SPEECH ENHANCEMENT IN FUNCTIONAL MRI ENVIRONMENT USING ADAPTIVE SUB-BAND ALGORITHMS

Venkat R. Ramachandran, Govind Kannan, Ali A. Milani, Issa M. S. Panahi

Department of Electrical Engineering, The University of Texas at Dallas, Richardson, USA

ABSTRACT

Very high level of acoustic noise in fMRI scanner rooms disrupts speech communication between the subject and the physician/researcher. Enhancing speech in such an environment is challenging due to the broadband and dynamic nature of the noise. Sub-band adaptive methods prove to be very effective in cancelling such noise. In this paper we present the results of using sub-band adaptive methods for enhancing speech corrupted by noise from a 3-Tesla fMRI scanner. We also observe that the performance depends on the synthesis filter bank structure.

Index Terms— functional magnetic resonance imaging, Speech enhancement, sub-band algorithms

1. INTRODUCTION

Functional MRI (fMRI) is an important tool for investigating the human brain function. During fMRI operations, researchers (in a control room) usually communicate verbally with the subject (in the scanner room) to give instructions and to monitor performance in languages or in other cognitive tasks. However, very high acoustic noise (noise levels greater than 130 dB SPL) generated by the scanner and the associated equipments overwhelms the subject's speech (as heard in the control room) and interferes with diagnosis and investigation [1]. This masking of speech is one of the impediments in current fMRI research. The background noise component must be considerably reduced to ensure reliable speech quality. In [2], we introduced a multi-band adaptive algorithm to address this issue of speech enhancement. Though effective in reducing the fMRI noise, the multi-band method suffers from considerable computational complexity when it comes to real-time implementation. It also introduces some distortion to the speech signal. Our recent investigations [4] have revealed that sub-band based adaptive algorithms are effective in addressing these two drawbacks and also give a superior performance. They are better suited for attenuating broadband noise like the fMRI acoustic noise. In this paper we report the results of using sub-band based methods to the fMRI speech enhancement problem. We also compare the performance of two sub-band methods ,with different synthesis structures , in terms of their effectiveness in enhancing the speech signal.

The organization of the paper is as follows. In Section 2 we introduce the general two-channel speech enhancement setup and qualitatively describe the effectiveness of sub-band adaptive filters. Section 3 is devoted to a more detailed description of the sub-band adaptive filter architectures. Section 4 describes the data acquisition setup used. Section 5 reports the results of simulation. Section 6 concludes the paper.

2. TWO-CHANNEL SPEECH ENHANCEMENT

Fig.1 shows the generic two channel adaptive speech enhancement method. This method assumes that the speech signal, $s(n)$ is additively corrupted by the noise, $b'(n)$. The adaptive filter, $W(z)$ tries to cancel $b'(n)$ by estimating it from a reference signal $b(n)$. $e(n)$ is the recovered speech signal. Various algorithms exist for updating the co-efficients (weights) of the adaptive filter $W(z)$ .The choice of the algorithms depends on application specific requirements like computational complexity, convergence rate, spectral characteristics etc. The most common weight updating algorithm is the least mean square (LMS) algorithm. However the transitions in signal energy due to the presence of voiced (high energy), unvoiced and silent regions (low energy) in speech lead to instability when using the conventional LMS algorithm. The solution is using the Normalized LMS (NLMS) algorithm where the normalization of step size leads to better stability.

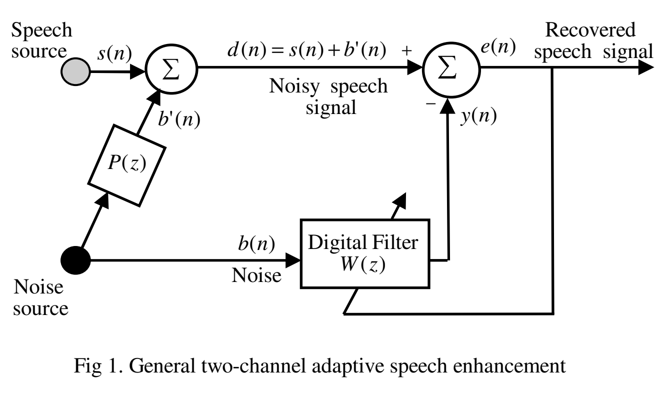

Fig 1. General two-channel adaptive speech enhancement

Effective cancellation of fMRI noise poses two challenges:

1. fMRI acoustic noise is inherently broadband (Fig.2), the estimation of which requires very high order $W(z)$. This leads to increased number of computations.
2. The fMRI acoustic noise has a highly dynamic spectrum which causes slow convergence.

