Low distortion speech enhancement

I.Y.Soon and S.N.Koh

Abstract: An innovative approach to speech enhancement is illustrated which minimises distortion to the underlying speech during the noise-reduction process. The key to this approach lies in the identification of whether the additive noise for a particular frequency component is constructive or destructive. Once this can be identified, both multiplicative and subtractive filters can be derived using the minimum mean-square error criterion. The optimal combination of the proposed multiplicative and subtractive filter is also shown.

1 Introduction

Most of the common noise-reduction algorithms make use of some form of attenuation filter in the transform domain. The transform used can be the Fourier transform (FT) [1–6], the Karhunen Loeve transform (KLT) [7] or the discrete cosine transform (DCT) [8]. The purpose of performing the transform in the first place is to facilitate distinguishing between speech and noise in the transform domain following which, some form of subtractive or attenuative type of filter is applied to each of the different coefficients. However, as there is a significant overlap between the speech and noise, the use of such filters very often results in distortion of the underlying speech signals. In fact, attenuative filters often require a compromise between speech distortion and residual noise. Some schemes tried to reduce speech distortion by introducing selective attenuation using the masking properties associated with the human auditory system [9, 10]. However, in this paper, a different approach is adopted.

The main reason behind the use of an attenuation filter is to reverse the additive noise process. However, when both noise and speech information exist in the same transform coefficient, the resultant magnitude of the addition may not be greater than that of clean speech. In the complex Fourier transform domain, the resultant magnitude of the addition of two complex amplitudes of noise and speech with different phases, may be smaller or greater than the original speech amplitude. In cases where the result is smaller, further attenuation will only lead to further distortion in the underlying speech signal. Therefore conventional schemes only reduce noise correctly when it is in phase with the speech signal, as in the case of the DFT, or has the same sign as the speech signal, as in the case of the DCT, or when only noise is present. When speech is present, the additive noise is constructive for half the time, resulting in a higher magnitude, and destructive otherwise, resulting in a lower magnitude. In the latter case, further attenuation will only lead to further distortion. Only when speech is absent will the attenuative filter function correctly to reduce the amount of noise present. However, the residual noise can have a musical tonal structure [4] that is unacceptable for most people.

Conventional noise-reduction filters are normally of two forms, the subtractive type [2, 7] or the multiplicative type [1, 3, 6]. In this paper, noise-reduction filters of both forms are derived through the minimisation of the mean-square error for both the constructive and destructive noise situation. For the constructive noise situation, the filtered coefficient will have a lower magnitude compared to the noisy coefficient, while for the destructive noise situation, the filtered coefficient will have a higher magnitude compared to the noisy coefficient. Their performances are compared theoretically and combined to form a new optimal filter.

The paper will also show how the decision as to whether the noise is constructive or destructive can be obtained from the noisy speech itself. The problem of musical tones when speech is absent is also reduced by the proposed algorithm. The proposed filter was implemented in both the DCT as well as the DFT domains, and the results derived are applicable for any real transforms such as DCT, KLT and wavelet transforms.

2 Multiplicative filter derivation

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

$$y(t) = x(t) + n(t) \tag{1}$$

Also, let the DCT 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. Statistically, $X(k)$, $Y(k)$ and $N(k)$ can be assumed to be zero mean Gaussian distributed random variables. Similarly, their relationship in the DCT transform domain is as follows:

$$Y(k) = X(k) + N(k) \tag{2}$$

Assuming a multiplicative filter is used and its output denoted by $\hat{X}(k)$,

$$\hat{X}(k) = W(k)Y(k) \tag{3}$$

IEE Proceedings online no. 20000323

DOI: 10.1049/ip-vis:20000323

Paper first received 9th August 1999 and in revised form 4th January 2000

The authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798

The objective is of course to obtain an expression for $W(k)$ which minimises the least mean-square error, D_m, defined as follows:

$$D_m(k) = E[(W(k)Y(k) - X(k))^2]$$
$$= (W(k)^2 - 2W(k) + 1)E[X(k)^2]$$
$$+ 2W(k)(W(k) - 1)E[X(k)N(k)] + W(k)^2E[N(k)^2] \quad (4)$$

The common assumption used is that $E[X(k)N(k)]$ is zero, following which the minimisation of $D_m(k)$ will result in the well known Wiener filter:

$$W(k) = \frac{\xi(k)}{\xi(k) + 1} \quad (5)$$

where

$$\xi(k) = \frac{E[X(k)^2]}{E[N(k)^2]} \quad (6)$$

Using the above filter, the resultant mean-square error can be shown to be

$$D_m^W(k) = \frac{E[X(k)^2]}{\xi(k) + 1} \quad (7)$$

The Wiener filter's gain is always less than one. However, it is highly possible that $Y(k)$ can be smaller than $X(k)$ if $X(k)$ and $N(k)$ are of different signs. Hence in this case, attenuation will only serve to distort the speech further. On the other hand, an attenuation filter is appropriate if $X(k)$ and $N(k)$ have the same signs.

