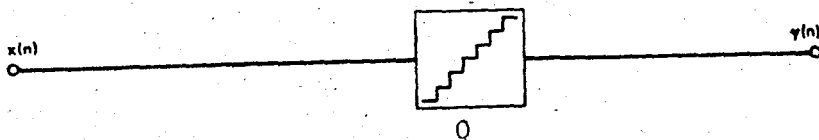


Figure 3.2. PCM encoder (After Noll [12]).

If a uniform quantizer is used 13-14 bits are necessary to accommodate the wide dynamic range of the speech signal. Therefore to

achieve a prescribed SNR over a wide input range with fewer bits, some form of companding is used. There are two widely used companding laws; namely the A-law used in Europe and the μ -law used in the USA, Canada and Japan. The μ -law compressor characteristics are shown in Fig.3.3 for various values of μ . The A-law characteristics are quite similar. Since these are logarithmic curves, these quantizers are also called as logarithmic quantizers or, log-quantizers.

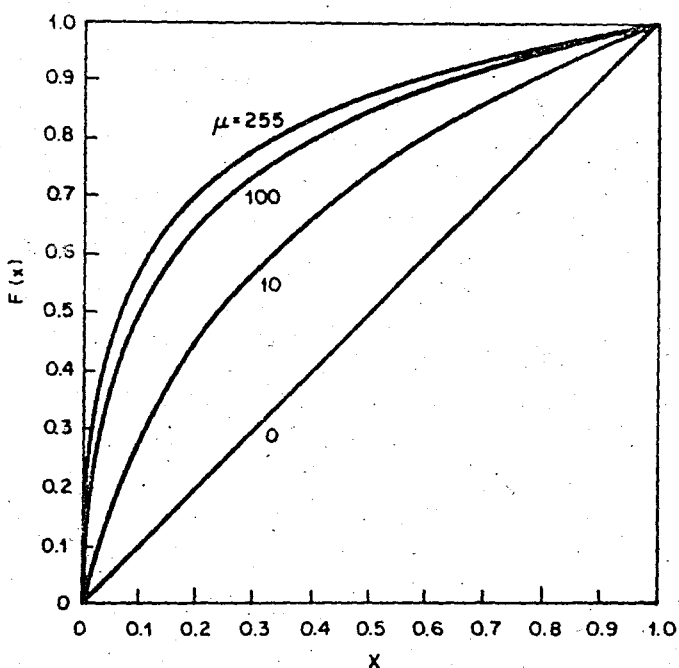


Figure 3.3. μ -law compressor curves (After Gersho [13]).

The input signal is first compressed at the transmitter and the compressed signal is quantized with a uniform quantizer. At the receiver the reconstructed signal is passed through a nonlinearity (expander) which is the inverse of the compressor curve. The overall effect is that of a quantizer with nonuniform spacings between reconstruction levels. In practice the logarithmic characteristics can be conveniently implemented by a piecewise linear all-digital method thereby avoiding the compressor-expander mismatch problem.

In this context it will be in order to mention other types of quantizers briefly. If the amplitude PDF of the input signal is known, it is possible to construct a quantizer (uniform or nonuniform) having minimum quantizing distortion according to some distortion measure. Usually the mean squared error is used as the distortion measure. The resulting quantizers are called PDF optimized or, in short, optimum quantizers [3,14]. If the input signal PDF matches the PDF that they were designed for, optimum quantizers provide somewhat greater SNR than log-quantizers. However, their idle channel noise is higher (i.e., first reconstruction level is greater) and dynamic range is smaller.

For nonstationary signals such as speech, the same performance can be achieved with fewer bits if the quantizer is adapted to changing signal levels. Note that this requires the quantizer to have memory as opposed to memoryless (or instantaneous) quantization methods discussed above. The idea in adaptive quantization is to decrease or increase the reconstruction level spacings according to some parameter of the signal (e.g. energy or instantaneous amplitude). To achieve the same effect one can envision a variable gain amplifier (whose gain is controlled by the signal) in front of a fixed quantizer. In adaptive quantization the exact quantizer characteristics is irrelevant. The important issue is the adaptation time constant for which there are two different approaches:

- a) Instantaneous adaptation,
- b) Syllabic adaptation.

Instantaneous adaptation is based on altering the quantizer levels at every sample whereas in syllabic adaptation the quantizer levels are altered at the syllabic rate, namely once in 10-40 ms.

Another important implementation problem is whether to base the adaptation algorithm on already received (Adaptive Quantization Backward, AQB) or yet to be transmitted (Adaptive Quantization Forward, AQF) samples. Block diagrams of AQF and AQB schemes are shown in Fig.3.4. Clearly, in case of AQF, it is necessary to transmit side information about the amplifier gain whereas AQB schemes do not require

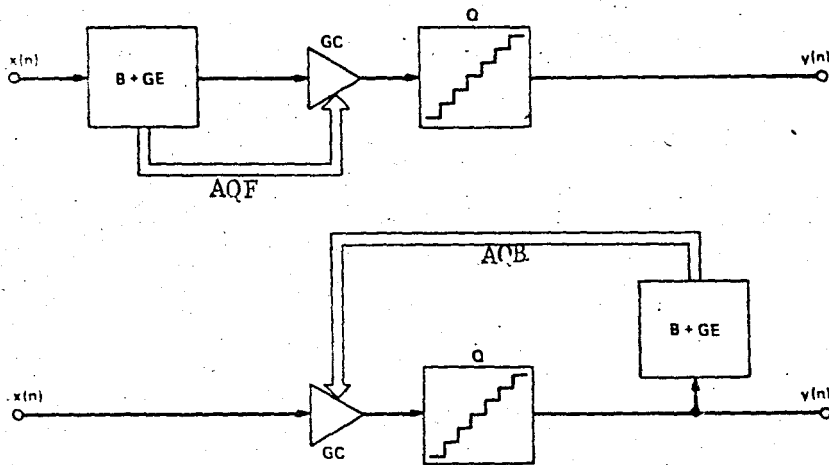


Figure 3.4. Forward adaptation (AQF) and backward adaptation (AQB). $B+GE$ = Buffer and gain estimation, GC = Gain control (After Noll [12]).

any side information. To keep the side information (which necessitates additional channel capacity) at a minimum, AQF schemes generally employ some sort of syllabic adaptation.

3.1.1. PCM with Forward Adaptive Quantization (PCM/AQF)

PCM/AQF is a natural extension of nonadaptive PCM. System block diagram is the same as Fig.3.4(a). In practice the speech signal is divided into frames of N_{SEG} samples each and buffered in the $B+GE$ block. Then an unbiased estimation of the variance of the frame is calculated as

$$G_N^{-2} = \alpha \frac{1}{\text{NSEG}} \sum_{j=1}^{\text{NSEG}} x^2(j) \quad (3.1)$$

where α is a coefficient which is varied to optimize the performance. The amplifier gain G_N is proportional to the inverse of the standard deviation estimated from the buffered frame.

3.2. Differential PCM (DPCM)

In PCM, every input sample is treated as a totally new and surprising event and except for amplitude distributions, no attempt is made to exploit the redundancy in the signal. Differential PCM (DPCM) attempts to exploit the correlation between input samples. The block diagram of a DPCM system is depicted in Fig.3.5.

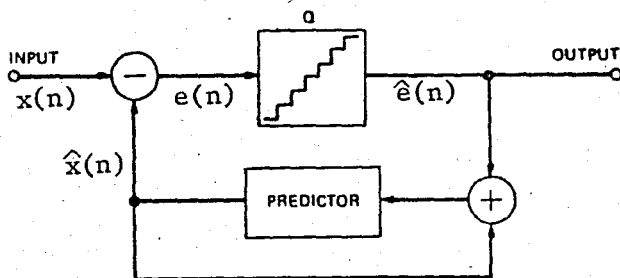


Figure 3.5. Differential PCM (After Noll [12]).

Instead of entering into the details of the theory of differential quantization, we shall qualitatively describe the operation of the system of Fig.3.5. The underlying idea in DPCM is to predict the current sample in some manner from past output samples. This predicted value is subtracted from the input sample and the difference (called the prediction error or simply error) is quantized

and transmitted. After receiving the quantized error sample the receiver adds to it its predicted sample (which is the same as the transmitter's) to get the output sample. Using the notation of Fig.3.5,

$$\begin{aligned} \text{SNR} &= \text{SQNR} + 10 \log_{10} \frac{\sigma_x^2}{\sigma_e^2} \\ &= \text{SQNR} + G_p \end{aligned} \quad (3.2)$$

where SQNR is the SNR of the quantizer. The second term in (3.2) is called the SNR improvement over PCM or prediction gain because had there been no prediction (straight PCM), total SNR would be the same as SQNR. If the prediction is "good" the variance of the error signal will be less than that of the input signal and the prediction gain G_p is positive.

The predictor is generally a linear predictor in the form

$$x(n) = \sum_{k=1}^p a_k x(n-k) \quad (3.3)$$

where p is the order of the predictor and a_k are the predictor coefficients. It is of interest to find the predictor coefficients in such a way that the prediction gain is maximized. It turns out that optimum predictor coefficients are the solution of the vector-matrix equation

$$\mathbf{a} = \mathbf{R}^{-1} \mathbf{r} \quad (3.4)$$

where

\mathbf{R} is the $p \times p$ matrix of the autocorrelation coefficients

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & \dots & r(p-1) \\ r(1) & r(2) & \dots & r(p-2) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ r(p-1) & r(p-2) & \dots & r(0) \end{bmatrix}$$

\mathbf{a} is the $p \times 1$ vector of predictor coefficients

$$\mathbf{a} = [a_1 \ a_2 \ \dots \ a_p]^T$$

\mathbf{r} is the $p \times 1$ vector of the autocorrelation coefficients

$$\mathbf{r} = [r(1) \ r(2) \ \dots \ r(p)]^T$$

There are various fast algorithms to calculate the vector \mathbf{a} [15].

3.2.1. Adaptive Quantization in DPCM

The principles of both forward (AQF) and backward (AOB) adaptive quantization mentioned in the context of PCM can be applied to adapt the DPCM quantizer. The block diagram of DPCM/AQF and DPCM/AOB schemes is shown in Fig.3.6.

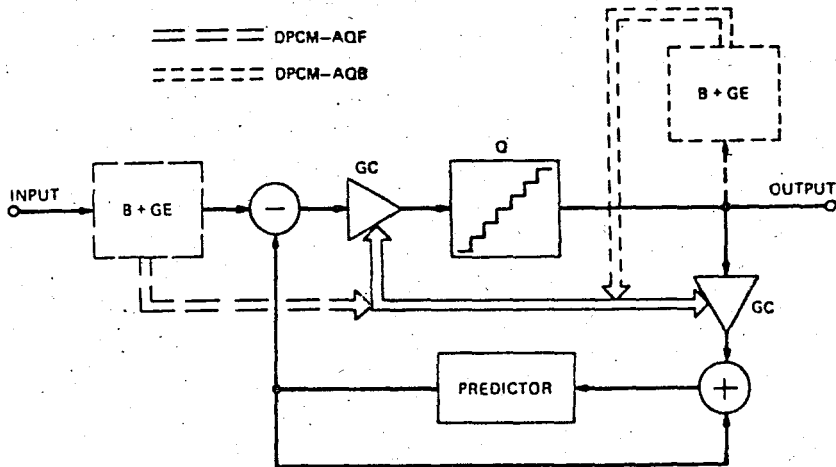


Figure 3.6. Adaptive quantization in DPCM (After Noll [12])

Evidently, the DPCM/AQF scheme requires to transmit side information about the quantizer status while DPCM/AOB doesn't. However, AQF schemes are, in general, more robust to the effects of channel errors (excluding errors in the side information, of course).

One AQB scheme widely used in DPCM is Jayant's one word memory algorithm [16]. This is an instantaneous adaptation algorithm which shrinks or expands the quantizer like an accordion according to the last reconstruction level occupied. Optimum step size multipliers for various number of quantizer levels are given in [16].

3.2.2. Adaptive Prediction in DPCM

Instead of a fixed predictor based on average signal statistics, we can adapt the predictor according to the short time behaviour of the signal. Thus, we get optimum instead of suboptimum performance. We shall call DPCM with an adaptive predictor as ADPCM. An ADPCM system is shown in Fig.3.7.

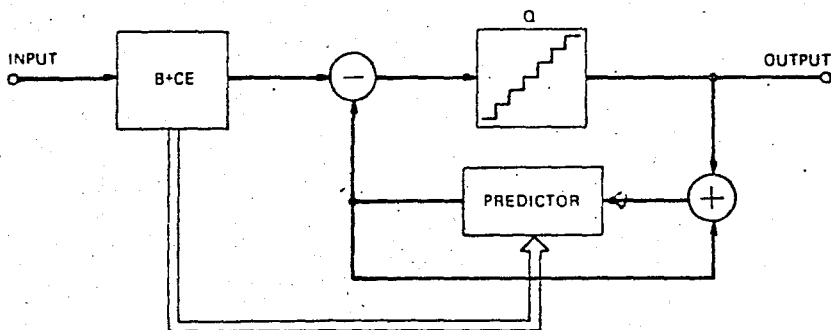


Figure 3.7. Adaptive prediction in DPCM (After Noll [12]).

In the system of Fig.3.7, a frame of input speech is first buffered in the B+CE (Buffer+Coefficient Estimation) block. Then, the autocorrelation coefficients and the predictor coefficients of the buffered signal are calculated. Although predictor adaptation brings considerable computational load, the results strongly favor adaptive predictors. SNR gain over PCM (G_p) versus predictor order is shown in Fig.3.8 for both fixed and adaptive predictors.

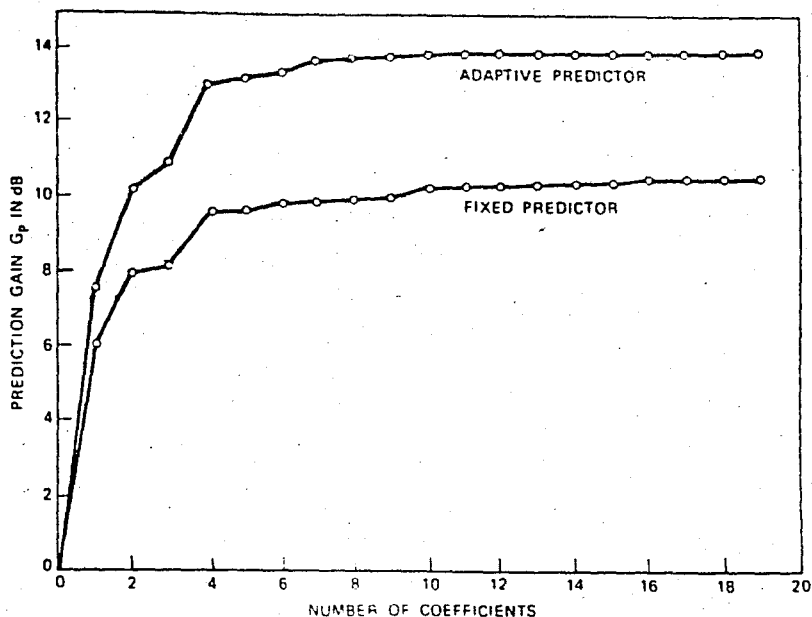


Figure 3.8. Prediction gain (G_p) vs. predictor order for fixed and adaptive prediction. Results are based on a single 2.3 second speech sample (After Noll [12]).

Prediction gain saturates at a predictor order of about four for both fixed and adaptive predictors. Also for $p > 4$ adaptive prediction has a prediction gain that is almost 4 dB greater than fixed prediction.

Another important point is that the curve for the fixed predictor shows the absolute maximum gains that can be achieved. If more than one speaker is considered, suboptimum predictor coefficients have to be used and this will lower the SNR. The adaptive predictor performance on the other hand, will almost remain the same.

The ADPCM system of Fig.3.7 uses a forward adaptation scheme and necessitates the transmission of side information. There are backward predictor adaptation schemes that require no side information. These systems employ sequential predictor adaptation algorithms instead of the block adaptive method of Fig.3.7 [17].

3.2.3. Adaptive Prediction and Adaptive Quantization in DPCM

The ultimate DPCM system is, of course, the one that combines the advantages of both predictor and quantizer adaptation. Fig.3.9 presents the block diagram of such a system.

