SIMULATION AND SUBJECTIVE EVALUATION OF AN ADAPTIVE

C = 1

DIFFERENTIAL ENCODER FOR SPEECH SIGNALS

by

Bruce Albert Hanson

B.Sc., University of Manitoba, 1975

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in the Department

of

Electrical Engineering We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

June 1977



Bruce Albert Hanson, 1977

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical Engineering

The University of British Columbia 2075 Wesbrook Place Vancouver, Canada V6T 1W5

Date June 21, 1977

This thesis describes the subjective analysis of a DPCM system featuring an adaptive quantizer.

ABSTRACT

The system is simulated on a digital computer and operated under variations in the sampling frequency and the number of available quantizer levels. The subjective performance of the system is judged using the isopreference method which presents test results in the form of isopreference contours drawn on a plane showing sampling frequency and number of quantizer levels as axes.

From these curves the minimum required channel capacity for a given subjective preference level is shown to occur when sampling is at the Nyquist rate. The previous statement applies when the quantizer output levels are naturally coded or entropy coded. The isopreference contours indicate implementation tradeoffs between the number of quantizer levels and the sampling frequency. The isopreference contours also show that odd level quantizers outperform even level quantizers when entropy coding is used.

Analytical measures of performance in the form of output signal-to-noise ratio (SNR) are obtained. Although correlation between curves of constant SNR and curves of constant subjective quality are evident, the SNR curves do not accurately reflect the results of subjective evaluation. A special experiment involving quantizer dc offset is described which indicates that SNR could not be used to compare speech samples containing large proportions of different types of noise.

Throughout the work, the digital channel between encoder and decoder is assumed noiseless.

i

TABLE OF CONTENTS

ABS	STRAC	т	• • • •	••	• •	•••	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	i
TAF	BLE O	F CONTEI	NTS	••	••	•••	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	ii
LIS	ST OF	TABLES	•••	• •	••	•••	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	iv
LIS	ST OF	ILLUST	RATIONS		•••	• •	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	v
ACK	NOWLI	EDGEMEN	TS	• •	••	•••	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	vii
I	INTR	ODUCTIO	м	• •	••	•••	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	1
	1.1 1.2 1.3	Differ Review Scope	ential 1 of Pre- of the 2	Encod vious Thesi	ing Wor s.	of k.	Spe	eec •	2h •	• •	•	• •	•	• •	• •	•	• ,	•	• • •	•	•	• • •	• •	1 2 3
II	DPC	M SIMUL	ATION .	• •	••		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	5
	2.1 2.2 2.3	Introd DPCM R The Co 2.3.1 2.3.2 2.3.3	uction eview an hn and M Introdu The Pro The Qua	nd Te lelsa uction edict antiz	rmin DPCM n. or er	iolo [-	AQE	s: s	Sys	te	• • •	• • • •	• • • •	• • • •	• • • •	• • • •	• • •	• • • •	• • • • •	• • • • •	• • • • • •	• • • •	• • • •	5 5 8 10 10
IIJ	: ទហ	BJECTIV	E EVALU	ATION	•	•••	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	17
	3.1 3.2 3.3	Introdu Subject The Iso 3.3.1 3.3.2	uction tive Tea oprefere Genera Scaling	sting ence 1 1 Des g Iso	Rat Meth crip pref	ion od otic	nale on ence	• • •	Con		ur	•	• • •	• • •	• • •		• • •	• • • •	• • • •	• • •	• • •	• • •	• • •	17 17 19 19 20
	3.4 3.5	Prepar. Subjec 3.5.1 3.5.2	ation o: tive Tes Test Fo Object:	f Spe st Pr ormat Lve o	ech oced f th	Mat lure ie J	eri • •	al • •	•	• • •	• • •	• • •	• • •	•	• • •	• • •	• • •	• • •	• • •	• • •	• • •	• • •	• • •	22 26 26 29
IV	RES	ULTS OF	SUBJEC	rive '	TESI	s.	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	31
	4.1 4.2 4.3 4.4	Introd Determ Presen Discuss 4.4.1 4.4.2	uction ination tation sion of Discuss Entrop Cl	of E of Te Test sion y Cod nanne	xper st F Res of I ing 1 Ca	ime lesu sult sop and ipac	enta ilts is oref l Mi city	il Ser	Is	sop	re Re	efe Con equ	to	our ed	· · · · ·	Co	• • •	:01 • •	175	•	• • •	• • •	•	31 31 33 35 35 35
		4.4.3	SNR CO	npari	sons												•				•	•	•	42

٠

	4.5	Comparisons of the Results with Previous Work $$.	•	•	•	••	48
V	CONC	CLUSION	•	•	•	•••	51
	5.1 5.2	Summary	•	•	•	•••	51 54
REF	ERENC	ES	•	•	•		55

;

LIST OF TABLES

TABLE

2.1	Experimentally obtained parameter values 15
	 (a) Quantizer thresholds
	SCALE and BIAS
2.2	Optimum scaling factors for Cohn and Melsa's delta modulator
4.1	A comparison of SNR values as derived from samples processed with and without a dc offset
5.1	Comparison of approximate SNR values for subj non-adaptive previous-sample feedback DPCM, and DPCM-AQB

LIST OF ILLUSTRATIONS

FIGURE

2.1	(a) A PCM system (b) A DPCM system	6
2.2	Block diagram of the Cohn and Melsa DPCM - AQB system. (a) Transmitter (b) Receiver	9
2.3	Quanitzer Formats. (a) even number of levels (b) odd number of levels	11
3.1	(a) A typical isopreference contour.(b) A typical psychometric curve	21
3.2	 (a) Normalized amplitude probability density of speech. Symetrical average of positive and negative data. (b) Power density spectrum of speech. 	24 25
3.3	Block representation of a paired comparison test set	27
4.1	 (a) An experimental psychometric curve. (b) Preference in units of normal deviates plotted against a varying parameter. The line is fit- ted by a least-squares method. 	32
4.2	DPCM - AQB isopreference contours. The test signals of each contour are marked "x". Two transitiv- ity test signals are marked "o". 95% confidence intervals are denoted by a bar through each ex- perimental point. SNR _{subi} values are given in	
	dB as are the SNR values which appear enclosed in brackets	34
4.3	Curves of constant bit rate. Superimposed are the isopreference contours (dashed) of Fig. 4.2	38
4.4	Subjective ratings of Fig. 4.2 plotted against their respective minimum required channel capacities	39
4.5	Curves of constant bit rate using entropy coding. Superimposed are the isopreference contours (dashed) of Fig. 4.2	40
4.6	Matrix of bit rate reduction coefficients r for entropy coding	43

FIGURE

(

4.7	Curves of constant signal to noise ratio. Superimposed are the isopreference contours (dashed) of Fig. 4.2
4.8	SNR vs. f/f curves for the 2, 3, 4, 5, 6 and 7 level quantizers studied. Also shown is
	a curve resulting from Jayant's one-bit mem- ory delta modulator

ACKNOWLEDGEMENTS

I am grateful to Teleglobe of Canada for support received under their fellowship program. Grateful acknowledgement is also given for the university research assistanships received from 1976 to 1977.

Special thanks are extended to Dr. R.W. Donaldson for his patience and help throughout the life of the project.

Thanks are also extended to Messrs. M. Koombes and A. MacKenzie for their technical assistance and to Ms. M.E. Flanagan for typing the manuscript.

I INTRODUCTION

1.1 DIFFERENTIAL ENCODING OF SPEECH

The miniaturization of digital components has enabled digital electronics to enter into all aspects of modern industry. The field of communications is no exception and the search to find simpler and more efficient methods to digitally code speech has been continuing for several years.

The system most commonly used to date is pulse code modulation (PCM). A variation of this method called differential PCM (DPCM) and the much more simply structured delta modulator (DM) are two common techniques being considered for improved A/D conversion. These schemes code the difference between the input signal and a system-generated predictor signal.

Two of the major problems encountered when using DPCM or DM are the introduction of noise through quantization and the determination of optimum stepsize to minimize that noise. If the step size is too small, then the quantizer will not be able to follow large changes in the input signal, while an overly large stepsize will introduce unwanted granular noise. By permitting either, or both, of the quantizer or predictor of such systems to be adaptive, the encoder is made self-adjusting to better suit the varying statistics of the input signal.

