Res. Lett. Inf. Math. Sci., 2005, Vol. **7**, pp 135-145 135
Available online at http://iims.massey.ac.nz/research/letters/

Kepstrum approach to real-time speech-enhancement methods using two microphones

J.JEONG & T.J.MOIR

Institute of Information & Mathematical Sciences
Massey University at Albany, Auckland, New Zealand

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 20cm broadside microphone configuration are implemented in real-time in a real environment, which is a typical indoor office with a moderate reverberation condition.

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 speech signal processing, a short segmented speech signal can be characterized by the output of a linear time- invariant system modeled by an acoustic transfer function and additive noise. Therefore, noise cancellation can be considered as a deconvolution problem. Moir and Barrett(Moir, 2003) have proposed a kepstrum (complex cepstrum) approach to minimum phase Wiener filtering of stationary processes and applied it to speech enhancement.Lim(Lim, 1979) has developed a new homomorphic deconvolution system, essentially the same as the logarithmic homomorphic deconvolution system except that the logarithmic and exponential operations are replaced with $(.)^r$ and $(.)^{1/r}$ operations.For a modified least-mean-squares (LMS) adaptive noise canceller, the approach uses small separation between two microphones with the use of voice activity detection (VAD), which gives favorable 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 and a modified two-microphone Griffith and Jim beamformer. 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 canceling and speech beamforming methods is introduced, which uses a much smaller number of LMS weights as most of the acoustic transfer function modeling is absorbed by the kepstrum front-end which is mostly FFT based.

2. Kepstrum Approach

The word ‘cepstrum’, an anagram of the word ‘spectrum’, has originally been defined by Bogert et al (1963) as the magnitude power spectrum of the logarithm of the power spectrum of the observed time series and the word ‘complex cepstrum’ has been used by Schafer (1969) for both the magnitude and phase spectra of the observed signal. On the other hand, the word ‘kepstrum’, quite similar to the word ‘cepstrum’, is collected from the first letters of the Kolmogorov Equation Power Series Time Response, and then adds the Latin singular ending ‘um’ to denote one kepstrum and its plural word ‘kepstra’ is also used to denote more than one kepstrum by adding the Latin plural ending ‘a’. The two methods, complex cepstrum and kepstrum give quite identical results for most purposes but seem to have been derived from a different theoretical backgrounds. The complex cepstrum method is essentially a practical approach to signal analysis based on the use of the discrete Fourier transform, using the fast Fourier transform algorithm in homomorphic signal processing. However, this method is theoretically based on quantities dependent both on sample length and statistical variation. The kepstrum method could be considered as an alternative approach to supply a surer theoretical foundation, not subject to statistical variation, providing the relationship that the complex cepstra is a truncated version of the kepstrum coefficients corresponding to the sample length chosen.

2.1 Estimation of acoustic transfer function by kepstrum analysis

Kepstrum (complex cepstrum) analysis 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.

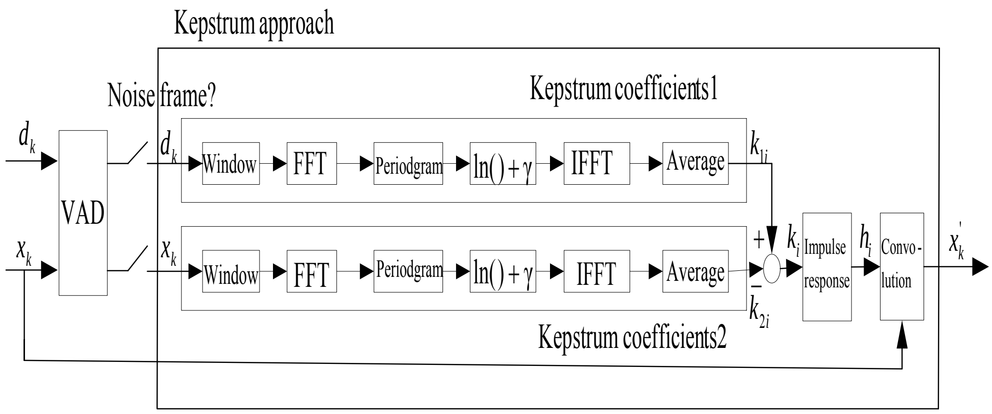

Figure 1: Block diagram for estimation procedure of acoustic transfer functions by the kepstrum method

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. Estimation of acoustic transfer function using kepstrum analysis is processed and illustrated in Fig. 1.

