10

Combined enhancement and adaptive transform coding of noisy speech

Y. Ephraim, M.Sc., D.Sc., and D. Malah, M.Sc., Ph.D.

Indexing terms: *Signal processing, Noise and interference*

Abstract: The paper deals with the problem of improving the performance of the adaptive transform coder which operates on noisy speech. We propose to estimate the short-time spectral amplitude (STSA) of the original speech and to utilise the noisy phase prior to the encoding process. The appropriate minimum mean-square error STSA estimator is derived and the system is examined in encoding speech which has been degraded by uncorrelated additive wideband spectrally flat noise. The above approach improves the quality of the encoded speech in the sense that the output noise level and the irregularities characteristic to the directly encoded noisy speech are reduced. However, the encoded speech loses some of its crispness.

1 Introduction

The adaptive transform coder (ATC) is a waveform coder which was found to be very efficient for encoding speech at rates of 7.2–16 kbit/s. At the rate of 16 kbit/s or above it gives toll quality, while at the rate of 7.2 kbit/s it results in communication quality [1–3]. The ATC quantises the speech spectral components in each analysis frame, in accordance with a dynamic bit allocation and a variable quantisation step size. Thus, the perceptually more important spectral components can be traced and better quantised. The bit assignment and the step size used for each spectral component are determined on the basis of knowledge of its variance. The variances of the spectral components are obtained from a parametric estimated spectrum of the speech signal in the analysed frame. The estimated parameters of the spectrum are encoded and transmitted as side information to the receiver.

While the ATC belongs to the class of waveform coders, which are supposed to be robust, it turns out that it is sensitive to background noise. Unlike other speech waveform coders (e.g. PCM, DPCM etc.), in which the input noise is just reflected to the reconstructed output speech, here the noise has an additional effect; it adversely influences the extraction of the parameters representing the speech short-time spectrum. This results in wrong bit allocation and quantisation step size, which cause further degradation of the reconstructed speech.

The structure of the ATC calls for a very convenient way for improving the quality of the reconstructed speech obtained under noisy environments. Specifically, since the encoding is already done in the frequency domain, we can estimate the perceptually important short-time spectral amplitude (STSA) of the speech signal, and use the noisy phase, prior to the encoding process. This STSA estimation reduces the level of the input noise which is reflected to the output of the encoder, as well as improving the estimation of the side information which depends solely on the STSA of the speech signal (Section 2).

We consider here a minimum mean-square error (MMSE) estimator of the amplitude (or magnitude) of the discrete cosine transform (DCT) of the original speech in the analysed frame, which is the transform used by the ATC. This has the perceptual significance of estimating the STSA of the original speech, since the DCT and the discrete Fourier transform (DFT) have, up to a constant, the same spectral envelope (see eqn. 5). The system developed here is examined in encoding speech which has been degraded by uncorrelated additive wideband spectrally flat noise.

The above enhancement approach is similar to that taken in Reference 4, where the amplitude of the DFT of the original speech is estimated. The emphasis in both papers is, however, different, as in Reference 4 we aimed at enhancing speech prior to its presentation to listeners, while here the enhancement is done in the encoding context. Our work is related to that of McAulay and Malpass [5], who consider improving the performance of the channel vocoder under noisy environment by using a maximum likelihood STSA estimator.

The paper is organised as follows. In Section 2 we briefly describe the ATC scheme used in this work. Then, in Section 3, we formulate the estimation problem and derive the MMSE STSA estimator. In Section 4 we describe the performance of the ATC operating on noisy speech with and without the above enhancement. In Section 5 we discuss the results obtained here, and briefly describe some experiments we have done to also reduce the quantisation noise level.

2 Adaptive transform coding

The ATC has been extensively investigated and several schemes were proposed [1–3]. In this paper we focus on the so-called 'speech-specific' ATC, which is well documented in Reference 3. Before describing that scheme, we briefly discuss the DCT and some of its properties. The DCT is used by the ATC rather than other transforms like the DFT, since, on the basis of a mean-square error (MSE) criterion, the DCT was found to be nearly optimal for speech encoding relative to the optimal Karhunen-Loeve transform [1, 6]. In addition, on an experimental basis, the DCT was found to perform better than the DFT [1]. Finally, the DCT reduces block-end effect problems [3].

