Mansoura Engineering Journal (MEJ),Vol.18.No. 4, Dec. 1993. E. 18

SPEECH CODING IN THE SEQUENCY DOMAIN

Fayez W. Zaki

Faculty of Engineering, Mansoura University, Egypt

تشفير الإشارات الصوتية بنظام التحويلات المتعامدة

خلاصة :

إن إرسال الإشارات الصوتية رقميا يهيىء الفرصة للحصول على سرية المعلومات بجودة مناسبة ولكن ذلك يتطلب إتساع فى حيز التردد لقناة المعلومات . و لذلك فإن إمكانية تطبيق نظام الإرسال الرقمى فى منظومات المستقبل تتطلب وجود طرق لخفض مدى النطاق الترددى لقناة المعلومات و قد وجد أن التحويلات التعامدية يمكن إستخدامها لخفض معدل المعلومات المرسلة للتعبير عن الإشارات الصوتية وذلك لأن قيمة متوسط المربعات لهذه التحويلات أقل بكثير من متوسط المربعات للإشارات قبل تحويلها أو بمعنى أخر فإن هذه التحويلات تتخلص من كمية كبيرة من الحشو المتواجد بالإشارات الصوتية و الذى يتطلب إرسالة إتساع فى النطاق الترددى .

فى هذا البحث تم إقتراح منظومة جديدة لإرسال الإشارات الصوتية رقميا بطريقة التحويلات المتعامدة مثل تحويلات والش و أيضا تحويلات هار وبعمل محاكاة على الحاسب الألى للمنظومة المقترحة أمكن الحصول على أصوات ذات جودة عالية عند معدل معلومات قدرة ٢٤ ك نبضة / ث .

ABSTRACT

Digital transmission of speech presents the opportunity to provide a more secure and bandwidth efficient service with a consistent quality. The feasibility and cost efficiency of future systems depends largely on the availability of low bit rate vocoders. Orthogonal transformations offer a means for reduction of the bit rate necessary for the transmission of speech signals. The ability to reduce the bit rate resides in the fact that the variance of the orthogonal function coefficients is far less than the variance of the speech itself.

Discrete Walsh and Haar transforms are examined for their capacity to reduce the bit rate necessary to transmit speech signals. The performance of these transforms when applied to bit rate reduction has been measured in a mean square error sense (signal-to-noise ratio). High quality speech reproduction is obtained at bit rate of 24 Kbps.

I-INTRODUCTION

Historically, speech coders have been divided into two broad categories, namely, waveform coders [1 - 7] and vocoders [8 - 13]. Waveform coders generally attempt to produce the original speech waveform according to some fidelity criteria whereas vocoders model the input speech according to a speech production and/or perception model. The encoder computes optimum model parameters (for a speech segment) which are then coded for transmission. The speech information is thus said to be "compressed" into the parameters. The parameters received at the decoder describe a model which is used to synthesise a perceptually close replica of the speech segment. If any of these parameters are corrupted, the synthesised segment will be perceptually different from the original. This suggests that model based low bit rate speech vocoders are inherently more sensitive to channel errors than waveform type coders. Generally, waveform coders have been more successful at producing good quality, robust speech, whereas, vocoders are more fragile and are more dependent on the validity of the speech production model. Vocoders, however, are capable of operating at much lower bit rates (2 - 5 kbps).

In order to reduce the bit rate of waveform coders, efforts have focused on taking greater advantage of speech production and speech perception models without making the algorithm totally dependent on these models as in vocoders [5]. A general category of coder algorithms which have been relatively successful in achieving this goal is the class of frequency

this way different frequency bands can be preferentially encoded according to perceptual criteria for each band, and quantising noise can be contained within bands and prevented from creating harmonic distortions outside of the band.

Two basic types of frequency domain coders are considered in literatures, namely, sub-band coders [14 - 18] and transform coders [19 - 24]. In the first case, the speech spectrum is partitioned into a set of, typically 4 - 8 sub-bands by means of a filter bank analysis [12]. In the second case, a block by block transform analysis is used to decompose the signal into frequency components. Both techniques attempt to perform some type of short time spectral analysis of the input signal, although the spectral resolution in the two methods is different. After encoding and decoding the frequency components are used to resynthesise a replica of the input waveform by either filter bank summation or inverse transform means.

In transform coding, the input signal is divided into time segments which are windowed by an analysis window. Each windowed time segment is transformed to the frequency domain by means of an M point discrete transform (e.g. Fourier, Cosine, Walsh, ... etc) to produce the sampled short-time spectrum. Synthesis is achieved by inverse discrete transforming each sampled short-time spectrum to obtain its time domain representation.

Walsh functions have been used in the processing of speech signals in several different ways: as a method of reducing the bandwidth occupied by the transmitted signals, as a tool for speech synthesis, and as a technique for automatic speech recognition. Important contributions in bandwidth compression were made by Campanella and Robinson [19], Shum et. al. [20] and Zelinski and Noll [21]. They showed that advantage could be taken of Walsh transform to remove some of the redundancy from transmitted speech. A bit rate of 48 kbps compared with 56 kbps conventional μ-law PCM coding was reported. Further reductions were reported later by Zelinski and Noll [22], and Tribolet and Crochiere [23]. Although they claimed a bit rate between 16 kbps and 9.6 kbps, this bit rate has been reached on the expense of higher instrumentation complexity and degraded speech quality.