Another of the difficulties faced in studying such systems is obtaining an accurate measure of performance. The most common mathematical approach has been to use mean square error or signal to noise ratios. However, the final test of a system used with a human observer as a sink is the subjective quality of the output signal. It is well known that purely analytical measures, such as signal to noise ratio, do not necessarily reflect system performance as perceived by human subjects.

The interest spurred by recently developed adaptive quantization schemes, along with the continuing need for subjective evaluation of such systems, has led to the subjective evaluation of a DPCM system featuring an adaptive quantizer presented in this thesis.

1.2 REVIEW OF PREVIOUS WORK

Various techniques for analogue to digital conversion in the form of delta modulation and differential PCM have been studied for several years. O'Neal [01, 02] has investigated the use of DM and DPCM onttelevision and Gaussian signals. McDonald [M1] has shown DPCM to be superior to PCM for speech applications.

It has been discovered that by allowing the quantizers of these systems to be self-adapting through the application of certain algorithms, improvements in signal reproduction are possible. DM has been of particular interest to many researchers because of its simplicity. Jayant [J1] has proposed an adaptive quantizer for a DM encoder using a one-bit memory. He has also conducted bit-sequence correlation studies on such a system [J4]. Tazaki et al [T1] have derived a set of equations which can represent several previously published formulas including Jayant's DM. Adaptive quantizers have also been applied to DPCM systems. Cohn and Melsa [C4], Qureshi and Forney [Q1], and Cummiskey et al [C6] have proposed and tested different algorithms for quantizer adaptation. The different types of encoding systems referred to above, have been brought together by Noll [N1] who has completed a comparative study of quantizing schemes for speech systems.

Many attempts have been made to analyse and measure the performance of coding schemes using a mathematical approach. Goodman [G3] has devised expressions for the quantizing noise in DM and PCM systems. Green-

stein [G5] has derived equations to calculate slope overload noise in delta modulators. More recently, Goldstein and Lui [G2] have derived equations describing the three basic types of quantization noise appearing in a DPCM system featuring an adaptive quantizer.

Others have approached the problem of performance evaluation using methods based on subjective perception. Donaldson et al [D1, D2, C2, Y1] have used extensive subjective testing for evaluating systems operating on speech signals. Their method of evaluation is based on the isopreference method first described by Munson and Karlin [M4]. Grether and Stroh [G6] on the other hand have successfully used a version of the category judgement method.

1.3 SCOPE OF THE THESIS

The purpose of this thesis is to investigate various aspects of the performance of a differential pulse code modulator utilizing an adaptive quantizer operating on speech signals. The two major parameters under study are the number of quantization levels and the sampling rate relative to the Nyquist sampling rate.

The model used for this study is presented and discussed in some detail in Chapter 2. Optimization of system parameters is also considered. For reasons of simplicity and reproducibility, simulation of the model is accomplished using a high level programming language on an IBM 370/168 computer.

Chapter 3 of the thesis describes subjective evaluation related to voice communication systems. A comparison of intelligibility testing, subjective testing, and analytical measures of performance is given. Following this an explanation of the isopreference method is presented. The chapter concludes with a description of the manner in which data was prepared and then presented for subjective evaluation.

Chapter 4 presents the results of the subjective tests. A plot of the isopreference contours as determined by analysis of the subjective test results is given. Following this presentation is a discussion of the contours, comparisons with signal to noise measurements and comparisons with previous relevant work. Also considered are the advantages of bit rate reduction schemes employing entropy coding.

Chapter 5 concludes the thesis with a summary of the work and its implications.

II DPCM SIMULATION

2.1 INTRODUCTION

This chapter presents a review of Differential Pulse Code Modulation (DPCM) and explains the terminology used in this and following chapter. A detailed description of the DPCM model used in this work is then given.

2.2 DPCM REVIEW AND TERMINOLOGY

The terminology presented in this section and throughout the paper will follow that of Noll [N1]. Noll also presents a good comparison of different quantizing schemes for those wishing further detail.

A pulse code modulation (PCM) system is shown in Figure 2.1a. (A version of this scheme using an 8-bit quantizer is now being used in the industry.) Operation of this system results in the input signal being band-limited, sampled at or just above the Nyquist rate, logarithmically quantized, and then coded for transmission. The receiver performs the reverse steps using an inverted quantizer.

In differential PCM (DPCM) (see Figure 2.1b) the addition of a feedback loop and adder effectively subtracts a predicted value, p_k , from the input sample s_k . Estimate p_k is generally a linear sum of past quantizer outputs; thus

$$p_{k} = \sum_{i=1}^{N} a_{k}^{(i)} \hat{s}_{k-i}^{(2.1)}$$

The resulting difference or error signal, e_k is quantized and transmitted. As the error signal is of lower redundancy than the original input signal, coding can generally be accomplished using fewer bits than a comparable PCM system. Conversely, quality could be improved for a given bit rate. This fact has been shown analytically and subjectively in many experiments



Fig. 2.1 (a) A PCM system (b) A DPCM system

6

[N2, D1, D2, G6, O2].

The receiver section of the DPCM system adds the received error e_{t} to the predicted value to arrive at the estimated sample \hat{s}_{t} .

A further improvement in signal reproduction has been introduced by using an adaptive predictor. That is, the predictor coefficients are modified according to some algorithm. This system is appropriately referred to as an adaptive DPCM (ADPCM) system. If adaptation of the predictor coefficients is generated from the original input signal then the scheme is referred to as forward prediction. That is, the predictor coefficients must be transmitted forward to the receiver as it has no knowledge of the adaptation strategy.

On the other hand if predictor coefficients are generated from the quantizer output the scheme is called backward prediction. In this, case parameter transmission is not required because the receiver has all the information needed to reproduce the required coefficients. Although this latter method has the attractive quality of not increasing the bit rate to accommodate predictor coefficient adaptation, it has been shown to be unsuitable when used on channels with high channel bit error probability [N2].

Reconstructed signal quality can be improved or the bit rate made lower by making the quantizer adaptive. This new system has been called a residual coder by Cohn and Melsa [C5]. However in keeping with the terminology of this thesis and that of Noll it will be referred to as an ADPCM - adaptive quantizer (ADPCM - AQ). Both backward adaptive quantization schemes (ADPCM - AQB) and forward (ADPCM - AQF) schemes are possible and different algorithms have been proposed and studied [C5, C7, J2, J3, M3, Q1].

A special case of the systems described above is the delta modulator (DM). The quantizer of this system contains only two levels and is of special interest because of its simplicity. It too has been improved through the use of adaptive quantizers (DM - AQ) [C8, J1, J3, S1]. Many studies concerning delta modulators have been carried out [C1, G3, G5, T1].

The system under study in this thesis is a DPCM - AQB incorporating an adaptive quantizer algorithm as derived by Cohn and Melsa [C5]. It represents one of the few systems devised to date which comes close to filling the three basic requirements; low bit-rate, good quality speech and relative simplicity of implementation.

2.3 THE COHN AND MELSA DPCM - AQB SYSTEM

2.3.1 Introduction

The system used in this work is modelled after the one presented by Cohn and Melsa [C5]. Their system is an ADPCM - AQB which attempts to estimate the standard deviation of the input signal and normalize it before quantization. The system is depicted in Figure 2.2. Note that all receiver variables maintain the same values as their transmitter counterparts as long as the channel is error free.

To simplify system implementation Cohn and Melsa's adaptive predictor has been replaced in our study with a linear time invariant predictor based on the immediately preceeding receiver output \hat{s}_k . Most predictor adaptation algorithms including the one presented by Cohn and Melsa involve much calculation. Furthermore, Qureshi [Q1] has shown that a system with a fixed predictor performs only 1 to 2 dB worse than the same system with an adaptive predictor. Although Cohn and Melsa reported a more appreciable difference of 4 to 5 dB it should be noted that the author's results without an adaptive predictor came to within 2-3 dB of Cohn and



Melsa's published results obtained using the adaptive predictor.

The bit rate reduction which is obtained by source coding the quantizer output signal has also been examined.

2.3.2 The Predictor

Assnoted earlier the predictor output p_k is formed from a linear combination of previous receiver outputs.

$$p_{k} = \sum_{i=1}^{N} a_{k}(i) \hat{s}_{k-i}$$
 (2.2)

In our study n = 1 and a_k was set to an experimentally determined optimum value of 0.8.

