



ELSEVIER

Speech Communication 24 (1998) 249–257

**SPEECH**  
COMMUNICATION

# Noisy speech enhancement using discrete cosine transform <sup>1</sup>

Ing Yann Soon <sup>a,\*</sup>, Soo Ngee Koh <sup>a</sup>, Chai Kiat Yeo <sup>b</sup>

<sup>a</sup> School of Electrical and Electronic Engineering, Nanyang Technological University, Block S2, Nanyang Avenue, Singapore 639798, Singapore

<sup>b</sup> School of Applied Science, Nanyang Technological University, Nanyang Avenue, Singapore 639798, Singapore

Received 19 June 1996; received in revised form 3 November 1997; accepted 30 March 1998

## Abstract

This paper illustrates the advantages of using the Discrete Cosine Transform (DCT) as compared to the standard Discrete Fourier Transform (DFT) for the purpose of removing noise embedded in a speech signal. The derivation of the Minimum Mean Square Error (MMSE) filter based on the statistical modelling of the DCT coefficients is shown. Also shown is the derivation of an over-attenuation factor based on the fact that speech energy is not always present in the noisy signal at all times or in all coefficients. This over-attenuation factor is useful in suppressing any musical residual noise which may be present. The proposed methods are evaluated against the noise reduction filter proposed by Y. Ephraim and D. Malah (1984), using both Gaussian distributed white noise as well as recorded fan noise, with favourable results. © 1998 Elsevier Science B.V. All rights reserved.

## Résumé

Cet article illustre les avantages apportés par l'utilisation de la Transformation Cosinus Discrète (DCT) par rapport à celle de la Transformée de Fourier Discrète (DFT) standard, pour le débruitage de la parole bruitée. On montre comment dériver un filtre MMSE à partir de la modélisation statistique des coefficients DCT. On montre également comment dériver un facteur de sur-atténuation basé sur le fait que, dans les signaux bruités, l'énergie de la parole n'est pas toujours présente à chaque instant ni dans chaque coefficient. Ce facteur de sur-atténuation est utile pour supprimer tout bruit résiduel musical. Les méthodes proposées ont été évaluées favorablement par rapport du filtre de réduction de bruit proposé par Ephraim et Malah (1994), en utilisant tant du bruit blanc gaussien que du bruit de ventilateur enregistré © 1998 Elsevier Science B.V. All rights reserved.

*Keywords:* Speech enhancement; MMSE amplitude estimation; Noise removal; Discrete cosine transform (DCT)

## 1. Introduction

It is often necessary to perform speech enhancement through noise removal in speech processing

systems operating in noisy environments. As the presence of noise degrades the performance of speech coders and voice recognition systems [1,2], it is therefore common to incorporate speech enhancement as a preprocessing step in these systems. The other important application of speech enhancement is to improve the perceptual quality of speech in order to reduce listener's fatigue. The additive noise may be due to the noisy envi-

\* Corresponding author. E-mail: eiysoon@ntu.edu.sg.

<sup>1</sup> Speech files available. See <http://www.elsevier.nl/locate/speech>.

ronment in which the speaker is speaking, or it may arise from noise in the transmission media. The latter problem is especially apparent in analogue mobile communications.

The topic on speech enhancement is widely researched and many speech enhancement algorithms [1–7] make use of the Discrete Fourier Transform (DFT) to make it easier to remove noise embedded in the noisy speech signal. This is often done as it is easier to separate the speech energy and the noise energy in the transform domain. For example, the energy of white noise is uniformly spread throughout the entire spectrum, but the energy of speech, especially voiced speech, is concentrated in certain frequencies. It will be shown in this paper, that the Discrete Cosine Transform (DCT) outperforms the DFT in terms of speech energy compaction.

Furthermore, most of these algorithms only attempt to modify the spectral amplitudes of the noise corrupted speech signal in order to reduce the effect of the noise component while leaving the noise corrupted phase information intact. It is of interest to note that in [3], the best estimate of the phase of the speech component was shown to be the phase of the corrupted signal itself. Hence the advantage of using a real transform, such as the DCT considered in this paper, is that the problem of not correcting for the phase will result in less severe consequences. More discussions on this can be found in Section 2.