Fig. 1, shows a general block diagram of a transform coder system. In which the input speech is buffered into short-time blocks of data and frequency domain transformed. The transformed coefficients or frequency components, as well as side information, are then quantised, coded and transmitted to the receiver. The side information may contain the variance of the transformed coefficients, the dominant coefficient number, the step size of the adaptive quantiser ....etc. At the receiver side, the coefficients are decoded and inverse transformed into blocks. These blocks are then used to synthesise the reconstructed speech signal by a concatenation of the blocks. From Fig. 1, it can easily be seen that the system complexity increases as the data block length increases. For example, if the block length is 256 samples, then 256 different quantisers with 256 different bit assignments and 256 different step sizes are required to quantise the transformed coefficients. Moreover, the side information are also computed, quantised, coded, and multiplexed with the coefficient codes. Hence, one may say, although, the bit rate for some transform coding systems is as low as 9.6 kbps [23], the computation complexity is enormously high. Therefore, the expected hardware implementation cost for such systems would defeat their bandwidth compression target.

The aim of this paper is to introduce a very simple transform coding system for speech based on either Walsh or Haar transformation. A general block diagram for the proposed system is shown in Fig. 2. This figure involves block transformation of buffered segments of the speech waveforms. Each segment is represented by a set of transform coefficients which are quantised and coded sequentially using a single quantiser designed to optimise the signal-to-noise ratio (SNR) depending on the variance (or mean square value) of the transformed coefficients as discussed in section III. At the receiver, the quantised coefficients are inverse transformed to produce a replica of the original input segment. Successive segments when joined, represent the input speech signal.

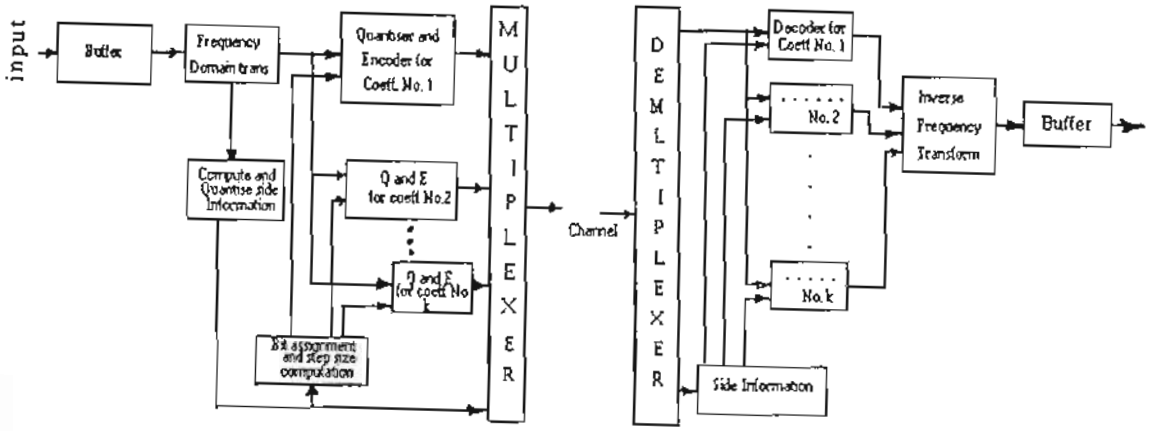
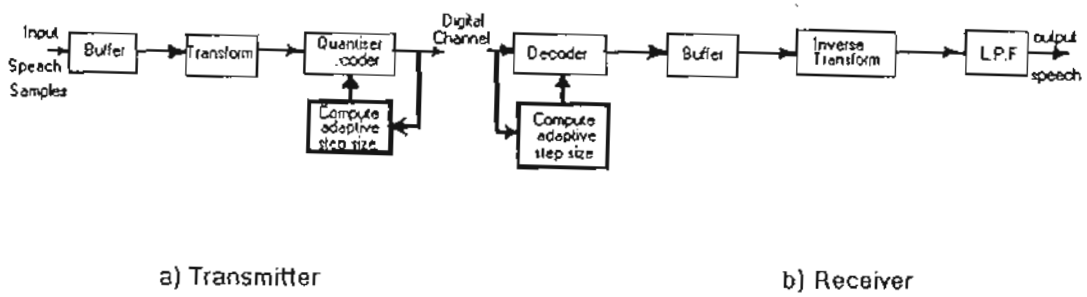


Fig. 1, Block Diagram of a General Transform Coding System.



a) Transmitter b) Receiver

Fig. 2, General Block Diagram for the Proposed Transform Coding System "The Bold Boxes are used Only for Adaptive Quantiser Case"

II-SEQUENCY DOMAIN TRANSFORMATIONS

The class of block transformatiom of interest for speech coding are time-to-frequency transformations. Since a primary goal is to attain the least audible coding noise, it is natural to control the quantisation noise by controlling its characteristics in the frequency domain. Particular time-to-frequency transforms, referred to as the Walsh and the Haar transforms, are found to be well suited for speech coding. These transforms are now introduced.