The sub-band architecture effectively reduces the computational burden by down-sampling. Moreover the dynamic spectral range that needs to be tracked is smaller in each sub-band which improves the convergence speed.

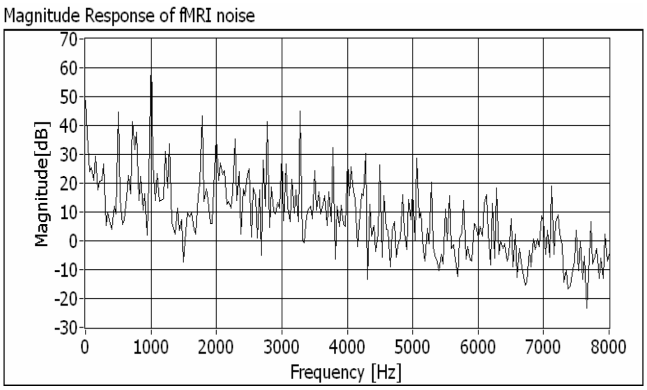

Fig 2. Magnitude response of the fMRI acoustic noise

3. SUB-BAND BASED ADAPTIVE SPEECH ENHANCEMENT

Sub-band adaptive filter structure for realizing $W(z)$ is based on the Delayless sub-band adaptive filter architecture introduced in [3]. The filter structure is shown in Fig.3. The reference $x(n)$ and the error signal $e(n)$ are filtered into M sub-bands using a uniform analysis filter bank that spans the entire bandwidth as shown in Fig. 4. In each sub-band, the signal is decimated by a factor $M/2$ and the NLMS algorithm computes the L sub-band filter weights for each sub-band. The details of the multi-rate filter bank implementation can be found in [3], [4] . The sub-band filter weights are transformed to the frequency domain and then appropriately stacked to construct the main full-band noise canceling filter. This is equivalent to passing the sub-band filter weights through a synthesis filter bank. We consider two weight transform and stacking methods which have different performance characteristics. The two methods are referred to as *FFT Method* and *FFT-2 Method.*

3.1. FFT Method

In this method the L sub-band filter weights are transformed into frequency domain using a L- point FFT. The full-band weights $W(l)$ (as DFT coefficients) are derived from the

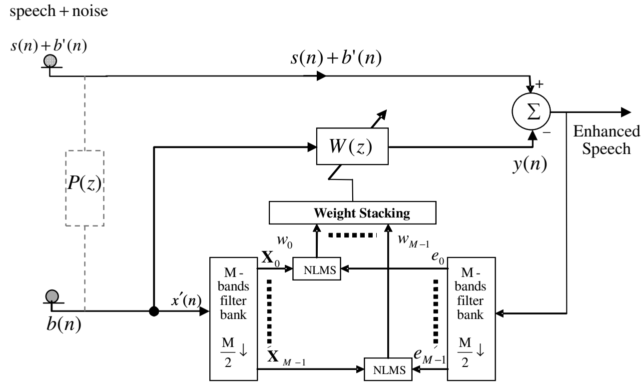

Fig 3. Two-channel speech enhancement with sub-band based adaptive filter

constituent sub-band weights $H_k(l)$ according to the following rules.

(I) for $l \in [0, N/2)$,

$$W(l) = H_{\lfloor lM/N \rfloor}\left(((l))_{2N/M}\right)$$

(II) for $l = N/2$, $W(N/2) = 0$

(III) for $l \in (N/2, N)$, $W(l) = W(N-l)^*$

where N is the length of the full-band adaptive filter $W(z)$.