Computer simulation of the model necessitated the threshold factor in Figure 2.2 being placed before the delay element of the predictor in odd level quantizers. The decision element's output

$$\hat{s}_{k} = 0 \qquad \text{if} \qquad z_{k} < .01 \qquad (2.3)$$
$$\hat{s}_{k} = z_{k} \qquad \text{if} \qquad z_{k} \ge .01 \qquad (2.4)$$

forced all low level outputs to equal zero. This threshold rule prevented \hat{s}_k from exponentially approaching zero in the case of a very low or zero level input. The problem was also solved by adding a small amount of noise to the input signal however, this solution was not used in this study.

2.3.3 The Quantizer

Two, three, four, five, six and seven output level quantizers were tested. Therefore, both odd and even level formats were needed (see Figure 2.3). For the quantizers considered all parameter values were symmetric.

Unlike standard PCM systems, the process of quantization must be broken into two sections, a quantization and an inverse quantization



(see Figure 2.2). Operation of the quantizer is straight forward. The input sample e_k is compared to the quantizer thresholds, $\sigma_k T_i$ and the range into which the sample falls is determined. This range specifies quantizer and transmitter output, q_k . The inverse quantizer receives the level q_k as its input and produces an output \hat{e}_k as defined by the product of the scaling factor $F(q_k)$ and the state variable σ_k which is described at a later point.

As the nonlinearities of the system makes mathematical optimization difficult, a random search was used to determine the optimum threshold and scaling factor values. The signal to noise ratio of the entire DPCM - AQB model was used as the optimization criterion It is typically measured in decibels (dB) and is computed by

SNR = 10
$$\log_{10} \frac{E[s^2]}{E[(s-\hat{s})^2]}$$
 (2.4)

Cohn and Melsa on the other hand optimized over the value $E[(e_k - \hat{e}_k)^2]$. As the SNR approach produced the same optimum values for the five and seven level quantizers as those published by Cohn and Melsa it suggests that both methods are equally valid. Tables 2.1a and 2.1b give the experimentally determined optimum values.¹

As the system was to operate under a range of sampling frequencies above the Nyquist rate, the quantizer parameters were reoptimized at two other sampling frequencies. Values obtained were very close to those obtained at the Nyquist rate. Also any change in SNR which resulted in using a reoptimized parameter was quite small, usually not much greater than 0.1 dB. It was therefore concluded that the model was stable over the range of sampling frequencies chosen and the initial values given in Tables 2.1a and 2.1b were maintained for all sampling frequencies under study.

¹All parameter values and data for graphs were obtained during or by repeating the stage of signal processing which resulted in the formation of the sample data base as described in Section 3.4.

The exception to the above statement occurred in the case of the two level quantizer or delta modulator. The optimum value of the scaling factor did vary with frequency and reoptimization was necessary at all sampling frequencies (see Table 2.2).

Adaptation of the quantizer is baseddon an estimate of the standard deviation of the quantizer input signal e_k . As the optimum threshold for quantizing a given variable varies linearly with the standard deviation of that variable [C5], dividing the input signal e_k by its standard deviation will result in a normalized signal with a standard deviation of unity. A quantizer with fixed thresholds can then be designed.

Alternately one can view the process as an attempt to keep the quantizer within operating range of the difference signal by a series of expansions and contractions.

The algorithm used for estimating the standard deviation of e_k is that described by Cohn and Melsa. It operates on two levels. For periods of unvoiced speech or silence a moving average of \hat{s}_k is used to estimate the standard deviation of e_k . As e_k is not available at the receiver an alternate signal must be used. Signal \hat{s}_k is used rather than \hat{e}_k as its SNR is better while at the same time its envelope tends to be very similar to that of e_k .

For voiced speech the standard deviation of e_k is very large at the beginning of a pitch period and the moving scaled average is no longer a good estimate. Therefore, a feature has been included to allow fast adaptation. Whenever either of the outermost quantizer levels occurs indicating a sharp increase in signal magnitude, the discrete standard deviation σ_k is significantly increased by a factor α (outermost level). This effectively pushes the quantizer levels out to accommodate high level

signals. If no further outer levels occur σ_k decays back to the scaled average. Accordingly σ_k is calculated by the equation:

$$\sigma_{k} = MAX (\alpha(q_{k}) \cdot \sigma_{k-1}, E[|\hat{s}_{k}|] / SCALE + BIAS)$$
 (2.5)

where:

 $\alpha(q_k)$ are the expansion-contraction coefficients, q_k is the quantizer output, $E[|\hat{s}_k|] / SCALE$ is the moving scaled average of \hat{s}_k , and BIAS maintains σ_k at a minimum or base value for low level signals.

The expansion-contraction coefficients, $\alpha(q_k)$ were obtained in the same way described earlier for the thresholds and scaling factors. Again results for the five and seven level cases matched those of Cohn and Melsa and remained essentially constant with changes in sampling frequency, (Table 2.1c). As delta modulation involves only two levels the expansioncontraction coefficients have been set equal to zero. Thus adaptation in the case of delta modulation depends only on the estimate of the standard deviation of e_{t_r} .

The values selected for BIAS and SCALE are also given in Table 2.1c. The difference between these and Cohn and Melsa's values may be explained by differences in the initial stages of data preparation, in particular the analogue to digital conversion.

The average $E[|\hat{s}_k|]$ is calculated from a moving window covering the one hundred samples preceding the one currently being processed.

$$E[|\hat{s}_{k}|] = \frac{\sum_{i=1}^{N} |\hat{s}_{k-i}|}{N} , \quad N = 100 \quad (2.6)$$

The above model was simulated on a digital computer and used to process all data for the thesis.

Table 2.1 Experimentally obtained optimum parameter values

- (a) quantizer thresholds
 (b) quantizer scaling factors
 (c) expansion-contraction coefficients plus SCALE and BIAS

		NUMBER	OF QUA	ŅTIZATI	ON LEVE	LS
	2	3	4	5	6	7
T 1	0	1.4	0	1.0	0	.5
^т 2	-	-	3.0	3.5	1.5	1.5
^{́т} з	-	-	-	-	3.5	3.5
			(a)			
	Table					
F(1)	2 ? 2 2	. 0	1.25	0	.75	0
F(2)	Table 2.2	-3.0	-1.25	-2.0	75	-1.0
F(3)	-	3.0	5.0	2.0	2.25	1.0
F(4)	-	-	-5.0	-5.25	-2.25	-2.0
F(5)	-	-	-	5.25	5,0	2.0
F(6)	-	-	-	-	-5.0	-4.5
F(7)	-	-	-	-	-	4.5
			(b)		1	
α(1)	0	.6	.30	.40	.50	.70
α(2)	0	1.35	.30	.80	.50	.80
α(3)	-	1.35	1.30	.80	.90	.80
α(4)	-	-	1.30	2.20	.90	.90
α(5)	-	-	-	2.20	1.70	.90
α(6)	-	-	-	-	1.70	2.30
α(7)	-	-	-	-	-	2.30
BIAS	1.0	1.0	1.0	1.0	1.0	1.0
SCALE	5.0	5.0	5.0	5.0	5.0	5.0

NUMBED OF OUNNET 7ATTON JEVELS

fs/fn	F(1)	F(2)	
.1	4.5	-4.5	
1.25	4.0	-4.0	
1.5	3.5	-3.5	
1.75	3.5	-3.5	
2	3.0	-3.0	,
2.25	3.0	-3.0	
2.5 2.75	3.0 2.75	-3.0 -3.0	
3	2.75	-2.75	
3.25	2.75	-2.75	
3.5	2.75	-2.75	

Table 2.2 Optimum scaling factors for Cohn and Melsa's delta modulatorn.

III SUBJECTIVE EVALUATION

3.1 INTRODUCTION

This chapter deals with subjective evaluation related to voice communication systems. A brief comment is first offered in Section 3.2 on the differences between evaluation of systems by the means of articulation and intelligibility tests, analytical means such as signal to noise power ratios, and subjective tests based on preference, such as the isopreference method. Section 3.3 of this chapter outlines isopreference testing as used in this paper. Section 3.4 then describes the phase of data preparation. The sentence used for evaluation purposes is presented, with arguments for its choice, followed by a description of the method involved in producing the test material. The section 3.5 which concludes the chapter describes the tests themselves.

3.2 SUBJECTIVE TESTING RATIONALE

A brief comment is in order concerning the evaluation of a system using articulation and intelligibility tests as opposed to using subjective tests based on preference, such as the isopreference method.

