1st International Conference on Sensing Technology
November 21-23, 2005 Palmerston North, New Zealand

Kepstrum Approach to Real-time Speech Enhancement Methods Using Two Microphones

Jinsoo Jeong and Tom J. Moir

Institute of Information and Mathematical Sciences,
Massey University, Albany Campus, Albany, New Zealand
{j.jeong, t.j.moir}@massey.ac.nz

Abstract

The objective of this paper is to provide improved real-time noise canceling performance by using kepstrum analysis. The method is applied to typically existing two-microphone approaches using modified adaptive noise canceling and speech beamforming methods. It will be shown that the kepstrum approach gives an improved effect for optimally enhancing a speech signal in the primary input when it is applied to the front-end of a beamformer or speech directivity system. As a result, enhanced performance in the form of an improved noise reduction ratio with highly reduced adaptive filter size can be achieved. Experiments according to a 20cm broadside microphone configuration are implemented in real-time in a real environment, which is a typical indoor office with a moderate reverberation condition.

Keywords: kepstrum, complex cepstrum, beamforming, adaptive noise canceller, VAD, NLMS

1 Introduction

The field of noise cancellation or noise reduction from speech has been developed into several approaches using 1) sensor arrays in the spatial domain, 2) noise statistics in the time and frequency domain and 3) deconvolution in the logarithmic (homomorphic) domain.

In a speech signal processing, a short segmented speech signal can be characterized by the output of a linear time invariant system modelled by an acoustic transfer function and additive noise. Therefore, noise cancellation can be considered as a deconvolution problem.

Moir and Barrett[1] have proposed a kepstrum (complex cepstrum) approach to minimum phase wiener filtering of stationary processes and applied it to speech enhancement.

For a modified least-mean-squares (LMS) adaptive noise canceller[2], the approach uses small separation between two microphones with the use of voice activity detection (VAD), which gives favourable results that reduce significantly the filter length required for noise cancellation and minimize the presence of reverberation.

In two-microphone beamforming approaches, we can use speech directivity or speech beamforming with one or two stage adaptive filters. This is called a hybrid method, benefiting from both a modified adaptive noise canceller[2] and a modified two-microphone Griffith and Jim beamformer[3].

The use of a large amount of tap weights in adaptive filters for a high performance in signal to noise ratio (SNR) results in complexity of computation and makes reliable processing limited in a real time implementation.

A new method, the kepstrum approach combined with modified adaptive noise cancelling and speech beamforming methods is introduced, which uses a much smaller number of LMS weights as most of the acoustic transfer function modelling is absorbed by the kepstrum front-end which is mostly FFT based.

2 Kepstrum approach

2.1 Estimation of acoustic function by kepstrum analysis

Kepstrum (complex cepstrum) analysis [4-6] is used to estimate the acoustic transfer functions between two microphone channels during noise periods only. As an efficient and robust method to identify acoustic transfer function in real-time implementation, its benefit comes from robustness on computational simplicity using the FFT.

The application techniques for estimation of the acoustic transfer function use small separation between two microphones with the use of a voice-activity detector (VAD) when speech is absent.

The estimation procedure using kepstrum analysis is explained as below and illustrated in Fig. 1.

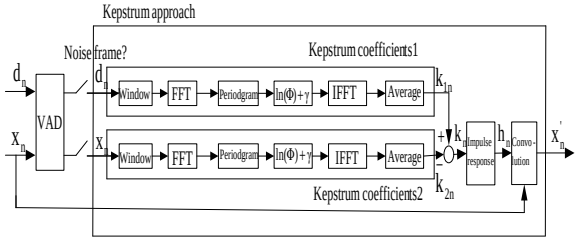

Figure 1: Block diagram for estimation procedure of acoustic transfer functions using the kepstrum method. (Window: Hanning, ln(Φ): natural log of periodogram, γ = Euler constant, 0.577215…)

2.1.1 Periodogram estimates

The conventional Welch method, which is a WOSA (weighted overlapped segment averaging) algorithm, has the computational attraction of using the discrete Fourier transform or FFT but it has limitations in real-time applications because it uses simple averaging methods. Therefore, the modified WOSA method is used in real-time processing for a faster response time because it is using a moving average at the end of each segment instead of a simple average at the end of the whole record[7]. The modified WOSA based auto and cross periodograms are processed from 50% overlapping hanning windowed 2048 FFTs as a discrete estimate of continuous power spectral density by straight batch (1, 2, 3) or smoothing methods (4, 5, 6).

