31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

A Multichannel Speech Enhancement Method For Functional MRI Systems Using A Distributed Microphone Array

Ali A. Milani†, Govind Kannan†, Issa M.S. Panahi†, Richard Briggs‡

***Abstract*— Multichannel speech enhancement has been shown to be an effective method to decrease speech distortion introduced during speech enhancement, especially in environments like MRI (magnetic resonance imaging) which have a distributed noise source. However, these methods suffer from high computational complexity which makes them almost impractical. The use of subband filtering has been suggested to reduce this complexity but the performance of the existing subband methods deteriorate as the number of subbands increases. In this paper we introduce a new multichannel speech enhancement algorithm based on subband adaptive filtering that works for higher number of subbands at a lower complexity. The real-world experiments demonstrate the performance of the new scheme in an MRI room.**

I. INTRODUCTION

Functional MRI (fMRI) is used for clinical purposes to study the haemodynamic response of the human brain and the spinal cord. During the scan, the subject responds to visual and audio cues by speaking over a microphone. The very high level of distributed acoustic noise generated during the scan by the three perpendicular gradient coils namely X, Y and Z, together with the noise generated by the ancillary equipments like helium fan, helium pump and air conditioning system, overwhelms the subject's speech and reduces the intelligibility of subject's speech response. This, disrupts the speech communication between the patient and researcher and is seen as a major impediment in fMRI research. fMRI acoustic noise (Fig. 1) is a wideband signal whose spectrum spans a broad range of frequencies with several strong high frequency harmonics with an eigenvalue spread of the order of 10^5. The special characteristics of the noise and environment impairs the performance of the conventional adaptive filtering methods and needs more suitable noise cancelation techniques.

The conventional approach for speech enhancement has been single channel method in which the speech and background noise are recorded concurrently by a single microphone. In this approach, the suppression of background noise also results in distortion in the desired speech signal [1]. To reduce this effect, multichannel speech enhancement techniques with microphone arrays have been suggested [2]. However the associated computational complexity remains a major challenge. The computational complexity in multichannel methods increases drastically with the number of input channels which indeed limits the number of usable channels. This computational burden can be reduced by using subband adaptive filtering.

Subband adaptive filtering can be exploited to cope up with adaptive filtering challenges for the distributed wideband fMRI acoustic noise [3]. This technique decomposes the signal spectrum into several subbands, allowing each subband to be handled separately. Therefore, the subband dynamic spectral range and eventually the subband eigenvalue spread decreases. This reduction, leads to more stability and robustness for adaptive filters.

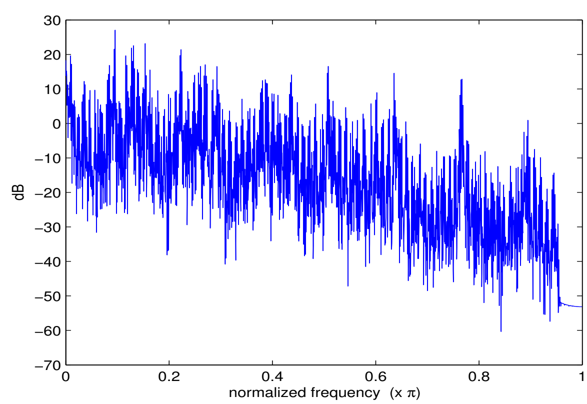


Fig. 1. The spectrum of fMRI acoustic noise with $f_s = 16kHz$.

In this paper we present a new multichannel subband adaptive filtering based technique extended from [4] that can be used in distributed microphone arrays. In this approach the output of each microphone is used to estimate the noise in the reference microphone which is close to the speaker and records the speech. The weights of all filters are computed and updated jointly in subbands. The inherent decimation and interpolation operations reduce the corresponding computational complexity. The proposed algorithm unlike similar approaches [5],[6] works very well at high number of subbands, due to which the computational complexity drastically reduces. The paper is organized as follows. Section II describes the problem and signal model. The proposed multichannel subband filtering algorithm is introduced in section III. Experimental results are presented in section IV and section V concludes the paper.

This work was supported in part by a subcontract from UT Southwestern Medical Center at Dallas, funded by the Department of Veterans Affairs through VA IDIQ contract number VA549-P-0027 awarded and administered by the Dallas, TX VA Medical Center. The content of this paper does not necessarily reflect the position or the policy of the Veterans Administration or the Federal government, and no official endorsement should be inferred.

Contacts: ali.a.milani@student.utdallas.edu, govind@student.utdallas.edu, issa.panahi@utdallas.edu; phone:972-883-4120; fax:972-883-2710.

