SPEECH DEREVERBERATION BASED ON LINEAR PREDICTION: AN ACOUSTIC VECTOR SENSOR APPROACH

M. Shujau, C. H. Ritz and I. S. Burnett

School of Electrical, Computer, and Telecommunications Engineering
University of Wollongong, Wollongong NSW Australia
ms970@uowmail.edu.au, critz@uow.edu.au

School of Electrical and Computer Engineering
RMIT University, Melbourne, VIC, Australia
ian.burnett@rmit.edu.au

ABSTRACT

This paper introduces a dereverberation algorithm based on Linear Prediction (LP) applied to the outputs of an Acoustic Vector Sensor (AVS). The approach applies adaptive beamforming to take advantage of the directional outputs of the AVS array to obtain a more accurate LP spectrum than can be obtained with a single channel or Uniform Linear Array (ULA) with a comparable number of channels. This is then used within a modified version of the Spatiotemporal Averaging Method for Enhancement of Reverberant Speech (SMERSH) algorithm derived for the AVS to enhance the LP residual signal. In a highly reverberant environment, the approach demonstrates a significant improvement compared to a ULA as measured by both the Signal to Reverberant Ratio (SRR) and Speech to Reverberation Modulation Energy Ratio (SRMR) for sources ranging from at 1m to 5m from the array.

Index Terms: Acoustic Vector Sensors, Dereverberation, Speech Enhancement

1. INTRODUCTION

The effect of reverberation on speech and audio plays an important role in making the sound more natural, but when the level of reverberation increases (beyond RT_{60}= 1s) it significantly degrades the quality of the speech signals in terms of intelligibility due to box effect and distant talker effect [1, 2]. Reverberation also leads to altering of the parameters derived for source-filter models of speech typically used in applications such as distant speech recognition and hence degrades their performance [2]. There are several methods that can be used for enhancing speech through dereverberation including beamforming methods; traditional speech enhancement methods; and blind system identification and equalization methods, where the acoustic impulse response of a room is identified blindly and then used to design an equalization filter [2]. These methods can also be categorized into single or multichannel approaches. This paper focuses on the latter methods, which generally result in significant performance gains compared to single channel approaches [2].

Estimation of a room impulse response or acoustic transfer function (ATF) is difficult to obtain accurately [2] as it depends on the geometry, the furniture in the room and materials used in the construction of the room. In contrast multichannel algorithms that do not rely on the ATF but instead are based on Linear Predictive Coding have been shown to be highly successful [3-5], including the Spatiotemporal Averaging Method for Enhancement of Reverberant Speech (SMERSH) [2] [6] that was designed for a Uniform Linear Array (ULA) of microphones. In this paper, a new multichannel dereverberation method based on adapting the SMERSH algorithm for an Acoustic Vector Sensor (AVS) is proposed. The AVS has three co-located velocity gradient microphones and one omni-direction microphone arranged orthogonally in an area occupying no more than 1cm^3 [7]. The use of gradient sensors allows for precise recording of directional sound component and has been shown to be affective for speech enhancement in additive noise and mildly reverberant environments [8-10] and here it is shown that adapting SMERSH to the AVS provides a superior technique for dereverberation in highly reverberant environments.

The SMERSH algorithm has three main stages: Delay and Sum Beamforming (DSB); Multichannel LP; and enhancing of the LP residual based on larynx signal modeling. The DSB stage relies on accurate Time Difference of Arrival (TDOA) to perform inter channel time alignment and becomes unreliable when the reverberation time increases above 0.18s [11]. In contrast, the approximate collocation of the microphones of the AVS avoids the necessity of time alignment and an alternative approach based on the Griffiths and Jim Generalized Side Lobe Canceller (GSC) beamformer is proposed [12, 13]. The final stage of SMERSH relies on an accurate LP model. Previous work has shown the gradient sensors of the AVS to provide a significant advantage in the accuracy of the LP model [7], and here a multichannel LP implementation for an AVS is derived.