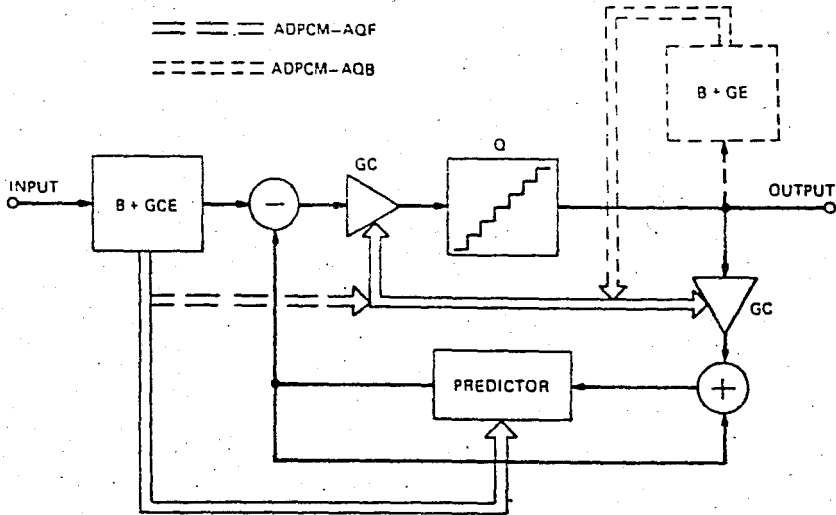


Figure 3.9. DPCM with adaptive prediction and adaptive quantization (After Noll [12]).

To be consistent with our notation we shall call such systems either as ADPCM/AQF or ADPCM/AQB depending on the type of quantizer adaptation. The number after the abbreviations DPCM and ADPCM will denote the predictor order, e.g. DPCM3/AQB stands for DPCM with a fixed third order predictor and a backward adaptive quantizer.

Before concluding this section we should note one final point about DPCM. To have the same prediction both at the encoder and the decoder, the prediction in DPCM is based on quantized output samples. Consequently, it is a noisy prediction. This is why the SNR gain readily saturates with increasing predictor order as shown in Fig.3.8.

This effect, called predictor-quantizer interaction manifests itself especially when the number of quantizer levels is less than 16. It is therefore not very meaningful to design a system with a 12th order predictor and a 2 bit quantizer.

3.2.4. Delta Modulation (DM)

Delta Modulation (DM) is a special case of DPCM with a 2 level (1-bit) quantizer. Since a 2 level quantizer is not adequate to provide a reasonable SNR, delta modulators operate at many times the Nyquist rate. At those rates the correlation between successive samples is much more pronounced and a 2 level quantizer is sufficient to achieve high SNR. Since only a single bit is transmitted for each sample, the sampling rate of delta modulators is equal to the bit rate.

To overcome the limitations of the basic (linear) DM, many adaptive versions have been introduced and DM, mainly because of its simplicity, has become a major competitor of PCM.

IV. THE ASYNCHRONOUS DELTA MODULATOR

The Asynchronous Delta Modulator (ASDM) is a modified delta modulator. In ASDM there is no system clock, hence the sampling instances are nonuniform. Actually, ASDM sampling instances are determined by the activity of the signal itself and the information in the signal is coded into the output pulses as well as the time intervals between these pulses. We shall refer to the latter as the inter-bit interval (IBI) sequence in the rest of this thesis. Obviously the decoder needs to know the inter-bit intervals as well as the transmitted pulse polarities to reconstruct the signal. The ASDM encoder, a typical input waveform, the reconstructed signal and the output pulses are shown in Fig.4.1.

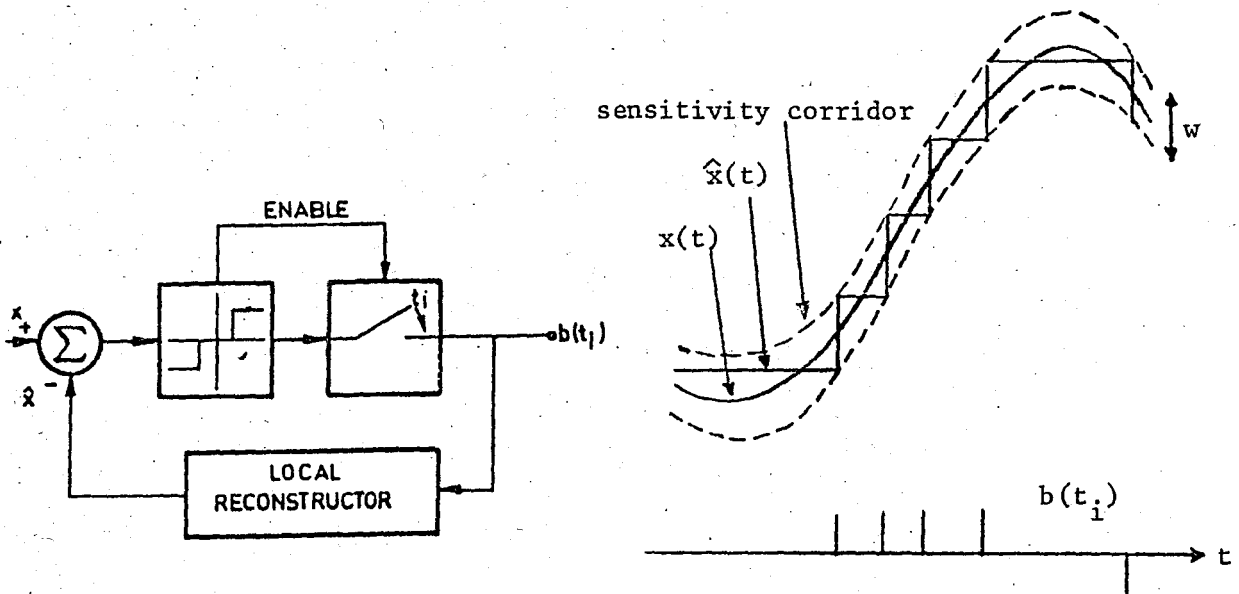


Figure 4.1. ASDM encoder and waveforms.

As shown in Fig.4.1, when the magnitude of the difference between the input and reconstructed signals exceeds a certain value, $w/2$, the coder transmits a bit indicating the direction of the change. Consequently, few samples are taken when the input signal is idle whereas the active regions of the signal are encoded with many samples. Since, in practice, many information sources such as speech and image waveforms contain idle regions with sudden bursts of energy, one expects this nonuniform sampling approach to result in a reduction of bandwidth. Indeed, it has been shown that variable rate sampling as in ASDM yields higher signal to noise ratios than fixed rate sampling when the input is nonstationary such as speech [18,19].

On the other hand, variable rate sampling necessitates a buffer to take up the slack between the synchronous channel and the asynchronous coder. In ASDM, the continuous inter-bit intervals have to be quantized, encoded and transmitted along with the bit polarity sequence to enable reconstruction at the receiver. The quantization and encoding of the IBI sequence forms the core of this thesis. The block diagram of a complete ASDM system is shown in Fig.4.2.

Another important property of ASDM is that it is not matched to any signal, i.e., it can, at least theoretically, encode any signal with the same performance. This is the sole property that prompted us to consider ASDM as an effective source coder both for speech and voiceband data signals. However, in practice the IBI quantization issue may prevent the achievement of comparable performance if the signals to be encoded have significantly different statistics.

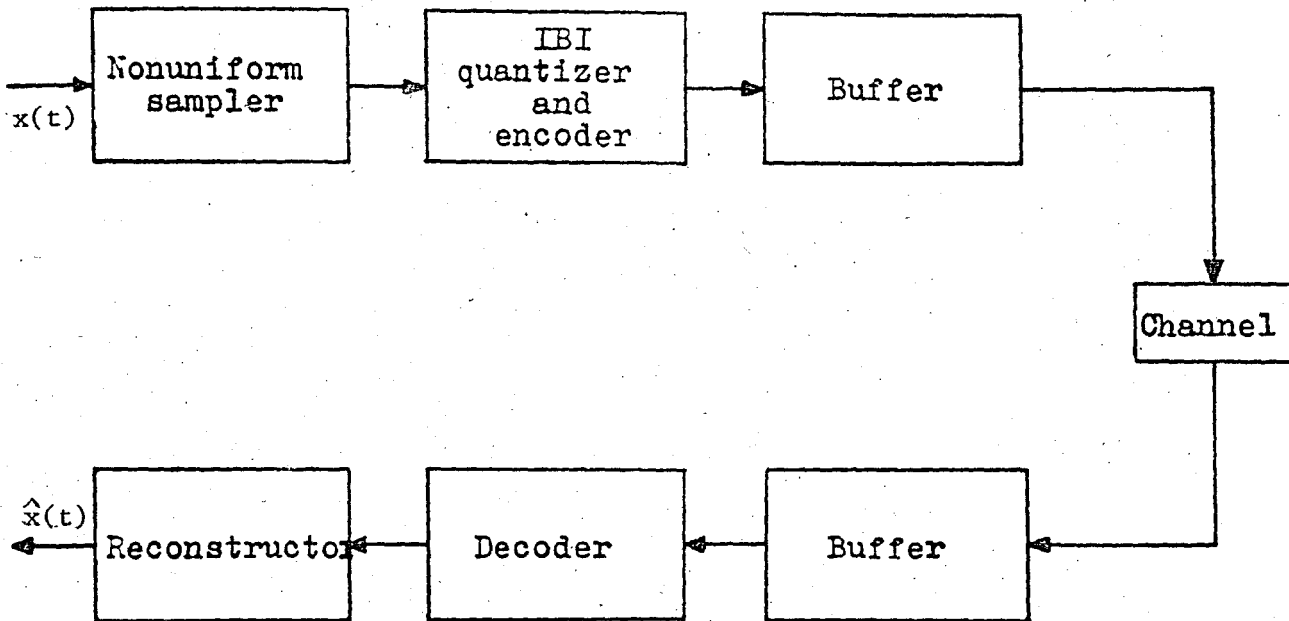


Figure 4.2. Complete ASDM system.

4.1. Analysis of Asynchronous Delta Modulator

Compared to synchronous delta modulation and its variations, ASDM is a much less documented source coder. Previous work on ASDM can be found in [20-24]. This section is intended to give an overview of ASDM and summarize the previous results. Although there are various modifications of the basic ASDM encoder [22], we shall only be concerned with the system of Fig.4.1 which is called corridor-ASDM or free running ASDM in the literature. The former name stems from the fact that one can envision a sensitivity corridor of width w around the input signal. The encoder transmits a pulse only when the reconstructed signal touches one of the corridor walls.

4.1.1. ASDM Sampling Statistics

ASDM can be visualized as an encoder which converts the original input process into another process (namely, the IBI process) before encoding. It is therefore of interest to study the statistics of the IBI sequence.

The IBI process is closely related to the level crossing problem of random processes. The sampling instants are nothing but the level crossings of the input process with equi-spaced levels l_n where

$$l_n = w/2 (2n-1) \quad n = 1, 2, \dots \quad (4.1)$$

The inter-bit intervals can be treated as the first passage times of these levels by the input process. However, these problems are very difficult to handle and analytic results exist only for some special input processes such as the Gaussian process [25]. Using Davenport's [26] experimental results, Sankur and Gungen [21] have derived the IBI PDFs for common speech models.

Another quantity of interest is the average number of crossings (or samples) per second which influences the channel bit rate of the ASDM encoder. This value can, of course, be estimated from the IBI PDF if it is known.

In this study we shall not be concerned more with the theoretical aspects of the sampling statistics. Interested readers are referred to [21] and [22] for details.

4.1.2. Signal to Quantization Noise Ratio

Before discussing the signal to quantization noise ratio (SQNR) it will be in order to define some parameters of the ASDM encoder. The corridor width, w , is generally chosen as a fraction of the rms value of the input signal. We define the k -factor as the ratio of the number of ASDM encoder samples to Nyquist samples over a suitable observation interval. Clearly, the k -factor depends on the input process as well as the corridor width w .

Although shown equal to w in Fig.4.1, the step size of the reconstructor can be assigned other values and may even be made adaptive. The optimum (SQNR maximizing) value of the step size as reported by Gungen [22] and confirmed by our simulations is about 90 percent of the corridor width. However, this value of the step size causes an increase in the number of average encoder samples and the improvement is small (less than 1 dB for all corridor widths we studied). We therefore used a step size of $0.99w$ in all our simulation studies.

Assuming no slope overload, i.e., the minimum sampling time of the encoder is small enough to respond to fast changing input signal sections without overload, the error signal may be assumed to be uniformly distributed between $-w/2$ and $w/2$. Then, the quantization noise power is given by:

$$\sigma_n^2 = w^2/12 \quad (4.2)$$

If w is expressed in terms of the rms value of the signal as

$$w = p \sigma_x$$

then, the signal to quantization ratio is given by

$$\text{SQNR} = \frac{12 \sigma_x^2}{w^2} = \frac{12 \sigma_x^2}{p^2 \sigma_x^2} = \frac{12}{p^2} \quad (4.3)$$

which constitutes an upper bound on SNR for the ASDM encoder.

The spectrum of the error signal depends on the corridor width and the input process and is reasonably flat in the frequency range of interest. The error spectra for various values of the corridor width are shown in Fig.4.3.

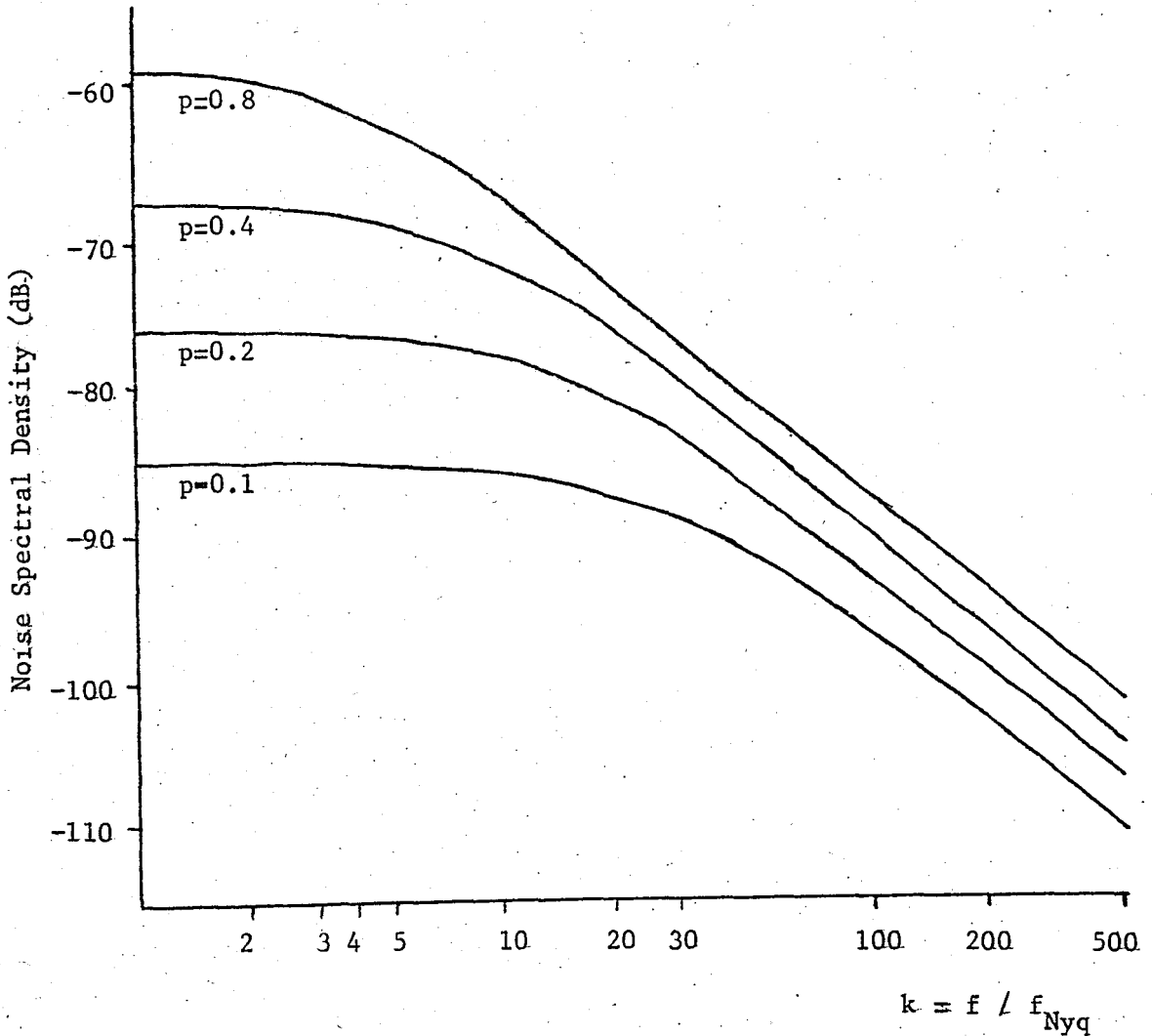


Figure 4.3. Quantization error spectra for various values of the corridor width (After Gungen [22]).

V. SIMULATION RESULTS

The simulation studies performed are presented in this chapter. Although remarked when necessary, the reader is reminded to interpret the results given in this chapter with caution because simulation results and field performance are different things. This is especially true for the speech case because we had to use an artificial signal and evaluation criteria were SNR based (not subjective).