$W(l)$ and $H_k(l)$ are the l th DFT coefficient of the full-band filter and the k-th sub-band filter respectively. $\lfloor . \rfloor$ denotes rounding to the nearest integer, superscript * denotes complex conjugation and $(a)_b$ denotes 'a modulus b'. The full-band filter co-efficients can be obtained by a N-point IFFT operation on the DFT coefficients , $W(l)$'s

In the case of FFT method, we observe nulls in the frequency response of the full-band filter at frequencies $(2l+1)\pi/N$ as seen in Fig.5. This degrades the performance of the speech enhancement system. To avoid the pass-band nulls, we can use the slightly more computationally complex FFT -2 method.

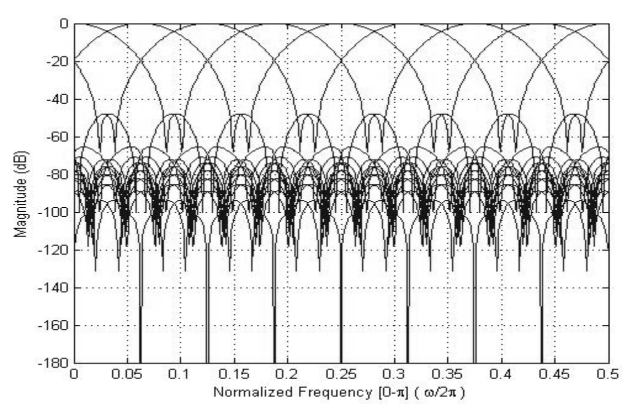

Fig 4. Frequency response of the analysis filter banks for 16 sub-bands

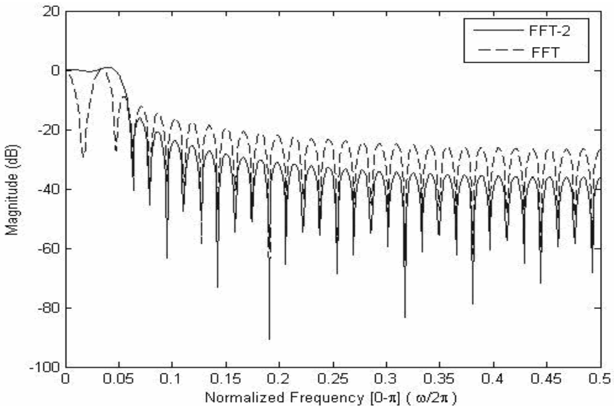

Fig 5. Frequency response of the first synthesis sub-band filter in the two stacking methods. There are no pass-band nulls in the FFT-2 method.

3.2. FFT -2 Method

In this method the L sub-band filter weights are transformed into frequency domain using a $2L$ - point FFT. The details of the stacking method are in [5]. The full-band weights $W(l)$ are derived as follows

(I) for $l \in [0, N)$,
$$W(l) = H_{\lfloor lM/2N \rfloor}\left((l)_{4N/M}\right)$$

(II) for $l = N$, $W(N) = 0$

(III) for $l \in (N, 2N)$, $W(l) = W(2N-l)^{*}$

where $W(l)$ and $H_k(l)$ are the l th DFT coefficient of the full-band filter and the k-th sub-band filter respectively. $\lfloor . \rfloor$ denotes rounding to the nearest integer and superscript * denotes complex conjugation. The full-band filter co-efficients can be obtained by $2N$-point IFFT operation on $W(l)$ and discarding the last N points

From the above steps, it can be seen that the FFT-2 method requires more computations due to the $2L$ - point FFT and $2N$ - point IFFT operations.

4. ACQUISITION OF ACOUSTICAL FMRI NOISE

The fMRI noise was recorded from a 3 Tesla Siemens Magnetom Trio. Fig.6. shows the experimental setup. The acoustic signal was recorded using a diffuse-field microphone that had an omnidirectional response and a good dynamic range. The microphone output was amplified using a preamplifier and conducted through 10 meters of shielded BNC cable to a custom bias voltage supply. Two minute segments of the analog data were digitized at 16 KHz using National Instruments PCI 4472 A/D board. LabVIEW 8.0 was used to control the data acquisition. Fig.7. shows the recorded fMRI acoustic noise.