$$\Phi_{m_1m_1}(i) = \frac{1}{N}\left|X_i\right|^2 \quad (1)$$

$$\Phi_{m_2m_2}(i) = \frac{1}{N}\left|Y_i\right|^2 \quad (2)$$

$$\Phi_{m_1m_2}(i) = \frac{1}{N}\left|X_i\right|\left|Y_i\right| \quad (3)$$

$$\Phi_{m_1m_1}(i) = \beta\Phi_{m_1m_1}(i-1) + (1-\beta)X_iX_i^* \quad (4)$$

$$\Phi_{m_2m_2}(i) = \beta\Phi_{m_2m_2}(i-1) + (1-\beta)Y_iY_i^* \quad (5)$$

$$\Phi_{m_1m_2}(i) = \beta\Phi_{m_1m_2}(i-1) + (1-\beta)X_iY_i^* \quad (6)$$

where $\Phi_{m_1m_1}$ and $\Phi_{m_2m_2}$ are the auto periodograms at each microphone 1 and 2 respectively and $\Phi_{m_1m_2}$ is the cross periodogram between microphone 1 and 2.

X_i and Y_i are the DFTs of the signals at each microphone 1 and 2 respectively and * indicates complex conjugate. β is forgetting factor (0< β <1) and frame number, *i*=0, 1,2,….*N*-1, where *N* is frame size. These auto and cross periodogram are applied to a VAD[8], which is comprised of the MSC (magnitude squared coherence) (7) and the TDOA (time difference of arrival) (8) function, used in this experiment.

$$\left|\gamma_{m_1m_2}(i)\right|^2 = \frac{\left|\Phi_{m_1m_2}(i)\right|^2}{\Phi_{m_1m_1}(i)\Phi_{m_2m_2}(i)} \quad (7)$$

$$R_{y_1y_2}(d) = F^{-1}\left[\psi(i)\Phi_{m_1m_2}(i)\right] \quad (8)$$

where F^{-1} denotes inverse fast Fourier transform and $R_{y_1y_2}(d)$ is the sampled generalized cross correlation[9] with weighting function, $\psi(i)$,

$$\psi(i) = \frac{\left|\gamma_{m_1m_2}(i)\right|^2}{\left|\Phi_{m_1m_2}(i)\right|\left[1-\left|\gamma_{m_1m_2}(i)\right|^2\right]} \quad (9)$$

2.1.2 Processing of kepstrum coefficients

The whole procedure above is repeated for each of the two microphones (Fig. 1). The kepstrum coefficients (k_{1n} and k_{2n}) are then found from the inverse of the natural logarithm of the auto-periodograms. By subtracting the two sets of kepstrum coefficients ($k_{1n} - k_{2n}$), we arrive at the kepstrum(k_n) equivalent to the ratio of the two acoustic transfer functions (since subtraction in logs is division in ordinary algebra). This difference in kepstrum coefficients from the two channels is then converted to an impulse response (h_n) by using the recursive formula(10)[6].

$$(n+1)h_{n+1} = \sum_{l=0}^{n} h_l(n+1-l)k_{n+1-l}\ ,\ n=0,1,2,.. \quad (10)$$

The reference signal (x_n) is then convolved with this impulse response (h_n) giving a new reference signal (x_n') produced during the noise periods.

2.2 Application to a existing speech enhancement methods

For the application to speech enhancement methods, such as the modified Griffith and Jim adaptive beamformer (Method III in section 3), the output (x_n') with TDOA delay is added with the primary input (d_n) so producing an enhanced speech signal and it is also subtracted from the primary signal (d_n) so producing a refined noise reference input.

3 Experiments and the test results

3.1 Experimental set-up

Experiments are processed by software implementation in LabVIEW in a real environment, which is typical office, indoor room with moderate reverberation condition (Fig. 2).

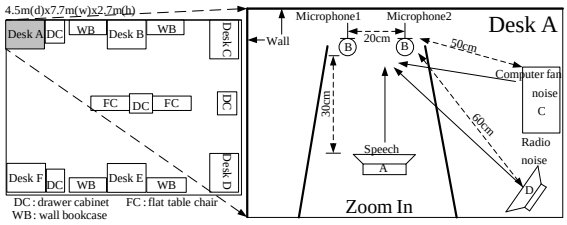

Figure 2: Experimental environment set-up (one speaker (A), two unidirectional electret condenser microphones (B) and two ambient noise sources - computer fan(C) and radio (D))

The speech signals are sampled using a standard internal sound card and two preamplifiers with unidirectional electret condenser microphones are used. The sampling frequency is chosen to be 22050Hz with 16 bits/channel, which gives quite a high quality performance as the Nyquist frequency bandwidth is around 11 kHz. Experiments according to 20 cm broadside microphone configuration are implemented.