In all simulations the input signal was scaled to have unity rms value. About half a second (4096 samples) of artificial speech was coded in all speech encoder simulations. For the data signal on the other hand, about 1/8 seconds was deemed adequate because the data signal characteristics do not show substantial variation with time.

This chapter starts with descriptions of the artificial speech and QPSK data signal characteristics and their generation. Then, simulated speech coders are briefly described and their performances with both speech and data signals are given. ASDM encoder simulations are discussed next in some detail. After presenting the performance results of ASDM with speech and data, results of entropy coding of the inter-bit intervals are given. The chapter concludes with simulation results of the buffer behaviour for various corridor widths.

5.1. The Artificial Speech Signal

Since the subject of this thesis is to evaluate the performance of a potential speech coder, we need digitized speech to run the simulation programs. However, because of hardware problems we had to use an artificial signal. Speech coders are usually tested using sinusoids or bandlimited Gaussian noise although their temporal and spectral characteristics are quite different from that of speech. We therefore searched for an artificial signal which closely mimics the temporal and spectral characteristics of speech. The work described below is based on two papers by Modena et.al. [27] and Billi and Scagliola [28].

5.1.1. The Artificial Speech Generation Model

The artificial speech generation model is based on the classical model of Fig.2.1, but there is a modification. Only the voiced excitation is considered because the effects of temporal adaptation strategies can be more easily observed with high energy, voiced segments. The block diagram of the model is shown in Fig.5.1.

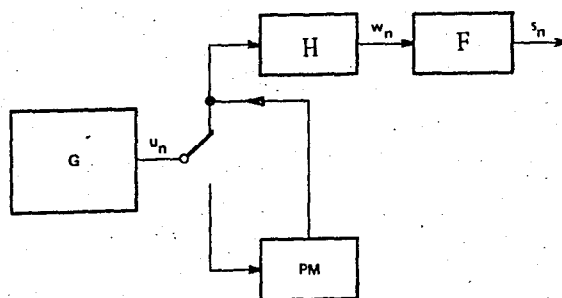


Figure 5.1. Block diagram of the artificial speech generation model (After Modena et.al. [27]).

The filter F is an all pole model of an average configuration of the vocal tract, and G is the generator of the excitation signal $u(n)$. Since the contributions of the glottal pulse, radiation and the vocal tract are all included in the filter F , in general the sequence $w(n)$ must have a white spectrum. This leaves us with periodic impulse like input signals because we are considering only voiced excitation. However, it is noted in [27] that using impulses as inputs to F results in unreasonably high amplitudes in the output signal at pitch epochs. Therefore, alternative excitation signals are used and since, in general, their spectrum is not white, H is a whitening filter.

In z -transform notation, with reference to Fig.5.1 :

$$W(z) = H(z)U(z)$$

$$S(z) = F(z)W(z)$$

$$H(z) = \frac{1}{|U(z)|}$$

and, after substitution

$$S(z) = F(z) \frac{U(z)}{|U(z)|}$$

5.1.2. The Glottal Excitation

The Glottal pulse is generated according to Rosenberg's model [29]. Specifically, the waveform used in our procedure is given by:

$$u(n) = \begin{cases} 0 & 1 \leq n \leq 32 \\ 1/2 [1 - \cos(\pi(n-32)/12)] & 33 \leq n \leq 44 \\ 1/2 [1 + \cos(\pi(n-44)/6)] & 45 \leq n \leq 50 \\ 0 & 51 \leq n \leq 64 \end{cases}$$

The synthesized glottal pulse has a rise time of 12 samples and a fall time of 6 samples and is shown in Fig.5.2. Its period is 64 samples and for the 8 kHz sampling rate we are working with, this corresponds to a pitch frequency of 125 Hz.

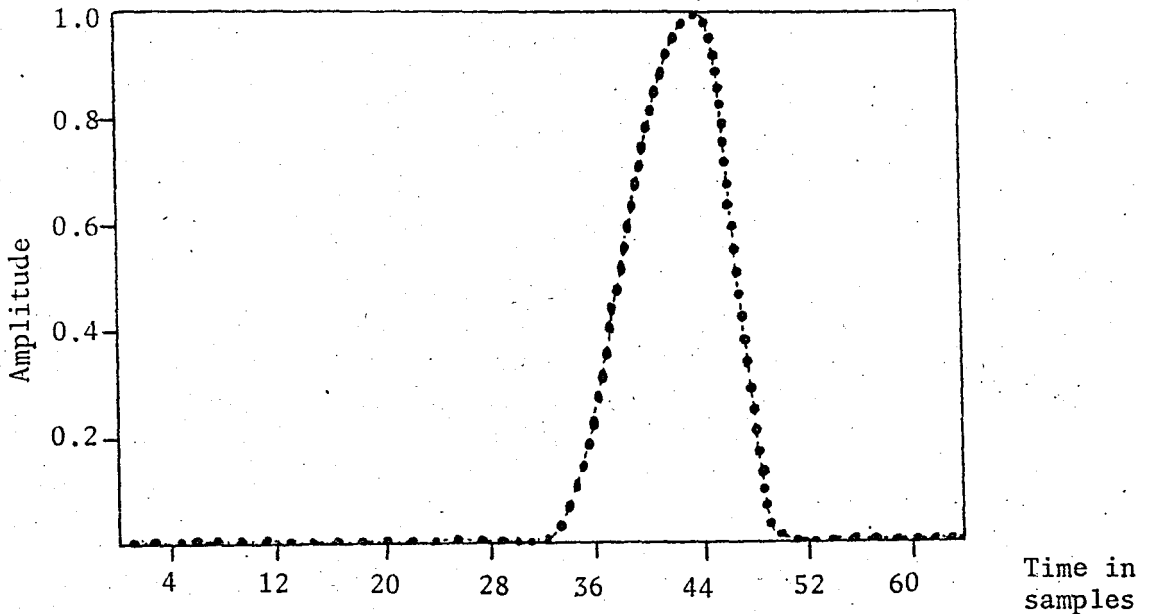


Figure 5.2. The glottal pulse waveform.

5.1.3. Glottal Pulse Whitening Filter

The glottal pulse is definitely a lowpass waveform as shown in Fig.5.3 and we need a filter (H in Fig.5.1) to whiten its spectrum without altering its phase structure.

A linear phase, symmetrical impulse response FIR design was considered. For such a filter,

$$h(n) = h(-n)$$

$$\begin{aligned} H(e^{j\omega}) &= \sum_{n=-N}^N h(n)e^{j\omega n} \\ &= h(0) + \sum_{n=1}^N 2h(n)\cos(\omega n) \end{aligned}$$

Noting that this expression is similar to a truncated Fourier series expansion with $h(n)$ being the series coefficients, we expanded the inverse of the magnitude of the glottal pulse spectrum in a Fourier series. Then, using a conventional trapezoidal integration routine, the series coefficients (or the impulse response of the filter) were determined. Determination of the filter order was rather empirical, based on the success of the inverse filter and the significance (magnitude) of series coefficients. We used a 31st order filter. The whitened glottal pulse spectrum is shown along with the original glottal pulse spectrum in Fig.5.3.

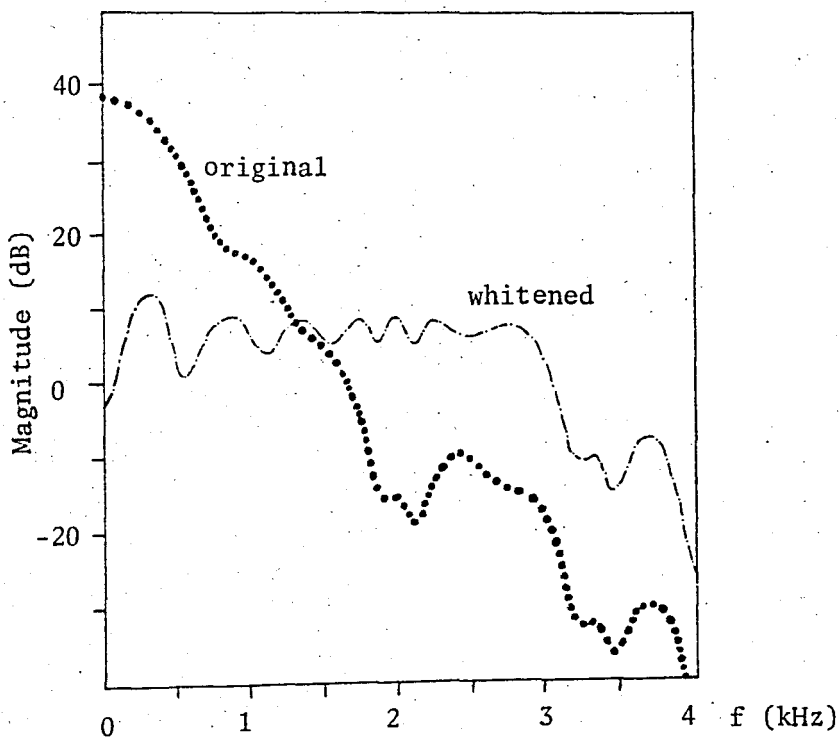


Figure 5.3. Spectra of the glottal pulse and the whitened glottal pulse.

Note that because the glottal pulse has very little energy at high frequencies, a 31st order filter is not truly sufficient to fill the ditch above 3 kHz. We shall comment on this later in the context of the vocal tract filter.

5.1.4. The Vocal Tract Filter

The average long term spectrum of 300-3400 Hz filtered and 8 kHz sampled speech of 10 speakers and its normalized autocorrelation function have been plotted in [27] and are reproduced in Fig.5.4. The numerical values of the autocorrelation function to be used to compute the vocal tract filter coefficients were obtained from Fig.5.4.(b). This is unfortunately an inaccurate procedure and we shall see its consequences very soon.

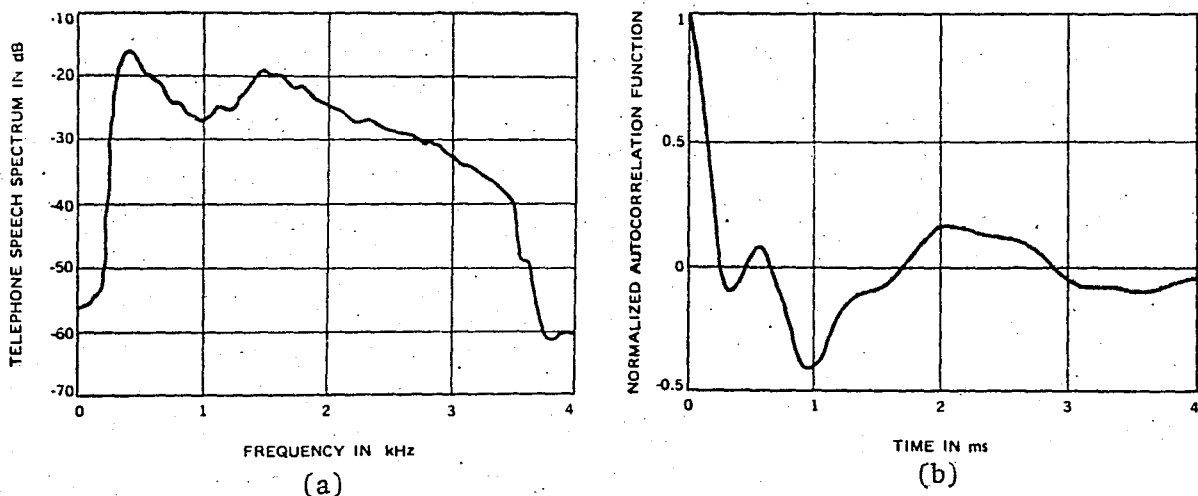


Figure 5.4. (a) Average long term spectrum and (b) normalized autocorrelation function of telephone speech of 10 speakers (After Modena et.al. [27]).

The coefficients of the all pole (LPC model) vocal tract filter were obtained by Levinson's algorithm [30]. Although Modena et.al. have used a 30th order filter, it was impossible to obtain a stable 30th order filter with our ~~inaccurate~~ autocorrelation values. Thus, a 16th order filter was deemed adequate. The modulus of the transfer function of the 16th order vocal tract filter is shown in Fig.5.5.

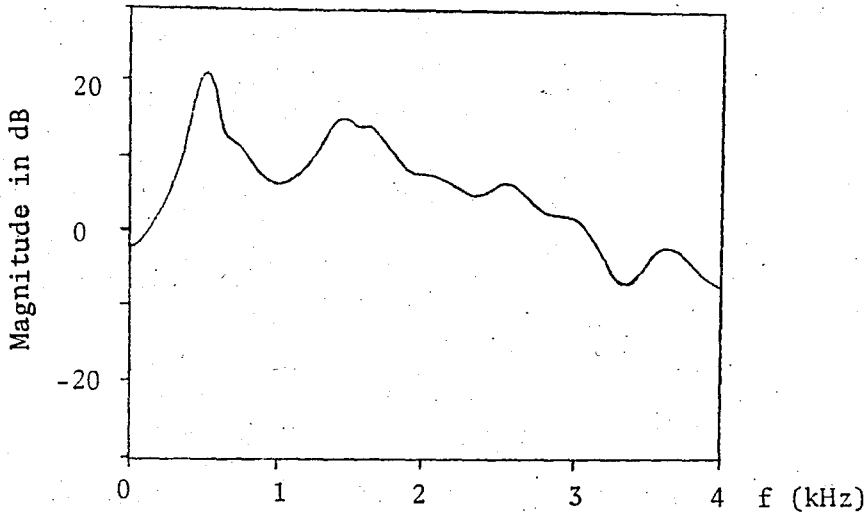


Figure 5.5. Modulus of the transfer function of the 16th order vocal tract filter.

When compared with the original spectrum of Fig.5.4.(a), the spurious high frequency response is evident. However this, when coupled with the poor high frequency response of the whitening filter, gives acceptable overall characteristics.

5.1.5. Pseudorandom Variation

The artificial signal as obtained from the combination of the three units described above is a purely periodic signal. This gives a line spectrum in the frequency domain and a poor amplitude histogram. In order to make the characteristics of the artificial signal more similar to that of speech, pseudorandom variation of the excitation amplitude and pitch period was attempted (the PM block in Fig.5.1). The time waveform of the artificial signal, its spectrum and amplitude histogram after pseudo random variation are shown in Figures 5.6, 5.7, and 5.8, respectively.

Apparently, there is a good match between Figs.5.4.(a) and 5.7, and the amplitude histogram is reasonably well behaved.

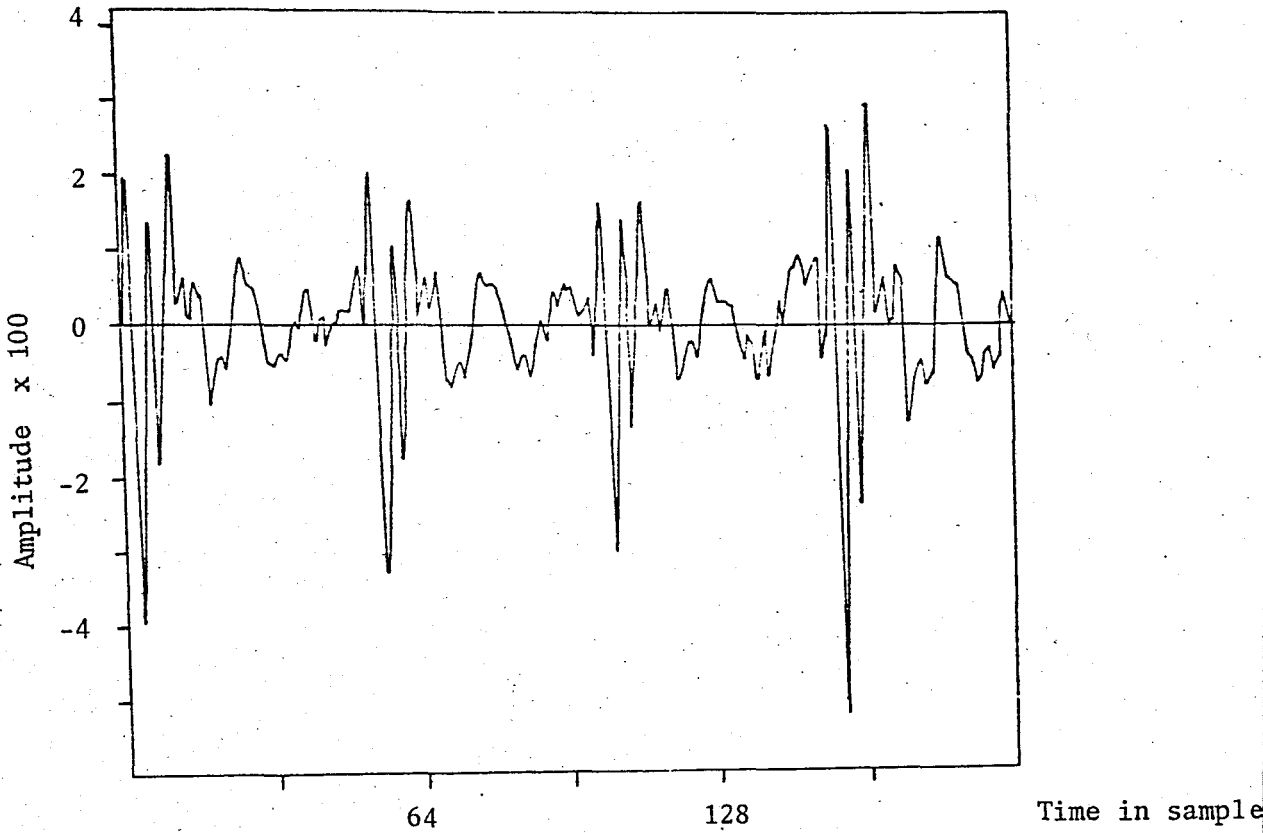


Figure 5.6. Time waveform of the artificial speech signal.