If an assumption is made that a sign detector can be used to estimate whether $X(k)$ and $N(k)$ are of the same sign or otherwise, different filters can be derived for the two different situations.

if $X(k)\ N(k) > 0$

$$E[X(k)N(k)] = E[|X(k)|]E[|N(k)|]$$
$$= \frac{2}{\pi}\sigma_X(k)\sigma_N(k) \quad (8)$$

if $X(k)N(k) < 0$

$$E[X(k)N(k)] = -E[|X(k)|]E[|N(k)|]$$
$$= -\frac{2}{\pi}\sigma_X(k)\sigma_N(k) \quad (9)$$

The above equations assume that $X(k)$ and $N(k)$ are independent of each other which is a reasonable assumption. Minimising $D(k)$ using eqns. 4, 8 and 9 results in the following:

if $X(k)\ N(k) > 0$

$$W(k) = \frac{\xi(k) + \frac{2}{\pi}\sqrt{\xi(k)}}{\xi(k) + \frac{4}{\pi}\sqrt{\xi(k)} + 1}$$

if $X(k)\ N(k) < 0$

$$W(k) = \frac{\xi(k) - \frac{2}{\pi}\sqrt{\xi(k)}}{\xi(k) - \frac{4}{\pi}\sqrt{\xi(k)} + 1} \quad (10)$$

In eqn. 10, the multiplicative factor is less than one for the case where the noise and speech amplitudes are constructive and more than one when they are destructive. These results are more correct intuitively.

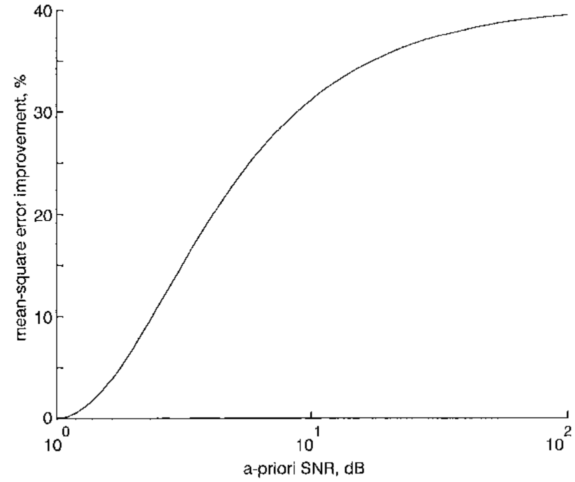

Fig. 1 *Improvement in MSE for multiplicative filter against Wiener filter*

Using this set of filters, the mean-square error, D_m^N, can be shown to be

$$D_m^N(k) = \frac{(\pi^2 - 4)(\xi(k) + 1)E[X(k)^2]}{\pi^2\xi(k)^2 + (2\pi^2 - 16)\xi(k) + \pi^2} \quad (11)$$

The percentage improvement in the mean-square error of the new filter versus the Wiener filter is defined as follows:

$$\delta D = \frac{D_m^W(k) - D_m^N(k)}{D_m^W(k)} \quad (12)$$

δD is plotted against $\xi(k)$ in Fig. 1. From the Figure, it can be seen that for high *a-priori* SNR, $\xi(k)$, a 40 percent improvement can be obtained.

3 Subtractive filter derivation

Another popular form of filtering in the transform domain is the subtractive filter used by Boll *et al.* The approach taken earlier can be repeated here. For subtractive filtering:

$$\hat{X}(k) = Y(k) - T(k) \quad (13)$$

where $T(k)$ is the threshold. The subtractive least mean-square error, $D_s(k)$, is given below:

$$D_s(k) = E[N(k)^2] - 2T(k)E[N(k)^2] + T(k)^2 \quad (14)$$

Minimising $D_s(k)$ with respect to $T(k)$ results in the following:

$$T(k) = E[N(k)] \quad (15)$$

which is not meaningful since $E[N(k)] = 0$ for the statistical assumption made whereby noise has a zero mean.