The periodogram is processed from 50% overlapping windowed 2048 point FFTs as a discrete estimate of continuous power spectral density by either a straight batch method

$$\Phi_x(i) = \frac{1}{N}\left|X_i\right|^2 \tag{1}$$

where the discrete-Fourier transform (DFT) is given by

$$X_i = \sum_{j=0}^{N-1} x_j W_N^{ji}, i = 0,1,2,...N-1 \tag{2}$$

and x_j is the time-series. Alternatively a smoother periodogram is estimated from the recursion

$$\Phi_x(m) = \beta\Phi_x(m-1) + (1-\beta)X_i X_i^* \tag{3}$$

where Φ_x is auto periodogram, β is forgetting factor (0< β<1) and frame number, m=0, 1,2,….N-1, where N is frame size. After taking the natural log of periodogram, it is found that there exists a bias equal in magnitude to minus Euler's constant γ=0.577215…(Wahba, 1980), which has to be removed. The whole procedure is repeated for each of the two microphones. The kepstrum coefficients are then found from the inverse of the natural logarithm of the periodograms. By subtracting the two sets of kepstrum coefficients we arrive at the kepstrum equivalent of 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 are then converted to an impulse response by using a recursive formula(M.T. Silvia, 1978). The reference signal is then convolved with this impulse response.

3. Experiments and test results

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

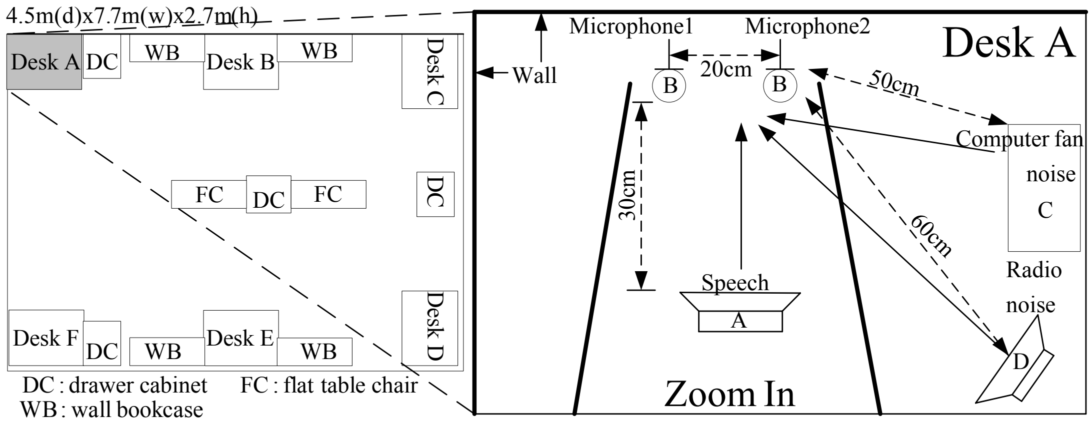

Figure 2: Experimental environment set-up (one speaker (A), two unidirectional electret condenser microphones (B) and two ambient noise sources -computer fan(C) and radio (D))

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. 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 D_1 and D_2 represent a time-delay and the TDOA (time-difference of arrival) algorithm uses the so-called Generalized Cross-correlation (GCC) method(Knapp, 1976).

Method I

This method is using two stage adaptive filters in cascade and a switching voice-activity detector (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.