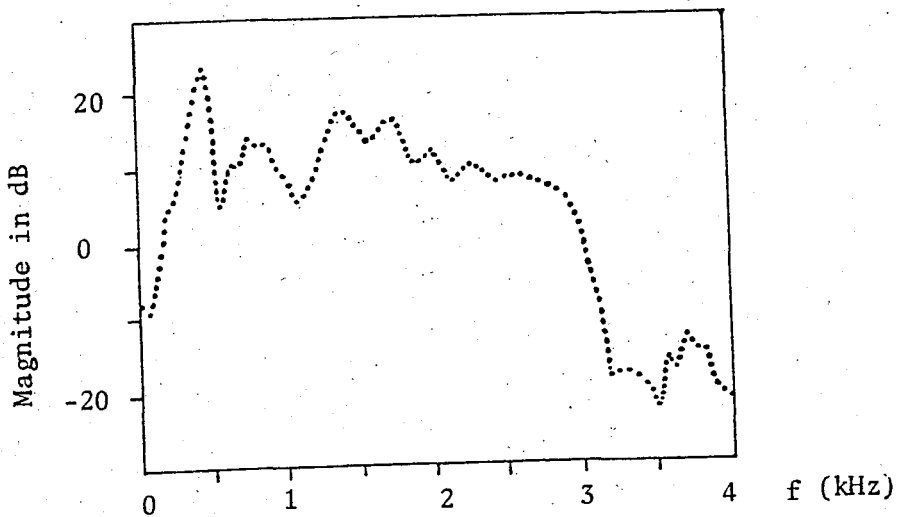


Figure 5.7. Spectrum of the artificial speech signal.

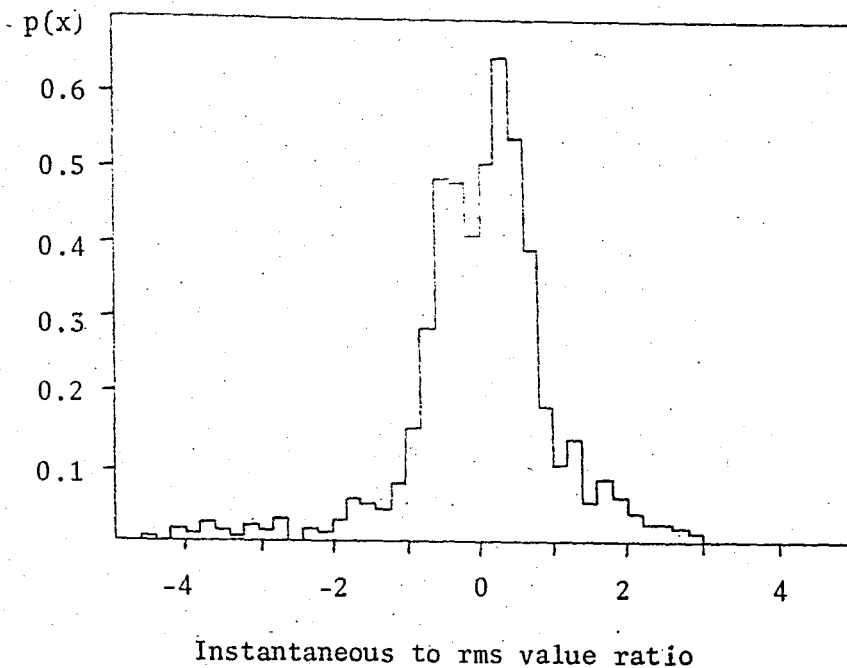


Figure 5.8. Amplitude histogram of the artificial speech signal.

5.1.6. Envelope Modulation

As a final improvement, it was decided to add a mathematically defined envelope to the signal which roughly resembles the alternation of voiced and unvoiced segments. A suitable envelope in the shape of a Hamming window has been derived in [28]. Specifically, the envelope used in our simulations is :

$$p(n) = \begin{cases} 0.54 - 0.46 \cos[2\pi(n-1)/(NH-1)] & 1 \leq n < NH/2 \\ 1 & NH/2 \leq n < NH/2 + NC \\ 0.54 + 0.46 \cos[2\pi(n-1-NC)/(NH-1)] & NH/2 + NC \leq n < NH + NC \\ 0.08 & NH + NC \leq n < NH + NC + NZ \end{cases}$$

where $NC = NH = 400$ and $NZ = 200$ samples.

Envelope modulation is obtained by multiplying the artificial signal with the mathematically defined waveform $p(n)$ given above. Adding the envelope significantly improves the already acceptable amplitude histogram especially at near-zero amplitudes. Because of its Hamming window shape, the envelope provides gradual transitions between low and high energy segments and causes almost no alteration in the signal spectrum.

As a final check on the -now enveloped- artificial speech signal, its autocorrelation coefficients are compared with the original coefficients of Modena et.al. (Fig.5.4.(b)). As revealed by Fig.5.9, the agreement is quite good.

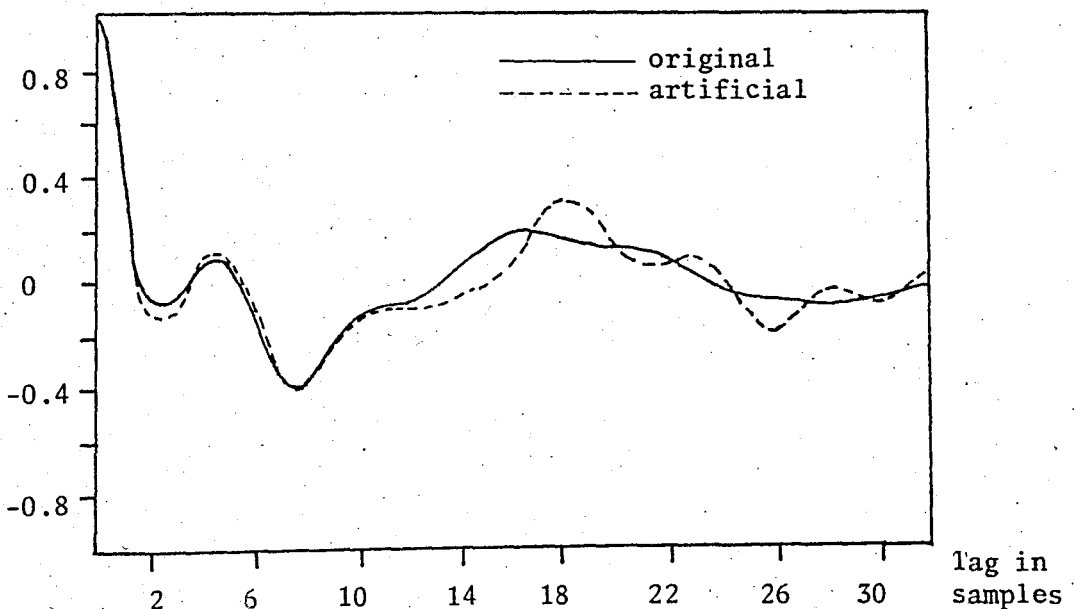


Figure 5.9. Autocorrelation coefficients of the artificial signal and original speech.

5.2. The Data Signal

In this section the characteristics of the simulated data signal are reviewed. As evidenced by Fig.2.6, most modems operating at rates above 1200 bits/second (bps) use some sort of PSK. Therefore, PSK was chosen as the modulation type. This also enables us to make quantitative comparisons between the performance of ASDM and O'Neal's coders [9].

The simulated data signal is a QPSK signal with a symbol rate of 1200 bauds which corresponds to a bit rate of 2400 bps. The carrier frequency was chosen as 1800 Hz. The signal is of the form

$$s(t) = \sum_n A E(t-nT) \cos[\omega_c(t-nT) + \phi_n] \quad (5.1)$$

where

A is the peak amplitude

E(t) is the envelope

T is the symbol period (1/1200 seconds)

ω_c is the carrier frequency (2π 1800 rad/sec)

ϕ_n is either 45, 135, 225 or 315 degrees.

The data envelope used is defined as

$$E(t) = \cos^2(\pi t/2T) \quad -T < t < T \quad (5.2)$$

Strictly speaking, this baseband waveform is not bandlimited, but has negligible content for $f > 1200$ Hz. This envelope has been used in Bell System model 201, 205 and 207 data sets [31].

We have assumed coherent detection, i.e., the frequency and phase of the carrier as well as symbol timing are known exactly at the demodulator. The block diagram of the whole system is given in Fig.5.10

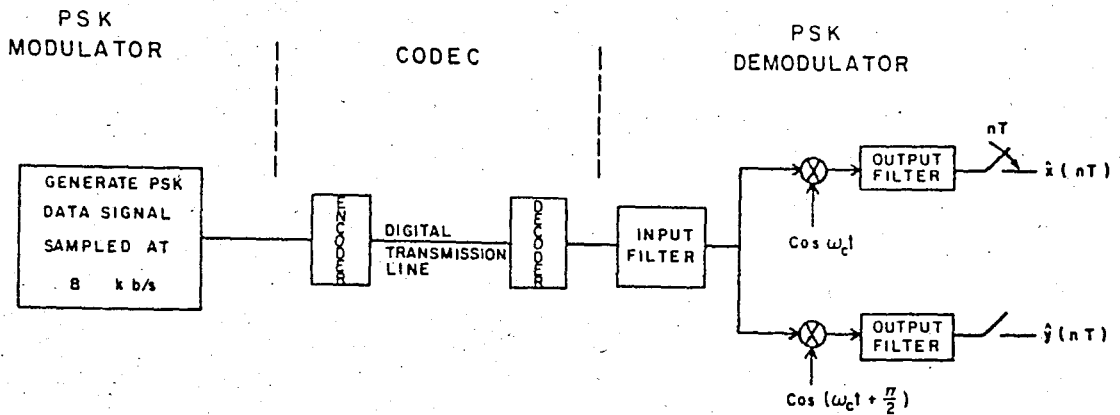


Figure 5.10. Simulation block diagram of the voiceband data encoding/decoding process (after O'Neal [9]).

The demodulated phase is given by

$$\hat{\phi}(nT) = \tan^{-1} \frac{\hat{y}(nT)}{\hat{x}(nT)} \quad (5.3)$$

We define the phase error

$$\delta(nT) = \phi(nT) - \hat{\phi}(nT) \quad (5.4)$$

as the difference of the transmitted and received phases. One problem arises because the demodulation process has to be carried out at the exact peaks of the data envelope. Since the sampling rate of the encoder is neither synchronous with, nor a multiple of the symbol rate, the two clocks slip with respect to each other. This effect is shown in Fig.5.11. The quantity Δ_n is a random variable which varies between 0 and 2π and is called the "sampling phase".

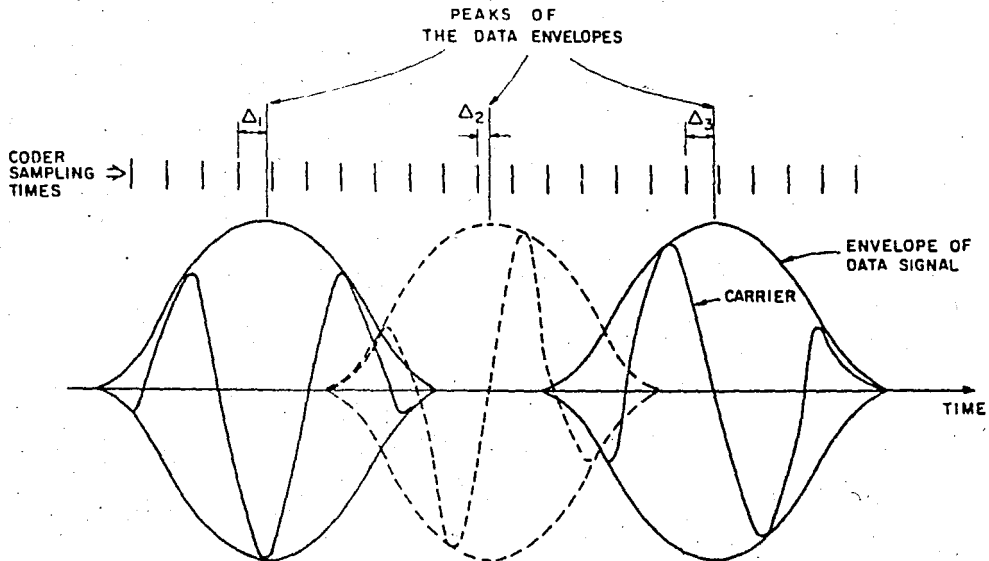


Fig.5.11. The sampling phase effect (After O'Neal [9]).

To enable demodulation at the exact peaks of the data envelope, interpolation between sample values must be made. We have used simple linear interpolation. O'Neal et.al. [9] have used $\sin x/x$ interpolation for 8 kHz sampled signals. Although we have tried $\sin x/x$ interpolation it brought less than 50 percent improvement over linear interpolation and computational issues motivated us to prefer the latter.

5.3. Speech Encoder Simulations

In order to be able to make quantitative comparisons between existing speech coders and ASDM, the following coders were simulated on the digital computer. For each encoder, 4 to 32 quantizer levels (2 to 5 bits) were considered which corresponds to transmission rates of 16 to 40 kbps.

- PCM : Nonadaptive μ -100 quantizer with 8 σ_x overloading.
- PCM/AQF : Frame length NSEG=32, optimum nonuniform Gauss quantizer.
- DPCM1/AQB : First order fixed predictor, Jayants' one word memory quantizer adaptation algorithm.
- ADPCM1/AQF : First order adaptive predictor, frame length NSEG=32, optimum nonuniform Gauss quantizer.
- ADPCM4/AQF : 4th order adaptive predictor, frame length NSEG=128, optimum nonuniform Laplace quantizer.
- ADPCM12/AQF : 12th order adaptive predictor, frame length NSEG=256, optimum nonuniform Gamma quantizer.

In all forward adaptive schemes, the need for additional channel capacity to transmit side information was not taken into account. Noll [12] gives an approximate formula to transform this additional channel capacity into an equivalent reduction in SNR. If each parameter is to be encoded with NADD bits/frame, then the equivalent loss in SNR is

$$\text{SNR}_{\text{loss}} = 6.02 \frac{\text{NADD (bits/frame)}}{\text{NSEG (samples/frame)}} \quad (\text{dB}) \quad (5.5)$$

5.3.1. Speech Encoding Performance of Simulated Encoders

The SNR based performance of the above encoders have been evaluated using the artificial speech signal and plotted in Fig.5.12 along with Noll's results [12]. Comparing our results with Noll's, we observe that the performances of PCM and PCM/AQF schemes are almost the

same while our results are inferior in all DPCM schemes. The difference is as much as 9 dB for the 16 kbps ADPCM12/AQF case. We attribute this difference to the rather low first autocorrelation coefficient of the artificial speech signal. Indeed, for a first order predictor the prediction gain is given by

$$G_p = \frac{1}{1-r(1)^2} \quad (5.6)$$

which corresponds to approximately 1.3 dB for the artificial speech whose first autocorrelation coefficient is about 0.5. On the other hand, the first autocorrelation coefficient of speech is assumed to be around 0.85 in the literature (this value is almost the same as that of Noll's speech data) which gives a prediction gain of 5.5 dB.