However, assuming a sign estimator is available to estimate the sign of $N(k)$, a more meaningful result can be obtained, i.e.

if $N(k) \geq 0$

$$T(k) = E[|N(k)|] = \sqrt{\frac{2}{\pi}}\sigma_N(k) \quad (16)$$

if $N(k) < 0$

$$T(k) = -E[|N(k)|] = -\sqrt{\frac{2}{\pi}}\sigma_N(k) \quad (17)$$

Using the above filter, the mean-square error, D_s^N, is given as follows:

$$D_s^N(k) = \left(1 - \frac{2}{\pi}\right)\sigma_N^2 \quad (18)$$

From the above equation, it can be seen that the noise power is reduced by more than 70%. The error expression is also independent of the signal energy.

This can be compared with the conventional subtractive filter implementation given as follows:

$$\hat{X}(k) = \text{sign}\,(Y(k))(|Y(k)| - T(k)) \quad (19)$$

For the above implementation, the mean-square error is given by

$$D_s^C(k) = \left(1 + \frac{2}{\pi}\right)\sigma_N^2 \quad (20)$$

There is no improvement at all for this implementation. In practice, improvement in the mean-square error can still be obtained using the half-wave rectification process that follows the subtraction.

4 Combined filter

The mean-square error expressions given in eqns. 11 and 18 are quite different. It is not apparent which expression is smaller. It would be ideal to combine the two filters so that the multiplicative filter is used only when it produces a lower mean-square error than the subtractive filter.

To find the condition that allows the switching of the filters, the following inequality must be solved:

$$D_m^N(k) < D_s^C(k) \quad (21)$$

$$(\xi(k) - 0.61)(\xi(k) + 2.6) < 0 \quad (22)$$

Since $\xi(k) \geq 0, \xi(k)$ should be less than 0.61 to apply the multiplicative filter. Otherwise, the subtractive filter should be applied. The result shows that the multiplicative filter is better than the subtractive filter for very low *a-priori* SNR. This will be the optimum combination of the two filters in the mean-square error sense. One of the most commonly reported problems of the subtractive filter is the musical like tones in the residual noise for low SNR regions, eg. the silence region. By switching to the multiplicative filter, there will be less impulse-like residual noise magnitudes in the filtered frame, hence reducing the musical-tone phenomenon.

5 Sign estimation

Some form of sign estimation is required before the low-distortion multiplicative and subtractive filters can perform their job. For this purpose, the short-term stationary feature of speech has to be taken into consideration. In a way, this is similar to the estimation of $|X(k)|$ except that accuracy is not important, as only a binary decision is required. It is not possible to determine whether the noise is constructive or destructive from a single received amplitude. However, the above filters are applied on overlapping frames, and if neighbouring frames are taken into consideration, it is possible to form a meaningful estimation.

If the speech information is assumed to be stationary over M (M chosen to be odd) frames, it can be assumed that the clean-speech magnitudes will be relatively constant. If the magnitude of the coefficient in the current frame is lower than the mean of the magnitudes of neighbouring frames, it will be assumed that the addition of the noise is destructive, and vice versa. Let the estimation of $|X(k)|$ be represented by $|\hat{X}_e(k)|$.

The combined filter thus becomes

if $\xi(k) < 0.61$ {Multiplicative}

if $|Y(k)| \geq |\hat{X}_e(k)|$

$$\hat{X}(k) = \frac{\xi(k) + \frac{2}{\pi}\sqrt{\xi(k)}}{\xi(k) + \frac{4}{\pi}\sqrt{\xi(k)} + 1} Y(k)$$

if $|Y(k)| < |\hat{X}_e(k)|$

$$\hat{X}(k) = \frac{\xi(k) - \frac{2}{\pi}\sqrt{\xi(k)}}{\xi(k) - \frac{4}{\pi}\sqrt{\xi(k)} + 1} Y(k)$$

ELSE {Subtractive}

if $|Y(k)| \geq |\hat{X}_e(k)|$

$$\hat{X}(k) = Y(k) - E[|N(k)|]$$

if $|Y(k)| < |\hat{X}_e(k)|$

$$\hat{X}(k) = Y(k) + E[|N(k)|] \quad (23)$$

The simplest estimation of $|\hat{X}(k)|$ is through the use of the mean or median of the corresponding magnitudes in the neighbouring M frames. This approach is only good if the clean-speech data is truly stationary over the chosen M frames, which is unlikely to be the case. A better approach, as proposed in [11], is to perform a first-order fit or even quadratic fit to perform the estimation, to account for the slow changing speech magnitudes, e.g. during a voiced speech period.

$$|\hat{X}_e(k)| = F((M+1)/2) \quad (24)$$

where $F()$ is the function for a first, second or even third-order least-square fit. It was found empirically that the second order fit is good enough for most purposes. Fig. 2 illustrates the estimated magnitude using the quadratic fit versus the clean-speech magnitude and noisy-speech magnitude of the 1 kHz Fourier coefficient for a speech