The Walsh functions form an ordered set of rectangular waveforms taking only two amplitude values +1 and -1 and are example of an orthonormal set of functions. These functions are defined over a limited interval T, known as the time base. Two arguments are required for complete definition, a time period t (usually normalised to the time base as t/T) and an ordering number n. The Walsh function is then written as WAL(n,t), and for most purposes a set of such functions is ordered in ascending value of the number of zero crossings found within the time base.

Every function f(t) which is integrable is capable of being represented by a Walsh series defined over the open interval (0,1) as:

$$x(t) = a_0 + a_1 WAL(1,t) + a_2 WAL(2,t) + \ldots \qquad (1)$$

where

$$a_k = \int_0^1 f(t)WAL(k,t)\ dt \qquad (2)$$

From this we are able to define a transform pair,

$$x(t) = \sum_{k=0}^{\infty} X(k)WAL(k,t) \qquad (3)$$

and

$$X(k) = \int_0^1 x(t)WAL(k,t)\ dt \qquad (4)$$

This definition applies to a continuous function limited in time over the interval $0 \leqslant t \leqslant 1$. For numerical use it is convenient to consider a discrete series of N terms set up by sampling the continuous functions at N equally spaced points over the open interval (0,1). In order that the properties of the continuous and discrete systems should correspond then N must equal to a power of 2, i.e., $N = 2^p$. The integration shown in Eq.(4) may then be replaced by summation, and using the trapezium rule on N sampling points x_i, the finite discrete Walsh transform pair can be written as:

$$X_n = (1/N) \sum_{i=0}^{N-1} x_i WAL(n,i) \quad ; n=0,1,2,\ldots\ N-1 \qquad (5)$$

and

$$x_i = \sum_{n=0}^{N-1} X_n WAL(n,i) \quad ; i=0,1,2,\ldots N-1 \qquad (6)$$

Since WAL(n,i) is symmetrical about the mid-point of the sequence, i=0,1,2,.... N-1 when n is even, and anti-symmetric when n is odd, then it follows that a sequence x_i will have a transform composed only of even order Walsh function coefficients if it is symmetric about its mid-point and be composed only of odd-order coefficients if the series is inversely symmetric.

The Haar functions also form a complete orthonormal function set of rectangular waveforms proposed originally by Haar [25]. The functions have several important properties, including the ability to represent a given function with few constituent terms to a high degree of accuracy. The amplitude values of these square waves do not have uniform value, as with

Walsh waveforms, but assume a limited set of values, 0, ± 1, $\pm\sqrt{2}$, ± 2, $\pm 2\sqrt{2}$, ± 4, etc. They may be expressed in a similar manner to the Walsh functions as: HAR(n,t).

If we consider the time-base to be defined as $0 \leq t \leq 1$, then we can write:

$$HAR(0,t) = 1 \quad ; \quad 0 \leq t \leq 1$$

$$HAR(1,t) = \begin{cases} 1 & ; \quad 0 \leq t < 1/2 \\ -1 & ; \quad 1/2 \leq t \leq 1 \end{cases}$$

$$HAR(2,t) = \begin{cases} \sqrt{2} & ; \quad 0 \leq t < 1/4 \\ -\sqrt{2} & ; \quad 1/4 \leq t < 1/2 \\ 0 & ; \quad 1/2 \leq t \leq 1 \end{cases} \qquad (7)$$

$$HAR(3,t) = \begin{cases} 0 & ; \quad 0 \leq t < 1/2 \\ \sqrt{2} & ; \quad 1/2 \leq t < 3/4 \\ -\sqrt{2} & ; \quad 3/4 \leq t \leq 1 \end{cases}$$

$$HAR(2^p+n,t) = \begin{cases} \sqrt{2^p} & ; \quad n/2^p \leq t < (n+1/2)/2^p \\ -\sqrt{2^p} & ; \quad (n+1/2)/2^p \leq t < (n+1)/2^p \\ 0 & ; \quad \text{elsewhere} \end{cases}$$

where $p = 1,2, \ldots$ and $n = 0,1,2,\ldots\ldots 2^p-1$

This allows a sequential numbering system analogous to that adopted by Walsh for his function series.

A given continuous function f(t) with the interval $0 \leq t \leq 1$ and repeated periodically outside this interval can be synthesised from a Haar series by

$$f(t) = \sum_{n=0}^{\infty} C_n HAR(n,t) \qquad (8)$$

where

$$C_n = \int_0^1 f(t)HAR(n,t)\, dt \qquad (9)$$

The discrete Haar transform and its inverse can be stated as

$$X_n = 1/N \sum_{i=0}^{N-1} x_i HAR(n,i/N) \quad ; \quad n=0,1,2,\ldots\ N-1 \qquad (10)$$

and

$$x_i = \sum_{n=0}^{N-1} X_n HAR(n,i/N) \quad ; \quad i=0,1,2,\ldots\ N-1 \qquad (11)$$

