Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 51(65), Fascicola 2, 2006

Acoustic Noise Reduction using an Improved Power Spectral Subtraction Method Based on Hartley Transform

Radu M. Udrea, Silviu Ciochină, Dragoş N. Vizireanu[1]

Abstract – We propose an improved spectral subtraction method for reducing acoustic noise added to speech in noisy environments like helicopter cockpit or car engine. Basic power spectral subtraction is modified using Discrete Hartley Transform to estimate cross-terms that are usually neglected. A large amount of memory storage and computational volume is saved using a real data transform. Experiments with speech affected by Gaussian and engine noise showed a better estimation of clean speech with the proposed method.

Keywords: speech enhancement, spectral subtraction

I. INTRODUCTION

There are many situations when speech has to be processed in the presence of undesirable background noise that degrades speech quality and intelligibility. A variety of speech enhancement methods capable to reduce background noise were studied in the literature [1]. Many of them are adaptive techniques that use a second microphone for noise-only capture [2]. But multiple input may not be always available because of environment or cost reasons.

Spectral subtraction method was extensively studied [3], [4] because it can suppress noise effectively from speech corrupted signal only. The approach used was to estimate the power frequency spectrum of the clean speech by subtracting the noise power spectrum from the noisy power spectrum. An estimate of the current noise spectrum is approximated using the average noise square-magnitude measured during non-speech activity.

Major disadvantages of implementing spectral subtraction method consist of the large amount of computations involved in this algorithm. Obtaining noise speech spectrum, subtracting noise spectrum components and returning in time domain are operations that require many memory and processing time.

This paper presents an optimized algorithm of spectral subtraction using Discrete Hartley Transform for computing signal and noise spectrum. Also, the noise reduction algorithm was modified for Hartley spectral domain. We first identify an accurate estimation of power spectrum for clean speech. Cross terms that usually are neglected when computing power spectrum can be estimated from the spectrum of input speech and estimated noise. This terms can be effective computed using relations between Discrete Fourier Transform and Discrete Hartley Transform [6]. Algorithm optimization resulted using a real transform instead of complex Fourier Transform. Second, the equivalence between the two spectral domains and algorithm modifications are established. Finally, algorithm implementation and experimental results are presented.

II. SPECTRAL SUBTRACTION

Spectral subtraction needs only noisy speech as input [1]. The standard algorithm consists in obtaining an estimate of the noise-free signal spectrum by subtracting an estimate of the noise spectrum from the input noisy signal spectrum.

The noise spectrum is obtained from the measured signal during non-speech activity. Several assumptions are necessary for developing the algorithm. The background noise is acoustically added to the speech. The background noise remains locally stationary to the degree that its spectral magnitude expected value prior to speech activity equals its expected value during speech activity. The algorithm requires a speech detector to determine presence of speech in noisy signal.

Assume that a speech signal $s(n)$ has been degraded by the uncorrelated additive noise signal $v(n)$:

$$x(n) = s(n) + v(n) . \quad (1)$$

Taking the Fourier Transform of $x(n)$ gives:

$$X(\omega) = S(\omega) + V(\omega) . \quad (2)$$

Power spectral relation resulted from above equation is:

$$\begin{aligned} |X(\omega)|^2 &= |S(\omega)|^2 + |V(\omega)|^2 + \\ &+ S(\omega)V^*(\omega) + S^*(\omega)V(\omega) \end{aligned} \quad (3)$$

[1] Politehnica University of Bucharest, Faculty of Electronics, Telecommunications and Information Technology
1-3 Iuliu Maniu Bd., 061071 Bucharest, e-mail mihnea@comm.pub.ro, silviu@comm.pub.ro, nae@comm.pub.ro.

where $S^*(\omega)$ and $V^*(\omega)$ are complex conjugates of $S(\omega)$ and $V(\omega)$ respectively.

Because in our system only the power of the input noisy signal $|X(\omega)|^2$ can be evaluated, the rest of terms are approximated by their average during non-speech activity period.