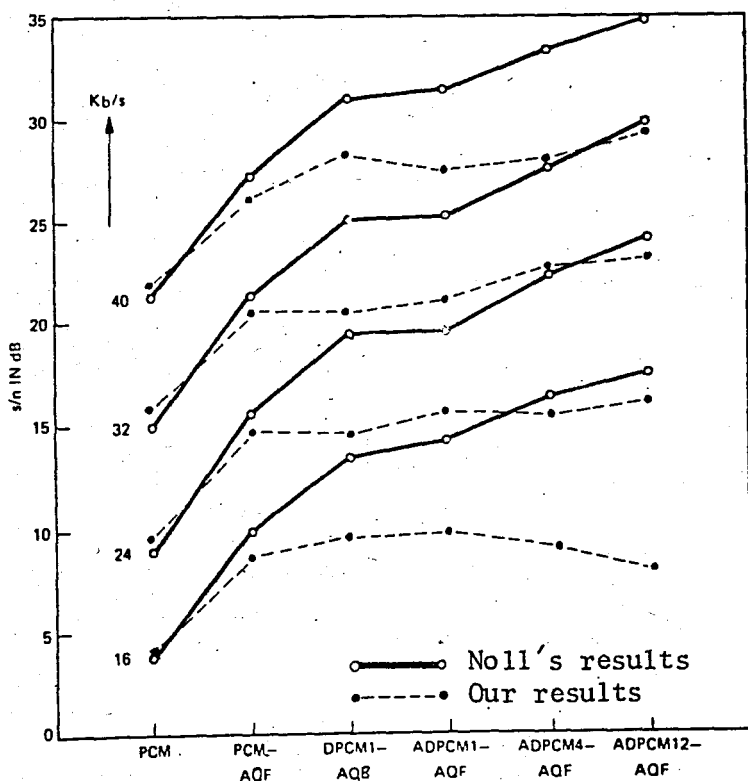


Figure 5.12. Performance of simulated speech encoders along with the results obtained by Noll [12].

Another important discrepancy between our and Noll's results is the performance of ADPCM12/AQF scheme. His results indicate improvement over ADPCM4/AQF at all bit rates while our results show a significant improvement over ADPCM4/AQF at only 40 kbps. We believe that in this aspect our results are equally trustable. A 12th order predictor should not bring much (if any) improvement over a 4th order predictor at 16 (and possibly 24 kbps) because of the very noisy prediction.

One of the interesting results in common is the good (and signal independent) performance of the very simple PCM/AQF scheme. Another is that at all bit rates almost nothing is gained by adapting a first order predictor.

5.3.2. Data Encoding Performance of Simulated Encoders

Errors in voiceband data may be caused by one or a combination of the following factors:

- Random noise
- Impulse noise
- Amplitude distortion
- Phase distortion
- Nonlinear distortion, etc.

The sensitivity of speech and data to these impairments are different. For example, speech is relatively insensitive to amplitude and phase distortion while data signals are quite fragile in the presence of these two impairments. In our simulations none of the above impairments were taken into account. If one or more of them are present, the performances of the encoders will degrade.

The encoders described before were used to encode (and decode) the QPSK data signal. Results for bit rates of 16, 24, 32 and 40 kbps are tabulated in Table 5.1. None of the simulated schemes made an error at any rate. An error is, of course, defined as a phase error greater than 45 degrees.

Type of coder	Simulation SNR	Maximum phase error	Variance of phase error	Mean phase error	SNR _e	Quantize levels
PCM	-0.1	32.3	148.6	1.8	10.4	4
	11.5	17.5	59.7	1.7	14.4	8
	14.2	11.1	27.8	2.8	17.7	16
	20.9	10.7	19.3	1.5	19.3	32
PCM /AQF	12.0	19.3	78.1	3.1	13.2	4
	15.5	10.4	27.1	0.3	17.8	8
	23.0	10.0	10.9	-0.4	21.8	16
	27.0	7.6	8.4	-0.1	22.9	32
DPCM1 /AQB	7.5	33.1	197.6	2.2	9.2	4
	13.4	20.2	33.6	0.6	16.9	8
	19.1	11.2	17.2	0.0	19.8	16
	23.8	9.5	15.0	0.7	20.4	32
ADPCM1 /AQF	12.2	18.0	32.5	-0.6	17.0	4
	15.6	11.2	21.6	0.9	18.8	8
	20.8	8.5	11.3	0.2	21.6	16
	27.5	7.2	6.8	0.2	23.9	32
ADPCM4 /AQF	11.4	13.2	23.1	0.0	18.5	4
	18.0	10.7	13.0	0.1	21.0	8
	22.1	9.4	10.4	0.2	22.0	16
	29.8	7.8	7.9	0.2	23.2	32
ADPCM12 /AQF	9.5	20.9	68.9	-0.7	13.8	4
	15.4	16.4	19.7	0.6	19.2	8
	20.0	7.3	7.9	0.2	23.1	16
	25.7	6.4	7.1	0.4	23.6	32

Table 5.1. Results of the data encoding performance of simulated speech coders. All values are in degrees except SNRs (dB).

In the DPCM1/AQF scheme the -fixed- predictor coefficient was chosen as the first autocorrelation coefficient of the artificial speech signal instead of that of data. However, since this coefficient

is low for artificial speech it is close to that of data and using the first autocorrelation coefficient of data brought little improvement. If the predictor coefficient is assigned a value of about 0.85, i.e., the first autocorrelation coefficient of real speech, the data encoding performance of the DPCM1/AQF scheme will become worse.

Although O'Neal et.al. [9] have not investigated the data encoding performances of exactly the same coders, it seems that there are no gross differences between his results and those presented here. In particular, note the poor performance of the DPCM1/AQB scheme. Although there is almost no difference between the performances of DPCM1/AQB and ADPCM1/AQF for speech, ADPCM1/AQF by far outperforms DPCM1/AQB when the data signal is coded. Apparently, the quantizer adaptation strategy accounts for this fact. O'Neal concludes that instantaneous quantizer adaptation algorithms designed for speech are not appropriate for data [6]. Because speech has a high crest factor, optimum quantizer step size for speech is too large for data. Especially the backward adaptation algorithms like that of Jayant's (which we used in our simulations of DPCM1/AQB) increase the step size a lot when the largest quantization level is exercised and decrease it a little when the smallest quantization level is exercised. For inputs with Laplacian type of PDFs this gives satisfactory results but for data the prediction error PDF is almost uniform and therefore the smallest and largest quantization levels are roughly equiprobable. Consequently, the instantaneous step size adaptation strategies designed for speech yield too large a step size and poor performance with data.

5.4. Simulation of ASDM

The ASDM encoder operates on a continuous input signal. We therefore linearly interpolated the 8 kHz sampled signals by 64. The minimum response time of the encoder is then $125 / 64 = 1.953125$ microseconds or, the maximum sampling rate is 512 kHz. The corridor width parameter, w , was varied between 0.1 and 1.0 and the step size was taken as $0.99w$ throughout the simulations. Whenever the IBI sequence was quantized, the resulting non-integer IBI values were rounded to the nearest integer. Unless stated otherwise, all signal to noise ratios were calculated based on the difference between the original and reconstructed signals at the Nyquist instants.

5.4.1. Average Sampling Rate

The average number of ASDM encoder samples per second are tabulated for speech and data in Table 5.2.

Corridor width	Average number of samples per second	
	Speech	Data
0.1	38125	92265
0.2	19070	46688
0.3	12516	30586
0.4	9438	23281
0.5	7594	18461
0.6	6250	16070
0.7	5274	12484
0.8	4602	11719
0.9	4117	10125
1.0	3789	8531

Table 5.2. Average number of ASDM samples per second for artificial speech and QPSK data.

Naturally, the more "active" data signal results in a larger number of samples per second. The average sampling rate versus corridor width has been plotted for speech in Fig.5.13. The curve for data is very similar except for a scale change of the vertical axis.

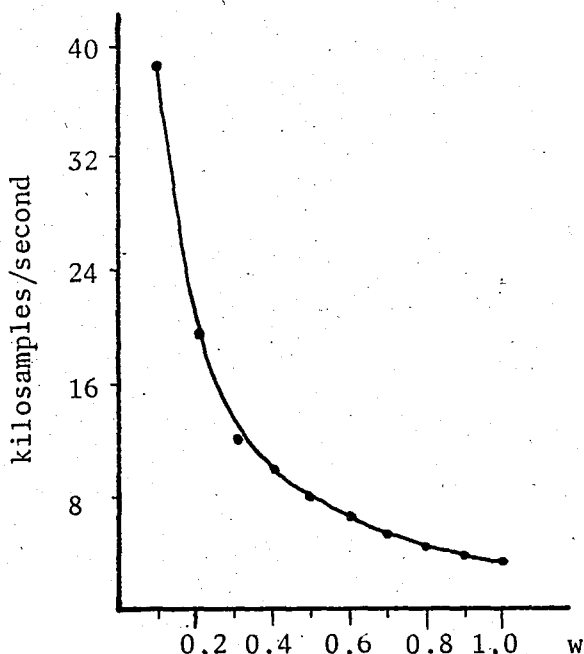


Figure 5.13. Average number of ASDM samples/sec vs. corridor width for the artificial speech signal.

We have observed that, at least in the range of interest, these curves form almost perfect hyperbolas in the form

$$w \bar{f} = c$$

where w is the corridor width as a fraction of the rms value of the input signal, \bar{f} is the average number of ASDM samples per second and c is a constant. The value of c , it seems, is determined by the input signal. If \bar{f} is expressed in kilosamples per second, for the artificial speech signal c comes out to be 3.757 with a standard deviation (over the 10 corridor widths considered) of 0.047.

For the data signal the value of c is 9.168 with a standard deviation of 0.303. Gungen's [22] data on bandlimited white Gaussian process gives c as 9.269 with a standard deviation of 0.030.

5.4.2. Sampling Statistics

The IBI PDFs of speech are shown for various corridor widths in Fig.5.14. One can observe that the PDFs spread out with increasing corridor width but always remain unimodal. They have rather steep rises and very long tails. Actually, because of the silent and low energy segments in speech, the inter-bit intervals may be pathologically long especially at large values of w . We have observed a dynamic range of more than 60 dB in speech IBIs. This may cause problems when the IBI sequence is quantized or runlength encoded. Positive and negative crossing inter-bit intervals yield identical PDFs.

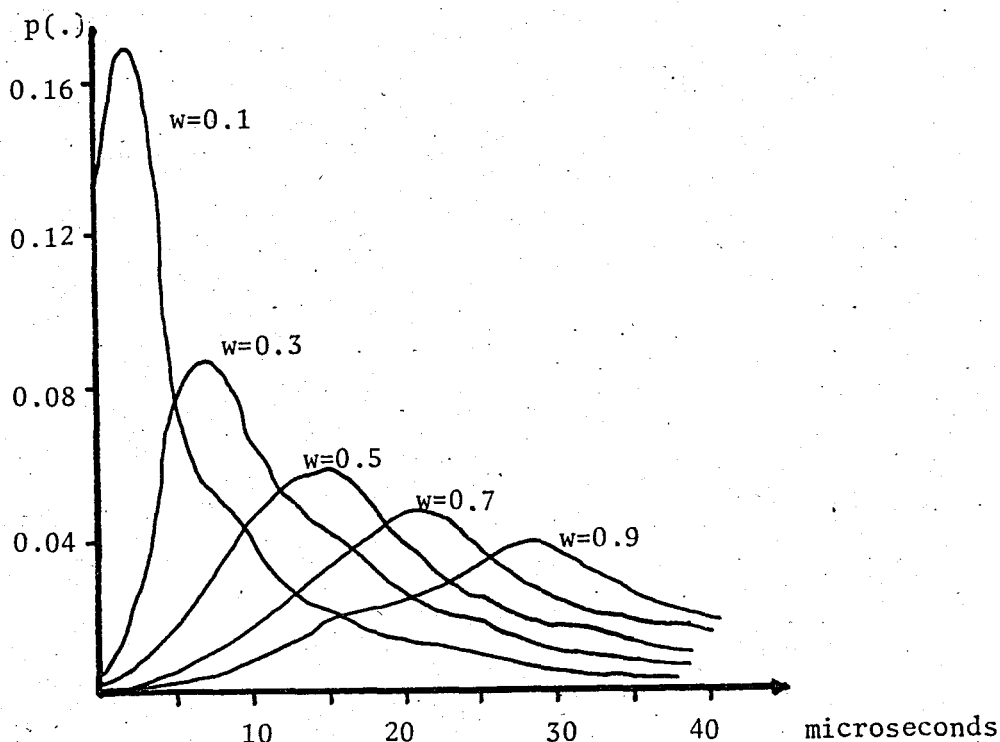


Figure 5.14. Inter-bit interval PDFs for speech.

The IBI PDFs of QPSK data, shown in Fig.5.15, exhibit similar behaviour in that they broaden with increasing w but remain unimodal. However, they are more distinct and do not spread to the same extent as those of speech with increasing w . This is, of course, consistent with the more structured nature of the data signal compared to speech. The dynamic range of the inter-bit intervals is considerably smaller for data than for speech and does not exceed 45 dB even when $w = 1.0$.

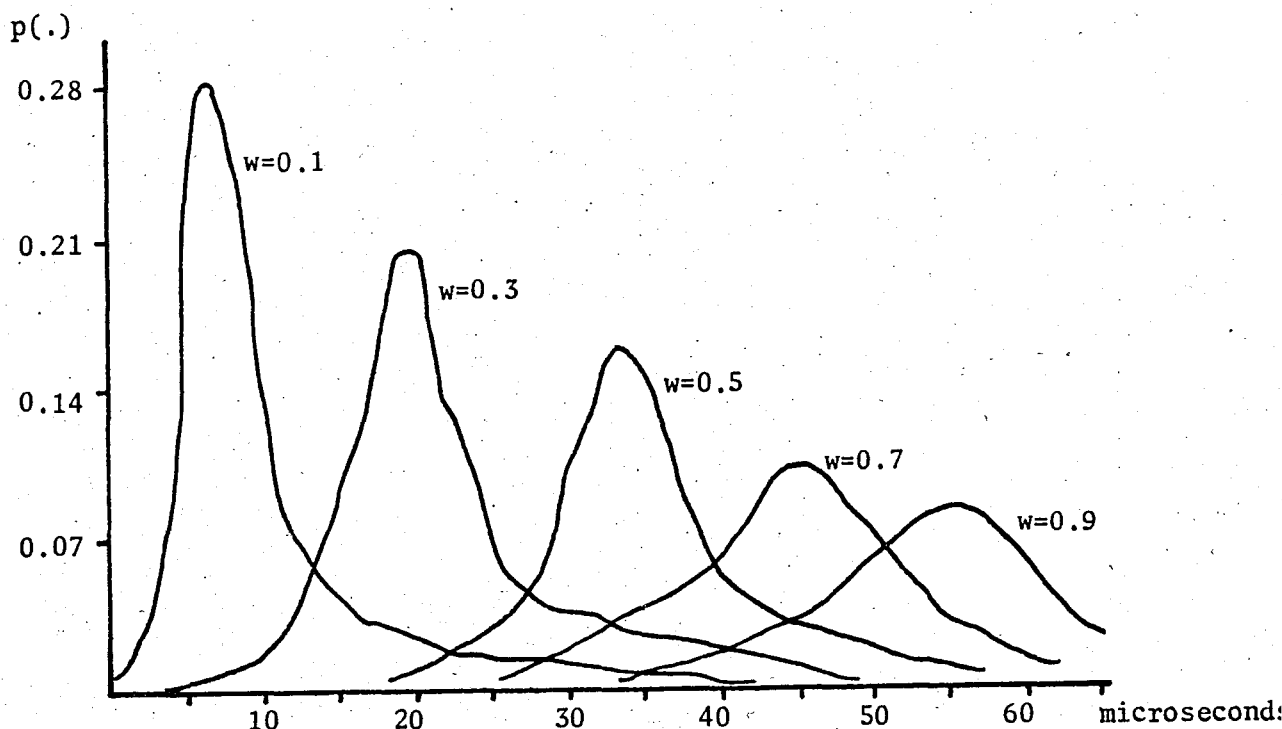


Figure 5.15. Inter-bit interval PDFs for QPSK data.

We have observed another interesting property of the PDFs which is common to both speech and data. For each signal, the ratio of the mean IBI value to the median is constant. For speech this ratio is about 3.132 with a standard deviation of 0.082 and for data it is about 1.345 with a standard deviation of 0.063.

Note that a high mean to median ratio implies long tails in the PDF and this ratio is considerably larger for speech than for data.

5.4.3. Speech Encoding with ASDM

The signal to noise ratio of ASDM with no IBI quantization is given by (4.3). Clearly, this constitutes an upper bound on the achievable performance because quantization of the IBI sequence reduces the SNR. At a corridor width of $w=0.1$ (4.3) gives an SQNR of about 30.8 dB. From Table 5.2, we see that at $w=0.1$ the average number of ASDM samples per second is 38125. Thus, even if we quantize each ASDM sample using a single bit and assume no IBI distortion, we can hardly obtain toll quality speech at 38 kbps. Consequently, because of the rather high average sampling rate at small values of w , it is unlikely that ASDM can compete with other coders in the toll quality range.

However, the average sampling rate rapidly drops with increasing w and ASDM can perform better than existing coders in the communications quality range. For example, at $w=1.0$ we have a theoretical maximum SNR of 12 dB and the average sampling rate is only about 4 kilosamples/second.