This paper is organized as follows: Section 2 describes the proposed method for dereverbration; Section 3 presents experiments investigating the performance of dereverbration. Conclusions are presented in section 4.

2. DEREVERBERATION BASED ON AN AVS

The system diagram of the proposed approach based on SMERSH is shown in Fig. 1. Dereverberation is achieved through processing of the LP residual and is based on the first stages of: multichannel LP analysis; beamforming; and a multichannel implementation of the Dynamic Programming Projected Phase-Slope Algorithm (DYPSA) for glottal closure instance modeling. The main modifications to SMERSH for the AVS are for the multichannel LP and beamforming stages.

2.1 Multichannel LP of Reverberant AVS recordings

Signals recorded in a reverberant environment can be modeled as multiple sources, arriving from different directions with varying energies and delays. In moderately reverberant environments signal with the highest energy is normally the direct component arriving at the receiver first, followed by the early and late reflections. For the AVS the recorded signals can be expressed as:

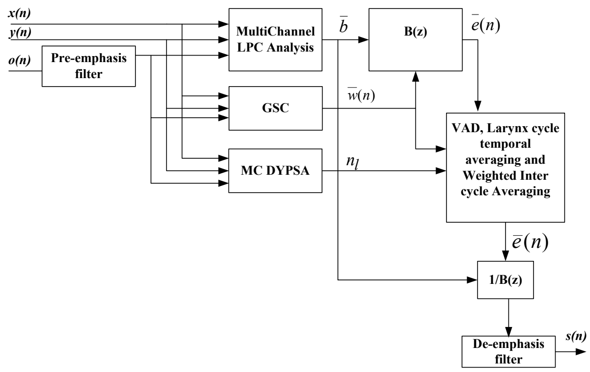


Figure 1: The block Diagram of the Proposed SMERSH algorithm

$$x(n) = g_x \cos(\theta_d)\, s_d(n,\theta_d) \otimes h_d(n,\theta_d) + \sum_{\theta_{r\neq}\theta_d} g_x \cos(\theta_r)\, s_r(n,\theta_r) \otimes h_r(n,\theta_r) \quad (1)$$

$$y(n) = g_x \cos(\theta_d)\, s_d(n,\theta_d) \otimes h_d(n,\theta_d) + \sum_{\theta_{r\neq}\theta_d} g_x \cos(\theta_r)\, s_r(n,\theta_r) \otimes h_r(n,\theta_r) \quad (2)$$

$$o(n) = g_o s_d(n,\theta_d) \otimes h_d(n,\theta_d) + \sum_{\theta_{r\neq}\theta_d} g_o s_r(n,\theta_r) \otimes h_r(n,\theta_r) \quad (3)$$

where $x(n), y(n)$ and $o(n)$ are the AVS channels, $s_d(n,\theta_d)$, and $s_r(n,\theta_r)$ are the direct component and the reflected components at the AVS, $h_d(n,\theta_d)$ and $h_r(n,\theta_d)$ are the ATFs of the direct and reflected components, respectively, and $g_{x,y}\cos(\theta_d)$ and $g_{x,y}\cos(\theta_r)$ are the gain of the gradient sensors and g_o is the gain of the omni directional sensor of the AVS. Here $\otimes$ is the convolution operator. The work presented here is only based on dereverberation hence the noise terms in the (1), (2) and (3) have been omitted. In this work, the AVS was restricted to the 2D space and hence only two gradient components are included in the analysis.