In this paper, the amplitude estimator for the DCT is obtained based on the assumption that both the noise and the original speech signal amplitudes can be modeled by zero mean Gaussian distributed random variables in the transform domain. This assumption is supported by the Central Limit Theorem as each transform coefficient is just a weighted sum of the speech samples. An improved amplitude estimator is also obtained by taking into account that speech is not always present in the noisy signal. This is obtained by incorporating a self adaptive estimator of the probability of speech presence based on the proposed statistical model. This improvement helps to reduce the residual noise present, which is musical in nature.

## 2. DCT versus DFT

DCT [8] is widely used in image compression because of its excellent energy compaction property. This is a useful feature for noise removal purpose too. If the speech energy can be concentrated predominantly into a few coefficients while the noise energy remains white, reduction of noise can be achieved easily. As it was shown in previous work on speech coding [9], DCT provides significantly higher energy compaction as compared to the DFT. In fact, its performance is very close to the optimum Karhunen Loeve Transform (KLT). The same result is also obtained for autoregressive models of speech signals. Although the KLT is optimum in energy compaction and it also has been applied successfully to suppress noise in [10], it is not popularly used. This is because there is no fast transform methods possible for KLT, leading to high computational requirements. Therefore, DCT which is a very good approximation to KLT for speech signals is used instead.

A simple experiment is devised to illustrate the superior energy compaction of DCT versus DFT. A clean speech is first divided into frames with 50% overlapping, and the transform is performed. The transformed coefficients are then sorted according to their magnitude and the  $n$  coefficients with the lowest energy set to zero. The speech is then reconstructed using the weighted overlap add technique [11]. The Mean Square Error (MSE) is computed and plotted against  $n$  for both the DFT and DCT in Fig. 1. It can be clearly seen that the DCT provides better energy compaction. Also, the DCT has the added advantage of higher spectral resolution than the DFT for the same window size. For a window size of  $N$ , the DCT has  $N$  independent spectral components while the DFT only produces  $N/2 + 1$  independent spectral components, as the other components are just complex conjugates. This is yet another point in favour of the use of the DCT.

Lastly, as mentioned earlier in the introduction, techniques using the DFT only attempt to correct the noisy amplitude but not the phase component. This actually results in an upper bound on the maximum improvement in SNR possible. If DCT is used, a higher upper bound is possible. The effect

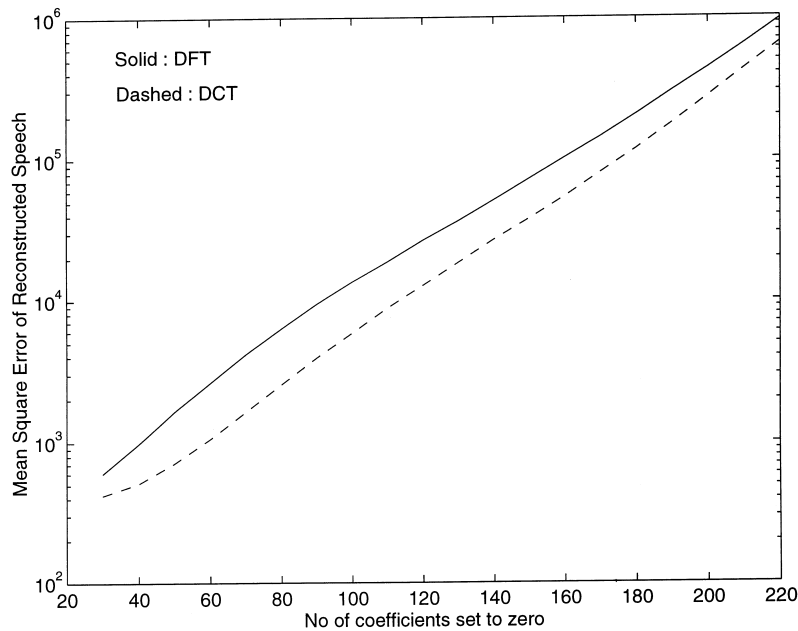


Fig. 1. Comparison of energy compaction.