In our simulations, the IBI sequence was first quantized using a μ -255 log-quantizer. 4 to 32 levels were considered and the results are plotted in Fig.5.16 along with the theoretical maximum SNR (no IBI quantization) curve. Note that 4 level log-quantizer gives unacceptable performance at all values of w . On the other hand, using more than 3 bits for the quantizer brings very little improvement. Assuming that we encode each sample using 3 bits, from Fig.5.16 and Table 5.2 we can see that one can obtain about 11 dB at 11.3 kbps (for $w=1.0$) and about 13 dB at 18.8 kbps (for $w=0.6$).

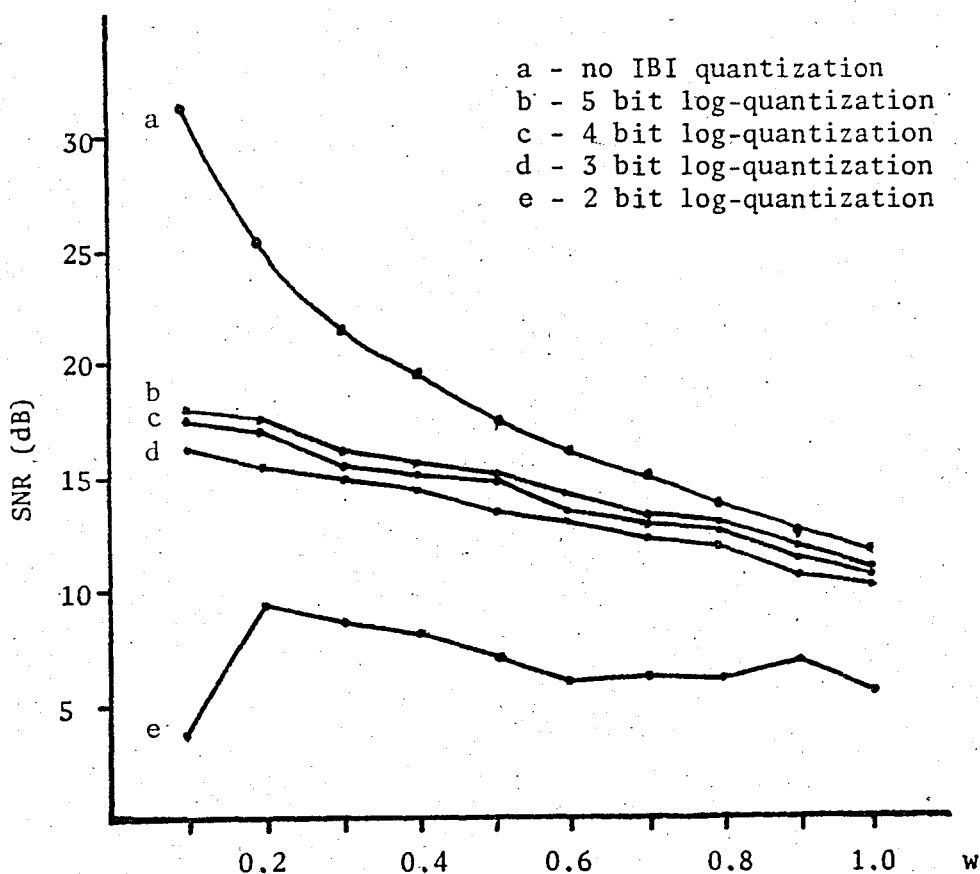


Figure 5.16. Speech encoding performance of ASDM with log-quantization of the inter-bit intervals.

At lower values of w , the performance does not justify the very high bit rates. Although the above SNR values are not high, one should contrast them with the speech encoder performance plots of Fig.5.12. Considering the results we have obtained, none of the simulated encoders provides an SNR greater than 10 dB at 16 kbps. Therefore, ASDM with 3-bit log-quantization of the IBI sequence is superior to all speech encoders we have simulated when $w > 0.6$.

An important issue we should mention now is that the ASDM SNR values were not found using the pairwise errors between the original and reconstructed signals. Because of the time quantization, if the

reconstructed signal is shifted, say by two samples, with respect to the input signal, a negative SNR may result although a human listener may not be able to discern the difference between the two.

We chose a simple way to get rid of this problem. To calculate the SNR the reconstructed signal was shifted in time with respect to the input signal and the cross-correlation between the two was computed at every lag. Some sort of time alignment was made (or, the maximum match between the two signals was found) at the lag where the cross-correlation is maximum and the SNR was calculated in the conventional manner. Note that this is an ad hoc procedure and perhaps spectral distance measures would have been more meaningful. Strictly speaking, we should call the SNR computed using the above method with a different name for clarity of notation. However, in the rest of the thesis it will be made clear from the context how the SNR is computed.

The small time scale perturbations, which, for lack of a better name we shall call the micro expansion/compressions, are shown in Fig.5.17 for the QPSK data signal.

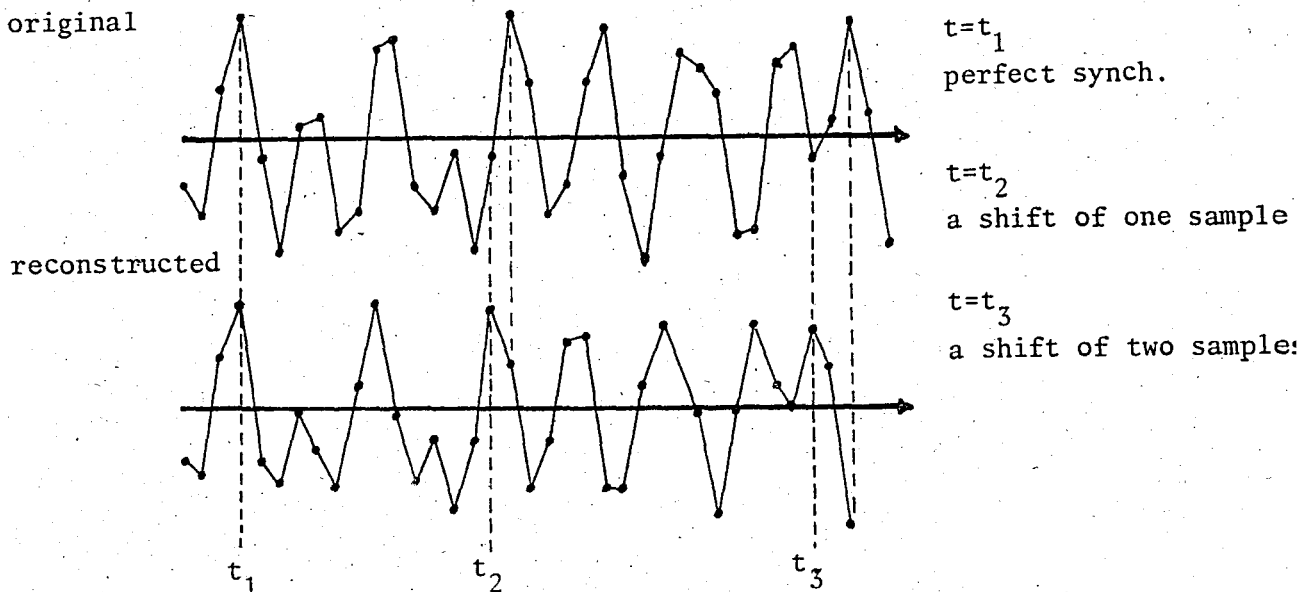


Fig.5.17. The micro expansion/compression effect.

For the speech signal for example, this effect may cause jitter in the pitch frequency. It has never been studied up to now and its degree of impairment is not known. Consequently, the ASDM SNRs given in this section may be higher, about the same, or lower than what subjective evaluations may suggest. The ultimate evaluation would have been, of course, to conduct listening tests and rate ASDM according to the results of these subjective measures.

The performance of ASDM with 2 to 5 bit optimum quantization of the IBI sequence is shown in Fig.5.18. There is surprisingly little difference between logarithmic and optimum quantization of the IBI sequence as far as the output SNR is concerned and the same comments for log-quantization apply exactly.

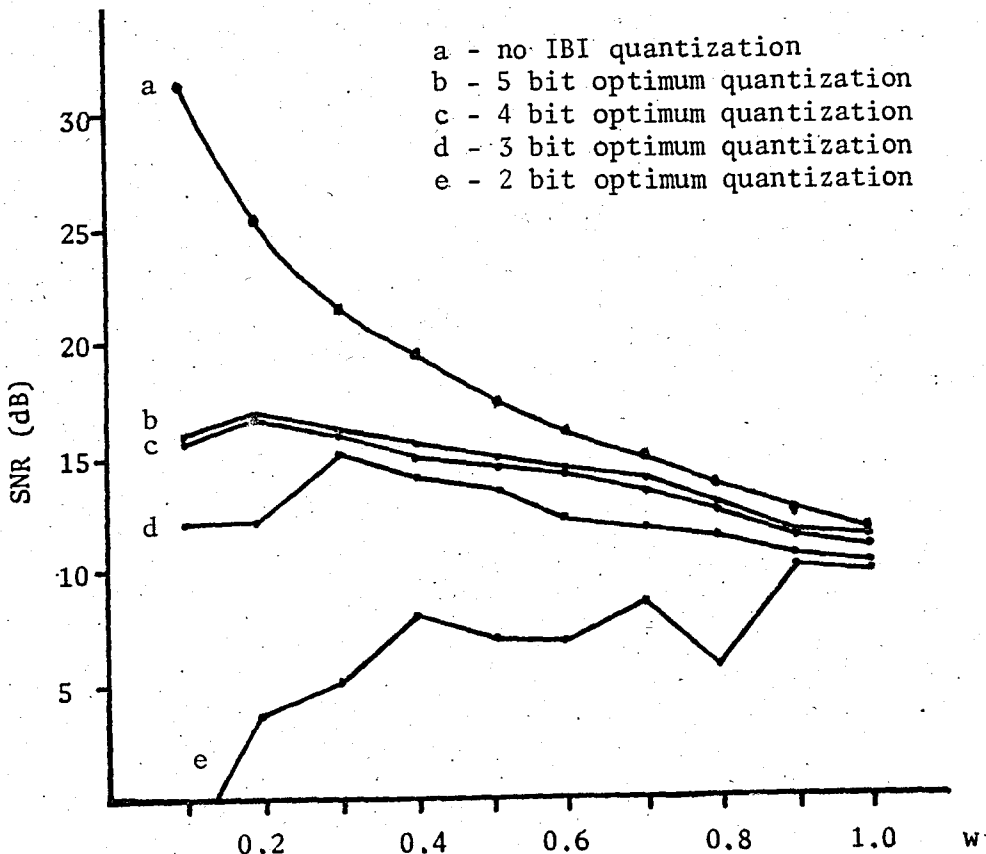


Figure 5.18. Speech encoding performance of ASDM with optimum quantization of the inter-bit intervals.

We have observed another interesting phenomenon in the context of optimum quantizers. The quantizer which minimizes the IBI quantization error (mean squared error) may not give the best output SNR. It seems that using the mean squared error as the distortion measure is not appropriate for the IBI sequence. This is evident from Fig.5.18 where adding another bit to the quantizer, although decreases the IBI distortion by about 6 dB, yields almost no SNR gain. However, we did not have time to study this fact in detail.

The evolution with corridor width of the reconstruction levels of a 2-bit optimum IBI quantizer for speech is shown in Fig.5.19.

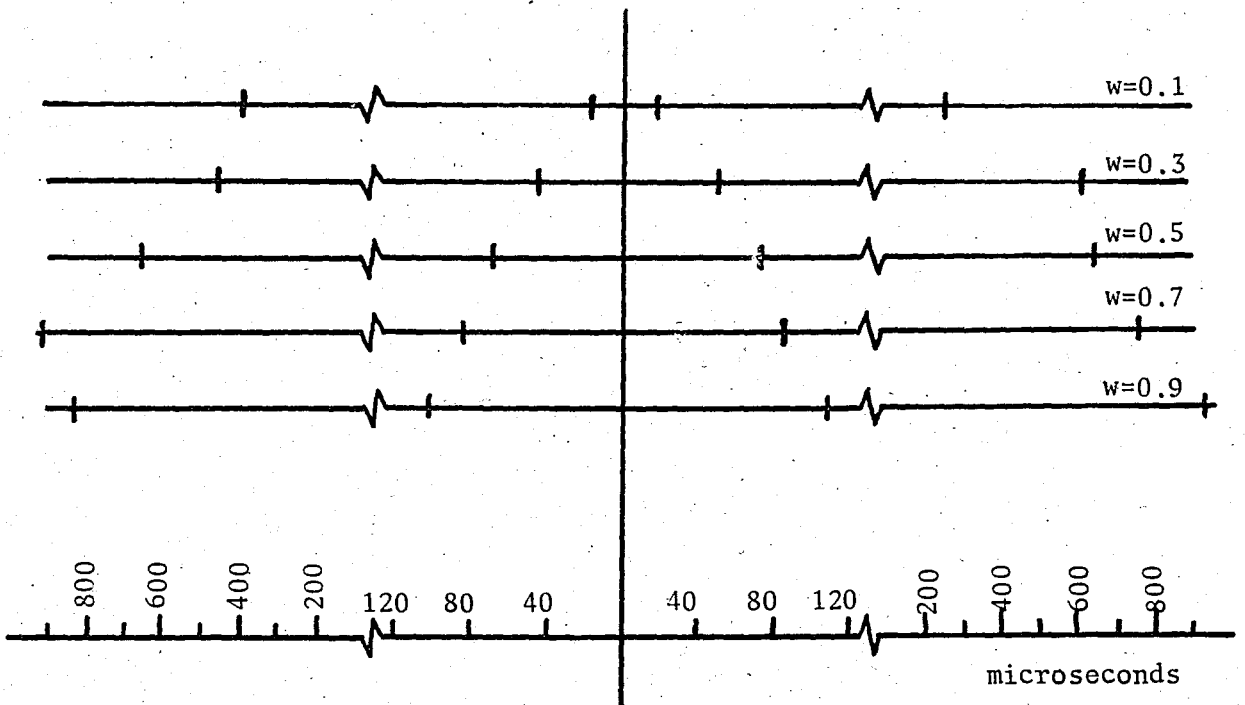


Figure 5.19. The reconstruction levels of a 2-bit optimum speech IBI quantizer for various values of w .

We then investigated the performance of ASDM with vector quantization of the IBI sequence. Vector quantization is a recent technique in which the samples are first collected in a block (or vector) of predetermined dimension and then compared with a set of previously stored reconstruction vectors and the binary index of the reconstruction vector which achieves minimum distance to the input vector is transmitted [32]. Vector quantization yields significant improvement over scalar quantization but it becomes computationally expensive with increasing vector dimension. With the aim of not losing the inherent simplicity of ASDM, vector dimensions of 2 to 4 have been considered. Computational considerations also influenced this choice. The SNRs obtained by vector quantization of the IBI sequence are shown in Fig.5.20 for various values of w and vector dimension.

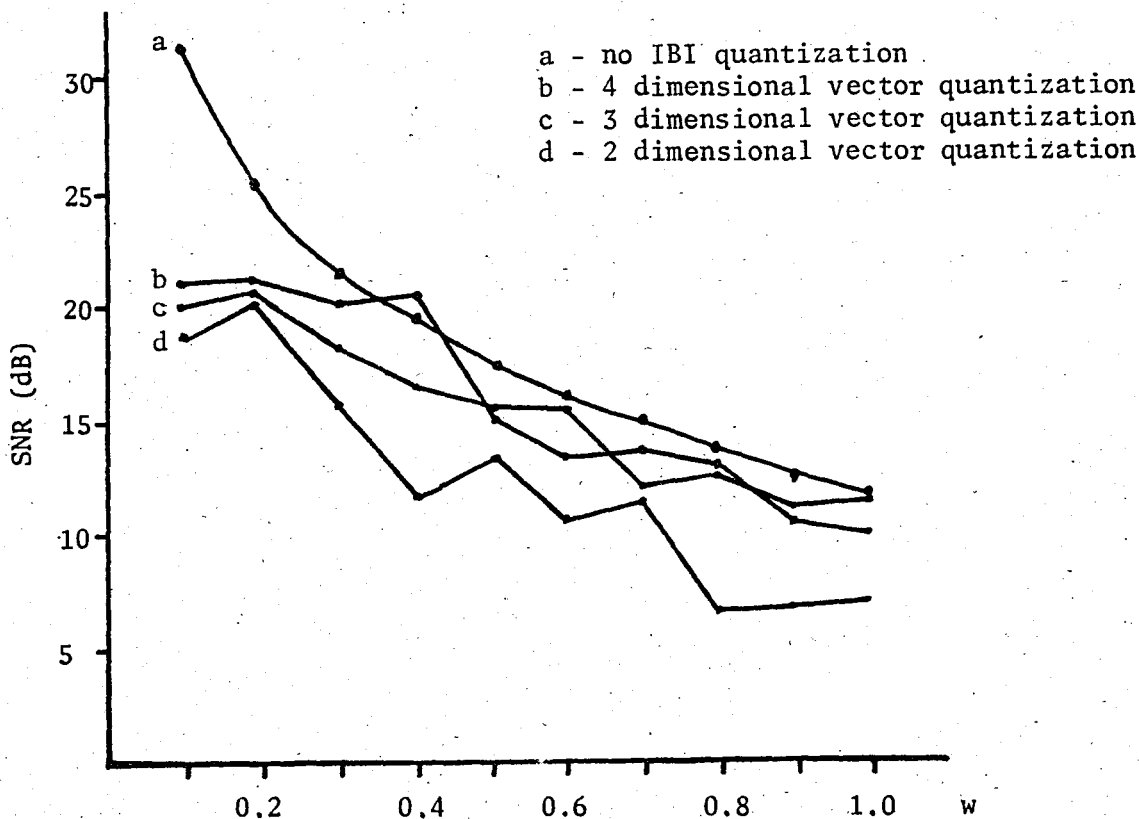


Figure 5.20. Speech encoding performance of ASDM with vector quantization of the inter-bit intervals.