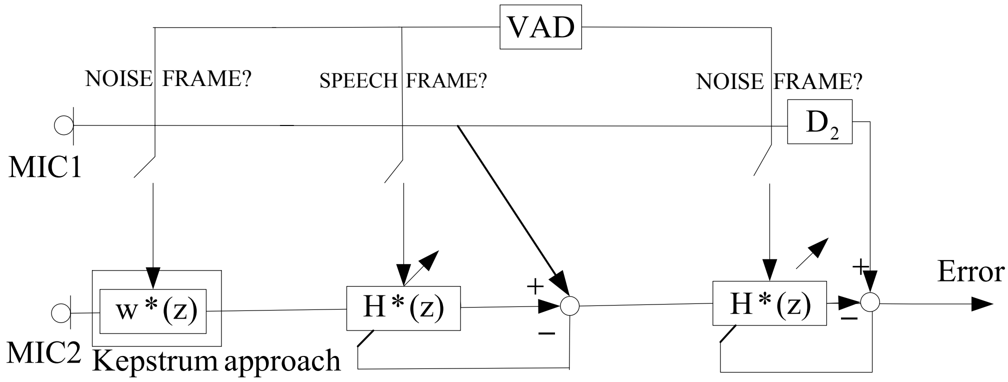

Figure 3: Block diagram of switching two-stage adaptive cascading filtering method with a kepstrum front-end.

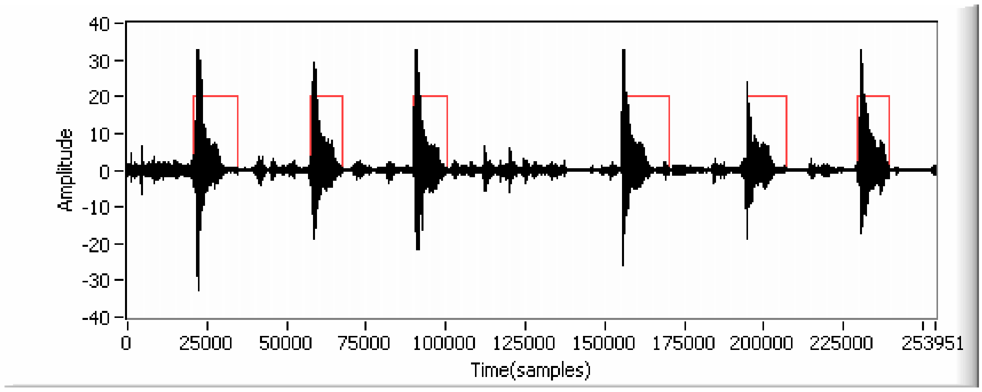

Figure 4: Test waveforms for switching two stage adaptive cascading filtering method with radio noise and speech . (kepstrum filter is switched on in mid sentence).VAD flag also shown.

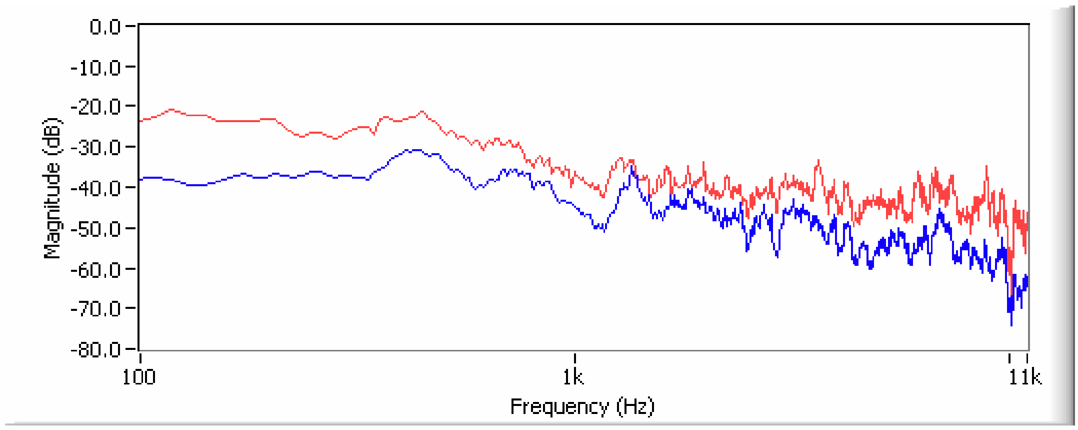

Figure 5: Average power spectra on stationary noise (computer fan) for switching two stage adaptive cascading filtering method without/with kepstrum front-end. (bottom line: kepstrum filter is on)

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.