The DCT of a real M-point sequence x_n is defined by

$$X_k = \sum_{n=0}^{M-1} x_n c_k \cos\,[(2n+1)\pi k/2M] \qquad k = 0, 1, \ldots, M-1 \quad (1)$$

where

$$c_k = \begin{cases} 1, & k = 0 \\ \sqrt{2}, & k = 1, 2, \ldots, M-1 \end{cases} \quad (2)$$

Paper 4171F (E8), first received 26th March and in revised form 18th July 1985

David Malah is, and Yariv Ephraim was formerly, with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel. Dr. Ephraim is now with the Information Systems Laboratory, Durand 145A, Stanford University, Stanford, CA 94305, USA

The inverse DCT is defined similarly as

$$x_n = \frac{1}{M}\sum_{k=0}^{M-1} X_k c_k \cos\,[2n + 1)\pi k/2M] \qquad n = 0, 1, \ldots, M-1 \quad (3)$$

The DCT can be efficiently calculated by using an FFT of M points [7]. However, it can also be obtained by applying a $2M$-point DFT on the sequence u_n, defined by

$$u_n \triangleq \begin{cases} x_n, & n = 0, 1, \ldots, M-1 \\ 0, & n = M, \ldots, 2M-1 \end{cases} \quad (4)$$

By so doing it can be shown that the DCT X_k and the DFT U_k *are related by*

$$X_k = c_k |U_k| \cos(\theta_k - \pi k/2\mathrm{M}) \quad k = 0, 1, \ldots, M-1 \quad (5)$$

where θ_k represents the phase of U_k. This form supplies an interesting spectral interpretation of the DCT. It shows that the DCT and the DFT have the same spectral envelopes (up to the constant c_k) [3].

The speech-specific ATC scheme is depicted in Fig. 1.

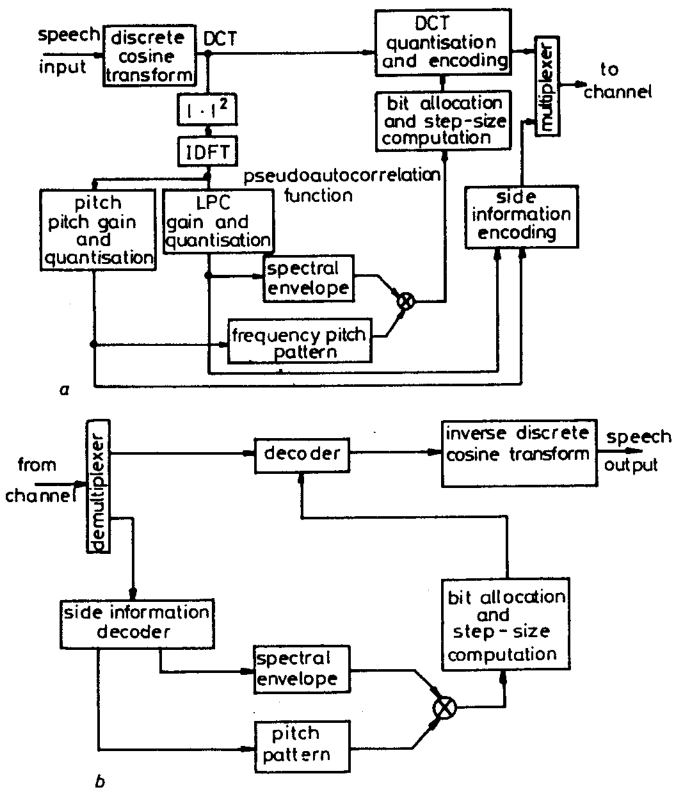

Fig. 1 *Block diagram of ATC*
a Transmitter *b* Receiver

Each frame of the input speech is first cosine-transformed. Then the parameters representing the estimated spectrum of the speech in the analysed frame are extracted and quantised. These parameters include the linear prediction coefficients (LPCs), the pitch period and the gain, which appear in the basic model of speech production. They are calculated from an estimate of the speech sample autocorrelation function, which is taken to be the inverse Fourier transform of the square of the DCT components. This estimate utilises the fact that the DCT and the DFT have the same spectral envelope (up to a constant).