III-QUANTISATION

Waveform coders quantise amplitude samples by rounding off each sample value to one of a set of several discrete values. In a B-bit quantiser, the number of these discrete amplitude levels is 2^B. A fundamental result in quantisation theory is that the quantisation error power is proportional to the square of the quantising step size, and since the step size is inversely proportional to the total number of levels for a given total amplitude range, a signal-to-quantisation-error ratio (SNR) can be defined that is proportional to 2^{2B}. In logarithmic units, SNR increases linearly with B. For PCM which is basically a quantiser of sampled amplitudes, the performance characteristic is of the form:

$$\text{SNR (dB)} = 6B - \theta \qquad (12)$$

where θ is a step size dependent parameter. In the derivation of Eq.(12) it is assumed that the range of the quantiser is aligned with that of the signal amplitudes at its input. This requirement is realised if the signal amplitudes do not exceed the overload points of the quantiser with any significant probability, and if all quantiser ranges are utilized in some equitable fashion. In practice, such quantiser input alignment is realised by one of two techniques, nonuniform or adaptive quantisation.

Nonuniform quantisation is characterised by fine quantising steps (and hence, a relatively small noise variance) for the very frequently occuring low amplitudes in the sequency domaim coefficients, while much coarse quantising steps take care of the occasional large amplitudes in the sequency domain coefficients. This problem was studied by Max [26] and later by Paez and Glisson [27] who reported their results in a form of tables that will be used in part of the current study.

While time-invariant nonuniform quantiser has been used as a traditional solution to the large dynamic range problem, better results can be obtained by noting that the large dynamic range of the sequency domain coefficients is a result of nonstationary nature of the speech signal, so that a truly optimal quantisation strategy is one that is also time-variable, or adaptive to the input signal. Adaptive quantisation utilizes a quantiser characteristic (uniform or nonuniform) that shrinks or expands in time like an accordion. Although sequency domain coefficients have a large dynamic range over a long period of time, their variance vary slowly enough to aid in the design of simple adaptation algorithms to keep track of these variations.

The basic idea of adaptive quantisation is to let the step size vary so as to match the variance of the input signal. Therefore, it is necessary to obtain an estimate of the time varying amplitude properties of the sequency domain coefficients. On one hand, if the variance is estimated from the input itself, the quantiser is called feedforward adaptive quantiser. On the other hand, if the variance is estimated from the output code word, the quantiser is called feedback adaptive quantiser. Feedback adaptive quantisers have the distinct advantage that the time-varying step size need not be transmitted since it can be derived from the sequence of code words. The disadvantage of such systems is increased sensitivity to errors in the code words (channel errors).

A feedback adaptive quantiser is used in part of the current study, in which case, the variance of the transformed coefficients was assumed to be proportional to the short-time energy of the coefficients. The short-time energy may be defined as the output of a low pass filter with input $\hat{X}^2(n)$ where $\hat{X}(n)$ is the decoded value of the received code word. Therefore,

$$\sigma^2_x(n) = \sum_{m=-\infty}^{\infty} \hat{X}^2(m)\ h(n-m) \qquad (14)$$

where $\sigma^2_x(n)$ is the variance of the decoded coefficients and h(n) is the impulse response of the low pass filter. The low pass filter used has a very simple impulse response given by

$$h(n) = \begin{cases} \alpha^{n-1} & ; \quad n \geq 1 \\ 0 & ; \quad \text{otherwise} \end{cases} \qquad (15)$$

Therefore,

$$\sigma_x^2(n) = \sum_{m=-\infty}^{n-1} \hat{X}^2(m) \; \alpha^{n-m-1} \qquad (16)$$

Eq.(16) can easily be simplified to the difference equation form given by

$$\sigma_x^2(n) = \alpha \, \sigma_x^2(n-1) + \hat{X}^2(n-1) \qquad (17)$$

where $0 < \alpha < 1$ for stability. The time-varying step size for the adaptive quantiser may now be given as:

$$\Delta(n) = \Delta_0 \, \sigma(n) \qquad (18)$$

and

$$\Delta_{min} \leq \Delta(n) \leq \Delta_{max}$$

A relatively constant signal-to-noise ratio over a dynamic range of 40 dB requires $\Delta_{max} / \Delta_{min} = 100$.

IV-SIGNAL-TO-NOISE RATIO

Voice quality of the recovered speech is usually judged by subjective quality tests. Unfortunately, these tests take much time and labour, and require a large number of trained listeners. Even though intelligibility is a substantially subjective matter, it is possible to use objective tests which are useful, if not ideal, indicators of intelligibility. The most common objective measure is the SNR defined as follows: The difference between samples of a coded waveform and the original input waveform is defined as the coding error. The square of this quantity averaged over an appropriate interval is termed the coding noise. The ratio of the average value of the square of the input signal to the coding noise is defined as the signal-to-noise ratio (SNR). The quantity is often expressed in dB as $10 \log_{10}$ SNR. For the sequency transform coding system reported here, the SNR is defined as:

$$\begin{aligned} SNR &= \sigma_s^2 / \sigma_q^2 \\ &= (\sigma_s^2 / \sigma_x^2) (\sigma_x^2 / \sigma_q^2) \\ &= G_T \; (SNR)_q \end{aligned} \qquad (19)$$

where σ_s^2, σ_q^2, and σ_x^2 are the variances of input signal, quantisation noise, and transformed coefficients respectively, G_T is a gain factor (greater than 1) obtained due to orthogonal transformation process, and $(SNR)_q$ is the signal-to-noise ratio of the quantiser used. In dB units, Eq.(19) is expressed as

$$SNR \; (dB) = G_T \; (dB) + (SNR)_q \; (dB) \qquad (20)$$

In the following section it will be shown that the SNR is increased by about 15 dB due to the redundancy removed by the orthogonal Walsh and Haar transformations. This may be equivalent to an 8th. order linear predictor which requires higher computation complexity as compared to sequency domain transformation.