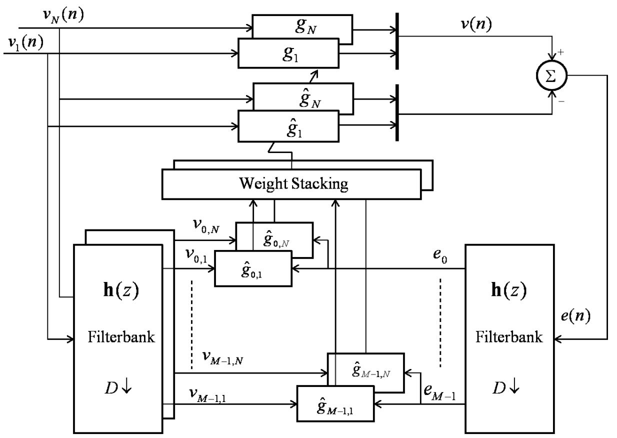


Fig. 2. Delayless multichannel subband adaptive speech enhancement structure.

II. Problem Description and Signal Model

In this paper we assume that there is a reference microphone close to the speaker whose recorded signal $y_r(n)$ has the speech plus the background noise $\nu(n)$ i.e.

$$y_r(n) = s(n) + \nu(n) \tag{1}$$

There are N other microphones, with unknown geometry, around the reference microphone that record the background noise $\nu_k(n)$. We assume that the noise signals are not coherent. $\nu(n)$ can be modeled as

$$\nu(n) = \sum_{k=1}^{N} \nu_k(n) * g_k \tag{2}$$

where $*$ stands for convolution, $\nu_k(n)$ is the signal recorded by the microphones and g_k is the acoustical signal transmission path. In order to estimate $s(n)$ we need to estimate the filters $\hat{g}_k$ of length L_g such that

$$\|e(n)\|^2 = \left\| y_r(n) - \sum_{k=1}^{N} \mathrm{v}_k^T(n)\hat{\mathrm{g}}_k \right\|^2 \tag{3}$$

is minimized where

$$\begin{aligned} \mathrm{v}_k(n) &= [\nu_k(n), \nu_k(n-1), \ldots, \nu_k(n-L_g+1)]^T \\ \hat{\mathrm{g}}_k &= [\hat{g}_{k,1}, \hat{g}_{k,2}, \ldots, \hat{g}_{k,L_g}]^T \end{aligned} \tag{4}$$

the summation in (3) can be written as

$$\|e(n)\|^2 = \left\| y_r(n) - \mathbf{v}^T(n)\hat{\mathbf{g}} \right\|^2 \tag{5}$$

where

$$\begin{aligned} \mathbf{v}(n) &= [\mathrm{v}_1^T(n), \mathrm{v}_2^T(n), \ldots, \mathrm{v}_N^T(n)]^T \\ \hat{\mathbf{g}} &= [\hat{\mathrm{g}}_1^T, \hat{\mathrm{g}}_2^T, \ldots, \hat{\mathrm{g}}_N^T]^T \end{aligned} \tag{6}$$

(5) can be minimized using a least mean square adaptive filter e.g. NLMS of length NL_g. In this case all filters $\hat{g}_k$ are jointly computed and recursively updated. Since the speech signal is highly non-stationary and impairs the weight adaptation, a voice activity detector (VAD) [7] may be used to halt the adaptation while speech is present. The impulse response g_k is usually long. Consequently, a large L_g results in a significant computational burden especially in the multichannel case. This can be resolved by employing subband adaptive filtering. The delayless subband adaptive filtering method employed in this paper is the DSAF algorithm developed in [4]. This algorithm is shown [4] to be superior in performance to other subband adaptive algorithms [5],[8].

III. Multichannel Subband Filtering

The delayless subband adaptive filtering (DSAF) algorithm involves the decomposition of the input noise (i.e. the reference signal) and error signal into subbands by analysis filter banks. The adaptive filters are computed in subbands and combined into a full band filter using a synthesis filter bank. Fig. 2 shows the delayless structured [4],[5] subband adaptive filtering scheme used in our distributed speech enhancement algorithm. In this method, all $\nu_k(n)$ for $1 \le k \le N$ and $e(n)$ are filtered into M subbands named $\nu_{k,m}(n)$ for $0 \le m \le M-1$ and $e_m(n)$ using the analysis filter bank $\mathbf{h}(z)$ with the decimation factor of D. $\mathbf{h}(z)$ is given by

$$\mathbf{h}(z) = [H_0(z), H_1(z), \ldots, H_{M-1}(z)]^T \tag{7}$$

where, $H_m(z)$ is transfer function of mth analysis bandpass filter with linear phase and length L_p. In each subband, (6) is computed using an adaptive algorithm such as NLMS i.e.