The LPC and the gain are used for estimating the speech spectral envelope, while the pitch period and the so-called 'pitch-gain' [3] are used for estimating the fine details of the spectrum, called the pitch pattern. The spectral envelope and the pitch pattern are combined (via multiplication) to produce the estimated parametric spectral magnitude of the speech in the analysed frame. The estimated parametric spectrum (i.e. the square of the estimated parametric spectral magnitude) is used in the transmitter for allocating the B bits available for encoding each frame, and the estimated parametric spectral magnitude is used for normalising the DCT components prior to their optimal quantisation.

The optimal bit allocation and quantisation of the DCT components are based on the observation that they are approximately Gaussian-distributed [1, 3]. The bit assignment rule results from the solution of the following optimisation problem:

$$\min_{B_k} \frac{1}{M}\sum_{k=0}^{M-1} w_k\, \varepsilon_k \quad (6)$$

subject to

$$\sum_{k=0}^{M-1} B_k = B$$

where B_k is the number of bits assigned to the kth DCT component, ε_k is the resulting distortion in that component and w_k is a positive weighting function. B_k and ε_k are related by the rate-distortion function of a Gaussian source:

$$B_k = \frac{1}{2}\log_2 \frac{\sigma_k^2}{\varepsilon_k} \quad (7)$$

where σ_k^2 is the variance of the kth DCT component. The optimal bit allocation and the resulting distortion are given by

$$B_k = \bar{B} + \frac{1}{2}\log_2 \frac{w_k \sigma_k^2}{\left[\prod_{l=0}^{M-1} w_l \sigma_l^2\right]^{1/M}} \quad (8)$$

$$\varepsilon_k = 2^{-2\bar{B}}\left[\prod_{l=0}^{M-1} w_l \sigma_l^2\right]^{1/M} w_k^{-1} \quad k = 0, 1, \ldots, M-1 \quad (9)$$

here $\bar{B} \triangleq B/M$. σ_k^2 is obtained as the kth component of the estimated parametric spectrum. A useful weighting function was proposed in Reference 3. It is given by

$$w_k = \sigma_{sk}^{2\gamma} \quad -1 < \gamma < 0 \quad (10)$$

where σ_{s_k} is the estimated parametric spectral magnitude without the pitch pattern, i.e. it is the estimated spectral envelope. Since ε_k is proportional to w_k^{-1} (eqn. 9), this weighting function results in a quantisation noise spectrum which follows the spectrum of the speech. As a consequence, low-energy spectral components will not be totally masked by the quantisation noise.

The quantisation of the kth DCT component is achieved by first normalising it by the kth sample of the estimated parametric spectral magnitude, and then using the optimal normalised quantisation step size for a Guassian source derived by Max [8].

At the receiver the bit stream is decoded, and the parametric spectrum is reconstructed from the side information. With this available spectrum, the receiver can follow the bit allocation and the DCT normalisation performed in the transmitter.

In this paper we examined the ATC at 12 and 16 kbit/s. The specific parameters of the coder used here are those recommended in Reference 3. Specifically, the transform size M is 256, the speech sampling rate is 8 kHz, the block overlap is 16 samples using a trapezoidal window, the maximal number of quantiser bits is four, the quantiser loading parameter (to multiply Max's quantisation step size) is 1.3 for 12 kbit/s and 1.5 for 16 kbit/s, the noise-shaping parameter $\gamma = -0.125$ and the number of LPCs is nine. In this paper the side information parameters are not quantised, but the number of bits needed for their quantisation (44) was taken into account. We refrained from quantising these parameters, as our main objective here is to examine the enhancement of the encoded noisy speech. In addition, very efficient algorithms for quantising these parameters are available [9, 10].

3 MMSE spectral amplitude estimator

In this Section we derive the MMSE estimator of the amplitude (i.e. the absolute value) of each DCT component of the original speech given the noisy DCT components. We use a statistical model which utilises asymptotic statistical properties of the DCT components. Specifically, we assume that the DCT components of the speech signal, as well as of the noise process, can be modelled as statistically independent Gaussian random variables. The Gaussian assumption is motivated by the central limit theorem, as each DCT component is a weighted sum of random variables. The statistical independence assumption results from the Gaussian model and the fact that the correlation between the DCT components reduces as the analysis interval length increases. It is satisfying to note that Zelinski and Noll arrived at the same statistical model for the speech signal on an experimental basis [1, 3]. This fact has already been utilised in Section 2, where we discussed the optimal bit allocation and quantisation of the speech DCT components. For the noise process one can obtain the above model or simply assume that the noise is Gaussian.