V-SIMULATION RESULTS AND DISCUSSION

The results presented here are based on some Arabic speech material prepared using C3MC25 microprocessor unit dedicated to speech applications. The system is built round a Texas Instruments TMS320C25 digital signal processor, running at 40MHz using stored program in a 64K x 8 EPROM, and having access to a maximum of 128K x 8 program RAM and 128K x 8 data RAM. Via an interface card installed in an 386 IBM compatible PC, we can download programs to the unit RAM, then run them. A cross Assembler is provided so that TMS320C25 code can be generated on the PC, for subsequent downloading. A monitor program is provided to simplify communication between PC and C3MC25 unit. As shown in Fig. 3, three inputs are provided ±5V , LINE (±640mV), and Microphone input (±300mV). The input is fed to the AD585 S/H which has a 3μs acquisition time, and is capable of sampling full scale signals at a frequency starts from 1KHz up to 100KHz (we used 8KHz) with 12-bit precision. Theinput to the S/H can be either via the input filter or the filter can be bypassed. The output of the S/H passes to the AD7572 A/D 12-bit converter, running from a 2MHz clock, giving a conversion time of 6.25μs. The analog input and resultant 12-bits of digital data produced can then be sampled by the TMS320C25 processor at a rate determined by the sample clock. This data can be stored and/or manipulated as desired, dependant on the program downloaded, results of the manipulated data may then be output or stored, again dependant on program. The data output by the TMS320C25 goes to the AD767 D/A converter, which has a 3μs settling time. The analog signal then appears at the ±5V output and the LINE output. As shown in Fig. 3, the signal from the Microphone, after being converted to digital, can be stored in the data RAM and then transfered to the PC where it can be recorded or manipulated. Since all our processing were conducted off line, the data was stored on the PC hard disk and the processed speech was sent from the PC to the C3MC25 unit for listening tests.

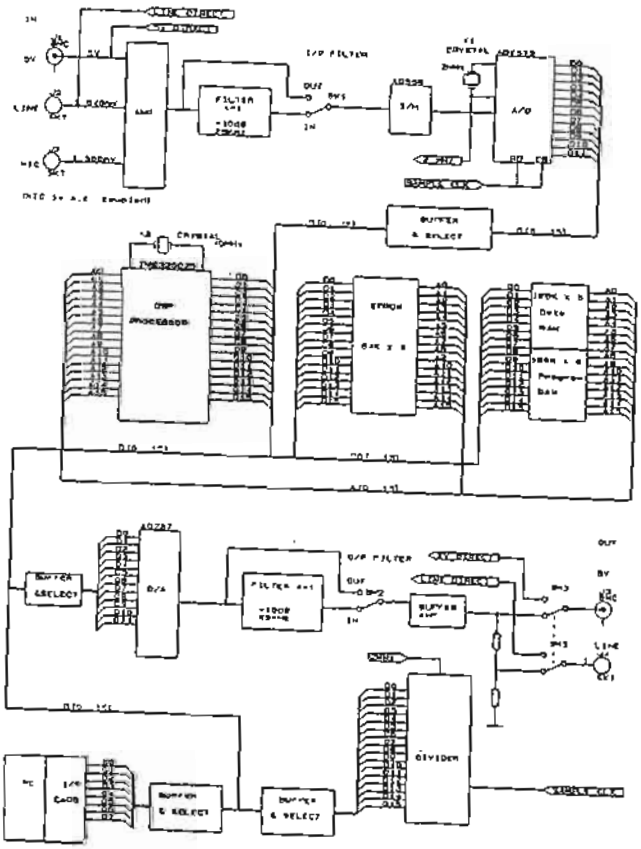


Fig. 3, Block Diagram of the C3MC25 Microprocessor Unit

Numerous computer simulation runs were carried out using Walsh and Haar transforms with linear, nonlinear, and adaptive quantisers. In all experiments, Walsh transform provided transformation gain G_T = 15 dB irrespective of the speech material, whereas the Haar transform provided an average G_T of 9.4 dB, this is shown in Table 1. Table 2, shows the signal-to-noise ratio of the quantiser $(SNR)_q$ for different types of quantisers and different number of bits/sample. It can be seen from Table 2 that adaptive 3-bit quantiser provides almost the same performance of the 4-bit nonlinear quantiser, and the 4-bit nonlinear quantiser provides almost the same performance of the 5-bit linear quantiser. In other words, the 3-bit/sample (24 Kbps data rate) adaptive quantiser provides an average SNR almost equal that. provided by the 5-bit/sample (40 Kbps data rate) linear quantiser. Moreover, listeners prefered the output speech from the system with adaptive quantiser than the output from the other systems especially during unvoiced sounds or silence periods. This is due to the fact that adaptive quantiser provides constant SNR whereas linear quantiser provides constant quantisation noise. Table 3, shows the total SNR in dB for each type of the quantisers as applied to both Walsh and Haar transforms. Also shown the total bit rate and the reduction in the bit rate from the original 96 Kbps.