For all vector dimensions, one bit per sample encoding was used. Therefore, the sampling rate of the ASDM encoder (Table 5.2) directly becomes the bit rate of the system.

As exemplified by Fig.5.20, even for moderate vector dimensions we obtain a SNR of more than 10 dB at about 4 kbps ! On the other hand, with $w=0.3$ and a vector dimension of 4 we obtain 20 dB at only 12.5 kbps which is a real improvement over existing coders (see Fig.5.12). Fig.5.20 is one of the major results of this thesis.

5.4.4. Data Encoding with ASDM

The data signal, as can be seen from Table 5.2, has a higher average sampling rate than speech. If the ASDM encoder is to operate with a fixed output bit rate, the corridor width w must be larger for data than for speech.

The micro expansion/compression effects mentioned before are present for data as well, and their impairments are much more severe. Because the data signal has a precise symbol timing, any expansions and compressions of the time axis may render the data transmission system useless. This is especially true for PSK in which the phase of the carrier has to be preserved by the coder.

Because of the strict timing requirements and objective evaluation criteria (a demodulator instead of human ear), computation of SNR at maximum cross correlation of the input and reconstructed signals is impossible with data. Therefore, all SNR values given in this section are ordinary SNRs, i.e., no time alignment was attempted.

The data encoding performance of ASDM with no IBI quantization is summarized in Table 5.3.

Corridor width	Simulation SNR	Maximum phase error	Variance of phase error	Mean phase error	SNR _e
0.1	31.1	6.7	3.3	0.2	27.0
0.2	25.5	9.3	6.6	-0.2	24.0
0.3	21.5	8.9	13.1	1.0	21.0
0.4	19.2	12.0	19.7	1.4	19.2
0.5	17.4	13.3	23.0	-0.8	18.5
0.6	15.7	12.1	29.5	0.5	17.4
0.7	14.1	19.3	49.2	1.0	15.2
0.8	12.8	16.0	55.8	1.0	14.7
0.9	12.0	21.0	72.2	0.5	13.6
1.0	11.3	27.6	88.6	0.5	12.7

Table 5.3. Results of the data encoding performance of ASDM with no IBI quantization. All values are in degrees except SNRs (dB).

The evolution with increasing corridor width of the reconstruction levels of a 2-bit optimum data IBI quantizer is shown in Fig.5.21. As for speech, the minimum mean squared IBI quantizer does not guarantee best performance.

The data encoding performance of ASDM using logarithmic, optimum and vector quantization of the IBI sequence are given in tables 5.4, 5.5 and 5.6, respectively.

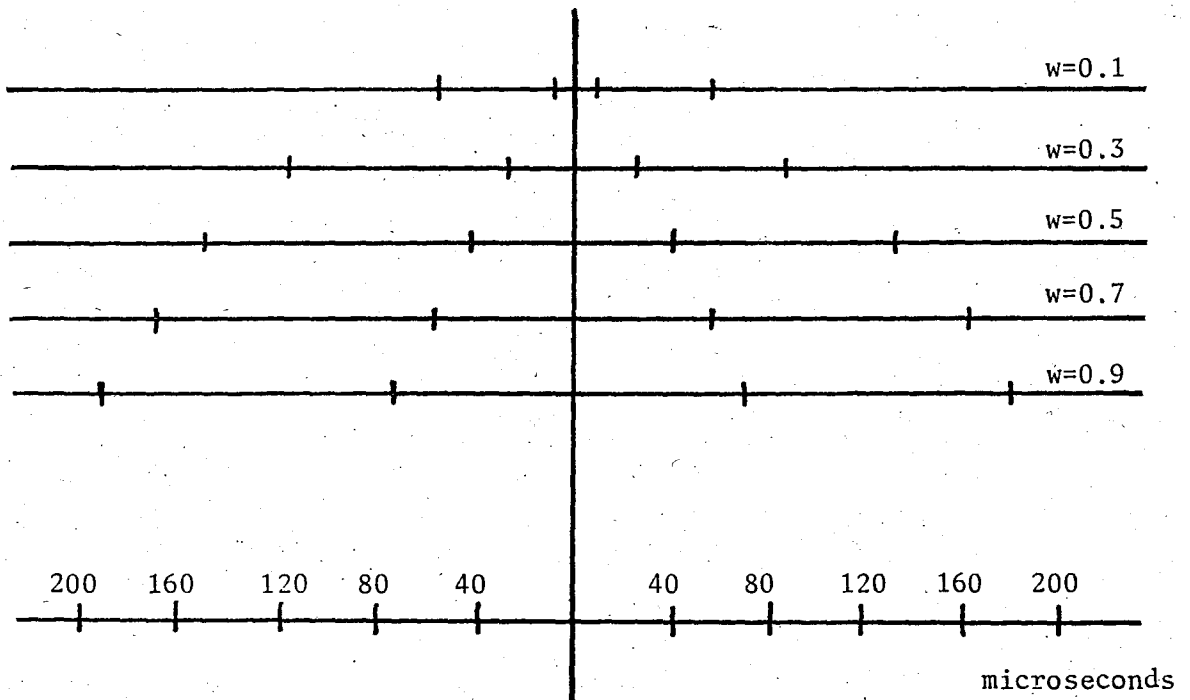


Figure 5.21. The reconstruction levels of a 2-bit optimum data IBI quantizer for various values of w .

As revealed by Table 5.4, logarithmic quantization of the inter-bit intervals does not give acceptable results until at 5 bits per sample where the performance becomes only marginally acceptable. Optimum quantization is only slightly better than logarithmic quantization and is still marginal. Also, since the reconstruction levels of the optimum quantizers for speech and data are different, either a dual mode or an adaptive IBI quantizer has to be used. Since the average sampling rate is also considerably high for data, these results imply that logarithmic and optimum quantization of the inter-bit intervals are not appropriate.

Corridor width	Simulation SNR	Maximum phase error	Variance of phase error	Mean phase error	SNR _e	Quantizer levels
0.1	5.5	140.1	1233.5	7.3	1.2	4
	9.3	72.6	208.3	3.4	9.0	8
	10.3	74.1	156.8	2.7	10.2	16
	11.1	33.3	88.7	1.8	12.7	32
0.2	8.6	159.3	535.1	2.1	4.9	4
	11.0	110.2	183.8	1.3	9.5	8
	11.3	31.9	98.5	1.8	12.2	16
	11.0	39.3	112.3	1.6	11.5	32
0.3	8.0	148.5	660.0	3.6	4.0	4
	10.7	30.6	91.9	1.4	12.5	8
	11.0	29.0	82.1	1.9	13.0	16
	11.8	36.2	71.0	2.8	13.6	32
0.4	7.8	143.4	485.6	4.1	5.3	4
	9.5	143.0	262.6	0.2	8.0	8
	10.9	28.6	101.8	0.5	12.1	16
	11.2	36.2	84.6	2.7	12.9	32
0.5	7.5	167.2	620.5	1.5	4.2	4
	9.6	147.1	395.2	2.3	6.2	8
	10.6	29.6	82.1	0.5	13.0	16
	11.5	25.0	55.8	-0.5	14.7	32
0.6	6.0	125.1	810.9	8.1	3.1	4
	9.4	142.4	344.3	3.2	6.8	8
	10.2	29.2	77.0	2.9	13.3	16
	10.9	23.5	52.9	1.7	14.9	32
0.7	5.6	143.2	859.4	11.2	2.8	4
	8.8	92.2	203.5	1.8	9.1	8
	9.5	37.5	141.2	0.8	10.7	16
	10.3	25.2	85.4	0.6	12.8	32
0.8	5.3	106.9	787.5	8.9	3.2	4
	9.1	47.2	164.1	2.8	10.0	8
	8.8	44.5	157.6	1.0	10.2	16
	9.9	38.0	108.1	2.6	11.8	32
0.9	4.7	165.6	1332.9	6.0	0.9	4
	8.3	154.0	272.6	1.8	7.8	8
	8.6	34.8	137.8	0.0	10.8	16
	9.6	30.0	85.4	-0.6	12.8	32
1.0	4.2	179.9	1898.8	12.5	-0.6	4
	8.0	163.3	344.7	0.6	6.8	8
	8.4	37.5	147.7	1.5	10.5	16
	9.1	30.3	111.6	-0.7	11.7	32

Table 5.4. Results of the data encoding performance of ASDM with 2 to 5 bit log-quantization of the IBI sequence. All values are in degrees except SNRs (dB)

Corridor width	Simulation SNR	Maximum phase error	Variance of phase error	Mean phase error	SNR _e	Quantize levels
0.1	10.0	35.9	102.5	1.1	12.0	4
	10.3	35.0	117.5	1.1	11.5	8
	10.3	35.0	123.8	1.4	11.2	16
	11.4	29.6	89.1	2.3	12.7	32
0.2	10.2	28.3	107.0	2.1	11.9	4
	10.9	34.4	123.0	1.0	11.3	8
	11.2	35.4	106.7	1.6	11.9	16
	10.9	41.2	138.6	2.3	10.7	32
0.3	10.1	37.3	113.9	1.8	11.6	4
	11.6	36.6	81.2	3.4	13.1	8
	11.9	34.8	84.2	3.6	12.9	16
	11.2	44.1	126.2	3.7	11.1	32
0.4	10.1	26.1	93.8	2.4	12.4	4
	11.6	35.5	114.8	2.0	11.6	8
	11.1	40.2	111.4	1.2	11.7	16
	11.4	36.2	79.0	2.8	13.2	32
0.5	10.9	23.1	85.3	-0.2	12.8	4
	11.6	35.3	74.9	1.7	13.4	8
	9.0	58.8	163.7	-0.2	10.0	16
	12.2	28.0	62.7	2.3	14.2	32
0.6	10.9	27.9	62.7	0.8	14.2	4
	11.4	28.7	83.4	3.5	12.9	8
	11.5	36.4	92.7	1.8	12.5	16
	11.8	29.8	62.3	2.2	14.2	32
0.7	9.0	33.2	114.7	0.9	11.6	4
	10.7	35.2	97.2	1.5	12.3	8
	10.6	37.3	111.1	1.7	11.7	16
	11.7	23.7	72.6	2.6	13.5	32
0.8	7.7	94.4	258.2	-3.2	8.0	4
	9.1	42.5	139.6	1.6	10.7	8
	8.2	106.1	288.8	0.5	7.5	16
	10.7	27.7	76.4	1.7	13.3	32
0.9	8.6	34.4	128.7	0.6	11.1	4
	7.5	152.0	879.9	-0.6	2.7	8
	7.5	147.7	813.7	0.6	3.0	16
	10.3	33.5	88.4	1.3	12.7	32
1.0	8.1	36.5	180.9	-2.3	9.6	4
	8.8	43.1	147.1	0.3	10.5	8
	8.2	41.1	154.0	0.6	10.3	16
	9.9	32.5	95.7	0.4	12.3	32

Table 5.5. Results of the data encoding performance of ASDM with 2 to 5 bit optimum quantization of the IBI sequence. All values are in degrees except SNR (dB).

Corridor width	Simulation SNR	Maximum phase error	Variance of phase error	Mean phase error	SNR _e	Vector dimension
0.1	-1.5	178.8	3504.9	9.3	-3.3	2
	3.1	172.9	2079.1	5.1	-1.0	3
	5.3	170.7	1658.3	4.8	0.0	4
0.2	7.1	174.3	1040.4	4.0	2.0	2
	2.2	172.8	1927.7	6.1	-0.7	3
	9.6	174.3	461.8	2.8	5.5	4
0.3	6.5	174.3	1355.8	7.5	0.8	2
	12.1	58.6	86.8	0.8	12.8	3
	7.1	115.8	442.3	1.4	5.7	4
0.4	3.5	172.9	1804.4	5.9	-0.4	2
	10.5	58.7	115.2	0.9	11.5	3
	8.9	174.3	415.9	1.8	6.0	4
0.5	3.1	172.9	1932.8	7.1	-0.7	2
	11.6	149.4	210.9	0.4	8.9	3
	10.9	172.8	289.2	0.3	7.5	4
0.6	2.8	172.7	2275.7	5.5	-1.4	2
	9.9	152.8	335.7	-0.1	6.9	3
	8.9	152.6	467.7	0.5	5.5	4
0.7	3.2	175.8	1959.6	5.2	-0.8	2
	11.6	62.9	94.9	-0.8	12.4	3
	8.9	174.3	428.8	2.6	5.8	4
0.8	3.0	179.6	2150.7	7.8	-1.2	2
	8.8	53.8	125.3	0.7	11.2	3
	11.2	143.6	205.6	1.0	9.0	4
0.9	1.7	179.1	2066.6	3.4	-1.0	2
	8.7	143.6	346.2	2.9	6.8	3
	11.3	115.8	172.6	0.4	9.8	4
1.0	1.1	172.9	2060.7	5.8	-1.0	2
	8.9	64.6	98.6	-0.1	12.2	3
	6.6	116.0	191.0	0.7	9.3	4

Table 5.6. Results of the data encoding performance of ASDM with vector quantization of the IBI sequence. 1 bit/sample vectors with dimensions 2 to 4 have been considered. All values are in degrees except SNRs (dB).

As for vector quantization, the results are worse than those obtained with logarithmic and optimum quantizers and do not seem to improve with increasing dimension.

In summary, we conclude that because of distortions of the time axis, ASDM is not a reliable coder for data, or more specifically PSK type data signals.

5.4.5. Entropy Coding

The IBI sequence can be further compressed by entropy coding of the quantizer outputs. We have, in our simulations, Huffman coded the optimum and vector quantizer outputs to reduce the bit rate. The reader is reminded that this is not the proper way of entropy coding, i.e., uniform quantizers yield lower output entropy than optimum quantizers thereby allowing higher compression ratios.

The quantizer output entropy H and average rate \bar{N} in bits per symbol of the Huffman code for speech are given in Table 5.7 for optimum and in Table 5.8 for vector quantizers, respectively.

Corridor width	2-bit optimum quantizer		3-bit optimum quantizer		4-bit optimum quantizer		5-bit optimum quantizer	
	H	\bar{N}	H	\bar{N}	H	\bar{N}	H	\bar{N}
0.1	1.157	1.531	1.611	1.767	2.229	2.312	2.816	2.863
0.2	1.170	1.531	1.703	1.824	2.806	2.820	3.450	3.482
0.3	1.246	1.557	1.897	1.966	2.958	2.968	3.803	3.848
0.4	1.308	1.579	1.918	1.991	3.045	3.062	3.976	4.012
0.5	1.304	1.577	2.014	2.085	3.183	3.217	4.078	4.112
0.6	1.317	1.581	2.103	2.167	3.301	3.345	4.156	4.189
0.7	1.328	1.587	2.116	2.177	3.327	3.371	4.262	4.298
0.8	1.360	1.602	2.161	2.225	3.384	3.418	4.317	4.353
0.9	1.362	1.603	2.247	2.325	3.474	3.522	4.365	4.397
1.0	1.401	1.618	2.290	2.379	3.403	3.434	4.393	4.431

Table 5.7. Results (in bits/sample) of Huffman coding of the optimum IBI quantizer outputs for speech.

Note that for small values of the corridor width w , the quantizer output entropies are small and increase with increasing w . This is in accordance with the spreading of the PDF with increasing corridor width. The entropy coding scheme yields a reduction in bit rate of about 20 to 70 percent depending on the corridor width and the number of quantizer levels.

Corridor width	vector quantizer of dimension 2		vector quantizer of dimension 3		vector quantizer of dimension 4	
	H	\bar{N}	H	\bar{N}	H	\bar{N}
0.1	0.744	1.182	1.216	1.480	2.830	2.877
0.2	0.644	1.152	1.585	1.656	3.029	3.043
0.3	0.587	1.139	1.889	1.930	3.069	3.124
0.4	0.703	1.188	1.711	1.803	3.347	3.391
0.5	1.328	1.482	1.876	1.907	3.297	3.327
0.6	0.904	1.247	2.123	2.148	3.478	3.513
0.7	0.746	1.199	2.263	2.289	3.540	3.586
0.8	0.953	1.288	2.387	2.424	3.322	3.351
0.9	1.042	1.317	2.069	2.103	3.511	3.559
1.0	1.191	1.376	2.411	2.469	3.453	3.482

Table 5.8. Results (in bits/vector) of Huffman coding of the vector quantizer outputs for speech.