The pressure gradient sensors in the AVS produce a direct representation of the particle velocity [14] the frequency responses of these microphones are different to that of the omni-directional microphone which is a direct representation of the pressure at the array [15]. The frequency responses of the two gradient sensors have a high-pass effect [15]. This high-pass effect can be assumed to be similar to the pre-emphasis filter which is required in applications such as linear prediction of speech. Hence, the pressure gradient sensors of the AVS can be assumed to introduce pre-emphasis like effect. When using the output from the omni-directional sensor with the outputs from the gradient sensors, the output from the omni-directional sensor is pre-emphasised such that the three channels have a similar frequency response. The pre-emphasis is performed according to [16]:

$$o(n) = o(n) - 0.96 \times o(n-1) \quad (4)$$

After the processing of the AVS channels with the proposed SMERSH algorithms the outputs are de-emphasised according to [16]:

$$s(n) = s(n) + 0.96 \times s(n-1) \quad (5)$$

where $s(n)$ is the output from the proposed SMERSH algorithm.

In the original SMERSH algorithm, multichannel LP is achieved based on the average autocorrelation method. Here, this is derived for the AVS. The linear prediction can be expressed as[2]:

$$\mathbf{X}_{o,x,y}(n) = \sum_{i=1}^{P} \mathbf{a}_{o,x,y,i}\mathbf{X}_{o,x,y}(n-i) + e_{o,x,y}(n) \quad (6)$$

here $\mathbf{X}_{o,x,y}(n) = [x(n), y(n), o(n)]^{\mathrm{T}}$ are the sampled signals of (1),(2) and (3), $e_{o,x,y}(n)$ is the LPC residual obtained from AVS channel, P is the prediction order and $\mathbf{a}_{o,x,y}$ is expressed as [2]:

$$\mathbf{a}_{o,x,y} = \boldsymbol{R}_{o,x,y}^{-1}\boldsymbol{r}_{o,x,y} \quad (7)$$

where

$$\mathbf{a}_{o,x,y} = [a_{o,x,y,1}\ a_{o,x,y,2} \cdots a_{o,x,y,p}] \quad (8)$$

and $\boldsymbol{R}$ and $\boldsymbol{r}$ are the autocorrelation matrix and first column of the autocorrelation matrix defined as [2]:

$$\boldsymbol{R}_{o,x,y} = \begin{bmatrix} r_{o,x,y,0} & r_{o,x,y,1} & \cdots & r_{o,x,y,p-1} \\ r_{o,x,y,1} & r_{o,x,y,0} & \cdots & r_{o,x,y,p-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{o.x,y,p-1} & r_{o,x,y,p-2} & \cdots & r_{o,x,y,0} \end{bmatrix} \quad (9)$$

The averaged autocorrelation function for the M channels of AVS can be described as [2]:

$$\widehat{\boldsymbol{R}} = \frac{1}{M}\sum_{m=1}^{M} R_m \quad (10)$$

and

$$\widehat{\boldsymbol{r}} = \frac{1}{m}\sum_{m=1}^{M} \boldsymbol{r}_m \quad (11)$$

where $\widehat{\boldsymbol{R}}$ and $\widehat{\boldsymbol{r}}$ are, respectively the $p \times p$ average autocorrelation matrix and $p \times 1$ average autocorrelation vector. By replacing the $\boldsymbol{R}$ and $\boldsymbol{r}$ in (7) with $\widehat{\boldsymbol{R}}$ and $\widehat{\boldsymbol{r}}$ from (10) and (11) the average LPC coefficient for the AVS channels is obtained.

2.2 Adaptive Beamforming

The spatiotemporal averaging of the speech signals is performed in the original SMERSH algorithm using a DSB, which requires time-alignment of the recorded signals prior to summation. However, time alignment is not require for the AVS as all the microphones in the array are approximately co-located [7]. This has two significant implications for SMERSH algorithm: 1) the need to apply GCC PHAT to time-align the signals is no longer required, hence the processing complexity is reduced; and 2) elimination of potential errors due to inaccuracies in determining the TDOA.