Articulation testing pertains to the comprehension of units of speech material consisting of meaningless syllables or fragments of speech. Intelligibility testing refers to the comprehension of phonetically balanced units of speech material such as meaningfull words, phrases or sentences [M2]. The two terms however are often confused as is the term 'articulation index'. For this reason the term articulation will be avoided. Instead the 'intelligibility index' will be defined as the percentage of units of speech correctly identified during an intelligibility test.

An interesting observation to make is that subjective tests do not necessarily reflect intelligibility. That is, as long as a high quality

signal is being used, the subjective tests, similar to the one described in this work, are essentially independent of intelligibility. This statement can be verified by considering the elements affecting intelligibility of a processed signal, the main two elements being filtering, and distortion caused by system noise. Concerning the tests carried out in this thesis a signal bandlimited from 200 to 3200 Hz retains an intelligibility index of approximately 90% [C6, K1]. It is also interesting to note that by using a closed set of test samples the effect of intelligibility losses due to system noise can be ignored. This is due to the fact that complete knowledge of the test material by the listeners removes the stipulation that the threshold of recognition of a word heard in noise be inversely proportional to the logarithm of its frequency of occurrence [W1]. Knowledge of the test material therefore has the effect of testing with samples having an apparent intelligibility index of 100% even though it may be some what less.

The preceeding discussion is presented not to cast doubt on the worthiness of subjective tests but rather to clarify the difference between a subjective rating such as an isopreference contour and an intelligibility score.

The converse of the above discussion states that while a signal may be 100% intelligible it may not possess, from a subjective point of view, the quality or naturalness of the original signal. It is for this reason that methods such as isopreference tests are necessary. Although points along an isopreference contour may conceivably possess different

intelligibility scores, one would expect that highly rated signals will reflect relatively high intelligibility indices in which case limits on intelligibility would be set by factors such as audio bandwidth.

A second argument for subjective tests arises from the inability of analytical methods, for example signal to noise power ratios, to reflect the signal quality as perceived by human subjects [C6, G4]. Increasingly, the practice has been to include some form of subjective testing of a system in its analysis. Various methods of subjective evaluation have been proposed and tested [M2, G6, M4]. The three main methods have been presented and discussed in "IEEE Recommended Practices for Speech Quality Measurements" [11].

3.3 THE ISOPREFERENCE METHOD

3.3.1 General Description

Originally proposed by Munson and Karlin [M4] the isopreference method has been studied and applied by numerous researchers [D1, R1, T2, Y1]. Because the method has been adequately described in numerous papers, only a brief description is given here.

The isopreference method assumes that the speech signals under test can be judged on the unidimensional scale of preference. This assumption allows a series of isopreference contours to be drawn on a plane whose axes are measures of the parameters under test. Points that lie on the contours are determined by a series of paired comparison tests presented in random order. A test signal, for example point A in Figure 3.1 a, is initially picked to define the subject quality of one curve. This signal, whose parameters are held constant is then compared to another signal with one varying parameter. As the parameter is varied a value is obtained for which all listeners show an equal preference for both signals. The two signals are then declared to be isopreferent. Whether parameter α or parameter β of Figure 3.1a is varied depends on the expected shape of the contour.

Since the parameters are varied by discrete amounts the results of the comparisons are generally expressed in proportions of subjects not preferring the test signal. The results are then plotted against the varying parameters. A smooth psychometric curve is then drawn through the experimental points as shown in Figure 3.1b. From this curve the abscissa corresponding to a proportion of one-half defines the value of the varying parameter that defines the isopreferent point. Repeating this process using various values of α and β results in an isopreference curve being drawn through the original test signal.

3.3.2 Scaling Isopreference Contours

As isopreference curves generally include points possessing different signal to noise ratios it is desirable to attach a common standard of quality to each curve. Various speech rating standards have been proposed and tested [D1, H1, R1, S2, S3].

The method of scaling generally used is that of comparing a test signal on each contour to a family of standard reference signals. These reference signals are generated by adding varying amounts of a degradation signal to a high quality signal. Paired comparison tests are then used to determine which reference signals are isopreferent to the test signals. The amount of degradation in the isopreferent reference signals, given by subjective signal to noise measure, is then attached to the isopreference contours from which the respective test signals were taken.

In this study the method of generating a family of reference signals introduced by Schroeder [S2] is used. The method produces ref-





Fig. 3.1 (a) A typical isopreference contour (b) A typical psychometric curve

erence signals defined by the equation

$$r_{\alpha}(t_k) = (1 + \alpha^2)^{-\frac{1}{2}} [s(t_k) + \alpha \cdot n(t_k)]$$
 (3.1)

where α defines the signal to noise ratio, SNR_{subi} , in dB via:

$$SNR_{subj} = 10 \log_{10} \alpha^{-2}$$
 (3.2)

The noise sample $n(t_k)$ is obtained by multiplying the signal sample $s(t_k)$ by a zero mean descrete stochastic process $e(t_k) = \pm 1$ which is uncorrelated with the signal.

Such a method was chosen over degradation using white gaussian noise, since the noise introduced by DPCM coding is signal dependent. As noted by Schroeder, degeneration with white gaussian noise does not result in the same subjective quality degradation, as quantization noise thereby making comparison more difficult when the two types of noise are compared.

3.4 PREPARATION OF SPEECH MATERIAL

The speech material chosen for the subjective evaluation of Cohn and Melsa's DPCM - AQB system was the pair of sentences "Joe took father's shoe bench out. She was waiting at my lawn". The pair was nominally low pass filtered at 3200 Hz in keeping with Cohn and Melsa's study and high pass filtered at 200 Hz to eliminate any low frequency noise such as 60 cycle hum.

The sentence "Joe ... lawn" was spoken by a thirty-eight year old male with a western Canadian accent. The sentence was repeated in an Industrial Acoustics Company model 1205-A quiet room. A full bandwidth recording was then obtained using a single track Scully 280 recorder operating at 15 i.p.s. with low noise Ampex 434 audio tape and a Bruël and Kjoer Type 2801 power amplifier and microphone set. Speech statistics of the sentence "Joe ... lawn" are given in Figure 3.2a. Chan and Donaldson [C2] have shown that the amplitude probability of the test sentence normalized with respect to its R.M.S. value is reasonably close to that of both Gaussian and Laplacian distributions, both of which have been suggested as models of the amplitude density of speech. Benson and Hirch [B1] have compared the spectrum of the sentence to samples of news and technical material and found them to be not significantly different (Figure 3.2b). The sentence "Joe ... lawn" can therefore be regarded as a reasonable representation of conversational speech.

Once recorded the sentence was played back and digitized using the system described by Chan [C3]. Eleven master samples were produced by repeatedly playing back the sentence and adjusting the effective sampling rate from 6,400 to 22,400 Hz at intervals of 3200 Hz. The sampled signals were uniformly quantized to twelve bits and stored on nine - track IBM compatible digital magnetic tape. The tapes were then transported to the IBM 370/168 facilities where processing was accomplished. All samples were initially normalized to a mean of zero thereby eliminating any d.c. bias introduced by the band-pass filters. Simulation of the DPCM - AQB algorithm was carried out in the PL/1 programming language. Each of the eleven master samples was processed six times using each of the six quantizers described in Chapter 2. This produced sixty-six samples which comprised the main data base.

The Nyquist - sampled master signal was then processed using the Schroeder algorithm described in Section 3.3.2. A family of standard reference signals were thereby obtained with SNR_{subj} values ranging from -2 to 34 dB in steps of 2 dB. These were added to the data base.

Samples to be used in the subjective listening tests were then transferred from the data base to other nine - track tapes in the order in



10₀ 25. Fig. 3.2 (b) Power density spectrum of speech. 1 SPEECH POWER DENSITY SPECTRUM .1 .01 BENSON & HIRSH DATA [C2] .001 .0001 L 01 10 .1 1 FREQUENCY (kHz)

which they were to appear on the analogue test tapes. These nine - track tapes were then returned to the facilities previously mentioned and passed back through digital to analogue converters and filters to produce the analogue test tapes. During analogue to digital conversion loudness was controlled by monitoring the record amplifier of the Scully tape deck.

3.5 SUBJECTIVE TEST PROCEDURE

3.5.1 Test Format

A total of seven one-half hour sessions were used to accumulate the data analyzed in Chapters 4. Each session consisted of paired comparison tests. The sentence "Joe took father's shoe bench out. She was waiting at my lawn" was used in all cases. All tests were conducted with the guidelines of the IEEE recommendations in mind [I1].