If $v(n)$ is uncorrelated with $s(n)$ then:

$$E\{S(\omega)V^*(\omega)\} = 0 \text{ and } E\{S^*(\omega)V(\omega)\} = 0 . \qquad (4)$$

The power spectral subtraction estimator results by replacing noise square-magnitude $|V(\omega)|^2$ with its average value taken during non-speech activity period:

$$|\hat{S}(\omega)|^2 = |X(\omega)|^2 - |\hat{V}(\omega)|^2 . \qquad (5)$$

where $|\hat{V}(\omega)|^2 = E\{|V(\omega)|^2\}$.

Based on the fact that human ear does not perceive phase modifications [5], the phase $\theta_X(\omega)$ of the input signal is used for reconstruction of the estimated signal spectrum:

$$\hat{S}(\omega) = [|X(\omega)|^2 - |\hat{V}(\omega)|^2]^{1/2} e^{j\theta_X(\omega)} . \qquad (6)$$

The block diagram of spectral subtraction algorithm is represented in Fig. 1. Input signal spectrum is obtained using Discrete Fourier Transform (DFT) over the windowed half-overlapped input data buffer. The magnitude spectra of the windowed data are calculated and subtracted by the noise spectra calculated during non-speech activity.

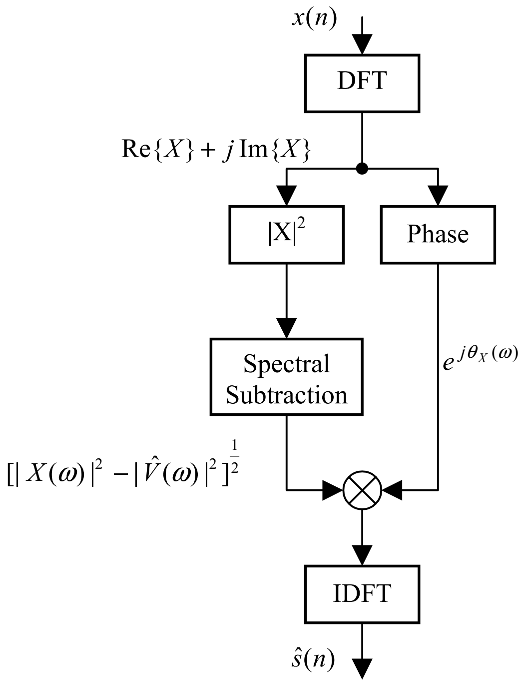


Fig. 1. Block diagram of spectral subtraction.

Then, time domain signal is obtained using Inverse Discrete Fourier Transform (IDFT) from the modified magnitude. Based on the relation (5) a number of secondary modifications can be implemented to reduce the spectral error resulted from the difference between real noise spectrum and its estimated average [1]. These include: magnitude averaging, half-wave rectification, nonlinear residual noise reduction and spectral over-subtraction.

One observation that results from the algorithm described above is that DFT returns the complex spectrum as real and imaginary parts. From this we have to compute magnitude and phase that involves a large number of computations for every spectral component. Notice that after spectral subtraction another number of computations are needed to obtain real and imaginary parts in order to perform IDFT.

III. THE IMPROVED ALGORITHM

In the literature, the cross-terms $E\{S(\omega)V^*(\omega)\}$ and $E\{S^*(\omega)V(\omega)\}$ are neglected based on the assumption that the additive noise $v(n)$ is uncorrelated with the speech signal $s(n)$. However, although these assumptions hold true in the statistical sense based on long term averaging, the assumption is not necessarily true for short-time estimates, as is the case with all the subtractive type algorithms, which are processed on a frame-by-frame basis. Figure 2 plots the values of the power spectrums of speech and noise with dashed lines and with solid line the cross terms $S(\omega)V^*(\omega) + S^*(\omega)V(\omega) = 2\,\mathrm{Re}\left[S(\omega)V^*(\omega)\right]$.

It can be clearly seen that the cross-terms are not negligible compared to the values of the power spectrum amplitudes of speech and noise.

By neglecting the cross-terms, will result an underestimate of the clean speech and thereby we will not completely suppress the noise. On accounting for the cross-terms it is possible to reduce the residual noise in the enhanced speech and thereby provide a better estimate of the clean speech.