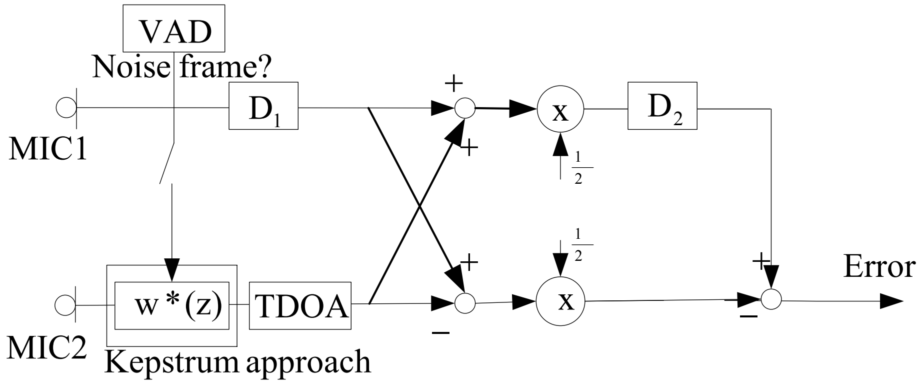

Figure 6: Block diagram of a modified G-J beamformer with a kepstrum and TDOA front-end.

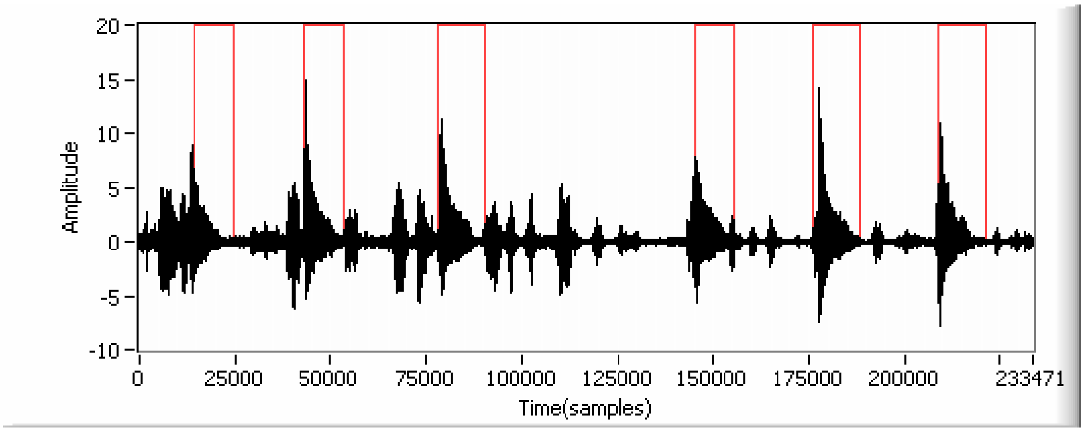

Figure 7: Test waveforms for modified G-J beamformer on radio noise and speech (kepstrum filter is switched on in mid sentence).VAD flag also shown.