$$\hat{\mathbf{g}}_m(n+1) = \hat{\mathbf{g}}_m(n) + \mu \frac{\mathbf{v}_m^*(n)e_m(n)}{\epsilon + \|\mathbf{v}_m(n)\|^2} \tag{8}$$

where

$$\begin{aligned} \mathbf{v}_m(n) &= [\mathrm{v}_{m,1}^T(n), \mathrm{v}_{m,2}^T(n), \ldots, \mathrm{v}_{m,N}^T(n)]^T \\ \hat{\mathbf{g}}_m(n) &= [\hat{\mathrm{g}}_{m,1}^T, \hat{\mathrm{g}}_{m,2}^T, \ldots, \hat{\mathrm{g}}_{m,N}^T]^T \end{aligned} \tag{9}$$

After weights are updated, each $\hat{\mathrm{g}}_k$ is reconstructed from $\{\hat{\mathrm{g}}_{0,k}, \hat{\mathrm{g}}_{1,k}, \ldots, \hat{\mathrm{g}}_{M-1,k}\}$ using a weight stacking method [4],[9].

A. Computational Complexity

The computational complexity of the subband filtering technique [4] is calculated by counting the number of real multiplications required for the following operations (a) decimation by $D = M/4$, (b) calculation of the $N(M/2+1)$ filter bank outputs, (c) operations for calculating updating $N(M/2+1)$ adaptive filters, (d) N weight stacking, and (e) calculation of all $\hat{\mathrm{g}}_k$.

Table I summarizes the computational complexity per input sample for the operations listed above. The total complexity is plotted in Fig. 3 versus the number of subbands M for NLMS for N=1, 2, and 4. As shown, the complexity reduces exponentially with the number of subbands. Since the method in [4] compared to those in [5] and [6] has lesser delay in its analysis filter banks, it works better with the higher number of subbands and results in lesser complexity.

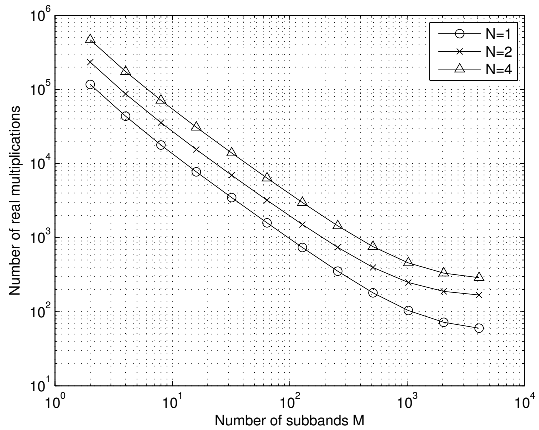


Fig. 3. Computational complexity for different number of subbands and for $N = 1, 2$ and 4.

Fig. 4. Real fMRI experiment used to record speech plus noise with a Siemens 3-Tesla MRI system. The targeted application is speech enhancement during fMRI experiment. $N = 2$.

IV. Experimental Results

To demonstrate the performance of the proposed algorithm we have recorded and compared speech enhancement in an MRI environment. As shown in Fig. 4 in a typical fMRI experiment i.e., 40 slices per 2 seconds of EPI (echo planar imaging) sequences, speech was recorded. In order to record the acoustic noise generated during the experiment, 6 reference microphones were used i.e., 3 inside and 3 outside the MRI bore. Fig.1 shows the spectrum of the recorded fMRI acoustic noise using a Siemens 3-Tesla MRI system. Since 8 kHz bandwidth is enough for speech, the noise was recorded at 64 kHz sampling rate and down sampled to 16 kHz to avoid any spectral aliasing. Each recording has 5 seconds of pure noise without speech followed by 5 seconds of speech plus noise. The first 5 seconds was used to estimate $\hat{g}_k$.

TABLE I

Computational complexity for the proposed method in terms of number of real multiplications per input sample.