of non-corrected phase on the speech is discussed in some detail in [12]. In [12], it was noted that if the phase is out by more than  $\pi/8$ , the speech becomes rough. If the phase is replaced by random noise uniformly distributed between  $-\pi$  to  $+\pi$ , a rough and completely unvoiced speech is obtained. On the other hand, if the phase is replaced by zero, the reconstructed speech sounds completely voiced and monotonous. Therefore it is not correct to view the phase as totally unimportant and especially for high levels of additive noise, the reconstructed speech quality will be affected. For DCT, the coefficients are real and can be considered to have a binary phase value. The phase will depend only on the sign of the coefficient. This provides a better degree of noise margin as unless the added noise changes the sign of the coefficient, the phase is unchanged. Therefore if strong speech energy is present in a particular coefficient, it is unlikely that the phase will be corrupted. If the noise energy is much higher than the speech energy in a particular coefficient resulting in an erroneous phase, the coefficient will be highly attenuated thus minimizing the effect of the erroneous phase. It is therefore likely that DCT would perform better than DFT.

To determine the upper bound of short-time amplitude estimation using the DFT as compared to using DCT, another experiment is carried out. In this experiment, Gaussian distributed white noise is added to a clean speech, which is then divided into 50% overlapping frames. The DFT is performed on the frames and the magnitude of the noisy frequency component is replaced by the minimum of the magnitudes of the clean frequency component and noisy frequency component. The reason behind this is that all transform based noise suppression are basically attenuation schemes. If the presence of noise results in a frequency component having a lower magnitude, it is unlikely that the noise suppression filter can increase the magnitude. The best estimate possible would be the noisy magnitude. However, if noise causes the amplitude of the frequency component to be higher, an ideal filter should lower the amplitude to the original speech amplitude. On the other hand, the noisy phase components are left unchanged. The ideally filter magnitude is then combined with the noisy phase components to obtain the ideal filtered frequency component. The optimal reconstructed speech signal is then reconstructed using the

weighted overlap add technique [11]. The process is repeated for different levels of noise added.

For DCT, a similar process is carried out. Instead of separating the frequency component into magnitude and phase as in the case of DFT, the frequency component is separated into sign and magnitude. The sign portion will be either positive or negative. The magnitude of the noisy frequency component is also replaced by the minimum of the magnitudes of the clean and noisy frequency components. Then the magnitude is recombined with the sign portion to obtain the optimal reconstructed speech using a similar scheme as here above.

The results can be seen in Fig. 2 which shows the global SNR of the reconstructed speech using DCT and DFT versus the global SNR of the noisy speech. The plots are approximately linear and it clearly shows that using DCT results in a higher upper bound for speech enhancement than using the DFT.

The one dimensional DCT is given as follows:

Forward transform of a sequence  $\{x(n), 0 \leq n \leq N - 1\}$  is given by

$$X(k) = \alpha(k) \sum_{n=0}^{N-1} x(n) \cos \left[ \frac{\pi(2n+1)k}{2N} \right],$$

$$0 \leq k \leq N - 1,$$
(1)

where

$$\alpha(0) = \sqrt{\frac{1}{N}}, \quad \alpha(k) = \sqrt{\frac{2}{N}} \quad \text{for } 1 \leq k \leq N - 1.$$
(2)

Inverse transformation is given by

$$x(n) = \sum_{k=0}^{N-1} \alpha(k) X(k) \cos \left[ \frac{\pi(2n+1)k}{2N} \right],$$

$$0 \leq n \leq N - 1.$$
(3)

More details about the DCT and its fast implementation can be found in [8,13,14].