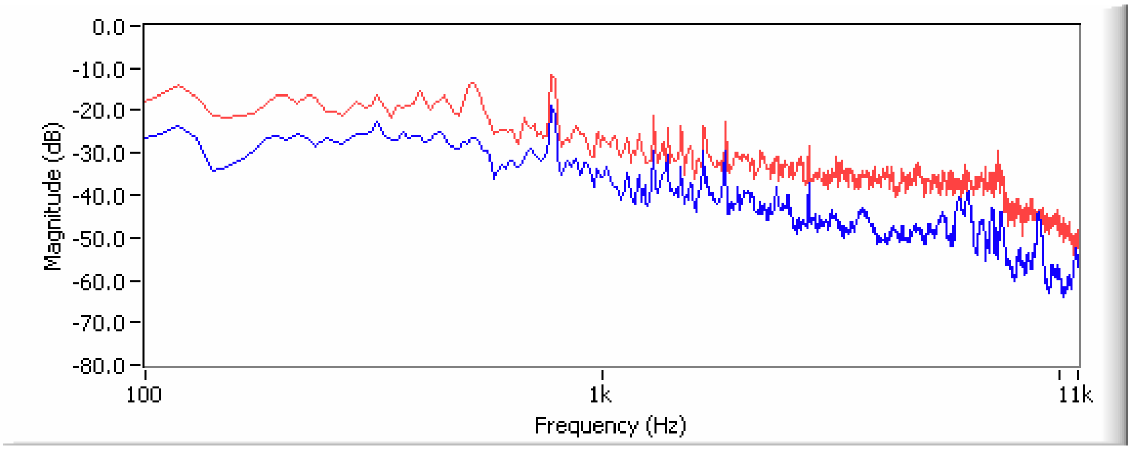

Figure 8: Average power spectra on stationary noise (computer fan) for a modified G-J beamformer without/with kepstrum front-end. (bottom line: kepstrum filter is on)

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.

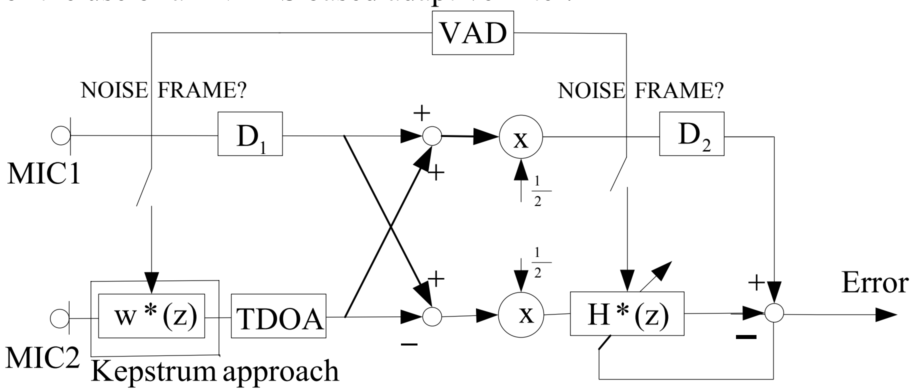

Figure 9: Block diagram of modified G-J adaptive beamformer with a kepstrum front-end.

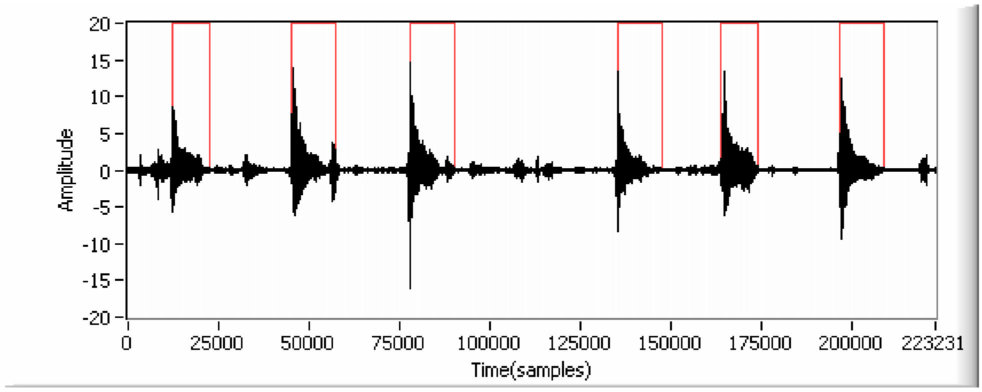

Figure 10: Test waveforms for modified G-J adaptive beamformer on radio noise and speech. (kepstrum filter is switched on in mid sentence).VAD flag also shown.