Because of the promising performance of vector quantization, it is of interest to examine Huffman coding of the vector quantizer outputs in detail. For $w=0.3$ and a vector dimension of 4, we have a SNR of 20 dB at a rate of 12.5 kbps. After entropy coding the quantizer outputs we obtain 3.124 bits per symbol (vector) instead of 4, and this reduces the bit rate from 12.5 to about 9.8 kbps. At lower SNRs the results are surprising. For $w=0.8$ and a vector dimension of 4, we obtain a SNR of about 13 dB at 3.9 kbps, which is well within the parametric coder range !

However, the Huffman codes are optimum for a given set of quantizer output probabilities and the results will degrade if these probabilities change.

Results of Huffman coding of the optimum data IBI quantizer outputs are given in Table 5.9. These results are similar to that for speech in the amount of compression offered. Because of the poor performance of vector quantization in the case of encoding data signals, vector quantizer outputs for data were not Huffman coded.

Corridor width	2-bit optimum quantizer		3-bit optimum quantizer		4-bit optimum quantizer		5-bit optimum quantizer	
	H	\bar{N}	H	\bar{N}	H	\bar{N}	H	\bar{N}
0.1	1.218	1.550	1.355	1.626	1.375	1.644	2.262	2.338
0.2	1.312	1.580	1.859	1.960	2.034	2.116	3.025	3.044
0.3	1.443	1.636	2.471	2.514	2.652	2.692	3.284	3.294
0.4	1.416	1.627	2.537	2.564	2.916	2.929	3.776	3.792
0.5	1.534	1.678	2.560	2.603	3.136	3.170	3.844	3.878
0.6	1.651	1.747	2.643	2.682	3.239	3.284	4.122	4.150
0.7	1.640	1.733	2.664	2.703	3.329	3.361	4.277	4.312
0.8	1.700	1.772	2.791	2.845	3.457	3.492	4.351	4.389
0.9	1.726	1.786	2.696	2.766	3.572	3.595	4.331	4.356
1.0	1.780	1.830	2.695	2.754	3.632	3.664	4.399	4.440

Table 5.9. Results (in bits/sample) of Huffman coding of the optimum IBI quantizer outputs for data.

5.4.6. Buffer Behaviour

In all asynchronous coding schemes a buffer is needed to output the asynchronous input information at a synchronous rate. The overflow and underflow of this buffer are critical problems and severely degrade the performance.

For ASDM, Sankur and Gungen [21] have derived the buffer overflow and underflow probabilities. In our work we investigated the behaviour of a finite length buffer for a 1/8 second long artificial speech sample. Results for various corridor widths are given in Fig.5.22. In all cases the synchronous channel rate was assumed to be more or less matched to the average sampling rate of the ASDM encoder.

Fig 5.22 is drawn assuming no quantization of the IBI sequence. If the quantization is "proper," the results are not significantly different. The vertical scale is in "words," i.e., if 3 bits is used to encode each sample, the scales should be multiplied by 3 to obtain the buffer size in bits.

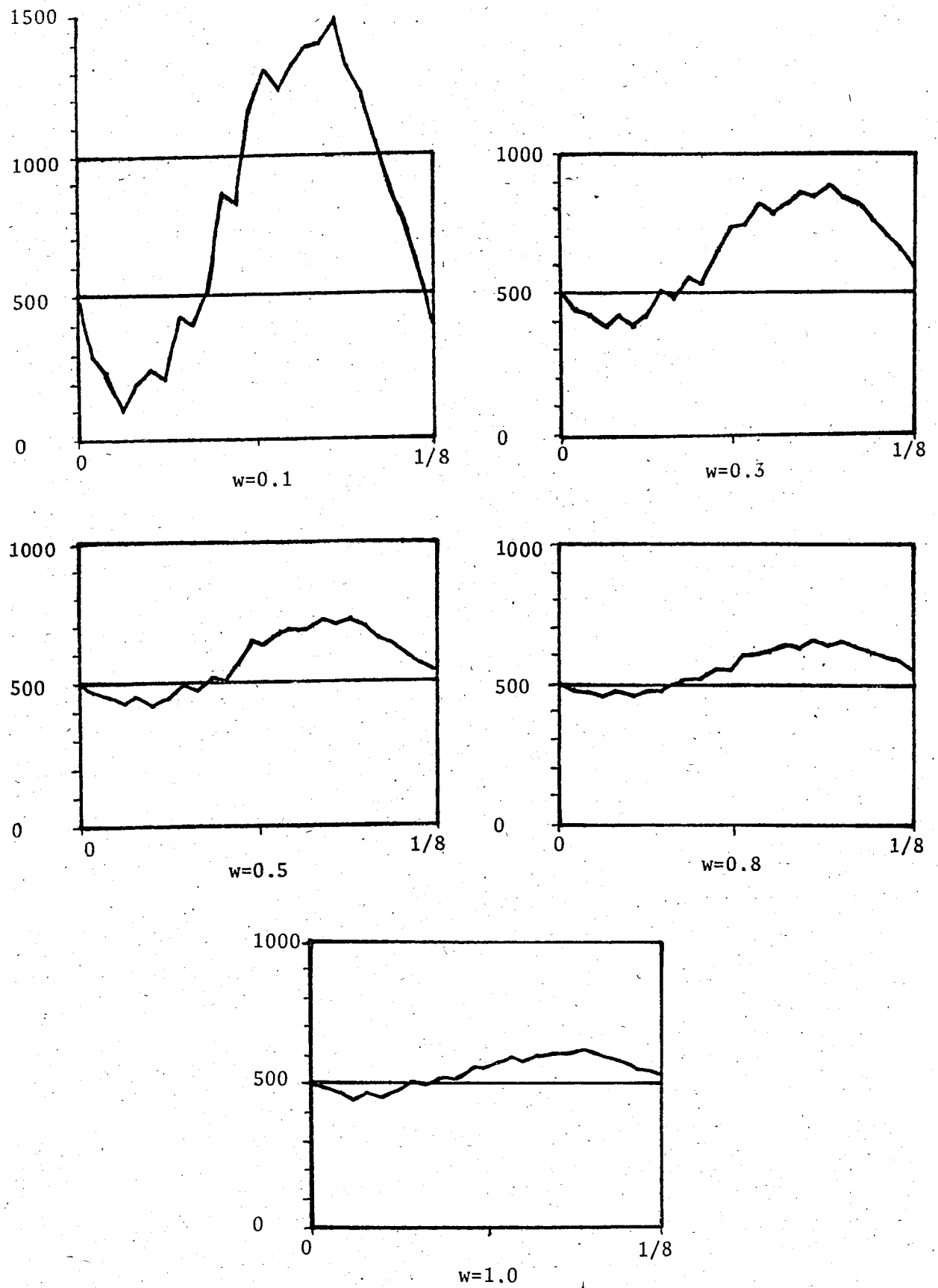


Figure 5.22. Buffer behaviour for various values of the corridor width w .

Observe that at small values of the corridor width the buffer is "wild" and a 1 kiloword buffer overflows when $w=0.1$ (and also 0.2). At these small values of w slight mismatches between the synchronous channel rate and average ASDM sampling rate cause the buffer to overflow or underflow.

As the corridor width is increased, the buffer is relaxed and a 1 kiloword buffer is apparently sufficient for corridor widths $w>0.5$.

Since the receiver buffer is the complement of the transmitter buffer (if the former is three quarters full the latter is three quarters empty and vice versa), the same buffer length will suffice for the receiver.

With the ever decreasing costs of memory chips however, overflow and underflow issues of finite length buffers have become trivial problems.

VI. CONCLUSIONS

In this study the performance of the asynchronous delta modulator has been investigated for both speech and PSK data signals and compared with other common speech coders.

ASDM is an effective source coder that is independent of the source statistics, i.e., as long as the inter-bit intervals are not quantized, it can encode any input signal with the same SNR which is dependent only on the selected corridor width. However, the average sampling rate and the inter-bit interval PDFs are dependent on the source statistics and may show significant variations if the statistics of the sources to be encoded are very different.

The signal dependence of the IBI PDFs is especially important when quantization of the IBI sequence is considered. We have observed that minimum mean squared error quantization of the inter-bit intervals not necessarily gives the highest output SNR.

For speech, logarithmic or optimum quantization of the IBI sequence with 3 or more bits can make ASDM a competitive coder in the communications quality range. With 1 bit/sample, 4-dimensional vector quantization ASDM has outstanding performance at medium and low bit rates.

For QPSK data on the other hand, performance of ASDM is not very promising. The consequence of time quantization is a warping of the time axis on a micro scale, i.e., a few samples. Presumably, this is not a very big problem for speech but for data signals which have a precise symbol timing, its effects are disastrous. We can state that when the inter-bit intervals are quantized, ASDM cannot preserve the "phase integrity" of the input signal which is the most important property that a coder should possess if it is to successfully encode PSK signals.

In our simulations errorless performance with data required at least 4 bit quantization of the IBI sequence. Since the average sampling rate is high for data, when at least 4 bits is used for each sample ASDM bit rate becomes too high to be of practical value. Vector quantization does not work well with data signals, either.

For speech, increasing the quantizer levels above 8 brings very little SNR improvement although the IBI quantization error drops by about 6 dB for each bit added. We attribute this to the fact that minimizing the squared IBI error may not result in the best output SNR. In this study we did not attempt to find a way to most effectively quantize the IBI sequence. To attack that problem analytical results between IBI distortion and output SNR should be established. However, IBI statistics are related to the extremely complex level crossing problems for which analytical results exist only for some very special input processes.

We did not attempt to differentially quantize the IBI sequence because the IBI sequence is almost totally uncorrelated for corridor widths greater than 0.3. Even for $w=0.1$, the first correlation coefficient is less than 0.5.

In this aspect ASDM may be viewed as a coder which tries to decorrelate the input process by converting it into an uncorrelated IBI process. However, other decorrelating coders, e.g., transform coders have an advantage in that they can decide which components of the decorrelated sequence are more "significant". Another decorrelating coder, DPCM, comes up with an "easier" sequence to quantize because the variance of the sequence to be quantized is less than that of the input signal. With ASDM, the dynamic range of the IBI sequence is 20 dB more than the input signal (can be even more with real speech) so we do not have an "easier" sequence to quantize. On the other hand although we feel that large IBI values are more important, deciding which components of the IBI sequence are more "significant" is a dangerous task with no analytic results at hand. Therefore the major problem, effective quantization of the IBI sequence remains unsolved.

This however, does not cast shadow on our results. Entropy coding of the quantizer outputs yields a further bit rate reduction of about 20 to 50 percent and enhances ASDM's performance as a medium band speech coder. Furthermore entropy coding, which is not very popular with synchronous coders because of its variable wordlength, can be effectively used with ASDM because ASDM is already an asynchronous coder.

The buffer size is not a very critical issue as long as the average coder rate and the channel rate are equal and a few kilobits of memory seems to be sufficient for almost all values of the corridor width.

Further research on ASDM might focus on effective quantization of the IBI sequence and try to establish the previously mentioned analytical results. Adapting the corridor width according to buffer occupancy must be investigated for two reasons :

a) It is the only way to cope with the very long silent periods (consequently, very long inter-bit intervals) present in speech.

b) Adapting the corridor width also modifies the IBI PDFs and limits their dynamic range. We believe that this will result in a sequence that is more suited to effective quantization.

Since the receiver buffer is the complement of the transmitter buffer another elegance of this technique is that the receiver can extract the necessary information from its own buffer, making the transmission of side information unnecessary.

REFERENCES

- [1] L.R.Rabiner and R.W.Schafer, Digital Processing of Speech Signals, Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [2] J.L.Flanagan, M.R.Schroeder, B.S.Atal, R.E.Crochiere, N.S.Jayant and J.M.Tribolet, "Speech coding," IEEE Trans. Commun., vol.COM-27, pp.710-737, April 1979.
- [3] M.D.Paez and T.H.Glisson, "Minimum mean squared error quantization in speech PCM and DPCM systems," IEEE Trans.Commun. vol.COM-20, pp.225-230, April 1972.
- [4] P.Noll, "Nonadaptive and adaptive differential pulse code modulation of speech signals," Polytech. Tijdschr., Ed.Electrotech.Electron., no.19, pp.623-629, 1972.
- [5] E.R.Kretzmer, "The evolution of techniques of data communication over voiceband channels," IEEE Commun.Soc.Mag., pp.10-14, Jan.1978.
- [6] J.B.O'Neal, Jr., "Waveform encoding of voiceband data signals," Proc.IEEE, pp.232-247, Feb.1980.
- [7] P.Noll, "Adaptive quantizing in speech coding systems," Proc.Int Zurich Seminar on Digital Commun., pp.B3(1)-(6), 1974.
- [8] J.B.O'Neal, Jr. and W.Stroh, "Differential PCM for speech and data signals," IEEE Trans.Commun., vol.COM-20, pp.900-912, Oct. 1972.
- [9] J.B.O'Neal, Jr., R.R.Koneru and J.P.Agrawal, "Digital encoding of phase shift keying voiceband data signals," IEEE Trans.Commun., vol.COM-28, pp.831-840, June 1980.
- [10] N.S.Jayant and P.Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video, Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [11] M.R.Schroeder, "Vocoders: Analysis and synthesis," Proc.IEEE, vol.54, pp.720-734, May 1966.
- [12] P.Noll, "A comparative study of various quantization schemes for speech encoding," BSTJ, vol.54, pp.1597-1614, Nov.1975.
- [13] A.Gersho, "Quantization," IEEE Commun.Soc.Mag., vol.15 pp.16-29 Sept.1977.
- [14] J.Max, "Quantizing for minimum distortion," IRE Trans.Inform. Theory, vol.IT-6, pp.16-21, March 1960.
- [15] J.D.Markel and A.H.Gray, Jr., Linear Prediction of Speech, New York: Springer-Verlag, 1976.

- [16] N.S.Jayant, "Adaptive quantization with a one word memory," BSTJ pp.1119-1144, Sept.1973.
- [17] J.D.Gibson, "Adaptive prediction for speech encoding," IEEE Acoust.Speech and Signal Processing Mag., vol.1, pp.12-26, July 1984.
- [18] J.J.Dubnowski and R.E.Crochiere, "Variable rate coding of speech" BSTJ, pp.577-600, March 1979.
- [19] C.K.Un and D.H.Cho, "Hybrid companding delta modulation with variable rate sampling," IEEE Trans.Commun., vol.COM-30, pp.593-599, April 1982.
- [20] T.A.Hawkes and P.A.Simonpieri, "Signal coding using asynchronous delta modulation," IEEE Trans.Commun., vol.COM-22, pp.346-348, March 1974.
- [21] B.Sankur and H.Gungen, "Signal coding properties of asynchronous delta modulator," Proc.Int.Conf.Acoust.Speech and Signal Proc., pp.1704-1708, Paris 1982.
- [22] H.Gungen, "Analysis of Asynchronous Delta Modulation Systems," M.S.Thesis, Bogazici University, 1978.
- [23] B.Sankur and W.Steenart, "Data compression of speech signals by variable rate sampling," Proc.GRETSI, pp.519-524, Nice 1983.
- [24] H.Inose, T.Aoki and R.Watanebe, "Asynchronous delta-modulation system," Elect.Letters, vol.2, no.3, pp.95-96, March 1966.
- [25] I.F.Blake and W.C.Lindsey, "Level crossing problems for random processes," IEEE Trans.Inform.Theory, vol.IT-19, pp.295-315, May 1973.
- [26] W.B.Davenport, Jr., "An experimental study of speech-wave probability distributions," JASA, vol.24, pp.390-399, July 1952
- [27] G.Modena, S.Sandri and C.Scagliola, "A mathematical model of an artificial signal for the testing of digital speech transmission systems," CSELT Rapporti Tecnici, vol.VI, pp.47-54, March 1978.
- [28] R.Billi and C.Scagliola, "Artificial signals and identification methods to evaluate the quality of speech coders," IEEE Trans. Commun., vol.COM-30, pp.325-335, Feb.1982.
- [29] A.E.Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," JASA, vol.49, pp.583-588, 1971.
- [30] J.Makhoul, "Linear prediction: A tutorial review," Proc IEEE, vol.63, pp.561-580, April 1975.
- [31] A.B.Carlson, Communication Systems, New York: McGraw-Hill, 1975
- [32] R.M.Gray, "Vector quantization," IEEE ASSP Magazine, vol.1, pp.4-29, April 1984.