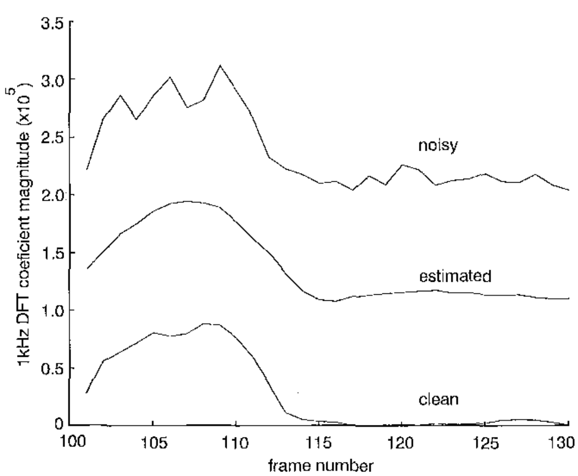

Fig. 2 *Sign estimator output (curves are offset for clarity)*

segment. The curves were offset with respect to each other by 10^5 for easy viewing. The number of frames used, M, is 11. Each frame contains 256 samples of speech data that are sampled at 8 kHz. Therefore in the time domain each frame will contains 32 ms of speech. The frames also overlap each other by 75%. It can be seen that the estimated magnitude is quite reasonable.

The main disadvantage of the sign-estimation technique is the inherent long delay time needed, especially when the number of frames used is in the range of 9 to 11. The problem of delay is acceptable for offline filtering but not suitable for real-time applications. One way of solving the problem is to use only the current and past frames for the curve fitting, which has been found to be still very acceptable. The resulting estimator in this case will be given by the equation

$$|\hat{X}_e(k)| = F(M) \quad (25)$$

6 DFT implementation

The filter discussed above can be implemented using the DFT or the DCT. The implementation of the DFT is very similar to that of the DCT, except that the phase has to be separated and only the spectral magnitude of the DFT is filtered. In the DFT domain, the relationship between the received amplitude, clean speech amplitude and noise amplitude can be described as follows:

$$Y(k)e^{j\theta_Y(k)} = X(k)e^{j\theta_X(k)} + N(k)e^{j\theta_N(k)} \quad (26)$$

Henceforth, eqn. 23 can be applied directly. Therefore, when using the DFT. $Y(k)$, $X(k)$ and $N(k)$ will only represent the magnitude of the Fourier coefficients. After filtering, the magnitude will have to be recombined with the phase to obtain the complex Fourier coefficient. The same notations will be followed in the remaining Sections.

7 Improved *a-priori* SNR estimation

All the filters developed above require the knowledge of ξ (*a-priori* SNR) which is not directly available and has to be estimated from the noisy speech itself. The most popular technique is the decision-directed approach of Ephraim and Malah [1]. This particular estimation technique has also been studied in detail by Scalart *et al.* [6]. It is described by the following equation:

$$\hat{\xi}_i(k) = \alpha\xi_{i-1}(k) + (1-\alpha)\max\left[0, \frac{Y(k)^2}{E[N(k)^2]} - 1\right] \quad (27)$$

where the parameter α is used to control the speed of the forgetfulness of the estimator. A low value of α will be suitable for rapidly changing speech regions, while a high value of α will be suitable for near stationary speech frames. Most authors use a fixed α value which is normally chosen to be in the range of 0.95 to 0.99. However, it is possible to deduce whether the speech frames are changing rapidly or not by computing the frame energy (FE). The common assumption is that noise is stationary and the noise energy does not change significantly from frame to frame. Hence, significant changes in frame energy from frame to frame must be due to the underlying speech changes. A possible formulation using only the previous frame energy is given below:

$$\alpha = \sqrt{1 - \frac{|FE_i - FE_{i-1}|}{\max(FE_i, FE_{i-1})}} \quad (28)$$

where

$$FE = \sum_k Y(k)^2 \quad (29)$$

8 Noise-only coefficients

The above filters are designed based on the situation that both noise and speech are simultaneously present. This situation will not be true during the silence periods that exist naturally during the pauses. Even when speech is present, the speech energy is not uniformly distributed in all the coefficients, and coefficients with low speech energy are dominated by the noise. For such situations the noise cannot be effectively removed using the above mentioned filters.

One of the more common methods of handling the silence period is through the incorporation of the probability of speech absence in the filter as in [1, 3, 6, 8, 12]. In the proposed algorithm, a method similar to that in [11] is adopted, as using information from multiple frames for a decision should be more accurate than using individual spectral magnitudes.