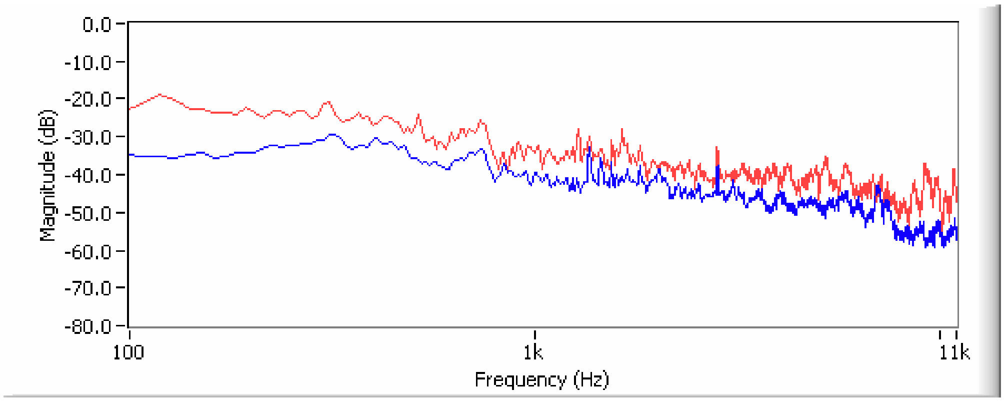

Figure 11: Average power spectra on stationary noise (computer fan) for a modified G-J adaptive beamformer without/with kepstrum front-end (bottom line: kepstrum filter is on)

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.

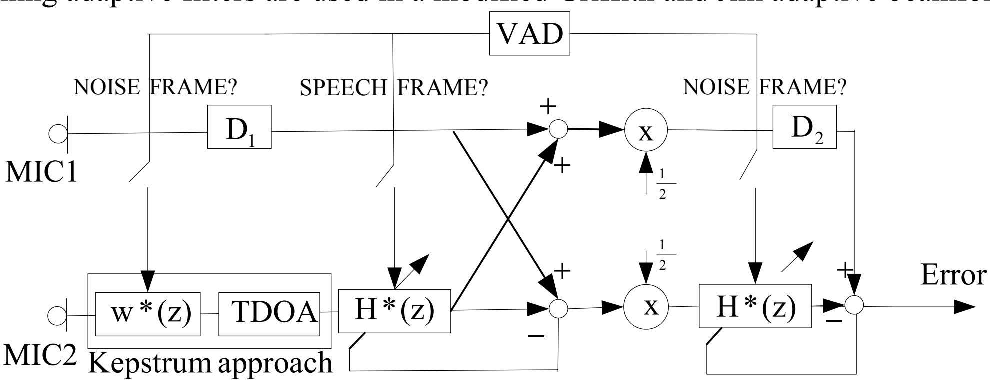

Figure 12: Block diagram of switching two-stage adaptive filters in a modified G-J adaptive beamformer with kepstrum and TDOA front-end steering.

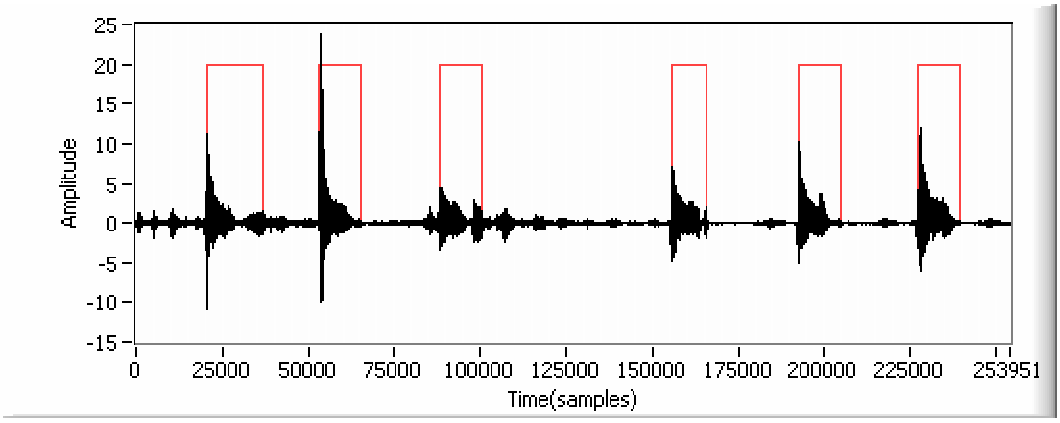

Figure 13: Test waveforms for switching two stage adaptive filters for a modified adaptive G-J beamformer on radio-noise and speech (kepstrum filter is switched on in mid sentence).VAD flag also shown.