Each paired comparison was presented in a set as shown in Figure 3.3. The first speech sample of a pair, designated as A, and the second, as B, were immediately repeated to form one set. Each sentence of a set was preceeded by a one second pause and each set was followed by a three second pause during which time the subjects could mark their decision, or preference, on supplied answer forms. A tone indicated the beginning of a new set. In the course of the tests each pair was presented a second time with samples A and B appearing in reverse order.

The half hour sessions were divided into two parts. Thirty-one sets were presented during the first fifteen minutes. A five minute break followed during which limited discussions were held on the topic under study.in an attempt to increase interest and eliminate fatigue. Twenty more sets were presented in the remaining ten minutes. The first set of each session was a familiarization set and was not included in the results. During this set participants were allowed to adjust their volume controls



Fig. 3.3 Block representation of a paired comparison test set.

27

from a preset medium as chosen from the results of the pilot test (see Section 3.5.2). No further adjustment in volume was allowed.

The sessions were conducted in a quiet classroom. The tapes were played back on the Scully 280 tape recorder, through an amplifier to nine individual volume controls. Sharpe HA - 10 - MK - II and Jensen model 220 stero headsets were used for the tests. Both headsets demonstrated similar frequency response curves and possessed -40 dB isolation characteristics.

Prior to the listening sessions the listeners were read the following instructions:

"The following speech samples are the result of a sentence which has been processed by a communications systems algorithm in which several parameters have been varied.

The samples will be presented in pairs. The first speech sample of each pair will be designated as A, the second as B. Each pair will be immediately repeated. A three second pause will follow to allow you to mark on the answer sheet which of the two speech samples you would prefer to listen to. A 'tone' will indicate the beginning of the next set of pairs.

In making your decision please ignore any clicks that may occur immediately before or after each speech sample. Also please try to ignore any volume differences. Please be as attentive as possible for a lack of concentration will lead to confusion. In the case of two samples being of equal preference, in your opinion, choose the second sample.

The sentence you will hear is 'Joe took father's shoe bench out. She was waiting at my lawn.'"

A total of eighteen listeners, fifteen male and three female, participated in the tests. All were university students ranging in age

from eighteen to thirty-two years and representing various cultural backgrounds. All had no previous experience in listening test and exhibited no hearing abnormalities as tested for in the pilot test, or other hearing abnormalities known to themselves. All listeners participated in every test.

3.5.2 Objective of the Tests

The seven test sessions were divided into three test aspects: a pilot test, a test for determining isopreference contours, and a test for rating the isopreference contours. The three tests comprised one, four, and two sessions respectively and spanned a period of four weeks.

The pilot test was run with three objectives in mind. The first objective was to choose a set of points that could be used for subsequent measurement of the subjective quality of the proposed isopreference contours. One of these points would appear in each of the paired comparisons of the next two test aspects. The second objective was to ensure that listeners were capable of consistently discerning speech quality of signals whose SNR values were within 3-6 dB of each other. The third objective was to permit subjects to select their individual volume settings. All volume controls would be initially preset to a single value as determined by the mean of these settings for all ensuing tests.

The test for determining the isopreference curves covered four sessions. The first session was used to discover the general characteristics of curves defined by the points chosen from the pilot test. The remaining three sessions were dedicated to defining precisely the isopreference contours. Results are presented in Chapter 4.

The third test aspect involved two sessions and utilized Schroeder's reference signals. As long as transitivity of subjective preference can be

assumed along the curves, and this is one of the basic assumptions of the isopreference method, then rating any point on a contour is equivalent to rating the whole contour. On this assumption, each of the test points as arrived at in the pilot test was compared to the reference signals to determine its isopreferent "mate". The value of SNR subj of the reference signal was then attached to the respective curve. Included in these tests were two extra points to test the transitivity assumption.

In this way a complete set of data was collected to which the isopreference analysis method could be applied.

IV RESULTS OF SUBJECTIVE TESTS

4.1 INTRODUCTION

The listening tests described in Chapter 3 resulted in an accumulation of data based on preference. This data was analyzed in order to determine isopreference contours. Section 4.2 outlines the method used to determine the value of a parameter which results in one signal being isopreferent to a test signal. The isopreference contours as determined by the data collected in this thesis are presented in Section 4.3. A comprehensive discussion of the results and their implications is given in Section 4.4 and 4.5.

4.2 DETERMINATION OF EXPERIMENTAL ISOPREFERENCE CONTOURS

The method outlined in Section 3.3.1 was used to determine those values of the independent system parameters for which a reconstructed speech signal is isopreferent to a test signal. After plotting the proportion of listeners not preferring the test signal a smooth curve can be drawn through the points which is assumed to be a cummulative normal curve relating the proportion, p, to the parameter values, L. (See Figure 4.1a.)

The Kolmogorov-Smirnov (K-S) goodness of fit test [L1] was used on all data to test the assumption of normality. The statistic used is the maximum absolute deviation of the experimental curve $F_n(x)$ to form the hypothesized curve $F_n(x)$ represented by D_n in equation 4.1.

$$D_n = \frac{SUP}{x} | F_n(x) - F_o(x) | \qquad (4.1)$$

All but a few of the cummulative curves resulting from the test data passed the K-S test at a significance level of .01.



35.*

Once the normality criterion had been justified the p values were transformed into measures of unit normal deviates, z. (See Figure 4.1b.) An approximately linear relationship between z and L resulted. A leastsquare solution using Müller-Urban weights was used to fit a straight line

$$y = a + bx \tag{4.2}$$

to the data points. The estimated mean \bar{x} and the estimated standard deviation S_x could then be obtained from (4.3) and (4.4).

$$\overline{\mathbf{x}} = -\frac{\mathbf{a}}{\mathbf{b}} \tag{4.3}$$

$$S_{\mathbf{x}} = \frac{1}{b} \tag{4.4}$$

This mean was then taken as the value of L which produced the isopreferent signal. The standard deviation was inserted in (4.5) to calculate the 95% confidence interval for the mean [L1]. Let μ represent the population mean of which \bar{x} is the estimate, where

$$\bar{\mathbf{x}} - \mathbf{t}_{\alpha} \frac{\mathbf{S}_{\mathbf{x}}}{\sqrt{\mathbf{n}-1}} < \mu < \bar{\mathbf{x}} + \mathbf{t}_{\alpha} \frac{\mathbf{S}_{\mathbf{x}}}{\sqrt{\mathbf{n}-1}}$$
(4.5)

The size of the confidence interval is given by $100(1 - \alpha)$ % where α is the significance level, t_{α} is a tabulated value corresponding to a t-distribution, and n is the sample size.

4.3 PRESENTATION OF TEST RESULTS

The isopreference contours as determined from the isopreference points obtained using the method described in Section 4.2 are presented in Figure 4.2. The two parameters which define the plane are the number of quantization levels, L, used in the DPCM - AQB quantizer, and the ratio of sampling frequency to the Nyquist frequency, f_s/f_p .



Fig. 4.2 DPCM - AQB isopreference contours. The test signals of each contour are marked "x". Two transitivity test signals are marked "o". 95% confidence intervals are denoted by a bar through each experimental point. SNR subj which appear enclosed in brackets.

The test signals as determined by the pilot test and used in the paired comparison tests to estimate the isopreference contours are marked by an "X". Beside each of these points is given the estimated SNR_{subj}, and its associated 95% confidence limits. Beneath these values in parentheses are the computed SNR's of those points. Because of the assumption of transitivity along the isopreference contours the SNR_{subj} of a test signal applies to all points on that curve. The SNR_{subj} of two other points marked "0" has also been determined. Their values support the assumption of transitivity and also indicate listener judgement consistency.

The points determined experimentally as being isopreferent to the test signals are marked ".". A bar through each point indicates the 95% confidence interval of the mean calculated using (4.5). The unit of measure for each interval is defined by the axis to which it is parallel.

The curves themselves were based on best visual fits to the data points. Constraints affecting their positioning were that they have the same general shape as neighboring curves, and that they be drawn close to points possessing small confidence intervals.

4.4 DISCUSSION OF THE TEST RESULTS

4.4.1 Discussion of Isopreference Contours

Several facts can be deduced from Figure 4.2. These will be presented in this and following sections.

As the sampling frequency is increased the correlation between samples is also increased. This increased correlation allows the adaptation strategy of the quantizer to better follow the input signal statistics resulting in a higher quality output signal. Figure 4.2 shows however,

that most of this improvement takes place in the first stages of frequency increase. Beyond this a saturation zone is encountered in which increasing the sampling frequency has a lesser effect on signal quality. The cause of this may be attributed to quantization noise. That is, any further gain made by increasing sample correlation is masked by the dominant quantizing noise.