Algorithms	Number of real multiplications
filter bank	$(N+1)4\log_2 M$
NLMS	$N(1+\frac{2}{M})(\frac{8L_g}{M})$
PFFT-1	$N\frac{4L_g}{M}[(2+\frac{4}{M})(\log_2\frac{4L_g}{M})+\log_2 L_g]$
PFFT-2	$N\frac{8L_g}{M}[(2+\frac{4}{M})(\log_2\frac{8L_g}{M})+\log_2 2L_g]$

Fig. 5 presents the effectiveness of the proposed algorithm in 4 different experimental setups. Fig.5(a) and Fig.5(b) show the spectrograms of the speech enhanced by the regular NLMS algorithm and proposed method when only one reference microphone inside the MRI bore is used. The filter length for NLMS and subband method that give the best results were 256 and 1024 respectively. The plots indicate the effectiveness of the proposed algorithm in the suppression of the background noise especially the harmonics of the fMRI acoustic noise [3].

Fig.5(c) shows the spectrogram of the enhanced speech for the proposed method when 3 reference microphones outside the MRI bore are used. Fig.5(d) shows the same plot when all 6 reference microphones are used. Comparing Fig.5(c) and Fig.5(d), we notice that the harmonics of the MRI acoustic noise are more suppressed in the second setup especially around 2kHz (highlighted with the rectangle) in which the speech components are more enhanced. The experiments show that the closer the reference microphones to the noise sources, the better the noise cancelation is.

V. Conclusion

In this paper a new multichannel subband adaptive filtering algorithm was introduced and tested in a distributed microphone array for speech enhancement. The proposed subband adaptive filtering algorithm enjoys lower delay in its analysis filter bank which enables it to operate in high number of subbands with much lower complexity. The resulting distributed algorithm does not need any information about the array geometry. An improvement in noise reduction is observed when the number of reference microphones close to noise sources increases.

References

[1] P. Loizou, *Speech Enhancement: Theory and Practice*. CRC Press, Boca Raton: FL, 2007.

[2] A. Khong, J. Benesty, and P. Naylor, "Sterophonic acoustic echo cancellation: Analysis of misalignment in the frequency domain," *IEEE transaction on audio, speech and language processing*, vol. 16, no. 5, pp. 957–968, Jul. 2008.

[3] A. A. Milani, I. Panahi, and R. Briggs, "LMS-based active noise cancellation methods for fMRI using sub-band filtering," in *Proceedings of 28th Annual IEEE EMBC conference*, New york, USA, Aug. 2006.

[4] A. A. Milani, I. M. Panahi, and P. Loizou, "A new delayless subband adaptive filtering algorithm for active noise control systems," *IEEE Trans. on audio, speech and language process.*, vol. 17, no. 5, Jul 2009.

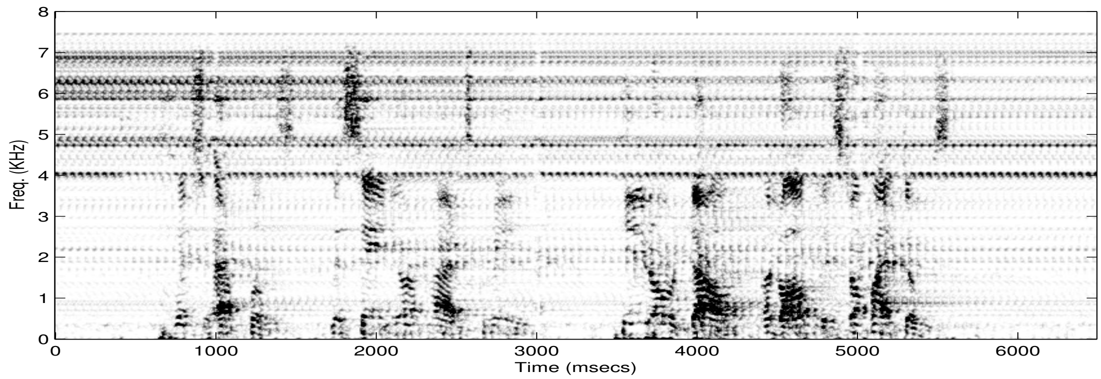
[5] D. Morgan and J. C. Thi, "A delayless subband adaptive filter architecture," *IEEE Transaction on Signal Processing*, vol. 43, no. 8, pp. 1819–1830, Aug. 1995.

[6] R. Merched and A. H. Sayed, "An embedding approach to frequency-domain and subband adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 48, no. 9, pp. 2607–2619, Sept. 2000.