Table 1, Transformation Gain G_T in dB

Walsh	Haar
15.052 Constant	9.366 average

Table 2, Average $(SNR)_q$ of the Quantiser

Quant. Type	No. of bits	Walsh	Haar
Linear	6	21.91	23.44
	5	15.62	16.92
Nonlinear	5	22.49	22.923
	4	13.613	16.78
	3	11.371	11.043
Adaptive	3	13.96	14.20

Table 3, Total SNR for Different Types of Quantisers

Quant. Type	No. of bits per Sample	Walsh	Haar	Total Bit Rate (Kbps)	Reduct. in Bit Rate (Kbps)
Linear	6	36.962	32.8	48	48
	5	30.672	26.285	40	56
Nonlinear	5	37.542	32.29	40	56
	4	28.665	26.145	32	64
	3	26.423	20.41	24	72
Adaptive	3	29.012	23.57	24	72

Fig. 4(a) shows the waveform of the original speech word /SOOD/ (male spe aker) and Figs. 4(b, c, and d) show the corresponding waveforms as obtained from Walsh system using 5-bit/sample linear, 4-bit/sample nonlinear, and 3-bit/sample adaptive quantisers respectively. It can easily be seen that the three waveforms in Figs. 4(b, c, and d) are similar during the voiced part between samples 2500 and 5300. However, the output from the 3-bit/sample adaptive quantiser is identical to the original speech during the unvoiced part between samples 2000 and 2500 also between samples 5300 and 5500. This is why listeners favoured the speech from the adaptive quantiser than that from linear and fixed nonlinear quantisers.

Fig. 5(a) shows the power spectral of the original speech for the Arabic voul / ĩ / (male speaker) as obtained by 512 point FFT, and Figs. 5(b, c, and d) show the corresponding power spectrals from the Haar system with 5-bit/sample linear, 4-bit/sample nonlinear, and 3-bit/sample adaptive quantisers respectively. It can be seen that the power spectrals in Figs.5 (b, c, and d) fit the original spectral density in Fig. 5(a) especially during the first three formants. The effect of quantisation noise is clear between the formants especially in the case of 5-bit/sample linear quantiser in which case the quantisation noise may mask the speech during low power regoins (e.g. between the formants and at the high frequencies).

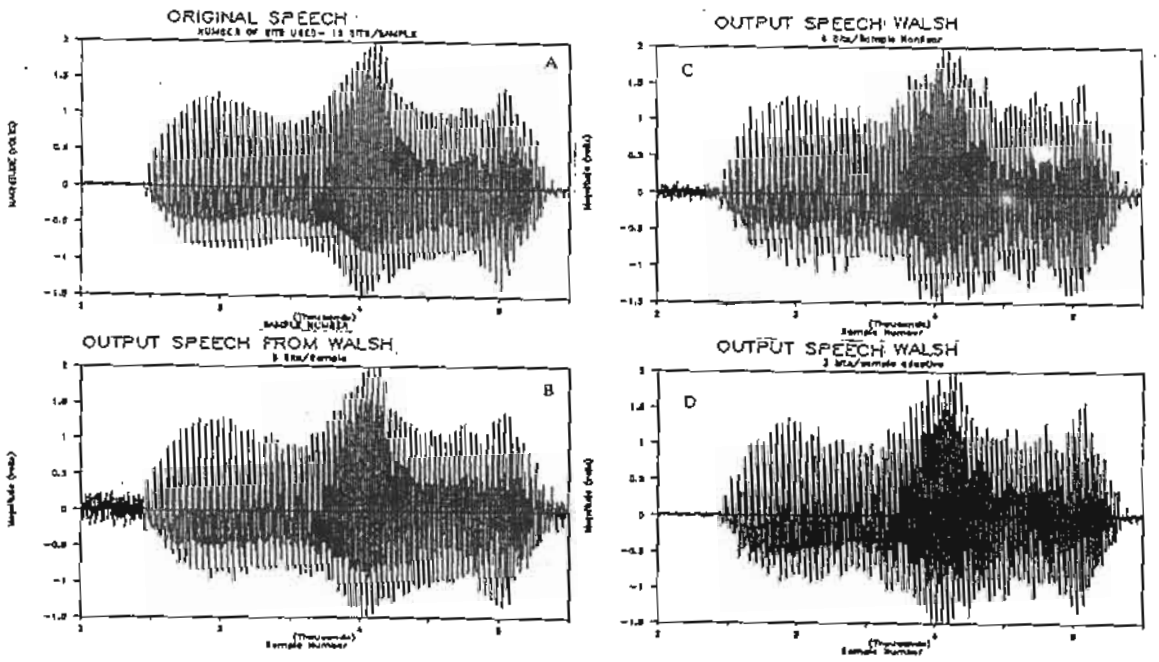


Fig. 4, a)Original Speech Waveform for /scod/ as Spoken by Male Speaker
b, c, and d) The Corresponding outputs from Walsh System