### 3. Minimum mean square error filter for DCT

Let the clean speech signal, noisy speech signal and the noise signal be denoted by  $x(t)$ ,  $y(t)$  and  $n(t)$ , respectively, and  $y(t)$  be given as follows:

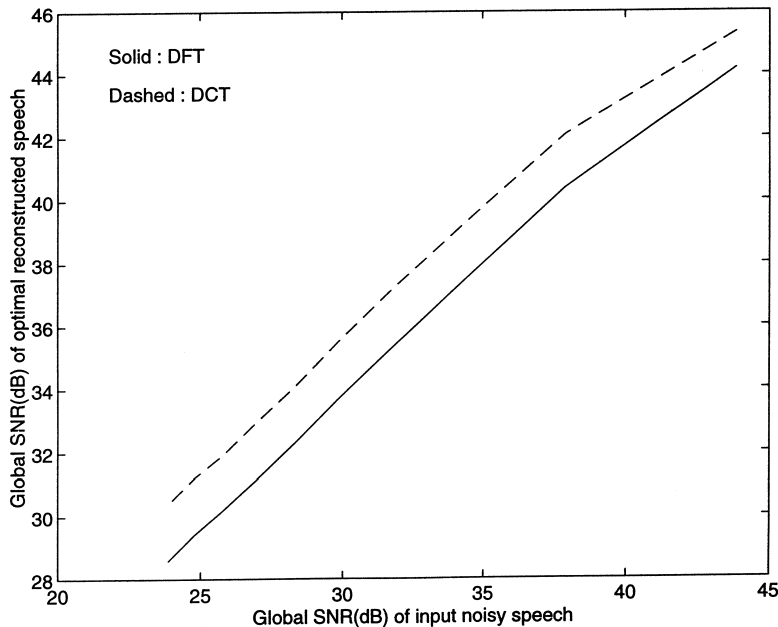


Fig. 2. Upper limits of enhancement using DCT and DFT.

$$y(t) = x(t) + n(t). \quad (4)$$

Also, let the transformed signals of the clean speech, noisy speech and noise be denoted by  $X(k)$ ,  $Y(k)$  and  $N(k)$ , respectively, where  $k$  denotes the position of the coefficient in the transform domain. With the assumption that the DCT transformed coefficients are statistically independent, the Minimum Mean Square Error (MMSE) estimated amplitude  $\hat{X}(k)$  can be obtained from  $Y(k)$  as follows:

$$\hat{X}(k) = E\{X(k) | Y(k)\}, \quad (5)$$

where  $E\{\}$  denotes the expectation operator. Eq. (5) can be rewritten, using Bayes' theorem, as

$$\hat{X}(k) = \frac{\int_{-\infty}^{\infty} a_k p\{Y(k) | a_k\} p\{a_k\} da_k}{\int_{-\infty}^{\infty} p\{Y(k) | a_k\} p\{a_k\} da_k}, \quad (6)$$

where  $p\{\}$  denotes the probability density function (PDF), and  $a_k$  is a dummy variable representing all possible values of  $X(k)$ .

Under the Gaussian distribution assumptions,  $p\{Y(k) | a_k\}$  and  $p\{a_k\}$  are given by the following equations:

$$p\{Y(k) | a_k\} = \frac{1}{\sqrt{2\pi\lambda_n(k)}} \exp\left\{-\frac{(Y(k) - a_k)^2}{2\lambda_n(k)}\right\}, \quad (7)$$

$$p\{a_k\} = \frac{1}{\sqrt{2\pi\lambda_x(k)}} \exp\left\{\frac{-a_k^2}{2\lambda_x(k)}\right\}, \quad (8)$$

where  $\lambda_x(k) = E\{|X(k)|^2\}$  and  $\lambda_n(k) = E\{|N(k)|^2\}$ . Substituting Eqs. (7) and (8) into Eq. (6),  $\hat{X}(k)$  can be easily shown to be given by

$$\hat{X}(k) = \frac{\xi(k)}{\xi(k) + 1} Y(k), \quad (9)$$

where

$$\xi(k) = \frac{\lambda_x(k)}{\lambda_n(k)}. \quad (10)$$

$\xi(k)$  is known as the a priori SNR by some authors. The above derivation shows that the Wiener filter is the MMSE amplitude estimator for the real transform case.