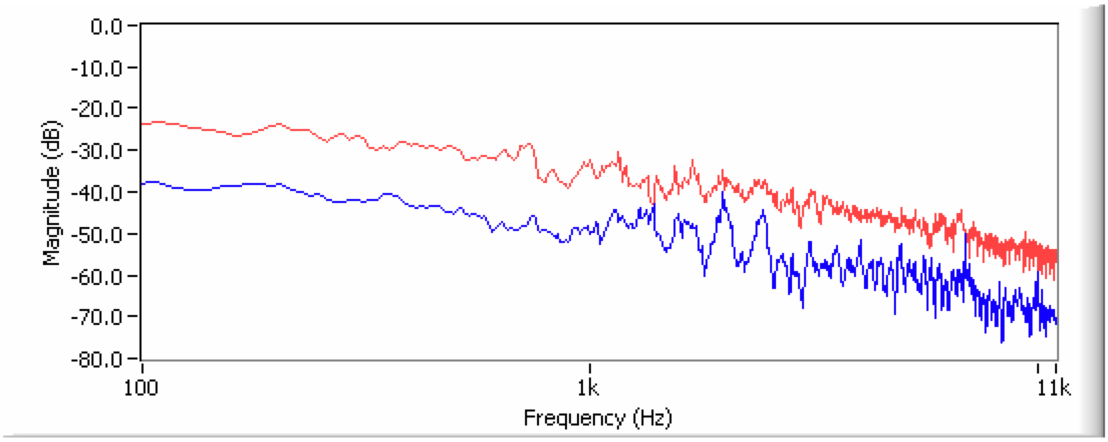

Figure 14: Average power spectra on stationary noise (computer fan) in switching two stage adaptive filters in modified G-J adaptive beamformer without/with kepstum (bottom line: kepstrum filter is on)

From the first set of test results, the kepstrum approach applied 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). The results show 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)
Method I	−28.72dB	−7.00dB	−30.62dB	−8.32dB	−24.55dB	−4.22dB
Kepstrum approach	−35.72dB		−38.94dB		−28.77dB	
Method II	−31.53dB	−7.08dB	−30.01dB	−11.03B	−31.24dB	−4.03dB
Kepstrum approach	−38.61dB		−41.04dB		−35.27dB	
Method III	−39.63dB	−5.10dB	−34.42dB	−6.86dB	−32.96dB	−3.91dB
Kepstrum approach	−44.73dB		−41.28dB		−36.87dB	
Method IV	−40.13dB	−4.15dB	−41.59dB	−6.82dB	−36.61dB	−1.92dB
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 / Method type	Test I		Test II	
	Filter size	Reduction ratio (%)	Filter size	Reduction ratio (%)
Method I	200	−75%	200	−75%
Kepstrum approach	50		50	
Method II	–	N/A	–	N/A
Kepstrum approach	–		–	
Method III	200	−95%	200	−95%
Kepstrum approach	10		10	
Method IV	200	−90%	200	−90%
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.

References

Knapp, C., Carter G. C. (1976). The Generalized Correlation Method for Estimation of Time Delay. *IEEE Transaction on Acoustics, Speech, and Signal Processing, ASSP-24*(4).

Lim, J. (1979). Spectral root homomorphic deconvolution system. *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on, 27*(3), 223-233.

M.T. Silvia, E. A. R. (1978). Use of the kepstrum in signal analysis. *Geoexploration, 16*, 55-73.

Moir, T. J., Barrett, J. F. (2003). A kepstrum approach to filtering, smoothing and prediction with application to speech enhancement. *Proc. R. Soc. Lond. A, 2003*(459), 2957-2976.

Wahba, G. (1980). Automatic smoothing of the log periodogram. *J. Amer. Stat. Assoc., 75*, 122-132.