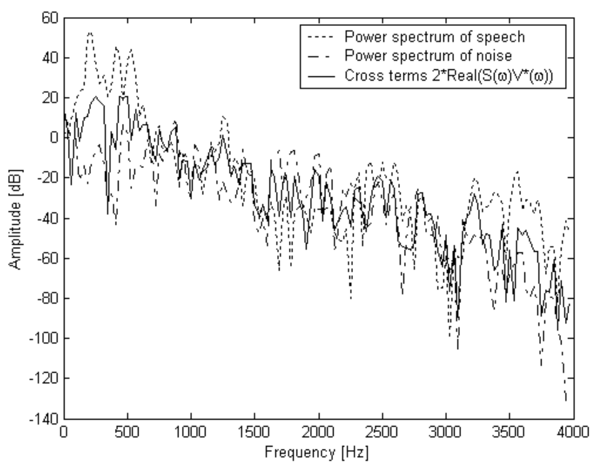


Fig. 2. Plots for power spectra of speech and noise along with the cross-term of speech and noise, over a 20 ms window.

Unfortunately, we do not have access to the clean speech and we can estimate only an average power spectrum of the noise. Therefore, in an attempt to approximate the cross-terms, using equation (2) results:

$$\begin{aligned} &X(\omega)V^*(\omega)+X^*(\omega)V(\omega)= \\ &=\left[S(\omega)+V(\omega)\right]V^*(\omega)+\left[S^*(\omega)+V^*(\omega)\right]V(\omega)= \\ &=S(\omega)V^*(\omega)+S^*(\omega)V(\omega)+2\left|V(\omega)\right|^2 \end{aligned} \quad (7)$$

If we replace the noise spectrum power $|V(\omega)|^2$ with its average estimate $|\hat{V}(\omega)|^2$ equation (3) becomes:

$$\begin{aligned} &|X(\omega)|^2=|S(\omega)|^2+|\hat{V}(\omega)|^2+ \\ &+X(\omega)V^*(\omega)+X^*(\omega)V(\omega)-2|\hat{V}(\omega)|^2 \end{aligned} \quad (8)$$

Still we cannot estimate the complex terms $V^*(\omega)$ and $V(\omega)$.

The idea of the proposed algorithm is to use for computations of short term spectrum, the Discrete Hartley Transform (DHT) which is a real transform defined by:

$$X_H(k)=\mathrm{DHT}\{x(n)\}(k)=\sum_{n=0}^{N-1}x(n)\,\mathrm{cas}\left(kn\frac{2\pi}{N}\right). \quad (9)$$

for $k=0$ to N,

and where $\mathrm{cas}(a)=\cos(a)+\sin(a)$.

The relation that can be determined between Discrete Hartley Transform and Discrete Fourier Transform (DFT) is:

$$X_H(k)=\mathrm{Re}\{X_F(k)\}-\mathrm{Im}\{X_F(k)\}. \quad (10)$$

where subscript H indicates DHT computed spectrum and subscript F indicates DFT computed spectrum

Also, because the input signal is a real sequence, symmetry of Fourier Transform gives:

$$\begin{aligned} &\mathrm{Re}\{X_F(k)\}=\mathrm{Re}\{X_F(N-k)\} \\ &\mathrm{Im}\{X_F(k)\}=-\mathrm{Im}\{X_F(N-k)\} \end{aligned} \quad (11)$$

These relations lead us to the following power spectral component relations:

$$\begin{aligned} &X_H^2(k)+X_H^2(N-k)= \\ &=2\left(\mathrm{Re}\{X_F(k)\}^2+\mathrm{Im}\{X_F(k)\}^2\right)=2|X_F(k)|^2 \end{aligned} \quad (12)$$

for $k=0$ to $\frac{N}{2}-1$.