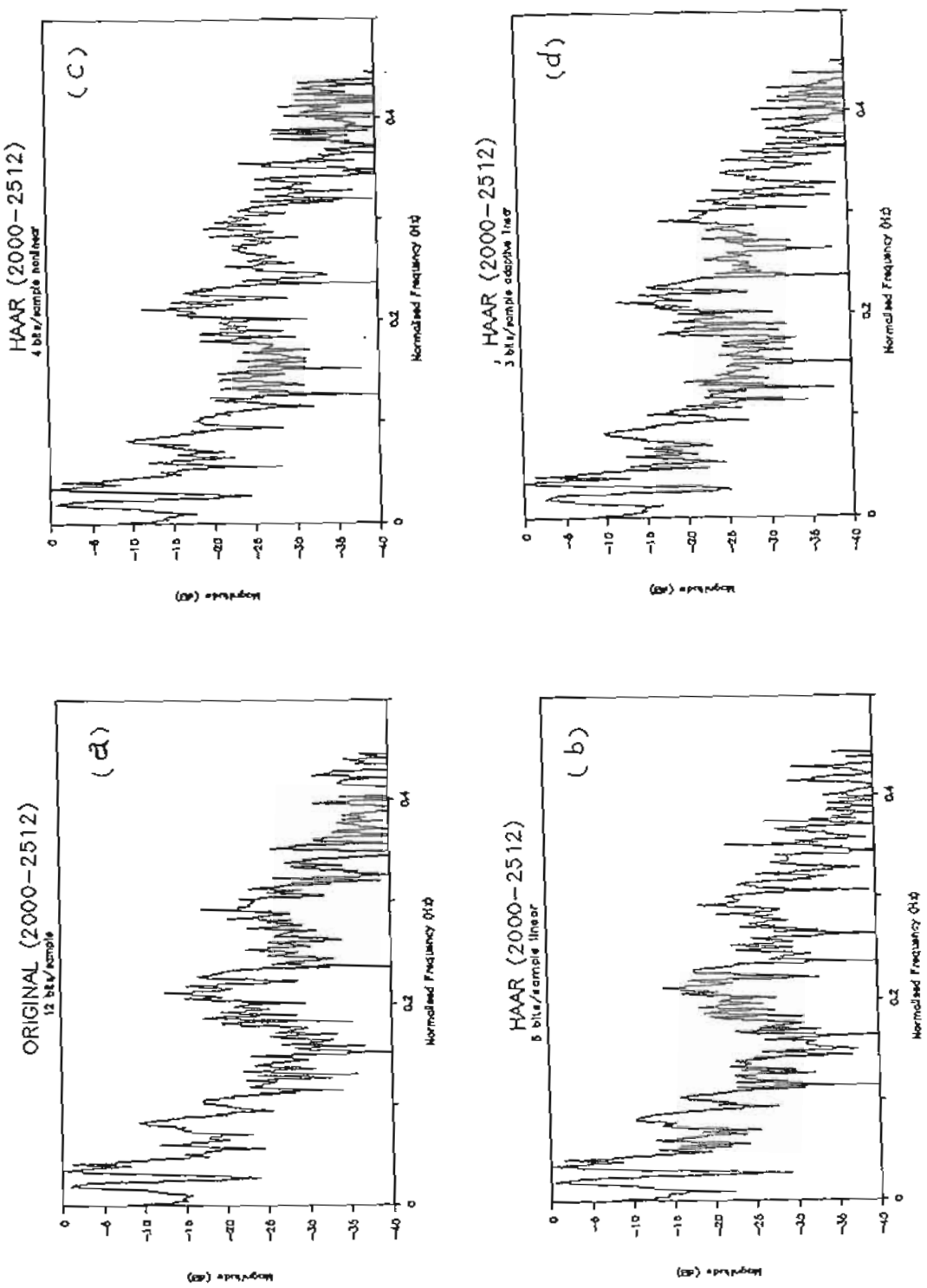


Fig. 5, a)Power Spectral Density of Original Speech Voul /ʕ/ (Male Speaker)

CONCLUSIONS

A very simple sequency transform speech coding system is introduced and discussed. Walsh and Haar transforms are used as redundancy removal transforms due to their inherent simplesity in carrying out transformation since this requires only addition and subtraction. These transformations may be realised in future applications using neural network techniques.

The introduced system provided good quality speech at data rates between 24 and 48 Kbps with a corresponding data reduction rate between 72 and 48 Kbps. The 24 Kbps system using adaptive quantiser may find its applications in commercial voice communication by mobile radio and/or satellite communication systems.

REFERENCES

1-Cattermole, K.W., "Principles of Pulse Code Modulation," Iliffe Books Ltd., London, and Elsevier Publishing Co., New York, 1969.

2-Crochiere, R.E., Rabiner, L.R., Jayant, N.S., and Tribolet, J.M., "A Study of Objective Measures for Speech Waveform Coders," Proc. Int. Zurich Seminar, on Digital Communications, Zurich, Switzerland, pp H1.1-H1.7, March 1978.

3-Jayant, N. S., "Waveform Quantisation and Coding," IEEE Press, New York, 1976.