To use the above formula, both  $\lambda_n$  and  $\lambda_x$  have to be known. The value of  $\lambda_n$  will be assumed to be

known in this paper. Methods for estimating  $\lambda_n$  are covered in some details in [4,5]. However, an accurate estimation of  $\lambda_x$  is more difficult to achieve. This paper uses the approach known as Decision Directed Estimation developed by Ephraim and Malah [3] to estimate  $\lambda_x$ . The superiority of this estimator is covered in some detail in [15]. The estimate  $\hat{\lambda}_x$  for  $\lambda_x$  is given by the following equation:

$$\hat{\lambda}_x(k) = \alpha \hat{\lambda}_x(k)_p + (1 - \alpha) \max\{Y(k)^2 - \lambda_n(k), 0\}, \quad (11)$$

where  $\max\{\}$  is the maximum function used to ensure that a non-negative value is obtained as an estimate.  $\hat{\lambda}_x(k)_p$  is the estimated value of  $\lambda_x$  in the previous frame, while  $\alpha$  is a constant which can be adjusted to achieve the best result.

The value of  $\alpha$  is set to 0.98 in the computer simulations of the filters. Smaller values of  $\alpha$  (e.g. 0.8) are found to result in a higher level of musical tone in the residual noise. On the other hand, if  $\alpha$  is set to 1, severe distortions in the speech signals were heard. This observation agrees with that in [3]. The effect of varying  $\alpha$  is discussed in detail in [16], which states that the value of  $\alpha$  has to be greater than 0.9 in order to counter the musical noise effect and 0.98 is considered a reasonable value for  $\alpha$ . The same value of  $\alpha$  is used in [15].

#### 4. Further reduction of residual noise using uncertainty of signal presence

The derivation of the MMSE filter in Section 3 is based on the assumption that speech signal energy is always present in the sampled speech data. However, it should be noted that even in the presence of speech, the signal energy is unlikely to be significant for all the transform coefficients. Insignificant signal energy can be treated as absence of speech. Furthermore, actual speech data also consist of periods of silence. Both [3] and [5] have taken note of this and modified their filters accordingly. It was emphasized in [5] that most noise filtering algorithms are inadequate when speech is absent, hence additional attenuation should be applied during periods in which speech is absent. The residual noise of commonly used

spectral subtraction, power subtraction and other algorithms tends to be musical in nature, and is considered to be very annoying to some users [6]. Using further attenuation as suggested by this section helps to reduce the residual noise.

One means of doing so is to scale the filter output down by the conditional probability of speech present given the received spectral amplitude,  $Y(k)$ . Let the input be represented by two states,  $H_0$  and  $H_1$ , where

$H_0$ : speech absent,

$H_1$ : speech present.

The modified filter output,  $A(k)$ , taking into account the conditional probability of speech presence given  $Y(k)$ , is then given by

$$A(k) = P(H_1|Y(k))\hat{X}(k). \quad (12)$$

The approach is logical, since when the value of  $P(H_1|Y(k))$  approaches one, the filter will revert back to the original filter. While when  $P(H_1|Y(k))$  approaches zero, the filter will produce a zero output. Using Bayes' theorem, the conditional probability is given as follows:

$$P(H_1|Y(k)) = \frac{P(H_1)P(Y(k)|H_1)}{P(H_1)P(Y(k)|H_1) + P(H_0)P(Y(k)|H_0)}. \quad (13)$$

The conditional probabilities  $P(Y(k)|H_1)$  and  $P(Y(k)|H_0)$  can be obtained from the Gaussian statistical model.

$$P(Y(k)|H_0) = \frac{1}{\sqrt{2\pi\lambda_n(k)}} \exp\left(-\frac{Y(k)^2}{2\lambda_n(k)}\right), \quad (14)$$

$$P(Y(k)|H_1) = \frac{1}{\sqrt{2\pi(\lambda_x(k) + \lambda_n(k))}} \times \exp\left(-\frac{Y(k)^2}{2(\lambda_n(k) + \lambda_x(k))}\right). \quad (15)$$

If an assumption is made that the probabilities of  $H_0$  and  $H_1$  are approximately equal, Eq. (13) can be simplified to the following:

$$P(H_1|Y(k)) = \frac{1}{1 + \exp\left(\frac{-\xi(k)Y(k)^2}{2(\lambda_n(k) + \lambda_x(k))}\right) \sqrt{(1 + \xi(k))}}. \quad (16)$$

The conditional probability derived above is used to further attenuate the estimated amplitude.