Expressed in terms of DFT equation (8) becomes:

$$\begin{aligned} &|X_F(k)|^2=|S_F(k)|^2-|V_F(k)|^2+ \\ &+X_F(k)V_F^*(k)+X_F^*(k)V_F(k) \end{aligned} \quad (13)$$

Based on relations:

$$\begin{aligned} &X_H(k)=\mathrm{Re}\{X_F(k)\}-\mathrm{Im}\{X_F(k)\} \\ &X_H(N-k)=\mathrm{Re}\{X_F(k)\}+\mathrm{Im}\{X_F(k)\} \end{aligned} \quad (14)$$

last terms from equation (13) can be expressed as:

$$\begin{aligned} &X_F(k)V_F^*(k)+X_F^*(k)V_F(k)= \\ &=X_H(k)V_H(k)+X_H(N-k)V_H(N-k) \end{aligned} \quad (15)$$

We can estimate Hartley transform of the noise from the current frame by its absolute value average taken from frames when there is no voice activity and use the sign of the DHT from the current frame.

$$\hat{V}_H(k)=\mathrm{sgn}\{X_H(k)\}\cdot E\{|V_H(k)|\} \quad (16)$$

Results the power spectral subtraction estimator:

$$\begin{aligned} &|\hat{S}_F(k)|^2=|X_F(k)|^2+|\hat{V}_F(k)|^2- \\ &-X_H(k)\hat{V}_H(k)-X_H(N-k)\hat{V}_H(N-k) \end{aligned} \quad (17)$$

for $k=0$ to $\frac{N}{2}-1$.

By replacing the DFT power spectrums with DHT like in relation (12) results an DHT based estimator:

$$\hat{S}_H(k)^2=X_H(k)^2+\hat{V}_H(k)^2-2X_H(k)\hat{V}_H(k) \quad (18)$$

A reduced arithmetic complexity results for spectral subtraction algorithm using Discrete Hartley Transform because, instead of complex number computations like phase and absolute value we use the real data DHT.

Time domain estimate of noise-free signal is obtained with inverse transform:

$$\hat{s}(n)=\mathrm{IDHT}\{\hat{S}_H(k)\cdot\mathrm{sgn}(X_H(k)\}(n). \quad (19)$$

where Inverse Discrete Hartley Transform (IDHT) is defined by:

$$\mathrm{IDHT}\{X_H(k)\}(n)=\frac{1}{N}\sum_{k=0}^{N-1}X_H(k)\,\mathrm{cas}\left(kn\frac{2\pi}{N}\right) \quad (20)$$

Since IDHT has the same formula like DHT we can use the same subroutine for both transforms, changing input data vectors accordingly.

IV. ALGORITHM IMPLEMENTATION AND EXPERIMENTAL RESULTS

The proposed power spectral subtraction algorithm was implemented on a 16bit fixed-point Freescale SC140 DSP. Two types of noise were used for experiments: Gaussian white noise low-pass filtered and engine recorded noise. The noise and speech signals are added and used as input. Input samples are stored into a buffer designed for overlap-add method. The input buffer is then windowed using Hanning window and the Discrete Hartley Transform is computed with a fast algorithm similar to FFT.

Since our algorithm uses a real transform, a large amount of DSP memory is saved in data storage because there is no need for imaginary part of data and spectrum like in the Fourier Transform based algorithm and because the IDHT is the same subroutine like the DHT. Also, the proposed algorithm reduces by 10% arithmetic complexity of spectral subtraction method because no phase and other complex calculation are needed.

Objective measurements, in terms of signal to noise ratio improvement, can be evaluated using results above. Output noise power is computed using norm of the difference between enhanced speech and originally input clean speech.

$$SNR_{out}(dB) = 10\log_{10}\frac{\sum_{n=0}^{n\max}|s(n)|^2}{\sum_{n=0}^{n\max}|\hat{s}(n)-s(n)|^2} \qquad (21)$$

In Table 1 is shown signal to noise ratio improvement in case of standard Power Spectral Subtraction and in case of the modified method that estimates cross-terms using Hartley Transform.

Table 1. Output signal SNR for standard algorithm and for modified method with different input SNRs.