The effect of reverberation on the residual signal is that in the absence of direct component in a section of speech the residual is primarily due to the reverberant components[4]. The system of Fig. 1 includes voiced / unvoiced detection and Larynx cycle temporal averaging based on the residual signal. The accuracy of these measurements is based on the effectiveness of the DSB in reducing the amount of reverberation in the speech signal. Hence, in this work, the DSB is replaced with the more robust Griffiths and Jim Generalized Side Lobe Canceller (GSC) beamformer [12, 13]. Theoretically, when the acoustic transfer function is known, the GSC is able to completely remove the reverberation [17, 18]. The beamforming operation is divided in to three main parts: a fixed beamformer; blocking matrix; and an adaptive filter. The fixed beamformer is normally a DSB but in this work a summing beamformer is used as all the microphones are co- located. The blocking matrix is a rejection filter that blocks the desired signal and passes the interference and noise signals. The adaptive filter suppresses the outputs from the blocking matrix based on the feedback from the output of the beamformer. The delayed signal from fixed beamformer is then subtracted from the output of the adaptive filter. One of the drawbacks of this beamformer is leaking of the signal from the blocking matrix. Several solutions have been proposed to limit signal leakage [13]. Here the improved version of the GJ beamformer described in [13] is implemented for beamforming the AVS outputs.

3. RESULTS

This section describes experiments to measure the accuracy of the multichannel LP modelling, reverberation in the recorded signals and dereverberation performance of the proposed algorithm.

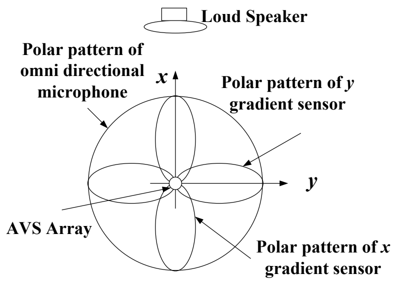


Figure 2**:** Experimental setup for Reverberant recordings

3.1 Experimental Setup

The experimental setup for the AVS and ULA are shown in Fig. 2. The AVS and the ULA are placed in front of the loud speaker such that the array is at zero degrees in azimuth to the loudspeaker as shown in Fig. 2. The recordings used in this Section are made in a room 4 m wide by 12 m long by 3 m high with concrete walls and only two doors. The ceiling and the floor are also concrete with minimal furniture such as chairs and a table. The RT_{60} for the room was found to be 3 seconds. The database of speech sources described in [19] was used. Recordings were made with the loudspeaker 1,2,3,4 and 5 m from the array. The microphone array is fixed such that the loud speakers are at zero degrees azimuth to the array as shown in Fig. 2. Recordings were sampled at 48 kHz and down-sampled to 16 kHz. The recordings were made with the AVS array of [7] and a ULA with 3 microphones with a separation of 21mm as described in [7]. The 3 microphone ULA was chosen to be comparable to the AVS array.

3.2 LP Spectral Characteristics of AVS outputs

The most widely used measure for the distortions between spectral envelopes between the two speech signals is the Itakura Saito Distance (ISD) [20]. It has been shown that the ISD can be used as an indicator for the subjective quality of speech. In [21], an enhanced version of the Itakura distance is presented, and it has been reported that if the ISD is less than 0.5 the difference MOS score is less than 1.6. The ISD between two signals can be expressed as [20, 22]:

$$ISD = \left\| \frac{f(\omega)}{\hat{f}(\omega)} - \log \frac{f(\omega)}{\hat{f}(\omega)} - 1 \right\| \tag{12}$$

where $f(\omega)$ is the spectral density of the original speech signal and $\hat{f}(\omega)$ is the spectral density corresponding to the test signal.

Fig 3 compares the ISD measure for signals recorded using an AVS and ULA from the database of Section 3.1. From the results it is seen that ISD following LP applied to a single omni-directional microphone recording is 4.3 dB at 1m, compared to 2.6 dB, 1.3dB for the x and y gradient microphones and 1.2 dB for the output from the multi-channel LP approach of Section 2.1. A similar trend is seen as the distance between the AVS or ULA to the source increases. This shows that the gradient channels of the AVS produce a much more accurate LP spectrum compared to that of the ULA and the spectrum of the multi-channel LP obtained from the AVS is much closer to the spectrum of the original clean speech signal. Hence, it is expected that the processing of the AVS channels with the proposed SMERSH algorithm will produce much better results than the ULA.