## 5. Results and discussions

A total of eight sets of speech data taken from the TIMIT database are used in our evaluation. Four of the sentences are spoken by female speakers while the remaining sentences are by male speakers. The duration of the sentences ranges from 5–10 s. The speech data used are sampled at 8 kHz and quantized to 16 bits.

The proposed enhancement algorithm are tested on the speech data corrupted by two different types of additive noise. The first type of noise is the widely used Gaussian white noise. It was reported that this type of noise is more difficult to remove than any other noise source. Attempts at removing white noise usually produces an irritating residual noise. The second type of noise added to the speeches were recorded fan noise.

The noisy speech data are then divided into frames each of which consists of 256 samples with an overlap of 192 samples with the neighbouring frame. Hanning windowing is then performed on each frame before it is enhanced individually. The final enhanced speech is reconstructed from the enhanced frames using the weighted overlap and add technique [11]. The overall block diagram of the filter is shown in Fig. 3.

The segmental signal to noise ratio (SEGSNR) is used as an objective test method in the evaluation of the speech enhancement schemes. SEGSNR is chosen over other objective measures such as Signal to Noise Ratio (SNR) and MSE, because it seems to have better agreement with listening tests. The SNR and MSE measures are normally dominated by the higher energy voice portion. On the other hand, the use of SEGSNR which is the average of all the SNR in non-overlapping segments gives the unvoiced portion its proper weighting.

$$\text{SEGSNR} = \frac{1}{n} \sum_n 10 \log_{10} \frac{\sigma_{fx}^2}{\sigma_{f(x-\bar{x})}^2}, \quad (17)$$

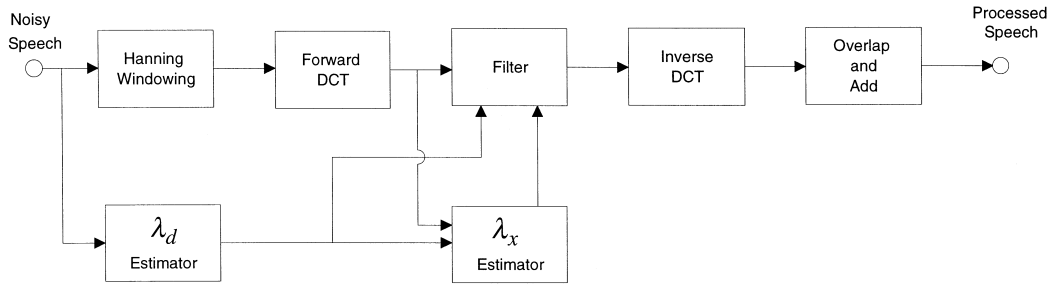


Fig. 3. Overall block diagram of noise reduction filter.

where  $n$  is the total number of non-silence frames. The classification of the frames is performed manually using the clean speech.

Both the DCT based speech enhancement filters described in Sections 3 and 4 are implemented and their results are compared with those of Ephraim and Malah filter (EMF) [3]. The DCT based Wiener filter based on Section 3 will hereafter be known as DCTF while that which is based on Section 4 will be known as DCTF2. The results are given in Tables 1 and 2.

A study of the results show that DCTF outperforms EMF in the objective test for all except one particular speech with low added noise power. The improvement in SEGSNR is significant especially for input SNR smaller than 0 dB. The superior performance is also more noticeable for actual recorded fan noise than white Gaussian noise. This is to be expected since the superior energy compaction property applies to both noise and speech. If the noise is also concentrated in fewer coefficients, it is also easier to suppress.