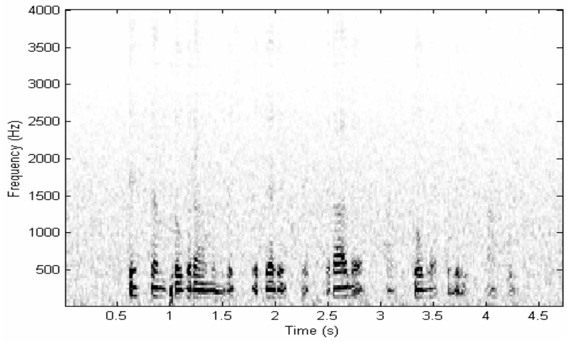
SNR_{in}	SNR_{out} (basic method)	SNR_{out} (modified method)
0 dB	10.98 dB	11.73 dB
-5 dB	3.6 dB	3.93 dB

Fig. 3 shows the spectrograms of the speech signal, the noise-speech signal and the noise-cleared speech signal obtained using our algorithm. Due to the low signal-to-noise ratio of the input signal (about -5dB) the reconstructed signal has still residual noise that can be perceived like musical tones with random frequency.

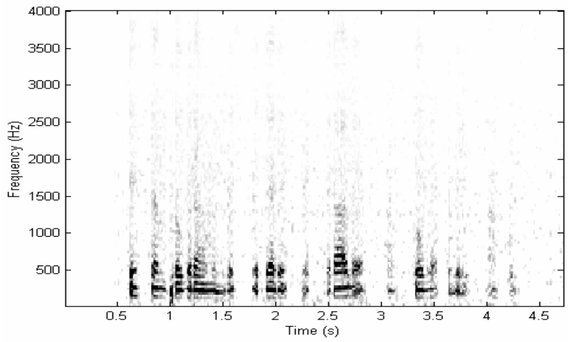
V. CONCLUSIONS

In this paper we presented a more accurate estimation of power spectrum for clean speech when affected by additive noise. Cross terms that usually are neglected when computing power spectrum can be estimated from the spectrum of input speech and estimated noise. These terms can be effective computed using relations between Discrete Fourier Transform and Discrete Hartley Transform. The noise reduction algorithm was modified for Hartley spectral domain.

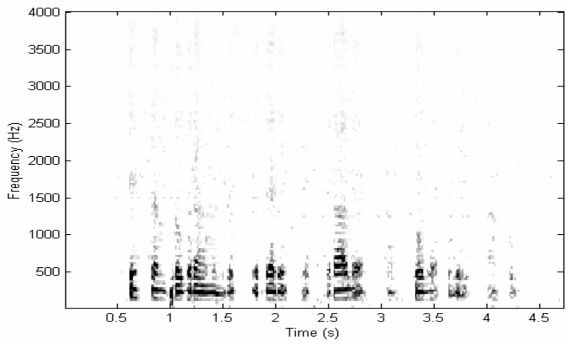
Experiments were effectuated with white noise or engine noise. Performances of the noise reduction algorithm were compared with the standard spectral subtraction algorithm.



(a)



(b)



(c)

Fig. 3. (a) Spectrogram of the speech corrupted by low-pass filtered white Gaussian noise, (b) Spectrogram of the enhanced speech using basic power spectral subtraction (c) Spectrogram of the enhanced speech using DHT modified spectral subtraction algorithm.

REFERENCES

[1] M. Berouti, R. Schwartz, J. Makhoul, “Enhancement of speech corrupted by acoustic noise”, Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc., pag. 208-211, Apr. 1979.

[2] L. Arslan, A. McCree, V. Viswanathan, “New methods for adaptive noise suppression,” ICASSP, vol.1, May 1995, pp. 812-815.

[3] Chin-Teng Lin, “Single-channel speech enhancement in variable noise-level environment”, IEEE Trans on Systems, Man and Cybernetics, Part A, Vol. 33, Jan. 2003, pp. 137–143.

[4] S. Kamath, P. Loizou, “A multi-band spectral subtraction method for enhancing speech corrupted by colored noise”, Proceedings of ICASSP-2002, Orlando, FL, May 2002.

[5] D.L. Wang and J.S. Lim, "The Unimportance of Phase in Speech Enhancement", IEEE Trans. On Acoustics, Speech, and Signal Processing, vol. 30, no.4, Aug. 1982, pp. 679-681.

[6] R. M. Udrea, S. Ciochină, “Implementation of an Optimized Algorithm for Acoustic Noise Reduction Based on Hartley Transform”, Proc. of the Symposium on Electronics and Telecommunications “Etc. 2000”, vol. I, Timişoara, Romania, Nov. 2000, pp. 223-227.