Let X_k, D_k, and Y_k denote, respectively, the DCT component of the speech, the noise and the noisy process. Owing to the statistical independence of the spectral components, the MMSE estimation of $|X_k|$ can be carried out from only Y_k. Therefore, the MMSE estimator of $|X_k|$ is given by (see Appendix)

$$\begin{aligned}|\hat{X}_k| &= E\{|X_k| \mid Y_0, Y_1, \ldots, Y_{M-1}\} \\ &= E\{|X_k| \mid Y_k\} \\ &= \int_{-\infty}^{\infty} |x_k| p(x_k \mid Y_k)\, dx_k \\ &= \frac{\xi_k}{1+\xi_k}\left[\Phi\left(\sqrt{\frac{v_k}{2}}\right) + \sqrt{\frac{2}{\pi}}\frac{1}{v_k}\exp\left(-\frac{v_k}{2}\right)\right]|Y_k| \end{aligned} \quad (11)$$

where

$$\Phi(x) \triangleq \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\, dt \quad (12)$$

$$v_k \geqslant \frac{\xi_k}{1+\xi_k}\gamma_k \quad \xi_k \triangleq \frac{\lambda_x(k)}{\lambda_d(k)} \quad \gamma_{kG} \triangleq \frac{|Y_k|^2}{\lambda_d(k)} \quad (13)$$

and $\lambda_x(k) \triangleq E\{|X_k|^2\}$, $\lambda_d(k) \triangleq E\{|D_k|^2\}$. ξ_k and γ_k are interpreted as the *a priori* and *a posteriori* signal/noise ratios (SNRs), respectively [5]. The MMSE estimator [11] is conveniently described by a gain function defined by

$$G(\xi_k, \gamma_k) \triangleq \frac{|\hat{X}_k|}{|Y_k|} \quad (14)$$

This gain function is described in Fig. 2 by means of parametric gain curves. The behaviour of these gain curves is similar to that of the gain curves obtained in Reference 4, where the amplitude of a DFT component is estimated. The explanation given there for the shape of the gain curves holds as well for the estimator discussed here.

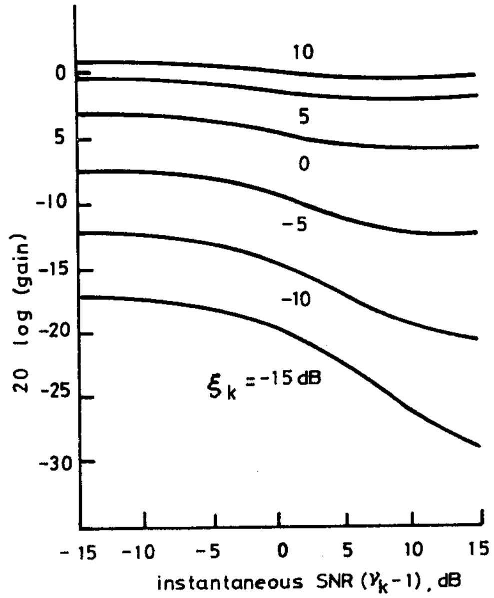

Fig. 2 *Parametric gain curves describing spectral amplitude estimator (eqn. 11) of DCT*

The estimate of the kth speech DCT component is obtained by combining the above MMSE amplitude [11] with the phase (i.e. the sign) of the kth noisy spectral component, that is

$$\begin{aligned}\hat{X}_k &\triangleq |\hat{X}_k| \frac{Y_k}{|Y_k|} \\ &= G(\xi_k, \gamma_k) Y_k \end{aligned} \quad (15)$$