The proposed method of thresholding is to use the interframe mean, $I(k)$, of a particular spectral magnitude to decide if it should be considered as a pure-noise amplitude.

$$I(k) = \frac{1}{M}\sum_i |X_i(k)| \quad (30)$$

If $I(k)$ is less than 2 times the average noise magnitude, $|T(k)|$, it is considered to be a noise-only amplitude. A more aggressive attenuation scheme as described in eqn. 31 is then applied to such a coefficient.
If $I(k) < 2^*|T(k)|$

$$\hat{X}(k) = \frac{\beta\xi(k)}{\xi(k)+1}\min(Y(k), I(k)) \quad (31)$$

The scheme is based on the Wiener filter, which is applied to the minimum of $Y(k)$ and $I(k)$. As $I(k)$ is the interframe mean, the scheme will help to reduce the spurious noise spikes that contribute to the musical tone structure of the residual noise. β is an additional attenuation factor that is set to 0.7 to further reduce the residual noise.

9 Results and discussions

The input speech is first divided into overlapping frames of length 256 samples with 75% overlap. Hanning windowing is applied. The frames are then transformed and the resulting spectral magnitude is passed through the filter as shown in Fig. 3. The values of G1, G2, G3 and T are given in eqns. 31 and 23, respectively. For the DFT case, the phase component is then recombined with the magnitude before the inverse transform. Finally the filtered speech can be reconstructed using the overlap and add technique [13].

For the evaluation of the filters, ten speech utterances comprising five male and five female speakers from the TIMIT database were used. They were sampled at 8 kHz

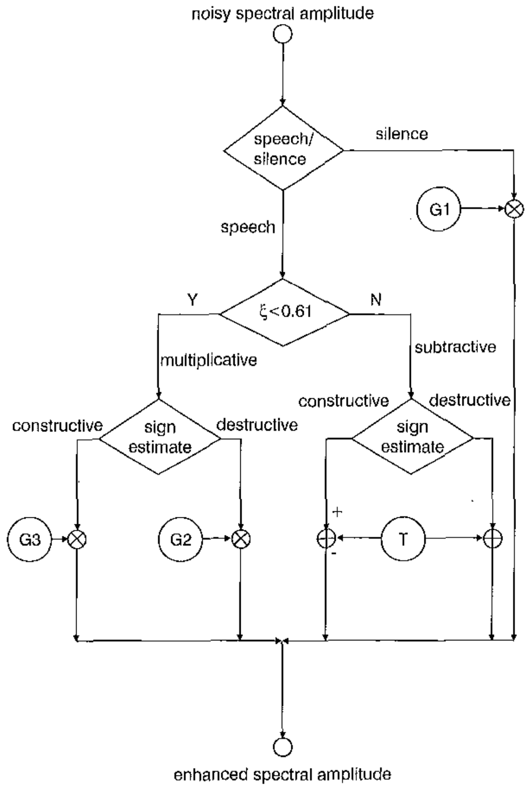

Fig. 3 *Flow chart of low-distortion filter*

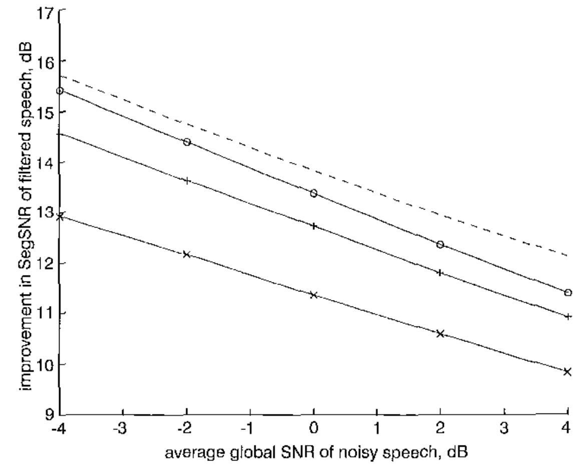

Fig. 5 *Results for additive white noise using DCT*

--·· IDCT
-O- DCT11
-+ DCT7
-x- DCT5

and linearly quantised to 16 bits. They were corrupted to different SNRs by the addition of three different types of noise. The first type was the standard Gaussian distributed white noise, while the second type of noise was the F16 noise from the NOISEX database. The third type of noise was the babble noise which is also obtained from the NOISEX database. The babble noise was recorded in a canteen with about a hundred people talking. It is highly non-stationary and speech-like in nature. The improvement in segmental signal-to-noise ratio (SegSNR) is used as an objective measure to evaluate the proposed filters. SegSNR is used as an evaluation tool because it correlates better with the mean opinion score (MOS), compared to the total SNR, and it is relatively simple to compute. The correlation factor is better than 0.8 according to [14].