3.2 Experimental methodology

The four methods are tested, one method uses a modified adaptive noise canceller and three methods use speech beamforming. A comparison of performance between existing methods alone and the ones with the kepstrum approach are shown in Table 1. The second test is to verify by just how much the filter size in existing methods can be reduced when the kepstrum approach is applied. It will be shown that the kepstrum approach with highly reduced adaptive filter size can achieve *almost the same performance as compared with the use of a large amount of adaptive filter weights* in an existing method.

In the following diagrams(Fig. 3, 6 ,9, 12), D_1 and D_2 represent time-delays. *K(z)* indicates kepstrum filter with 64 kepstrum coefficients and $H_1(z)$ and $H_2(z)$ are NLMS adaptive filters, each with 200 filter weights. A robust VAD based on the MSC(18) and the TDOA(19) function is used[8].

3.2.1 Method I

This method is using two stage adaptive filters in cascade and switching VAD. The kepstrum approach is applied to the front end of the first adaptive filter so its output provides a more refined noise reference input to the second adaptive noise canceller.

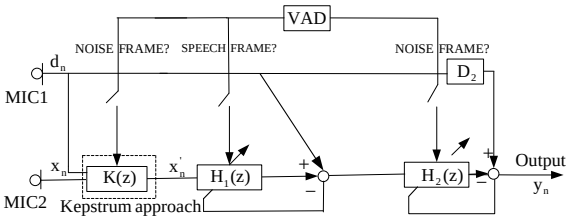

Figure 3: Block diagram of switching two stage adaptive cascading filters method with kepstrum front-end. (kepstrum is activated during noise periods only and reference output is transferred to x_n' from x_n)

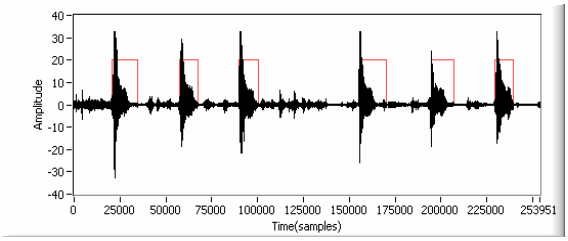

Figure 4: Test waveforms for switching two stage adaptive cascading filters method on radio and speech (kepstrum filter is switched on in mid sentence).VAD flag also shown.

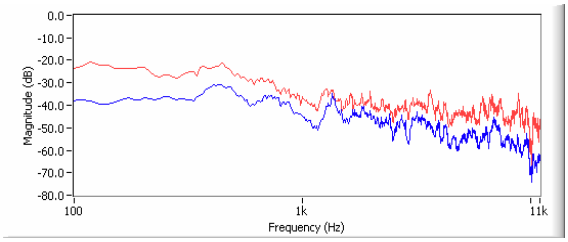

Figure 5: Average power spectra on stationary noise (computer fan) only in switching two stage adaptive cascading filters method without/with kepstrum (bottom line: kepstrum filter is on)

3.2.2 Method II

The objective is to verify the performance of kepstrum approach, which is applied to the front end of the time-difference of arrival (TDOA) steering mechanism as part of a modified Griffiths and Jim beamformer.

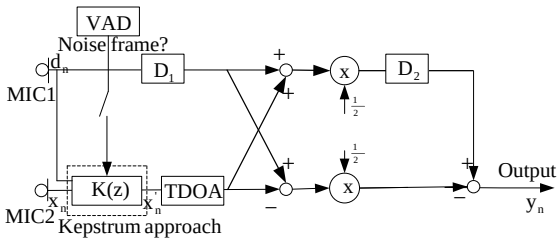

Figure 6: Block diagram of a modified G-J beamformer with a kepstrum and TDOA front-end.

(kepstrum is activated during noise periods only and reference output is transferred to x_n' from x_n)

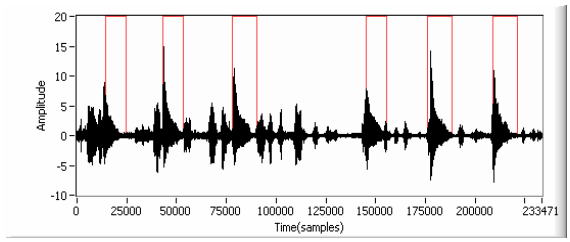

Figure 7: Test waveforms for modified G-J beamformer on radio and speech (kepstrum filter is switched on in mid sentence).VAD flag also shown.