To implement the above estimator, the noise variance $\lambda_d(k)$ and the *a priori* SNR ξ_k should be known. In the experiments we carried out here we examined stationary noise and estimated its variances once only from an initial non-speech interval 640 ms in duration. The *a priori* SNR is a slowly varying parameter due to speech quasistationarity. Its value in each analysis frame is estimated by the 'decision-directed' estimator proposed in Reference 4. This estimator is given by

$$\hat{\xi}_{k,n} = \alpha \frac{|\hat{X}_{k,n-1}|^2}{\lambda_d(k, n-1)} + (1 - \alpha P[\hat{\gamma}_{k,n} - 1] \quad (16)$$

where $\xi_{k,n}, |X_{k,n}|$, $\lambda_d(k, n)$ and $\gamma_{k,n}$ are, respectively, the *a priori* SNR, the speech spectral amplitude, the noise variance and the *a posteriori* SNR of the corresponding kth DCT component in the nth analysis frame. $\hat{}$ denotes 'estimated value', and $P[.]$ is an operator defined by

$$P[x] = \begin{cases} x, & x \geqslant 0 \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

whose function is to prevent $\hat{\xi}_{k,n}$ from being negative if $(\gamma_{k,n} - 1)$ is negative. α is an averaging parameter whose value is determined on the basis of listening tests. We have found, for example, that, for the voiced sentence 'we were away a year ago', the best values are $\alpha = 0.94$ for 10 dB SNR and $\alpha = 0.85$ for 5 dB SNR.

Fig. 3 shows the estimation procedure of the speech DCT components from the noisy speech. In the proposed algorithm for encoding noisy speech, the block labelled 'discrete cosine transform' in Fig. 1 is replaced by the system described in Fig. 3.

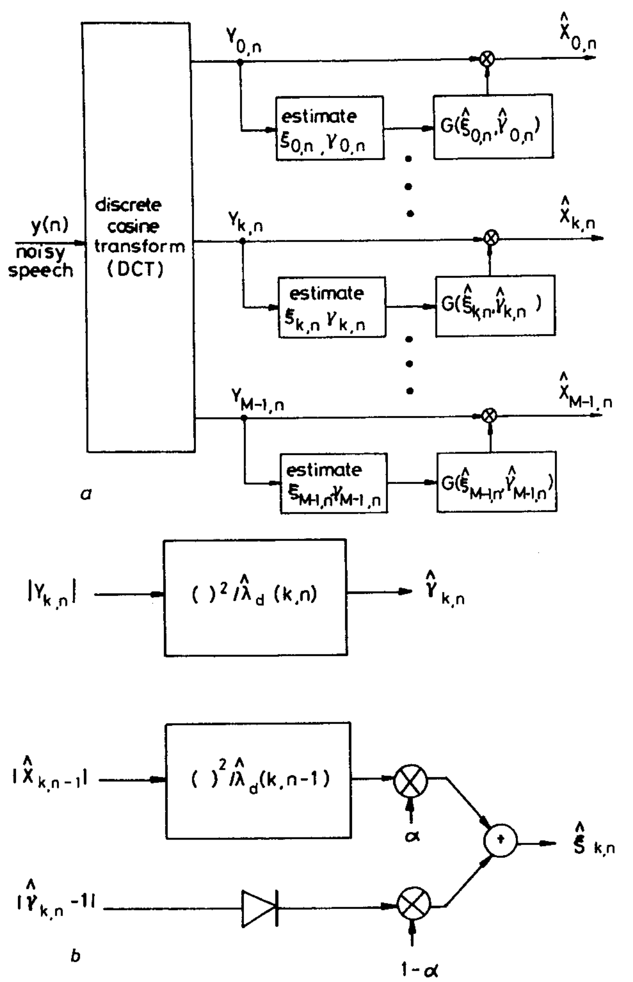

Fig. 3 *Block diagram for estimation process of DCT components*

4 Performance evaluation

The STSA estimator [11] was applied in the 'speech-specific' ATC described in Section 2, and the system was examined in encoding noisy speech at 12 and 16 kbit/s. Speech signals which were degraded by uncorrelated additive wideband spectrally flat noise, with SNR of 10 and 5 dB, were examined.

We now demonstrate our results by focusing on the analysis of the simulation outcomes obtained for the 16 kbit/s encoding rate and the 5 dB input SNR. For that case, subjective listening tests were carried out with seven listeners (the authors and five students). The following three sentences, each spoken by a male and a female, were examined:

(*a*) A lathe is a big tool.
(*b*) An icy wind raked the beach.
(*c*) Joe brought a young girl.