To the designer, saturization zones of this type mean a limitation or lower bound on parameter values. For example, to obtain a subjective quality of 25 dB only quantizers with five or more levels need be considered.

The plot also shows that certain tradeoffs are possible between the number of quantization levels and the sampling rate. For example, the much more easily implemented delta modulator could replace a five level quantizer simply by oversampling at 3.25 times the Nyquist rate to obtain a subjective performance of approximately 13.5 dB. One consideration that may detract from carrying out such a replacement may be bit rate considerations. This aspect is presented in the following section. 4.4.2 Entropy Coding and Minimum Required Channel Capacity

Most studies to date on DPCM and ADPCM systems involve quantizers for which

$$L = 2^{b}$$
(4.6)

where L denotes the number of quantization levels and b is normally an integer equal to the number of bits required to code each input sample. For equiprobable quantizer output levels, such a scheme results in a minimum bit rate or minimum required channel capacity of

$$C = b \cdot f_s$$
 bits per second (4.7)

On the basis of (4.7) the bit rate resulting from all combinations of quantization levels and sampling frequency has been calculated. The solid curves of Figure 4.3 represent paths of equal bit-rate. Superimposed on these curves are the isopreference contours of Figure 4.2. As any two curves, one from each set, intersect at only one point it becomes obvious that the minimum bit rate for a given preference level occurs when sampling is at the Nyquist rate, or just far enough above the Nyquist rate to ensure integer values of L. The only exception to this rule may apply in very low quality regions where the two sets of curves become almost parallel.

These minimum required channel capacities as estimated from Fig. 4.3 are plotted against SNR_{subj} values in Fig. 4.4. The bars through the points indicate the 95% confidence intervals. The shaded region of the graph is bounded on one side by a line fitted to the lower four points by the least-squares method, and on the other side by the minimum obtainable bit-rate for the value of parameters covered by this study.

If bit rate and preference level are held constant in Figure 4.3, the only design implementation tradeoffs appear in the low quality region of the plot. Outside of this region it would be necessary to design around a quantizer which is positioned at the intersection of the given bit rate and preference curves.

It is possible to reduce bit rates by employing coding schemes which take advantage of the fact that quantizer output levels are not normally equiprobable. Coding schemes such as these, developed around quantizer statistics, are referred to as entropy coding. Cohn and Melsa have proposed such a source coding scheme to reduce the bit rate of their system.







Fig. 4.4 Subjective ratings of Fig. 4.2 plotted against their respective minimum required channel capacities.



Fig. 4.5 Curves of constant bit rate using entropy coding. Superimposed are the isopreference contours (dashed) of Fig. 4.2. The lower limit on the number of bits required to encode a quantizer output level can be obtained by calculating the entropy of the quantizer output samples [G1].

$$H(L) = \sum_{i=1}^{N} p(L_i) \log_2(\frac{1}{p(L_i)})$$
(4.8)

where $p(L_i)$ is the probability of occurrence of output level L_i . Note that correlations in adjacent output samples have been ignored, since such correlations are minimized by quantizing the difference signals. Contours of constant bit rate as defined by the entropy of the source are indicated by the solid curves of Figure 4.5. Superimposed are the isopreference contours from Figure 4.2.

As with the curves of constant bit rate without entropy coding, it is seen that for a given preference level the minimum bit rate occurs when sampling at or near the Nyquist rate minimum required channel capacities have been plotted in Figure 4.5 as previously described. Comparison of this line and the line without entropy coding indicates that for SNR_{subj} values greater than 4 dB entropy coding will result in a saving of bit rate and that the magnitude of this saving increases with increasing SNR_{subj}.

Entropy coding presents the designer with several options. For example, a system requiring an SNR of approximately 16 dB with a bit rate of 16 k bps results from using either a three- or five-level quantizer. This example also reveals that for a given bit rate, entropy coding produces better speech quality when implemented with quantizers having an odd number of levels than with neighboring even level quantizers.

To give a clearer picture of the amount of bit rate reduction possible through the use of entropy coding techniques for this system, a

¹The frequency with which quantizer levels occurred during the processing of the master samples to form the data base (Section 3.4) were used to calculate the entropy of the source.

matrix of reduction coefficients, r, has been calculated and is shown in Figure 4.6 where

$$r = \frac{\text{bit rate (entropy coding)}}{\text{bit rate}}$$
(4.9)

DM has the special property of always being entropy coded and therefore posses a coefficient of unity. Deleting the unique case of DM, dividing the remaining matrix into two sections and averaging over each section

yields:
$$E[r] = .7$$
 for $f_s/f_n > 2.0$
(4.10)
 $E[r] = .8$ for $f_s/f_n \le 2.0$

In other words the bit rate, on the average, can be reduced from between 20% to 30% by employing coding based on the entropy of the quantizer output levels. Melsa and Cohn have recorded an entropy of 1.37 bits using Nyquist rate sampling and a five level quantizer with an adaptive predictor. The result was a reduction coefficient of r = .62 compared to r = .78 without the adaptive predictor. It seems that a further savings in bit rate can be accomplished at the expense of the complexity involved in adaptive prediction.

A general trend indicated by the reduction coefficient matrix is an increase of magnitude of r with increasing f_s/f_n .

4.4.3 SNR Comparisons

The ultimate measure of performance of a speech digitization scheme is the subjective quality as perceived by a human listener. Other measures of performance can only be used to indicate subjective quality. One of the most common of these is the signal to noise power ratio given by

3.5r 1.0 .62 .68 .60 .71 .72 1.**0** .64 .70 .76 .73 .63 3.0 1.0 .66 .71 .77 .64 .75 E[r] = .7.77 1.0 .66 .72 .65 .75 2.5 1.0 .68 .74 .65 .76 .78 <u>fs</u> fn .<u>6</u>7 1.O .66 .77 .74 .78 2.0 .7°0 1.0 .76 .68 .79 .79 1.0 .78 .80 .81 .7**3** .71 1.5 .76 1.0 .80 .83 .83 .73 E(r) = .8.77 1.0 .89 .81 .91 .90 1.0 1.0 .80 .87 .78 .88 .86 П 2 5 3 4 6 NUMBER OF QUANTIZATION LEVELS 1 Fig. 4.6 Matrix of bit rate reduction coefficients r for entropy coding.

SNR = 10
$$\log_{10} \frac{E[s^2]}{E[(s - \hat{s})^2]}$$
 (4.10)

where s represents the original input signal and \$ is the output or reconstructed signal. Although a useful guide to measuring relative performance between signals containing varying amounts of a characteristic noise type, SNR does not necessarily reflect a system's subjective performance when different noise types are present.

An experiment has been conducted to clearly indicate this inconsistancy. Table 4.1 presents the SNR values computed for two sets of data labelled A and B. The master samples were processed to present a diagonal crossection of the plane defined by the two parameters L and f_g/f_m .

		SNR	SNR	
L	fs/fn	A	B	Table 4.1 A comparison of
		ac = 0	dc = 20	SNR values as derived from
2	1.0	4.91	5.14	Shit vardeb ub derryed from
•				samples processed with and
3	1.5	12.66	12.71	without a do offset
4	2.0	15.95	15.96	without a de offset.
-				
5	2.5	17.91	17.94	
6	3.0	21.17	21.19	
7	3.5	23.32	23.40	

To produce the results under section B each signal was preconditioned by adding a dc offset of 20 units¹ before processing. Those signals in group A were not altered before processing by the DPCM - AQB system. The SNR values would indicate that no difference existed between the two sets of processed data². The dc offset, however, produced relatively high energy

¹ These units are defined by the 12 bit quantization process described in Chapter 3.

²The recording process removes any dc offset present in the signal so that all signals are cleared of dc offsets before subjective evaluation.

limit cycles [C4] in the signals of set B resulting in an audible and often disturbing ringing sound. Results of comparisons included in the subjective tests previously described revealed a strong preference for those signals not including the dc bias. These findings confirm that SNR does not always reflect a signals's subjective quality, particularly when more than one form of noise is present.