Table 1  
Segmental SNR tests for white noise corrupted speech

White noise added speech	Segmental SNR (dB)			
	Unprocessed	EMF	DCTF	DCTF2
fd	6.271	11.927	11.817	11.270
fb	2.897	8.614	9.102	8.739
fa	1.772	7.708	8.077	7.52
fc	1.182	8.06	9.407	9.483
mb	3.744	7.78	7.824	7.357
ma	-2.735	3.28	4.032	3.849
mc	-5.02	1.874	2.438	2.241
md	-10.166	-0.0686	1.929	2.087

Table 2  
Segmental SNR tests for fan noise corrupted speech

Fan noise added speech	Segmental SNR (dB)			
	Unprocessed	EMF	DCTF	DCTF2
fa	-1.045	11.342	13.688	13.318
fd	-2.782	10.337	12.930	8.739
mc	-13.76	-0.5062	3.373	3.592
md	-15.309	-2.737	1.941	2.449
fc	-15.631	-1.057	5.103	6.188
mb	-19.176	-4.681	0.867	1.728
fb	-20.08	-5.101	-1.566	2.749
ma	-21.994	-6.99	-0.04	0.95

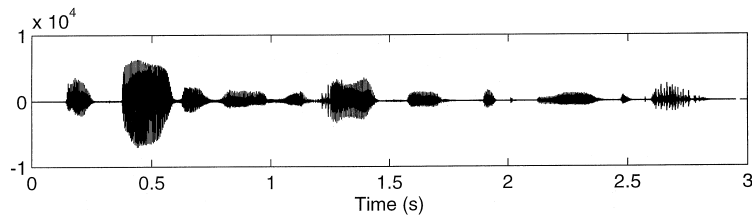


Fig. 4. Clean speech fb.

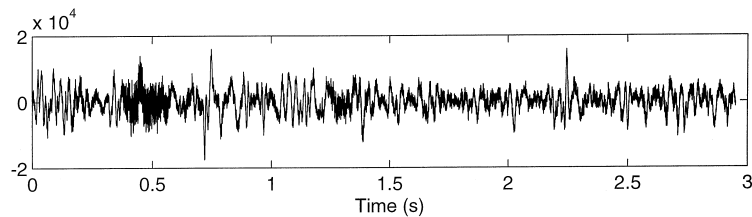


Fig. 5. Fan noise corrupted speech.

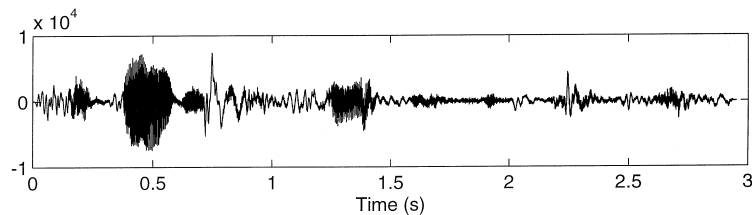


Fig. 6. EMF restored speech.

The results based on SEGSR shows that both the DCTF and DCTF2 outperform EMF, especially for the case of fan noise. However, for cleaner speeches, DCTF is better, while for very noisy speech, the additional attenuation provided by DCTF2 gives a better result. EMF does not produce any musical tones in the residual noise which is approximately white. However, the remaining background noise is still significant. The output from the DCTF has significantly less residual noise than that produced by the EMF. This is especially the case for high noise situations. However, the residual noise using the DCTF sounds slightly less uniform. DCTF2 introduces slightly more distortion for high SNR speech, but as the noise power increases, it becomes superior to the DCTF. Generally, speech processed using DCTF2 sounds much cleaner for higher degree of white noise, as

compared to DCTF or EMF. However, differences between DCTF and DCTF2 are less noticeable for the added fan noise.

Fig. 4 illustrates the original clean speech segment from the file fb. Fig. 5 shows the same speech segment with fan noise added. Figs. 6–8 show the results of enhancement using the different algorithms (EMF, DCTF, DCTF2). The  $y$  axis of Figs. 4–8 are the amplitudes of the signal while the  $x$  axis units are time in seconds. The audiofiles<sup>2</sup> <http://www.elsevier.nl/locate/specom>.<sup>2</sup> are made available for listening. The audiofiles have been converted to 8 bit mulaw sun format from the original 16 bit raw form.

<sup>2</sup> See <http://www.elsevier.nl/locate/specom>.

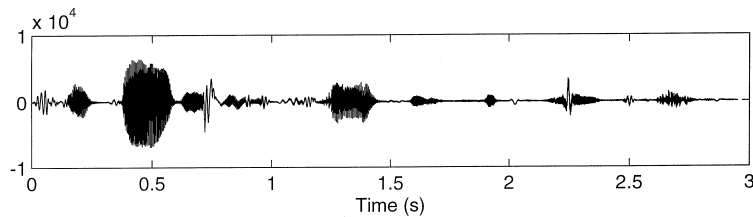


Fig. 7. DCTF restored speech.