3.3 Measuring the Amount of Reverberation

There are two approaches to measure dereverberation: a channel based measure (Direct to Reverberant Ratio (DRR)) and a signal based measure (Signal to Reverberant Ratio (SRR)). The former is based on the reverberating system impulse response and hence is suitable for measuring DRR when the system impulse response is known or can be calculated [23]. The SRR is an approach used when the effect of the dereverberation algorithm cannot be characterized in terms of impulse response [23]. In [23], it was shown that, with the correct normalization, SRR is equivalent to DRR. In this work, SRR will be used as a measure of reverberation for an objective measure and is expressed as:

$$SRR = 10\log\frac{\|s(n)\|^2}{\|s_r(n)\|^2} \tag{13}$$

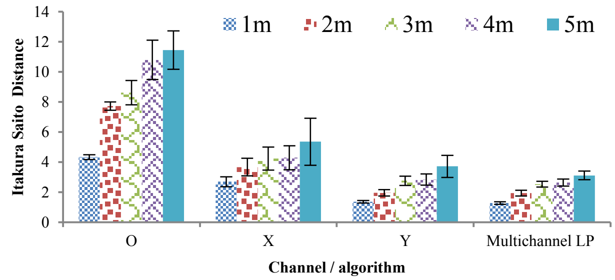
where $s_r(n)$ is a delayed version of the source signal $s(n)$. The SRR gives an indication of the amount of reverberation but it does not indicate if the filtering has reduced the perceptual quality. The method used for perceptual evaluation is to use a panel of listeners who rate the quality on a pre defined scale such as Mean Opinion Score (MOS). In [24] MOS tests were used to determine the effect of coloration and reverberation decay tail. A objective measure know as the Speech to Reverberation Modulation Energy Ratio (SRMR), that gives an accurate representation of subjective listening test based on the effects of coloration, reverberation tail effects and the overall quality and intelligibility is presented in [25, 26]. The SRMR is expressed as [25]:

$$SRMR = \frac{\sum_{k=1}^{4} \bar{\varepsilon}_k}{\sum_{k=5}^{K^*} \bar{\varepsilon}_k} \tag{14}$$

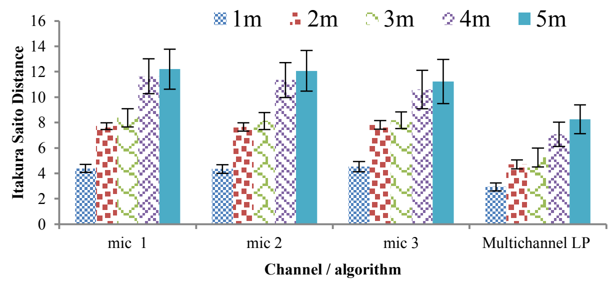
where $\bar{\varepsilon}_k$is the average modulation energy for the k^{th} modulation filter and K^* in this work is set at 8. In this work SRMR will be used to evaluate the performance of the proposed algorithm in addition to the SRR.

3.4 Results from SRR Measures

The dereverberation method described in Section 2 and the original SMERSH algorithms were applied to the recordings of the AVS and the ULA. In Figs 4,5,6 and 7, original is the clean speech signal, unfiltered is the recording made in the reverberant room before processing, SMERSH is the output from the original SMERSH algorithm and proposed is the output from the proposed algorithm. In the case of the ULA the original SMERSH algorithm is tested against the proposed algorithm with only the GSC



a)



b)