The other cases of either 10 dB SNR or 12 kbit/s encoding rate were examined similarly by the authors alone, who, however, arrived at similar conclusions to those described below.

The directly encoded and the enhanced-encoded noisy sentences were presented in that order through earphones to the listeners. The test was blind for the five students, as they did not know whether the material was directly encoded noisy speech or the enhanced-encoded speech. They were asked to judge the quality of the material they have heard on the basis of its clarity, crispness, the level of the residual noise and its nature. In addition, they were asked to give their preference in comparing the directly encoded noisy speech with the enhanced-encoded noisy speech. Three of the listeners were unfamiliar with the original speech material. The following description is a summary of the subjective evaluation of the listeners.

When the ATC was operated on the noisy speech but no enhancement performed, a noisy reconstructed speech results. In addition, it has some noticeable irregularities which strongly influence its quality and intelligibility. These irregularities are probably a result of using a wrong bit allocation and an incorrect quantisation step size, owing to the poor estimate of the speech spectrum from the noisy input speech. By applying the STSA estimator [11] to the above ATC system, the quality of the reconstructed speech is improved in the sense that the noise level and the above-mentioned irregularities in the encoded noisy speech were reduced. However, the encoded speech loses some of its crispness. Nevertheless, all the listeners preferred the enhanced-encoded noisy speech rather than the directly encoded noisy speech. The three listeners who were unfamiliar with the original material also indicated that the enhanced-encoded noisy speech is better understood than the directly encoded noisy speech; however, even then the material was not fully understood. Two listeners pointed out that the enhancement is more effective for the sentences spoken by the male speaker.

Fig. 4 demonstrates by means of waveform plots the above results. Table 1 gives the SNR and the segmental SNR values (in the time and frequency domains) for the various cases examined here. The segmental SNR measured in the log-spectral domain is also provided.

Table 1: Measured SNR for encoded noisy speech at 16 kbit/s

	Time domain SNR, dB		Frequency domain SNR, dB	
	SNR	Seg. SNR	Seg. SNR	Log. Seg. SNR
Encoded noiseless Speech	11.2	11.0	26.3	23.6
Encoded noisy speech	7.3	5.0	15.5	13.1
Enhanced-encoded noisy speech	8.6	6.8	18.5	16.3
Noisy speech (no coding)	5.0	3.3	12.8	9.3

An interesting observation realised in this research is that the encoding itself (without enhancement) has some filtering effect on the noisy input speech. This effect is demonstrated in Table 1 by the higher SNR of the directly encoded noisy speech than that of the input noisy speech. This seemingly odd phenomenon can be explained on the basis of the following three facts: (i) the bit allocation for the noisy spectral components is proportional to their power (eqn. 8), (ii) we use here wideband spectrally flat noise, which means that each of its spectral components has approximately the same power, and (iii) for the range of encoding rates examined here, some of the noisy spectral components are not encoded as zero bits are assigned to them. Now, owing to the wideband spectrally flat noise used here, the spectrum of the noisy speech is similar to that of the original speech, except that it is raised by the approximately constant noise spectrum. Therefore, the noisy spectral components for which zero bits are assigned are those corresponding to low-energy spectral components of the original speech. Hence, the removal of those spectral components has an insignificant effect on the total energy of the signal, but reduces the noise power by an amount proportional to the number of noisy spectral components for which zero bits were assigned. These two effects result in the above increase of the SNR. It should, however, be clear that the application of the above enhancement procedure results not only in a greater reduction of the noise power in the encoded speech compared to that obtained as a byproduct of the encoding process (see Table 1), but, more importantly, it also reduces the irregularities characteristic to the directly encoded noisy speech.