4-Flanagan, J.L., Schroeder, M.R., Atal, B.S., Crochiere, R.E., Jayant, N.S., and Tribolet, J.M., "Speech Coding," IEEE Trans. Comm., Vol. COM-27, No.4, pp710-737, April 1979.

5-Cohn, D.L. and Melsa, J.L., "The Residual Encoder- An Improved ADPCM System for Speech Digitisation," Proc. Int. Comm. Conf., San Francisco, pp30-26 to 30-31, June 1975.

6-Cummisky, P., Jayant, N.S., and Flanagan, J.L., "Adaptive Quantisation in Differential PCM Coding of Speech," Bell Sys. Tech. J., Vol.52, pp1105-1118, September 1973.

7-Zaki, F.W., " Sequentially Adaptive Differential Pulse Code Modulation Using Adaptive LSP Filters," Mansoura Engineering Journal (MEJ), Vol. 16, No.2, pp E1-E18, Dec. 1991.

8-Dudley, H., "The Vocoder," Bell Labs. Record, Vol. 17, pp122-126, 1939.

9-Atal, B.S., and Schroeder, M.R., "Predictive Coding of Speech Signals," Bell Sys. Tech. J., Vol. 49, pp1973-1986, October 1970.

10-Estiban, D., Galand, C., Maudvit, D., and Menez, J., " 9.6/7.2 kbps Voice Excited Predictive Coder (VEPC)," Proc. 1979 IEEE Int. Conf. on Acoust., Speech, and Signal Processing, Tulsa, OK, pp307-311, April 1978.

11-Flanagan, J.L., and Golden, R.M., "Phase Vocoder," Bell Sys. Tech. J., Vol. 45, pp1493-1509, 1966.

12-Gold, B., and Rader, C.M., "The Channel Vocoder," IEEE Trans. Audio, Electroacoust., Vol. AU-15, No.4, pp148-160, December 1967.

13-Tribolet, J.M., and Crochiere, R.E., "A Vocoder-Driven Adaptation Strategy for Low Bit Rate Adaptive Transform Coding of Speech," Proc. 1978 Int. Conf. on Digital Signal Processing, Florence, Italy, pp638-642, Sept. 1978.

14-Crochiere, R.E., Webber, S.A., and Flanagan, J.L., "Digital Coding of Speech in Sub-Bands," Bell Syst. Tech. J., Vol. 55, pp1069-1085, Oct. 1976.

15-Barabell, A.J., and Crochiere, R.E., "Sub-Band Coder Design Incorporating Quadrature Filters and Pitch Prediction," in 1979 Proc. IEEE Int. Conf. Acoust., Speech, Signal Proc., Washington, DC, pp530-533, April 2-4, 1979.

16-Crochiere, R.E., "On The Design of Sub-Band Coders for Low Bit Rate Speech Communications," Bell Syst. Tech. J., Vol. 56, pp747-770, May-June 1977.

17-Crossman, A.H., and Fallside, F., " Multipulse Excited Channel Vocoder," IEEE Int. Conf. Acoust., Speech, Signal Proc., pp1926-1929, 1987.

18-Holmes, J.N., "The JSRU Channel Vocoder," IEE Proc. F, Vol. 127, pp53-60, 1980.

19-Campanella, S.J., and Robinson, G.S., "A Comparison of Orthogonal Transformation for Digital Speech Processing," IEEE Trans. Comm. Tech., Vol. COM-19, No. 6, pp1045-1049, Dec. 1971.

20-Shum, F.Y., Elliott, A.R., and Brown, W.O., "Speech Processing With Walsh Hadamard Transforms," IEEE Trans. Audio and Electroacoust., Vol. AU-21, No.3 pp172-179, June 1973.

21-Zelinski, R. and Noll, P.," Adaptive Transform Coding of Speech Signals," IEEE Trans. Acoust. Speech, Signal Proc., Vol. ASSP-25, pp299-309, Aug. 1977.

22-Zelinski, R. and Noll, P., "Approaches to Adaptive Transform Speech Coding at Low Bit Rates," IEEE Trans. Acoust. Speech, Signal Proc., Vol. ASSP-27, No. 1, pp89-95, Feb. 1979.

23-Tribolet, J.M., and Crochiere, R.E., "Frequency Domain Coding of Speech," IEEE Trans. Acoust. Speech, Signal Proc., Vol. ASSP-27, No. 5, pp512-530, Oct. 1979.

24-Haddad, R.A., and Akansu, A.N., "A New Orthogonal Transform for Signal Coding," IEEE Trans. Acoust. Speech, Signal Proc., Vol. ASSP-36, No. 9, pp1404-1411, Sept. 1988.

25-Haar, A., "Zur Theorie der Orthogonalen Funktionen Systeme," Math. Annal., Vol. 69, pp331-371, 1910.

26-Max, J., "Quantising For Minimum Distortion," IRE Trans. Inf. Theory, Vol. IT-6, pp7-12, March 1960.

27-Paez, M. D. and Glisson, T. H., "Minimum Mean-Square-Error Quantisation in Speech PCM and DPCM Systems," IEEE Trans. Comm., Vol. COM-20, pp225-230, April 1972.