For comparison purposes, the results are plotted together with the minimum mean square error filter (MMSE) by [1] with the probability of speech absence set to 0.2. The results for the DCT and DFT implementions are shown separately in Figs. 4 and 5 for the white noise, Figs. 6 and 7 for the F16 noise and Figs. 8 and 9 for the babble noise. The proposed filter implemented using DFT is denoted as DFTx while that using DCT is denoted as DCTx. The suffix x represents the number of frames, M, used for the sign estimation. The results in the Figures were obtained for $M=5$, 7 and 11. For evaluating the sign estimation algorithm, the ideal results (IDFT for DFT and IDCT for DCT) were obtained using clean speech magnitude. The sign estimator is implemented using a second-order fit for all values of M for ease of comparison.

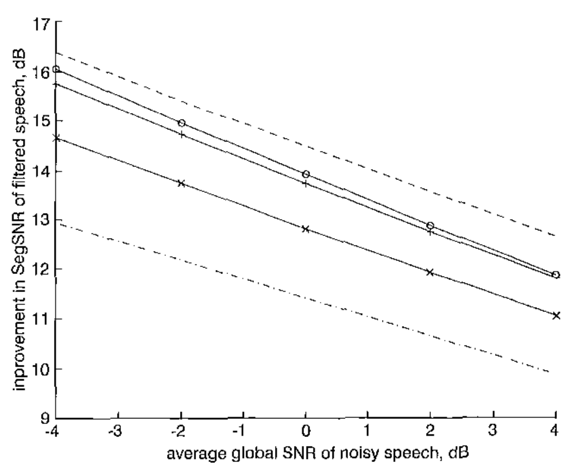

Fig. 4 *Results for additive white noise using DFT*

·--- IDFT
·O- DFT11
-+· DFT7
-x- DFT5
·-·- MMSE

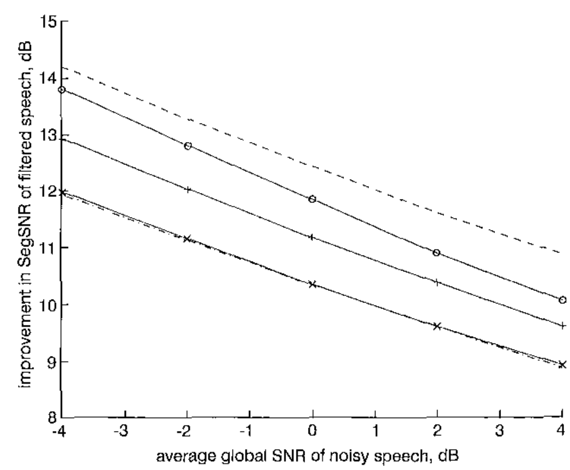

Fig. 6 *Results for F16 noise using DFT*

--· IDFT
·O· DFT11
-+- DFT7
-x· DFT5
·-·- MMSE

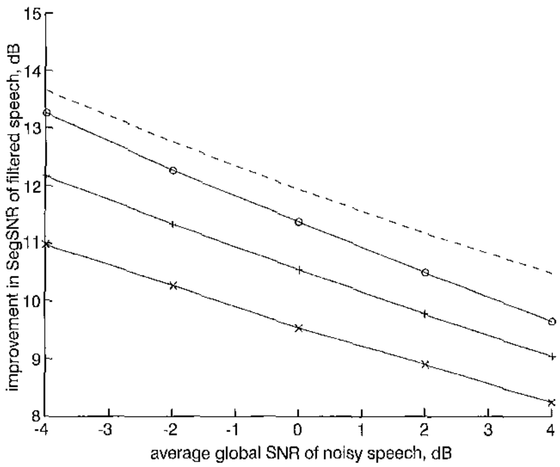

Fig. 7 *Results for F16 noise using DCT*

--·· IDCT
-O- DCT11
··+·· DCT7
-x- DCT5

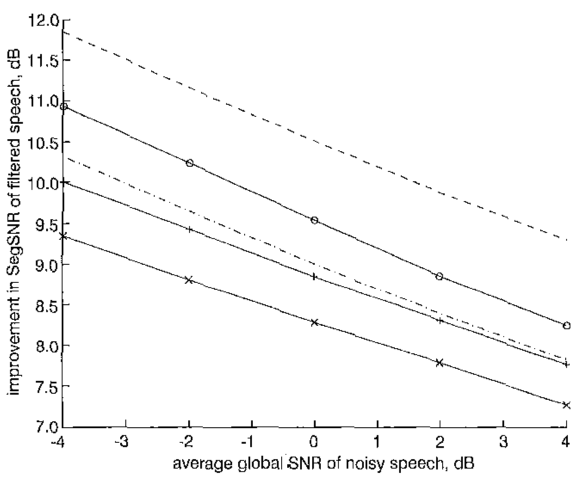

Fig. 8 *Results for babble noise using DFT*

-··· IDFT
-O· DFT11