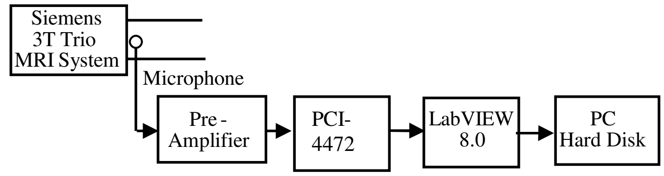

Fig 6. LabVIEW data acquisition setup

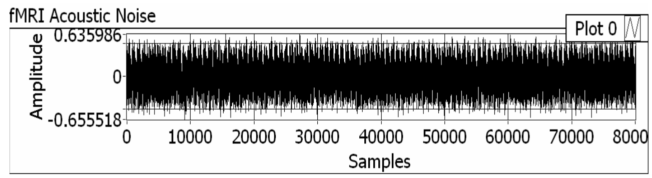

Fig 7. fMRI acoustic noise

5. SIMULATION AND RESULTS

To evaluate the performance of the two sub-band methods, 10 sentences (5 uttered by male speakers and 5 uttered by female speakers) from Noisy Speech Corpus (NOIZEUS) database were used. The original speech waveforms, sampled at 25 kHz, were down sampled to 16 kHz and used for our simulations. The recorded fMRI acoustic noise modified by a transfer function $P(z)$ was added to the speech signals. For $P(z)$, we have used an actual acoustic transfer function which is a 25-pole IIR filter given in [6]. The time waveforms of the clean speech signal and noisy speech signal are shown in Fig.8 and Fig.9 respectively. The magnitude response of $P(z)$ is shown in Fig.10.

The Signal to Noise Ratio (SNR) of the noisy speech signal was set at -10 dB. SNR is calculated as follows,

$$SNR = 10\log(S/E) \text{ dB} \quad (1)$$

where S is the clean speech signal power and E is the noise power.

In the simulation, an adaptive filter of length 512 was used. The adaptive filter was initially trained with 32000 samples (2 seconds) of noise before adding speech. NLMS algorithm was used to update the filter weights. A step-size of 0.5 was used during training phase to ensure fast convergence and reduced to 0.001 when speech was added. The time domain results of speech enhancement using FFT and FFT-2 methods for the 8 sub-band case are shown in Fig.11 and Fig.12 respectively. The noise level can be seen to reduce considerably in both methods.

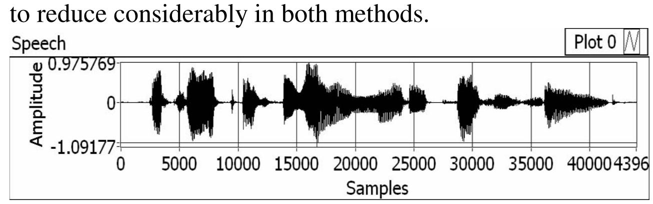

Fig 8. Clean speech signal

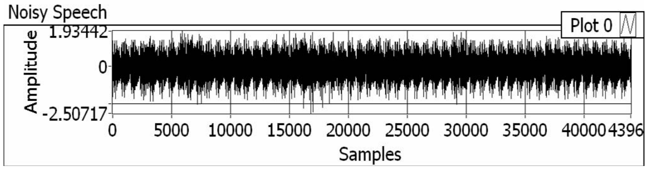

Fig 9. Noisy speech signal of -10 dB SNR

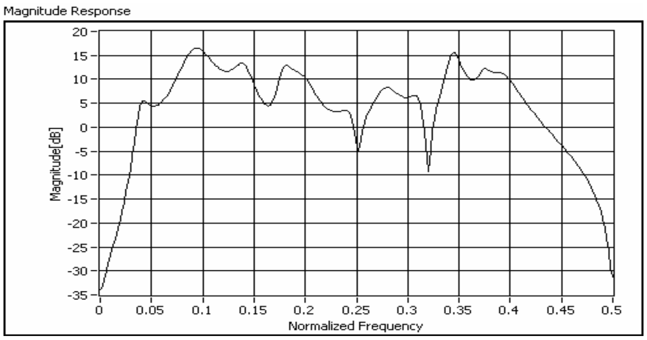

Fig 10. Magnitude Response of *P(z)*

The performance of the algorithms was assessed by measuring the improvement in SNR defined as

SNR improvement = SNR(Enhanced speech) – SNR(Noisy speech)

In calculating the SNR for the enhanced speech, *E* in (1) now represents the power of difference between the clean and the enhanced speech signals. The initial convergence period of 2 seconds was discarded when calculating the SNR. The SNR improvement was calculated for all the 10 sentences. Average SNR improvement and the relative computational complexity are tabulated in Table 1.