Figure 3: Itakura Saito Distance measure for a) AVS b) ULA

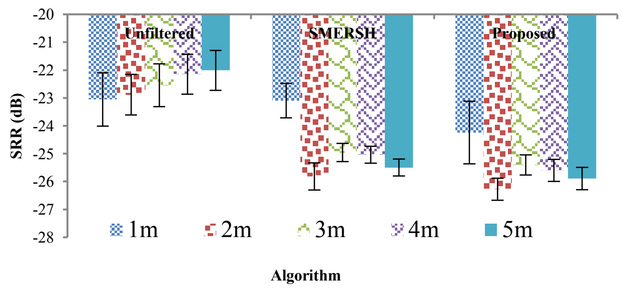


Figure 4: SRR results for the ULA array

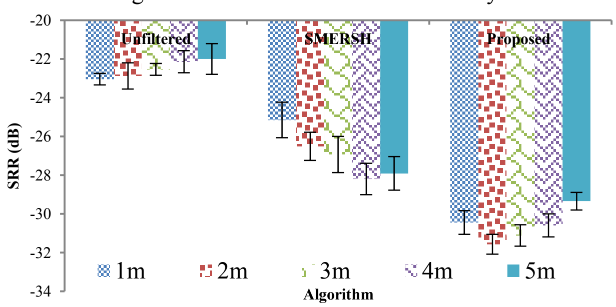


Figure 5: SRR results for the AVS array

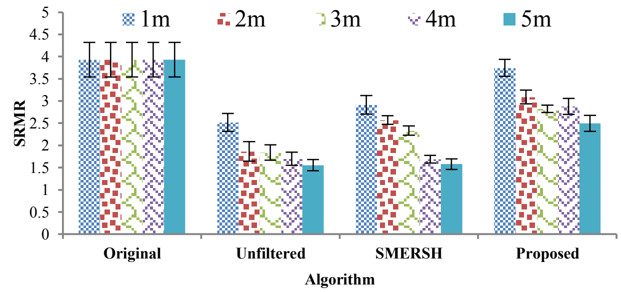


Figure 6: SRMR for AVS array

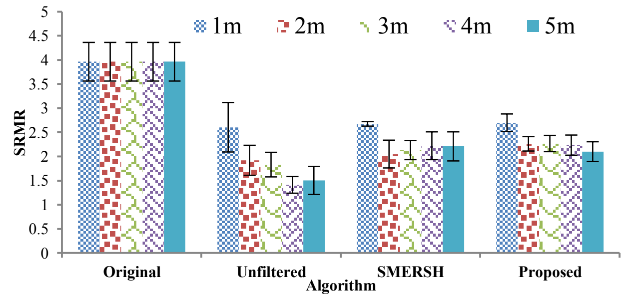


Figure 7: SRMR for ULA array

changed. Since the microphones in the ULA are spatially located GCC PHAT has to be used in order to get the TDOA estimates as described before. The result for SRR of the dereverberation using the proposed method is presented in Fig. 4 and Fig. 5 for the ULA and the AVS respectively. From the results it can be seen the proposed method performs better than the original SMERSH algorithm for both arrays.

The results for the ULA are of special importance as these results show the effect of using a more robust beamformer inside the SMERSH algorithm. When the original SMERSH algorithm is applied to the ULA the difference in SRR between the unprocessed and processed recording at 1m is 0.03 dB, and the performance improves as the distance is increased to 3.5 dB at 5m. When the proposed beamformer is introduced into the SMERSH algorithm improvement in terms of the difference in SRR between the processed and unprocessed is 1.1 dB for 1m and the difference increases to 3.8 dB for 5m; hence by introducing the beamformer there is a 10 fold improvement in the results at 1m. The poor results in the ULA are due to the inaccuracy in determining the TDOA at close proximity to the source [7]. In [7] it was shown that when a the ULA was used for DOA estimation which is based on the accurate estimation of TDOA, the ULA produces higher errors in DOA estimation at 1m.

The results for the AVS (Fig. 5) show that, on average, the proposed method has an improvement in terms of difference in SRR between the unprocessed recordings to the output of the proposed algorithm of 7 dB for separations of 1m to 5m. For the original SMERSH algorithm the improvement in SRR at 1m is 2 dB and the difference in SRR increases to 5 dB at 5m for the AVS. These results show that the accurate estimation of the TDOA and LP coefficients in combination with the GSC beamformer improve the performance of the SMERSH algorithm considerably.