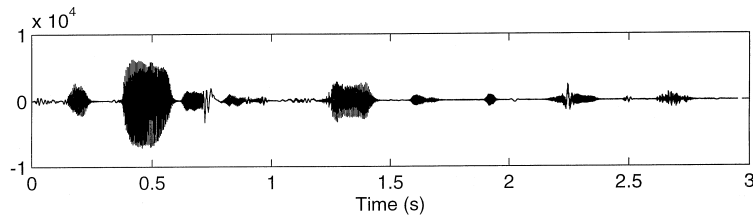


Fig. 8. DCTF2 restored speech.

## 6. Conclusions

This paper clearly shows the feasibility of using the DCT for noise reduction. The use of DCT based filters generates better results as compared to a more complex technique by Ephraim and Malah. The improvement is also more noticeable for fan noise than for white noise. Although the algorithm is simulated with the DCT, it should also be applicable to other types of real transform such as the Wavelet transform.

## References

- [1] M.R. Sambur, N.S. Jayant, LPC analysis/synthesis from speech inputs containing white noise or additive white noise, *IEEE Trans. Acoust. Speech Signal Process.* ASSP-24 (1976) 484–494.
- [2] B.H. Juang, Recent developments in speech recognition under adverse conditions, *Proceedings of the International Conference on Spoken Language Process*, November 1990, Kobe, Japan, pp. 1113–1116.
- [3] Y. Ephraim, D. Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, *IEEE Trans. Acoust. Speech Signal Process.* ASSP-32 (1984) 1109–1121.
- [4] S.F. Boll, Suppression of acoustic noise in speech using spectral subtraction, *IEEE Trans. Acoust. Speech Signal Process.* ASSP-27 (1979) 113–120.
- [5] R.J. McAulay, M.L. Malpass, Speech enhancement using a soft-decision noise suppression filter, *IEEE Trans. Acoust. Speech Signal Process.* ASSP-28 (1980) 137–145.
- [6] M. Berouti, R. Schwartz, J. Makhoul, Enhancement of speech corrupted by acoustic noise, *IEEE Proc. ICASSP*, Vol. 1, April 1979, pp. 208–211.
- [7] E. Munday, Noise reduction using frequency-domain non-linear processing for the enhancement of speech, *Br. Telecom. Technol. J.* 6 (2) (1988) 71–83.
- [8] N. Ahmed, T. Natarajan, K.R. Rao, Discrete cosine transform, *IEEE Trans. Comput.* C-23 (1974) 90–93.
- [9] R. Zelinski, P. Noll, Adaptive transform coding of speech signals, *IEEE Trans. Acoust. Speech and Signal Process.* 25 (1977) 299–309.
- [10] Y. Ephraim, D. Malah, A signal subspace approach for speech enhancement, *IEEE Trans. Speech and Audio Process.* 3 (1995) 251–266.
- [11] R.E. Crochiere, A weighted overlap-add method of short time Fourier analysis/synthesis, *IEEE Trans. Acoust. Speech Signal Process.* ASSP-28 (1980) 99–102.
- [12] P. Vary, Noise suppression by spectral magnitude estimation – Mechanism and theoretical limits, *Signal Processing* 8 (1985) 387–400.
- [13] W.H. Chen, C.H. Smith, S.C. Fralick, A fast computational algorithm for the discrete cosine transform, *IEEE Trans. Commun.* COM-25 (1977) 1004–1009.
- [14] M.J. Narasimha, A.M. Peterson, On the computation of the discrete cosine transform, *IEEE Trans. Commun.* COM-26 (6) (1978) 934–936.
- [15] P. Scalart, J. Vieira Filho, Speech enhancement based on a priori signal to noise estimation, *Proceedings ICASSP*, Vol. 2, 1996, pp. 629–632.
- [16] O. Cappe, Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor, *IEEE Trans. Speech and Audio Process.* 2 (2) (1994) 345–349.