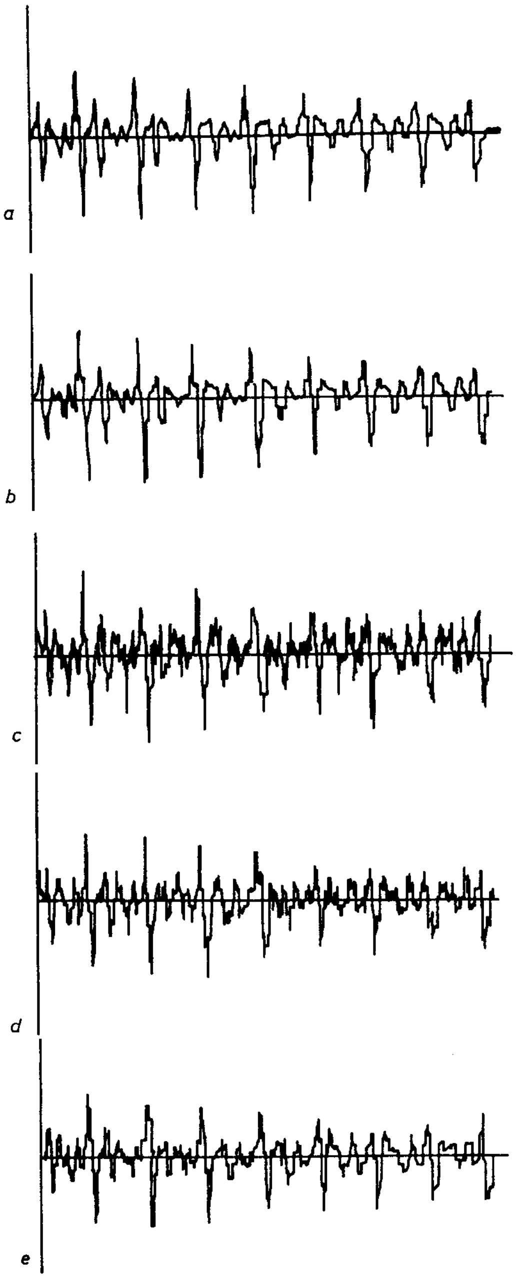

Fig. 4 *Waveform plots for 16 kbit/s encoding rate and 5 dB input SNR*
a Original speech (64 ms)
b Encoded original speech
c Noisy speech
d Directly encoded noisy speech
e Enhanced-encoded noisy speech

5 Discussion

In this paper we examine the ATC under a noisy environment. It was noted that its performance significantly degrades when the input speech is noisy. However, the application of an MMSE STSA estimator prior to the encoding process improves its performance. As the ATC is already operated in the frequency domain, the above enhancement method is well suited to its structure and can be easily implemented. For example, a look-up table which contains a finite number of samples of the multiplicative enhancing gain function can be used in a similar manner to that described in Reference 4.

It is worthwhile noting that in the course of this work we also examined the possibility of improving the ATC performance when it operates on clean speech. Specifically, we examined three alternative approaches. In the first, we reduced the quantisation noise level in each quantised DCT component by estimating the speech DCT component from the quantised (and therefore noisy) one. In the second alternative, we decorrelated the speech and the quantisation noise in each DCT component by using dithered quantisation [12, 13]. In the third alternative, we unified the above two approaches and estimated the speech DCT component from the dithered quantised one. The first and third approaches are, of course, reasonable only if the quantiser used is not optimal. Here, however, this is the case, since a uniform quantiser is used. In both cases the estimated speech DCT component, given the quantised component, is the centroid of the partition cell of the quantiser to which the quantiser component belongs [14]. We note that here we estimate the speech DCT component rather than its absolute value, since in this application the sign of the DCT component is known exactly, and both approaches are identical.

Unfortunately, the above three approaches do not noticeably improve the ATC performance. A possible explanation is that, on the basis of the MSE criterion, the expected improvement is limited by that which can be obtained by using Max's nonuniform optimal quantiser. The reason is that Max's quantiser was optimised for both the partition cells and the reproduction values, where in our case the partition cells are identical and are given, and the reproduction values alone are optimised for those given partition cells. As can be seen from Fig. 5 of Reference 8, the nonuniform quantiser can reduce the MSE (in comparison with the uniform quantiser) by, at most, 20% if the number of bits is less than or equal to four (as in the discussed case).