SNR values have been determined for the parameters under study to allow a more detailed comparison with subjective preference. Curves of constant SNR are drawn in Figure 4.7. Superimposed are the isopreference contours of Figure 4.2. Certain similarities and differences are evident. First, both sets of contours have the same general shape, indicating some degree of correlation. Second, the largest proportion of increase with respect to sampling rate both in preference and SNR levels, occurs in the bottom portion of the plane. In fact Figure 4.8 suggests that 75% of observed improvement in SNR occurs by the time the sampling rate has doubled the Nyquist rate. Thirdly, both sets of contours show a positive change in quality with increasing sampling rate, f. Here the similarity ends. The flatter isopreference curves indicate that more subjective gain is possible by increasing f than the SNR curves would indicate. The slope of the SNR curves steepens quickly as soon as f_s/f_n is increased beyond 2, while the slope of the preference curves change more gradually. The difference may be explained by the ears sensitivity to the type of noise being eliminated at this level.

The question left unanswered is "When can SNR be used and how effective is it as a measure of subjective quality?" SNR is always a good measure of signal reproduction. When the signals are produced for human



Curves of constant signal to noise ratio. Fig. 4.7 Superimposed are the isopreference con-tours (dashed) of Fig. 4.2.



resulting from Jayant's one-bit memory delta modulator.

listening it would appear that some distortions are more disturbing than others, so that SNR could only be used as a relative measure, and then only when it is judged that the dominant noise types perturbing the signals being compared are of the same general type.

4.5 COMPARISONS OF THE RESULTS WITH PREVIOUS WORK

The purpose of previous sections has been to study a DPCM - AQB system utilizing the Melsa and Cohn adaptive quantizer. It would be useful at this point to make comparisons between the present work and that of other researchers.

Isopreference curves from Chan and Donaldson [C2] indicate that for a DPCM system utilizing a 2 bit quantizer, an SNR_{subj} rating near 2 dB can be expected while for a 3-bit quantizer, a value of about 8 dB can be reached. These values were taken when sampling at the Nyquist rate. Figure 4.2 reveals lower limits of 4 and 18 dB for the two cases sited, therby indicating a definite improvement in subjective quality between DPCM coding with and without adaptive quantization. It should be pointed out that their reference signal "Joe ... lawn" was effectively bandlimited to 4 kHz and sampled at 8 kHz while our tests utilized the same signal bandlimited to 3.2 kHz and sampled at 6.4 kHz.

The lower bandlimited reference signal was used in our work to give it the same characteristics as the signals under test. This was done to simplify the task of signal comparison. An interesting observation from Figure 4.2 is that the SNR values associated with test signals sampled close to 6.4 kHz compares very closely to the SNR_{subj} values of their isopreferent feference signals. This would tend to confirm that signal degeneration by Schroeder's technique (Chapter 3) represents quite well the noise introduced by DPCM coding. Furthermore the consistency of SNR_{subj} values in the oversampled region, where SNR is no longer a good indicator of subjective quality, supports the assumption that Schroeder's reference signals present a valid means of comparing the subjective quality of different processing systems which introduce signal dependent noise.

Another study, by Goldstein and Lui [G2], has investigated the operating characteristics of a DPCM - AQB system using an adaptive quantization. scheme similar to the one described by Cummiskey et al [C7]. Goldstein and Lui's system operating on a flat band-limited Gaussian signal displayed a linear relationship between SNR and log (f_s/f_n) . For R-C shaped Gaussian signals their mathematically derived equations again predicted a linear relationship while simulation results suggest a slight leveling off at high SNR values.

Jayant's one-bit memory DM [J1] operates using an adaptation algorithm similar to the one used by Goldstein and Lui. Operation of Jayant's DM when applied to voice signals supports Goldstein's results and indicates that the general characteristics displayed by Goldstein and Lui may hold for speech samples. The SNR vs. log (f_s/f_n) plot for Jayant's DM - AQB over the operating range considered in this thesis, has been determined and is also presented in Figure 4.8. Although better performance can be expected at low frequencies, the Cohn and Melsa curve quickly flattens while Jayant's curve continues to demonstrate a linear relationship. We note that the difference in behaviour between our results and those of Jayant [J1] and Goldstein and Lui [G2] are probably due to differences in the quantizer adaptation algorithms.

Although performing well both at and just above the Nyquist rate,

Cohn and Melsa's DPCM - AQB system seems not to allow for optimum performance at highly oversampled rates. It is suggested that speech studies on adaptive quantization schemes such as the one proposed by Cummiskey et al. may result in much flatter isopreference curves thus resulting in improved performance and interesting design tradeoffs between the quantizer structure and sampling frequencies.

V CONCLUSION

5.1 SUMMARY

In this thesis the subjective quality of a DPCM system featuring a quantizer adaptation algorithm proposed by Cohn and Melsa [C5] has been investigated. The system operated on high quality speech samples and was subject to controlled variations in the sampling frequency relative to the Nyquist rate, and in the number of quantization levels. Because low bit rate applications were of particular interest, the sampling rate and quantizer structure were bounded at 3.5 times the Nyquist rate and at seven levels, respectively. Simulation of the system was carried out on a digital computer while all subjective test results were evaluated according to the isopreference method.

The subjective tests resulted in a plot of isopreference contours being drawn on a plane whose ordinate was defined by the ratio of sampling frequency to the Nyquist frequency, and whose abcissa was defined by the number of quantization levels. The curves revealed that increases in subjective quality resulting from increases in sampling frequency relative to the Nyquist frequency became minimal after a ratio of two had been reached. This result indicated that gains made by the resulting increase in sample correlation were being masked by other elements such as quantizer noise.

Plots of bit rate resulting from various combinations of sampling frequency and quantizer structures were obtained, both with and without entropy coding. It was determined that the minimum required channel capacities for a given subjective preference level occurred when sampling at the Nyquist rate. Implementation tradeoffs between the number of quantization levels and the sampling frequency became apparent from the isopreference

contours. However, no real design options resulted from coding schemes which assigned an equal number of bits to each quantizer level under constant bitrate constraints. On the other hand, use of entropy coding showed that at suboptimal bit rates a given subjective preference level could be attained using different combinations of sampling frequency and quantizer structures. Also, for a given bit rate, it was found that odd-level quantizers out performed even level quantizers when entropy coding was employed.

Comparisons with calculated SNR values indicated a general correlation between curves of constant subjective quality and curves of constant SNR. The SNR curves did not however, accurately describe the subjective test results and the conclusion was drawn that SNR could not generally be used as a precise measure of subjective quality. A special experiment was conducted to show that in particular, SNR could not be used to compare speech samples containing large proportions of different types of noise.

The subjective test results were also compared with results of others' work. On the basis of Schroeder's [S2] speech quality standard signals it was determined that DPCM using an adaptive quantizer out performed a fixed quantizer DPCM system. This improvement can easily be seen in Table 5.1 where SNR_{subj} values for PCM and DPCM have been taken from Chan and Donaldson [C2] and Yan and Donaldson [Y1]. The performance of another adaptive quantization scheme studied by Goldstein and Lui [G2] was also compared to the results of this work. It was suggested that the adaptation scheme of this thesis, although performing very well at low ratios of sampling rate to Nyquist rate, did not perform optimally at higher ratio values.

		DPCM – AQB				
∦ of	CHAN		YANN	YAN _F		
Quantization Bits	$W_{v} = 3.2$ f = 6.4	$W = 4$ $f_s = 8$	W = 4 f = 8 s	$W = 4$ $f_s = 8$	W = 3.2 f = 6.4	W = 3.2 f = 8
2	2	4	(1.68)	(5.07)	4	6.5
3	8	10 0	12.5	11	18	20
4	13	16	18	17		
5	20	24	24	25		
6	25		(25.05)	(27.3)	-	

Table 5.1 Comparison of approximate SNR subj values for nonadaptive previous-sample feedback DPCM, and DPCM - AQB. Results for DPCM are from Chan and Donaldson [C2] and Yan and Donaldson [Y1]. (Values in brackets represent the limiting values of the respective graphs.

- N natural binary coding.
- F folded binary coding.
- W bandwidth.)

5.2 SUGGESTIONS FOR FURTHER WORK

The results of this study further confirm the economy-efficiency compromise obtainable using a differential encoder with an adaptive quantizer. Further testing and development is necessary before such systems are proven acceptable for use in modern communications systems.

Effects of digital channel transmission errors must be considered both from the point of view of overall subjective effects as well as system error propagation. It remains to optimize the quantizer adaptation strategy for various values of f_s/f_n and L. This need for optimization suggests that other adaptive algorithm, such as the one presented by Goldstein and Lui, should be subjectively tested on speech samples to fully understand their operating characteristics.