3.5 Results from SRMR Measures

The results for the SRMR for the AVS and the ULA are shown in Fig. 6 and 7. From the results it can be seen that the original signal has a SRMR of 3.9 while the unfiltered signal has a SRMR of 2.4 a difference of 1.4 at 1m. The results of the AVS show that when the SMERSH algorithm is applied to the AVS output at 1m, a SRMR value of 2.9 is obtained with a difference of 1. The results show that when the proposed algorithm is applied to the AVS output the SRMR value is 3.74 which is difference of 0.18. The SRMR values for both SMERSH and the proposed algorithm decreases as the distance increase and for 5m the SMERSH algorithm and the unfiltered have the same SRMR values. But the proposed algorithms have a SRMR value of 2.4 which is better than the SMERSH and the unfiltered. For the ULA when the SMERSH with GSC is applied there is an improvement of 0.1 over the unfiltered at 1m and increases to 0.6 at 5m. In the case of the original SMERSH algorithm a similar result is obtained with SRMR of 0.1 at 1m increasing to 0.7 at 5m. Since the differences are too close it can be concluded that there is no significant improvement in terms of SRMR measure for both the algorithms for the ULA. A possible cause of these errors is due to miss alignment of speech due to in accuracy of estimated TDOA which was discussed in section 3.4.

4. CONCLUSIONS

The results presented in this paper have shown that by using an AVS array and introducing a robust beamformer to the well known SMERSH algorithm, effective dereverberation of speech sources can be achieved. From the results, it can be seen that by applying the proposed changes to the SMERSH algorithm it is possible to dereverberate speech in adverse conditions (RT_{60} > 2s) without requiring knowledge of the ATF. These results have also shown that multichannel recordings from a co-located array can be successfully used for speech dereverberation. The results show that there is an improvement of 7dB for a source at 1m in SRR for the proposed algorithm. Furthermore, SRMR test have shown that the performance of the proposed algorithm is comparable to a state of the art dereverberation algorithm with a SRMR score of 3.74 for 1m and 2.4 for 5m.

Acknowledgement: This project was partially supported by the Australian Research Council Grants DP0772004 and DP1094053.The authors also thank Emanuel Habets for providing the code that was used in this work for calculating the SRR.

5. REFERENCES

[1] K. Kinoshita and T. Nakatani, "Speech Dereverberation Using Linear Prediction," *NTT Technical Review 9,* No: 7, 2011.

[2] P. A. Naylor and N. D. Gaubitch (Eds), *Speech Dereverberation*, Springer, 2010.

[3] N. D. Gaubitch, P. A. Naylor and D. B. Ward, "On the use of linear prediction for dereverberation of speech," In *proc. of International Workshop on Acoustic Echo and Noise Control (IWAENC2003)*, Kyoto, Japan, pp. 99–102, 2003.

[4] B. Yegnanarayana and P. Murthy, "Enhancement of reverberant speech using LP residual signal," *Speech and Audio Processing, IEEE Transactions on,* vol. 8, pp. 267-281, 2000.

[5] B. Yegnanarayana, S. R. Prasanna and K. Sreenivaasa Rao, "Speech enhancement using excitation source information," *Acoustics, Speech, and Signal Processing (ICASSP), IEEE International Conference on*, pp. I-541, 2002.

[6] M. Thomas, N. D. Gaubitch, J. Gudnason and P. A. Naylor, "A practical multichannel dereverberation algorithm using multichannel DYPSA and spatiotemporal averaging," *Applications of Signal Processing to Audio and Acoustics*, *IEEE Workshop on,* New Paltz, NY, pp. 50-53, 2007.