6 References

1 ZELINSKI, R., and NOLL, P.: 'Adaptive transform coding of speech signals', *IEEE Trans.*, 1977, **ASSP-25**, pp. 299–309
2 FLANAGAN, J.L., SCHROEDER, M.R., ATAL, B.S., CROCHIERE, R.E., JAYANT, N.S., and TRIBOLET, J.M.: 'Speech coding', *ibid.*, 1979, **COM-27**, pp. 710–737
3 TRIBOLET, J.M., and CROCHIERE, R.E.: 'Frequency domain coding of speech', *ibid.*, 1979, **ASSP-27**, pp. 512–530
4 EPHRAIM, Y., and MALAH, D.: 'Speech enhancement using a minimum mean square error short-time spectral amplitude estimator', *ibid.*, 1984, **ASSP-32**, pp. 1109–1121
5 MCAULAY, R.J., and MALPASS, M.L.: 'Speech enhancement using a soft-decision noise suppression filter', *ibid.*, 1980, **ASSP-28**, pp. 137–145
6 HUANG, J.J.Y., and SCHULTHEISS, P.M.: 'Block quantization of correlated Gaussian random variables, *ibid.*, 1963, **CS-11**, pp. 289–296
7 NARASIMHA, M.J., and PETERSON, A.M.: 'On the computation of the discrete cosine transform', *ibid.*, 1978, **COM-16**, pp. 934–936
8 MAX, J.: 'Quantizing for minimal distortion', *IRE Trans.*, 1960, **IT-6**, pp. 7–12
9 MARKEL, J.D., and GRAY, A.H. Jun.: 'Linear prediction of speech' (Springer-Verlag, New York, 1976)
10 GRAY, A.H. Jun., GRAY, R.M., and MARKEL, J.D.: 'Comparison of optimal quantization of speech reflection coefficients', *IEEE Trans.*, 1977, **ASSP-24**, pp. 4–23
11 GRADSHTEYN, I.S., and RYZHIK, I.M.: 'Table of integrals, series, and products' (Academic Press, 1980)
12 SCHUCHMAN, L.: 'Dither signals and their effect on quantization noise', *IEEE Trans.*, 1964, **COM-12**, pp. 162–165
13 LIM, J.S., and OPPENHEIM, A.V.: 'Reduction of quantization noise in PCM speech coding', *ibid.*, 1980, **ASSP-28**, pp. 107–110
14 LLOYD, S.P.: 'Least squares quantization in PCM', *ibid.*, 1982, **IT-28**, pp. 129–137

7 Appendix

We derive here the MMSE estimator [11]. Applying Bayes's rule and then using the symmetry of the Gaussian probability density function of X_k give

$$|\hat{X}_k| = \int_{-\infty}^{\infty} |x_k| \, p(x_k \mid Y_k) \, dx_k$$

$$= \frac{1}{p(Y_k)} \int_{-\infty}^{\infty} |x_k| \, p(Y_k \mid x_k) p(x_k) \, dx_k$$

$$= \frac{1}{p(Y_k)} \int_{0}^{\infty} x_k [p(Y_k \mid x_k) + p(Y_k \mid -x_k)] p(x_k) \, dx_k \tag{18}$$

From the Gaussian statistical model we assumed, we have

$$p(x_k) = \frac{1}{\sqrt{2\pi\lambda_x(k)}} \exp\left(-\frac{x_k^2}{2\lambda_x(k)}\right) \tag{19}$$

$$p(Y_k) = \frac{1}{\sqrt{2\pi(\lambda_x(k) + \lambda_d(k))}} \exp\left(-\frac{Y_k^2}{2(\lambda_x(k) + \lambda_d(k))}\right) \tag{20}$$

$$p(Y_k \mid x_k) = \frac{1}{\sqrt{2\pi\lambda_d(k)}} \exp\left(-\frac{(Y_k - x_k)^2}{2\lambda_d(k)}\right) \tag{21}$$

On substituting eqns. 19–21 into eqn. 18 and rearranging the resulting expression, we obtain

$$|\hat{X}_k| = \sqrt{\frac{2}{\pi}} \frac{1}{\lambda_k} \exp\left(-\frac{v_k}{2}\right) \int_0^{\infty} x_k \exp\left(-\frac{x_k^2}{2\lambda_k}\right) \times \cos h\left(\frac{x_k Y_k}{\lambda_{d(k)}}\right) dx_k \tag{22}$$

where λ_k satisfies

$$\frac{1}{\lambda_k} = \frac{1}{\lambda_x(k)} + \frac{1}{\lambda_d(k)} \tag{23}$$

Now, using eqn. 3.562.4 of Reference 11 to solve the integral in eqn. 22 and rearranging the resulting expression, we obtain the estimator [11].