--+-- DFT7
-x- DFT5
·-·- MMSE

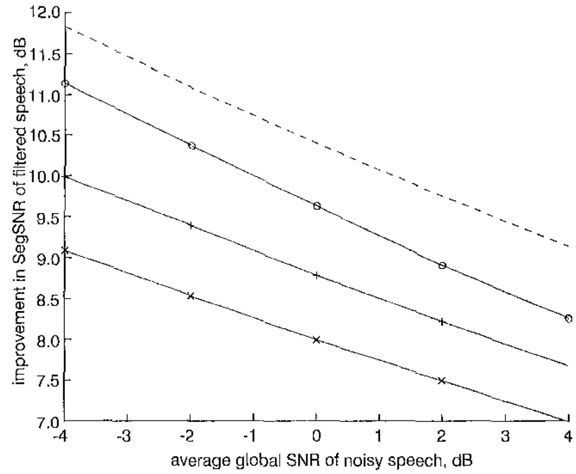

Fig. 9 *Results for babble noise using DCT*

-·· IDCT
-O· DCT11
··+·· DCT7
-x DCT5

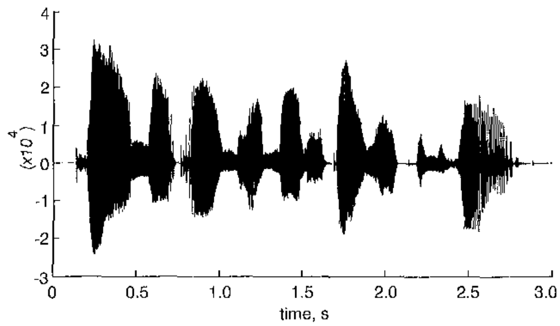

Fig. 10 *Clean speech (Jane may earn more money by working hard)*

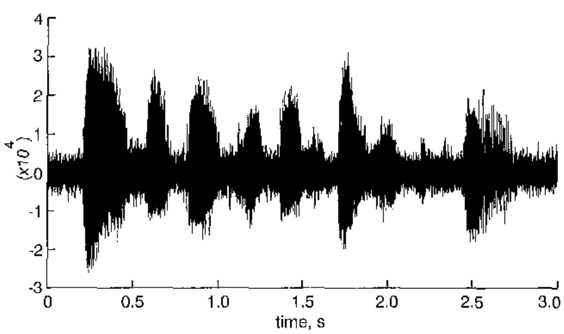

Fig. 11 *Noisy speech (white noise)*

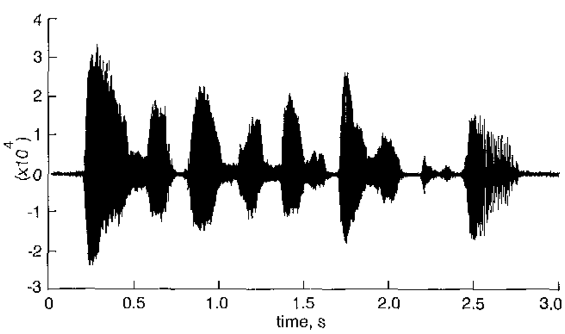

Fig. 12 *MMSE filtered speech*

From the graphs, one can conclude that the proposed algorithm works best for white noise and is relatively less effective for babble noise. The problem can be attributed to the sign estimator not functioning effectively as the babble noise is speech-like and exhibits frame to frame correlation. For white noise, both DFT11 and DCT11 are very close to IDFT and IDCT, respectively, testifying that the sign estimator gives valid results. It can also be seen that higher values of M produce better results and this fact is consistent for all types of noise. Both DFT11 and DCT11 outperform the MMSE filter for all types of noise. However, smaller values of M resulted in a filter which is less effective than the MMSE filter for the babble noise case. The peformances of the DCTx filters are quite close to those of the DFTx except for the white noise case, in which the DFTx filters give slightly better results.