[7] M. Shujau, C. H. Ritz and I. S. Burnett, "Designing Acoustic Vector Sensors for localisation of sound sources in air," *17th European Signal Processing Conference (EUSIPCO 2009)*, Glasgow, Scotland, 2009.

[8] H. Jwu-Sheng and Y. Chia-Hsing, "Adaptive signal blocking for generalized sidelobe canceller using matched filter array," *Acoustics, Speech and Signal Processing. ICASSP, IEEE International Conference on*, pp. 2597-2600, 2008.

[9] M. E. Lockwood and D. L. Jones, "Beamformer performance with acoustic vector sensors in air," *The Journal of the Acoustical Society of America,* vol. 119, pp. 608-619,2006.

[10] M. Shujau, C. H. Ritz and I. S. Burnett, "Linear Predictive Perceptual Filtering For Acoustic Vector Sensors: Exploiting Directional Recordings For High Quality Speech Enhancement," *Acoustic Speech and Signal Processing, (ICASSP)*, *IEEE International Conference on*, Praque, pp. 5068-5071,2011.

[11] B. Champagne, S. Bédard and A. stéphenne, "Performance of time-delay estimation in the presence of room reverberation," *Speech and Audio Processing, IEEE Transactions on,* vol. 4, pp. 148-152,1996.

[12] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *Antennas and Propagation, IEEE Transactions on,* vol. 30, pp. 27-34,1982.

[13] O. Hoshuyama, A. Sugiyama and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *Signal Processing, IEEE Transactions on,* vol. 47, pp. 2677-2684,1999.

[14] D. Davis and E. Patronis, *Sound system engineering*, Focal Press, 2006.

[15] J. Eargle, *The microphone book*, Focal Press, 2004.

[16] J.R.Deller, J. G. Proakis and HL. Hansen, *Discrete-Time Processing Of Speech Signals*, New York, Macmillan Publishing Company, 1993.

[17] E. Habets, J. Benesty, I. Cohen, S. Gannot and J. Dmochowski, "New Insights Into the MVDR Beamformer in Room Acoustics," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 18, no. 1, pp. 158-170, 2010.

[18] J. Benesty, J. Cheng , Y. A. Huang and J. Dmochowski "On Microphone-Array Beamforming From a MIMO Acoustic Signal Processing Perspective," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 15, no. 3, pp. 1053-1065,2007.

[19] M. Shujau, C. H. Ritz and I. S. Burnett, , "Seperation of Speech Sources Using An Acoustic Vector Sensor," *Multimedia Signal Processing (MMSP), IEEE 13th Internatinal Workshop on*, Hanzhou, pp 1-6,2011.

[20] J. Benesty, M. M. Sondhi and Y. A. Huang (Eds) *Springer Handbook of Speech Processing*, Springer, 2007.

[21] G. Chen, S. N. Koh and I. Y. Soon, "Enhanced Itakura measure incorporating masking properties of human auditory system," *Signal Processing,* vol. 83, no. 7, pp. 1445-1456, 2003.

[22] R. Gray,A. Buzo, A. Gray Jr. and Y. Matsuyama, "Distortion measures for speech processing," *Acoustics, Speech and Signal Processing, IEEE Transactions on,* vol. 28, no.4, pp. 367-376,1980.

[23] P. A. Naylor, N. D. Gaubitch, and E. A. P. Habets, "Signal-based performance evaluation of dereverberation algorithms," *Journal of Electrical and Computer Engineering,* p. 1, 2010.

[24] J. Y. C. Wen, N. D. Gaubitch, E. A. P. Hebets, T. Myatt and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," 2006.

[25] T. H. Falk, C. Zheng and W. Chang, "A Non-Intrusive Quality and Intelligibility Measure of Reverberant and Dereverberated Speech," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 18, no. 7, pp. 1766-1774, 2010.

[26] M. Jeub, M. Schafer, T. Esch and P. Vary, "Model-based dereverberation preserving binaural cues," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 18, no. 7, pp. 1732-1745, 2010.