# of sub-bands	SNR improvement for FFT method (dB)	SNR improvement for FFT-2 method (dB)	Relative Computational Complexity α
2	35.7	35.7	2.3
4	32.0	32.7	1.59
8	36.1	41.96	0.64
16	29.26	33.38	0.31
32	28.63	35.91	0.17

Table.1 Sub-band filtering – Performance Comparison

$$\alpha = \frac{\text{computation time using sub-band}}{\text{computation time with no sub-band}}$$

We observe very good noise cancellation in both the methods. The best performance is obtained for 8-sub-bands. The computational complexity decreases exponentially for increasing number of sub-bands. α is averaged over FFT and FFT-2 since negligible differences exist in their computational complexities. FFT-2 method outperforms FFT method in terms of SNR improvement due to absence of pass-band nulls as mentioned earlier at the cost of a slightly higher computation time.

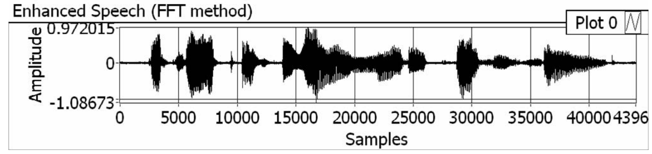

Fig 11. Recovered speech signal for 8 sub-bands with FFT method

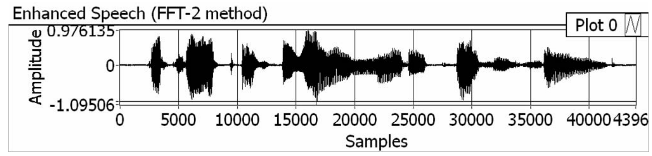

Fig 12. Recovered speech signal for 8 sub-bands with FFT-2 method

6. CONCLUSION

We have shown that sub-band adaptive filters are very effective for fMRI speech enhancement problem. We also establish the effect of weight stacking methods on the performance. FFT-2 method is observed to provide superior performance. Real-time implementation of the method is currently under progress.

7. ACKNOWLEDGMENT

Supported by a subcontract from the Epidemiology Division, Department of Internal Medicine, UT Southwestern Medical Center at Dallas under grant no. DAMD17-01-1-0741 from the U.S. Army Medical Research and Material Command. The content of this paper does not necessarily reflect the position or the policy of the U.S. government, and no official endorsement should be inferred.

8. REFERENCES

[1] M. Ravicz, J. Melcher, and N. Kiang "Acoustic Noise during Functional Magnetic Resonance Imaging", Journal of Acoustic Society of America, 108(4), October 2000.

[2] V. Ramachandran, I. M. S. Panahi, et al, "Multi-band Speech Enhancement for Functional MRI" ,ICASSP-2006 at Toulouse, France.

[3] D. Morgan and J. Thi, "A Delayless Subband Adaptive Filter Architecture", IEEE Transactions on Signal Processing, Vol. 43, No. 8, August 1995.

[4] A. Milani, I.M.S. Panahi, R. Briggs, "LMS-based Active Noise Cancellation Methods for fMRI using Sub-band Filtering", 28th Annual Inter. Conf, IEEE EMBS, New York, Aug.-Sept., 2006.

[5] J. Huo, S. Nordholm and Z. Zang, "New Weight Transform Schemes for Delayless Subband Adaptive Filtering," Global Telecommunications Conference, GLOBECOM'01, vol.1, pp. 197 – 201, 2001

[6] S. M. Kuo and D. R. Morgan, "Active Noise Control Systems - Algorithms and DSP Implementations", John Wiley & Sons, Inc. 1996.