A time plot of one of the noisy speech utterances and its DFT11 filtered outputs are shown in Figs. 10 to 14. The lower residual noise during the silence period can be

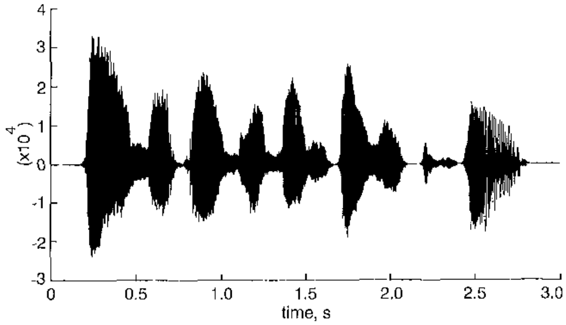

Fig. 13 *DFT11 filtered speech*

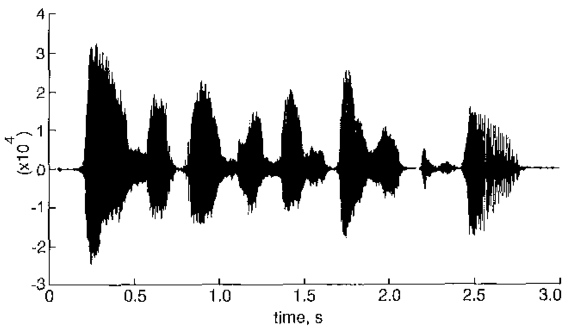

Fig. 14 *DCT11 filtered speech*

observed quite easily. On listening to the enhanced speeches, both the proposed filters are found to produce lower residual noise and noticeably less speech distortion in some speech segments. In addition, the speech energy in the proposed scheme is maintained, unlike other schemes which reduce the speech energy.

The proposed algorithm is therefore very good at minimising both the speech distortion and residual noise; the two factors which are normally traded off against each other in other algorithms.

10 References

1 EPHRAIM, Y., and MALAH, D.: 'Speech enhancement using a minimum mean-square error. short-time spectral amplitude estimator', *IEEE Trans. Acoust. Speech Signal Process.*, 1984, **32**, pp. 1109–1121

2 BOLL, S.F.: 'Suppression of acoustic noise in speech using spectral subtraction,', *IEEE Trans. Acoust. Speech Signal Process.*, 1979, **27**, pp. 113–120

3 McAULAY, R.J., and MALPASS, M.L.: 'Speech enhancement using a soft-decision noise suppression filter,', *IEEE Trans. Acoust. Speech Signal Process.*, 1980, **28**, pp. 137–145

4 BEROUTI, R.S.M., and MAKHOUL, J.: 'Enhancement of speech corrupted by acoustic noise'. Proceedings of IEEE ICASSP, pp. 208–211, 1979

5 MUNDAY, E.: 'Noise reduction using frequency-domain non-linear processing for the enhancement of speech', *Br. Telecom. Technol. J.*, 1988, **6**, (2), pp. 71–83

6 SCALART, P., and FILHO, J.V.: 'Speech enhancement based on a a priori signal to noise estimation'. Proceedings of IEEE ICASSP, 1996, **2**, pp. 629–632

7 EPHRAIM, Y., and MALAH, D.: 'A signal subspace approach for speech enhancement', *IEEE Trans. Speech Audio Process.*, 1995, **3**, pp. 251–266

8 SOON, I.Y., KOH, S.N., and YEO, C.K.: 'Noisy speech enhancement using discrete cosine transform', *Speech Commun.*, 1998, **24**, pp. 249–257

9 TSOUKALAS, D.E., MOURJOPOULOS, J.N., and KOKKINAKIS, G.: 'Speech enhancement based on audible noise suppression', *IEEE Trans. Speech Audio Process.*, 1999, **5**, pp. 497–514

10 VIRAG, N.: 'Single channel speech enhancement based on masking properties of the human auditory system', *IEEE Trans. Speech Audio Process.*, 1999, **7**, pp. 126–137

11 ERWOOD, A.E., and XYDEAS, C.S.: 'Reduction of acoustic noise in speech signals by pre-processing and spectral subtraction'. Proceedings of the IERE International Conference on *Digital processing of signals in communications*, Loughborough, England, 1981

12 SOON, I.Y., KOH, S.N., and YEO, C.K.: 'Improved noise suppression filter using self adaptive estimator of probability of speech absence', *Signal Process.*, 1999, **75**, pp. 151–159

13 CROCHIERE, R.E.: 'A weighted overlap-add method of short-time Fourier analysis /synthesis', *IEEE Trans. Acoust. Speech Signal Process.*, 1980, **28**, pp. 99–102

14 WANG, S., SEKEY, A., and GERSHO, A.: 'An objective measure for predicting subjective quality of speech coders', *IEEE J. Sel. Areas Commun.*, 1992, **10**, pp. 819–829