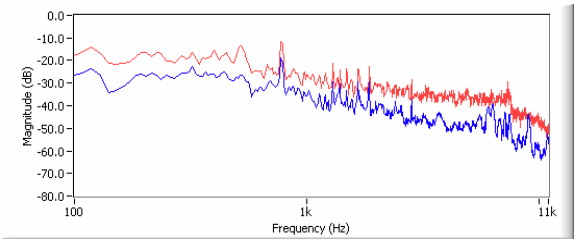

Figure 8: Average power spectra on stationary noise (computer fan) only in modified G-J beamformer without/with kepstrum(bottom line: kepstrum filter is on)

3.2.3 Method III

The kepstrum approach is applied to the front end of a TDOA modified Griffith and Jim adaptive beamformer. This method should give better performance than method II because of the use of an NLMS based adaptive filter.

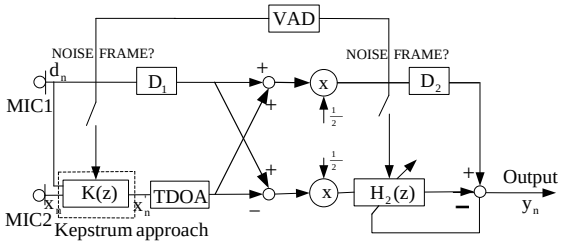

Figure 9: Block diagram of modified G-J adaptive beamformer with a kepstrum front-end. . (kepstrum is activated during noise periods only and reference output is transferred to x_n' from x_n)

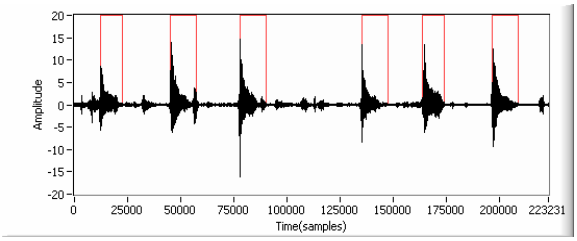

Figure 10: Test waveforms for modified G-J adaptive beamformer on radio and speech (kepstrum filter is switched on in mid sentence).VAD flag also shown.

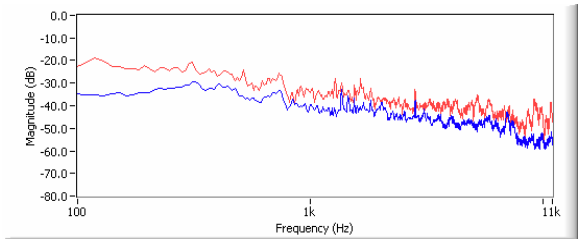

Figure 11: Average power spectra on stationary noise (computer fan) only in modified G-J adaptive beamformer without/with kepstrum front-end (bottom line: kepstrum filter is on)

3.2.4 Method IV

The kepstrum approach is applied to the front end as previously but two NLMS based switching adaptive filters are used in a modified Griffith and Jim adaptive beamformer.

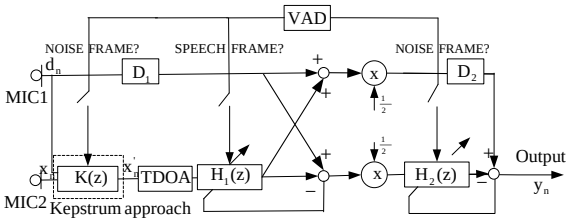

Figure 12: Block diagram of switching two stage adaptive filters in a modified G-J adaptive beamformer with kepstrum and TDOA front-end steering. . (kepstrum is activated during noise periods only and reference output is transferred to x_n' from x_n)

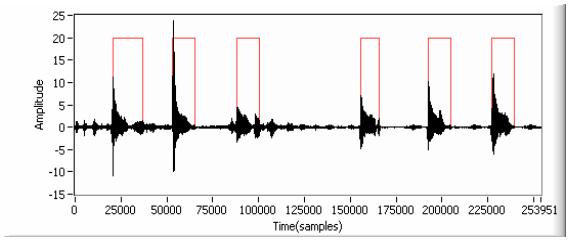

Figure 13: Test waveforms for switching two stage adaptive filters in modified adaptive G-J beamformer on radio and speech (kepstrum filter is switched on in mid sentence).VAD flag also shown.