REFERENCES

- B1 R.W. Benson and I.J. Hirsh, "Some variables in audio spectrometry", J. Acoust. Soc. Am., Vol. 75, pp. 499-505, May 1953.
- Cl J.C. Candy, "Limiting the Propagation of Errors in One-Bit Differential Codec's" BSTJ, Vol. 53, pp. 1667-1677, October 1974.
- C2 D. Chan and R.W. Donaldson, "Subjective Evaluation of Pre and Post Filtering in PAM, PCM and DPCM Voice Communication Systems", <u>IEEE</u> <u>Trans. on Communications Technology</u>, Vol. 19, pp. 601-612, October 1971.
- C3 D. Chan, "Optimal pre- and postfiltering of noisy sampled signals particular applications to PAM, PCM and DPCM communication systems", Ph.D. Thesis, Department of Electrical Engineering, University of British Columbia, Vancouver, B.C., Canada, August 1970.
- C4 T.A.C.M. Classen, W.F.G. Mecklenbraüker, and J.B.H. Peek, "Effects of quantization an overflow in recursive digital filters", <u>IEEE</u> Trans. on ASSP, Vol. 24, pp. 517-529, December 1976.
- C5 D.L. Cohn and J.L. Melsa "The residual encoder an improved ADPCM system for speech digitization" in Conf. Rec. IEEE Int. Conf. Commun., San Francisco, Calif., pp. (30-26) - (30-30), 1975.
- C6 R.E. Crochière, S.A. Webster, and J.L. Flanagan, "Digital coding of speech in subbands", BSTJ, Vol. 55, pp. 1069-1085, October 1976.
- C7 P.C. Cummiskey, N.S. Jayant, and J.L. Flanagan, "Adaptive quantization in differential PCM coding of speech", <u>BSTJ</u>, Vol. 52, pp. 1105-1118, September 1973.
- C8 P. Cummiskey, "Single integration adaptive delta modulation", <u>BSTJ</u>, Vol. 54, pp. 1463-1474, October 1975.
- Dl R.W. Donaldson, and D. Chan, "Analysis and subjective evaluation of differential pulse code modulation voice communication systems" IEEE Trans. on Communication Technology, Vol. 17, pp. 10-19, Feb 1969.
- D2 R.W. Donaldson and R.J. Douville, "Analysis, subjective evaluation, optimization and comparison of the performance capabilities of PCM, DPCM, ΔM, and PM voice communication systems." IEEE Trans. on Communication Technology, Vol. 17, pp. 421-431, August 1969.
- F1 J.L. Flanagan, <u>Speech Analysis</u>, <u>Synthesis</u>, and <u>Perception</u>, <u>Springer-</u> Verlag, Berlin·Heidelberg·New York, 1972.
- G1 R.G. Gahlager, Information Theory and Reliable Communication, Wiley, 1968.
- G2 L.H. Goldstein and B. Lui, "Quantization Noise in ADPCM Systems", IEEE Trans. on Communications, Vol. 25, pp. 227-237, February 1977.

- G3 D.J. Goodman and L.J. Greenstein, "Quantizing noise of ΔM/PCM encoders", BSTJ, Vol. 52, pp. 183-204, February 1973.
- G4 D.J. Goodman, B.J. McDermatt and L.H. Nakatani, "Subjective evaluation of PCM coded speech", BSTJ, Vol. 55, pp. 1087-1110, Octo1976.
- G5 L.J. Greenstein, "Slope overload noise in linear delta modulation with gaussian inputs", BSTJ, Vol. 52, pp. 387-422, March 1973.
- G6 C.B. Grether and R.W. Stroh, "Subjective evaluation of differential pulse code modulation using the speech goodness rating scale", <u>IEEE</u> Transaction on Audio, Vol. 21, pp. 179-184, June 1973.
- G7 J.P. Guilford, Psychometric Methods, McGraw-Hill, 1936.
- H1 M.M.L. Hecker and C.E. Williams, "Choice of references conditions for speech preference tests", <u>J. Accoust. Soc. Am.</u>, Vol. 39, pp. 946-952, May 1966.
- Il "IEEE Recommended practice for speech quality measurements", <u>IEEE</u> Trans. on Audio Eng., Vol. 17, pp. 227-246, September 1969.
- J1 N.S. Jayant, "Adaptive Delta Modulation with a one-bit memory", <u>BSTJ</u>, Vol. 49, pp. 321-342, March 1970.
- J2 N.S. Jayant, "Adaptive Quantization with a one-word memory", <u>BSTJ</u>, Vol. 52, pp. 1119-1144, September 1973.
- J3 N.S. Jayant, "Step size transmitting differential coders for mobile telephony", BSTJ, Vol. 54, pp. 1557-1582, November, 1975.
- J4. N.S. Jayant, "On the correlation between bit sequences in consecutive delta modulation of a speech signal, <u>BSTJ</u>, Vol. 53, pp. 937-949, May -June 1974.
- K1 K.D. Kryter, "Methods for the calculation and use of articulation index", J. Acoust. Soc. Am., Vol. 34, pp. 1689-1697, 1962.
- L1 B.W. Lindgren and G.W. Elrath, <u>Introduction to Probability Statistics</u>, The MacMillan Co., 1969.
- M1 R.A. McDonald, "Signal to Noise and Idle channel performance of differential pulse code maodulation systems - particular applications to voice signals", <u>BSTJ</u>, Vol. 54, pp. 1123-1152, September 1966.
- M2 P. Milner, "Advantages of experienced listeners in intelligibility testing", <u>IEEEETrans. on Audio Electroacoustics</u>, Vol. 21, pp. 161-165, June 1973.
- M3 O. Mitra, "Mathematical analysis of an adaptive quantizer", <u>BSTJ</u>, Vol. 53, pp. 867-898, May-June 1974.
- M4 W.A. Munson and J.E. Karlin, "Isopreference methods for measuring speech quality", J. Acoust. Soc. Am., Vol. 34, pp. 762-774, June 1962.

- N1 P. Noll, "A comparative study of various quantizing schemes for speech encoding", BSTJ, Vol. 54, pp. 1597-1614, November 1975.
- N2 P. Noll, "Effects of channel errors on the signal-to-noise performance of speech encoding systems", <u>BSTJ</u>, Vol. 54, pp. 1615-1636, Nov 1975.
- 01 J.B. O'Neal, Jr. "Delta modulation quantizing noise analytical and computer simulation results for gaussian and television input signals", BSTJ, Vol. 45, pp. 117-142, January 1966.
- 02 J.B. O'Neal, Jr. "Predictive quantizing systems (differential PCM) for the transmission of television signals", <u>BSTJ</u>, Vol. 45, pp. 689-722, May-June 1966.
- Q1 S.U.H. Qureshi and G.D. Forney, Jr. "A 9.6/16 kb/s speech digitizer" in Conf. Rec., IEEE Int. Conf. Commun, San Francisco, Calif, pp. (30-31) - (30-36), 1975.
- R1 E.H. Rothauser, G.E. Urbanek and W.D. Pachl, "Isopreference method for speech evaluation", <u>J. Acoust. Soc. Am.</u>, Vol. 44, pp. 408-418, February 1968.
- S1 H.R. Schindler, "Linear, nonlinear, and adaptive delta modulation" IEEE Trans. on Communications, Vol. 22, pp. 1807-1822, November 1974.
- S2 M.R. Schroeder, "Reference signals for speech quality studies", <u>J.</u> Acoust. Soc. Am., Vol. 44, pp. 1735-1736, December 1968.
- S3 J. Swaffield and D.L. Richards, "Rating of speech links and performance of telephone networks", Proc. IEEE, Vol. 106-B, pp. 65-76, March 1959.
- T1 S.Tazaki, H. Osawa, Y. Shigematsu, "A useful analytical method for Discrete adaptive delta modulation", <u>IEEE Trans. on Communications</u>, Vol. 25, pp. 193-200, February 1977.
- T2 W.H. Tedford, Jr. and T.V. Frazier, "Further study of the isopreference method of circuit evaluation", <u>J. Acoust. Soc. Am.</u>, Vol. 39, pp. 645-649, April 1966.
- W1 Woodward and Schlosberg, Experimental Psychology, Vol. 1, Holt-Rinehart Winston (1938) 1972.
- Y1 J. Yan and R.W. Donaldson, "Subjective effects of channel transmission errors on PCM and DPCM voice communications systems", <u>IEEE Trans. on</u> Communications, Vol. 20, pp. 281-290, June 1972.