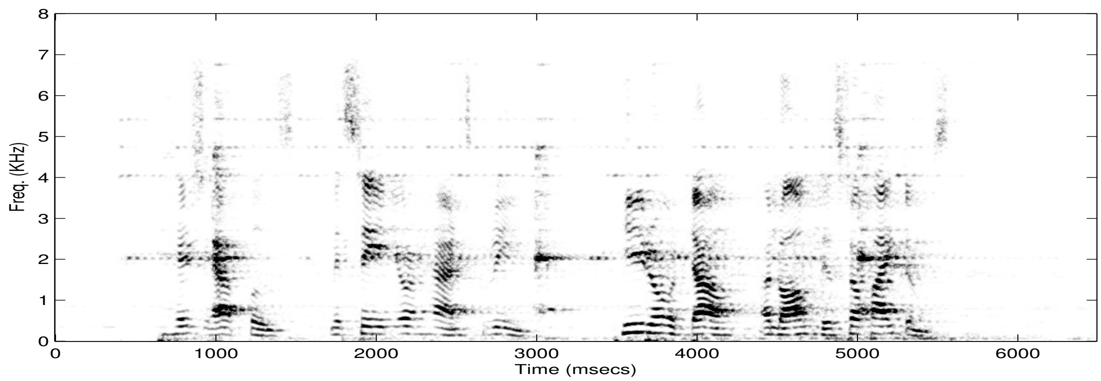
[7] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *Signal Processing Letters, IEEE*, vol. 6, no. 1, pp. 1–3, Jan 1999.

[8] A. H. Sayed, *Fundamentals of adaptive filtering*. New York: Wiley, 2003.

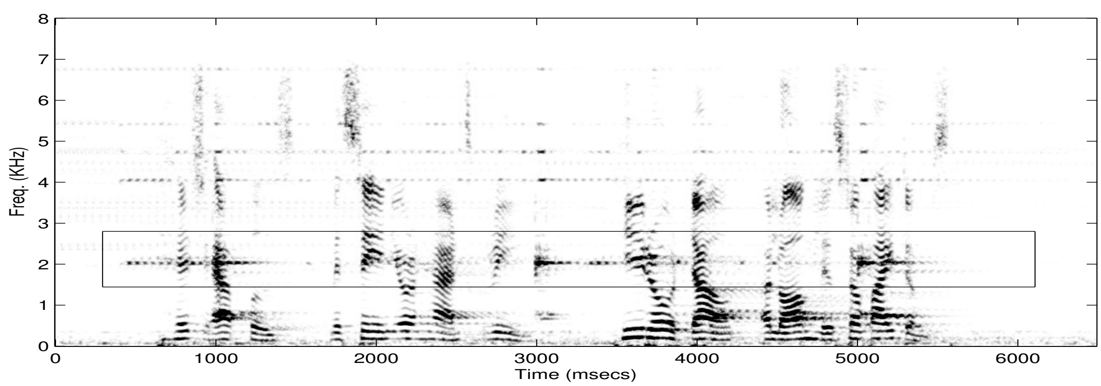
[9] J. Huo, S. Nordholm, and Z. Zang, "New weight transform schemes for delayless subband adaptive filtering," in *Proceedings of GLOBECOM'01*, New york, USA, Nov. 2001.



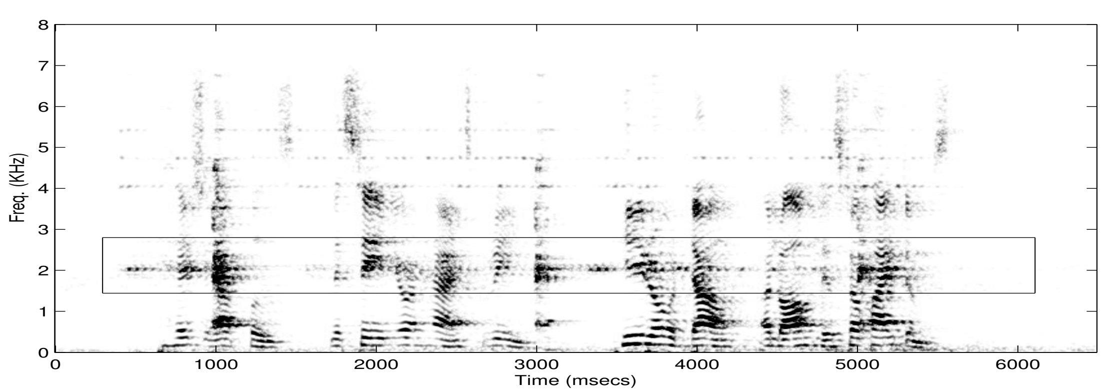
(a)



(b)



(c)



(d)

Fig. 5. Noisy speech and the enhanced speech, (a) non-subband NLMS, with 1 reference microphone inside, and proposed subband adaptive filtering with (b) 1 reference microphone inside, (c) 3 reference microphones outside, and (d) all 6 reference microphones inside and outside the MRI bore.