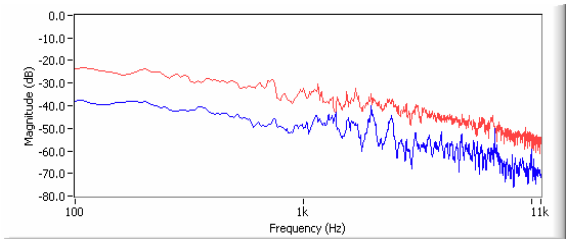

Figure 14: Average power spectra on stationary noise (computer fan) only in switching two stage adaptive filters in modified G-J adaptive beamformer without/with kepstum (bottom line: kepstrum filter is on)

3.3 Summary

From the first test results, the kepstrum approach to all four existing methods shows a quite remarkable noise reduction ratio as shown in Table 1. The higher performance in noise reduction ratio can be achieved by increasing the number of kepstrum coefficients. The second test results are shown in Table 2 and this indicates that the kepstrum approach is more applicable to speech beamforming methods (method III and IV) than the modified adaptive noise canceling method (method I). It shows that with method III and IV the number of weights can be reduced in size by up to 90%~95% in the second adaptive cascaded filter.

Table 1: Results based on Test I: stationary (computer fan), Test II: nonstationary (radio) noise, Test III: above noises with speech

Test type / AverageRMS noise power / Method type	Test I		Test II		Test III	
	Average noise power (dB)	Noise reduction ratio (dB)	Average noise power (dB)	Noise reduction ratio (dB)	Average noise power (dB)	Noise reduction ratio (dB)
1) Method I	−28.72dB	−7.00dB	−30.62dB	−8.32dB	−24.55dB	−4.22dB
Method I with kepstrum approach	−35.72dB		−38.94dB		−28.77dB	
2) Method II	−31.53dB	−7.08dB	−30.01dB	−11.03B	−31.24dB	−4.03dB
Method II with kepstrum approach	−38.61dB		−41.04dB		−35.27dB	
3) Method III	−39.63dB	−5.10dB	−34.42dB	−6.86dB	−32.96dB	−3.91dB
Method III with kepstrum approach	−44.73dB		−41.28dB		−36.87dB	
4) Method IV	−40.13dB	−4.15dB	−41.59dB	−6.82dB	−36.61dB	−1.92dB
Method IV with kepstrum approach	−44.28dB		−48.41dB		−38.53dB	

Table 2: The second test results on application of 64 kepstrum coefficients to each existing method which commonly have 200 weights in the second adaptive filter (except method II).

Test type / Filter size(H_2) / Method type	Test I (based on stationary noise)		Test II (based on nonstationary noise)	
	Filter size (H_2)	Reduction ratio (%)	Filter size (H_2)	Reduction ratio (%)
1) Method I	200	−75%	200	−75%
Method I with kepstrum approach	50		50	
2) Method II	–	N/A	–	N/A
Method II with kepstrum approach	–		–	
3) Method III	200	−95%	200	−95%
Method III with kepstrum approach	10		10	
4) Method IV	200	−90%	200	−90%
Method IV with kepstrum approach	20		20	

4 Conclusions

It can be concluded that application of the kepstrum approach gives improved results when it is applied to the front end of a speech directivity or speech beamforming system. Moreover the kepstrum front-end gives a dramatic reduction in the number of weights used for latter stage adaptive filtering or beamforming. This is an obvious advantage for real-time processing though the extra computational overhead of the kepstrum part itself must also be catered for.

5 References

[1] Moir, T.J., Barrett, J. F., "A kepstrum approach to filtering, smoothing and prediction with application to speech enhancement," *Proc. R. Soc. Lond. A*, vol. 2003, pp. 2957-2976, (2003).

[2] Harrison, W. A., Lim, J. S. and Singer, E., "A new application of adaptive noise cancellation," *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on*, vol. 34, pp. 21 - 27, (1986).

[3] Griffiths, L. J., Jim, C. W., "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on antennas and propagation*, vol. AP-30, pp. 27-34, (1982).

[4] Bogert, B. P., Healy, M.J.R. and Turkey, J.W., "The quefrency analysis of time series for echoes," presented at *Proc. Symp., Wiley, New York*, (1963).

[5] Schafer, R. W., *Echo removal by discrete generalized linear filtering*, Res. Lab. Electron. MIT, Tech. Rep., 466, (1969).

[6] Silvia, M.T., Robinson, E. A., "Use of the kepstrum in signal analysis," *Geoexploration*, vol. 16, pp. 55-73, (1978).

[7] Allen, J. N., Berkley, D.A. and Blauert, J., "Multi-microphone signal-processing technique to remove room reverberation from speech signals," *J. Acoust. Soc. Am.*, vol. 62, pp. 912-915, (1977).

[8] Agaiby, H., Moir, T. J., "A robust word boundary detection algorithm with application to speech recognition," presented at *Digital signal processing proceedings, 1997, 13th international conference*, (1997).

[9] Knapp, C., Carter G. C., "The generalized correlation method for estimation of time delay," *IEEE Transaction on Acoustics, Speech, and Signal Processing*, vol. ASSP-24, (1976).