Multi-Band Speech Enhancement for Functional MRI

V. Ramachandran(1), I. M. S. Panahi(1), Y. Hu(1), P. C. Loizou(1), R. W. Briggs(2), S. R. McCaslin(3)
Department of Electrical Engineering, The University of Texas at Dallas, Richardson, Texas(1)
Department of Radiology, The University of Texas Southwestern Medical Center, Dallas, Texas(2)
National Instruments Inc., Austin, Texas(3)

ABSTRACT

We present an adaptive enhancement method for speech signal corrupted by the 3 Tesla (3T) functional Magnetic Resonance Imaging (fMRI) acoustic noise. The noisy speech signal is filtered into several frequency bands and noise cancellation is applied to each band independently. We use Normalized Least Mean Squares (NLMS) algorithm for updating the co-efficients of the adaptive filters in both wideband and in multi-band implementations. The NLMS algorithm is used to ensure faster convergence and stability. The multi-band method shows a better performance than the conventional wideband method resulting in a better noise reduction. It is shown that the distortion introduced by the band-pass filters limits the overall speech enhancement when the number of bands increases beyond an optimal value.

1. INTRODUCTION

MRI instruments allow noninvasive mapping of structure and function in the intact human body and are used regularly for diagnostic and research studies. During fMRI scans, researchers often need to communicate with the subject to give instructions and to monitor performance in language or other cognitive tasks. However, the very high (≥130 dB) acoustic noise, generated by the MRI scanner makes such communication difficult [1]. This background acoustic noise has to be removed or reduced for reliable communication. The noisy signal can be acquired using a single sensor (microphone) or multiple sensors (microphone array). In speech enhancement, multi-microphone methods give a better Signal to Noise Ratio (SNR) than a single microphone setup due to the availability of more information.

In this paper, we present two-channel speech enhancement. The subject's speech and background acoustic noise produced by the scanner are recorded simultaneously. The acoustic fMRI noise is wide-band non-stationary signal whose amplitude, frequency, phase, and propagation velocity vary with time. Therefore the speech enhancement system must be adaptive to be able to adjust to these variations. Adaptive speech enhancement using a sub-band approach for Gaussian noise has been studied in [2]. Acoustic control of noise has been discussed in [3–5]. The multi-band filtering approach taken in this paper decomposes the wideband input signal into an optimum number of band-limited signals and applies NLMS algorithm to obtain improved speech enhancement. Improvement in the speech quality using fixed multi-band filters [6] has motivated our work. We analyze this approach and adapt it to obtain an optimum number of multi-band filters for the speech enhancement in a 3T fMRI scanner system.

Section 2 gives a brief description of the method for recording fMRI noise in real-time. In Section 3, we present the multi-band method for speech enhancement. Section 4 shows the simulation results that were obtained for the proposed method with an optimum number of multi-band filters.

2. ACQUISITION OF fMRI NOISE

We recorded the fMRI noise from a Siemens 3T Magnetom Trio. The block diagram of the experimental setup for data acquisition is shown in Fig.1. A diffuse-field microphone (designed to have flat response when signal arrives simultaneously from all the directions) was used for measurement of the acoustic noise. A pre-amplifier amplifies the microphone outputs and the signal was conducted through 10 meters of shielded BNC cable to custom bias voltage power supplies located in the control room. The diffusion type microphone used a 12V power supply. The cable shields were tied to the power supply ground. One-minute segments of the amplified microphone output were digitized at 16 kHz with a National Instruments PCI 4472 A/D board and streamed to the hard disk of a Windows XP computer running LabVIEW 7.0.

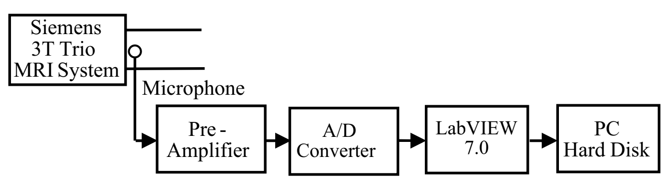

Fig. 1. Experimental setup for fMRI noise acquisition

3. MULTI-BAND FILTERING

Fig. 2 shows the block diagram for general two-channel enhancement. The desired speech signal *s(n)* is assumed to be present in only one channel corrupted by the background noise *b'(n)*.The second channel has the reference noise signal *b(n)*. The adaptive filter tries to model the transfer function, *P(z)* between the two inputs. The filter output *y(n)* becomes an estimate of only the noise present in *d(n)* and the output *e(n)* becomes an estimate of *s(n)*.

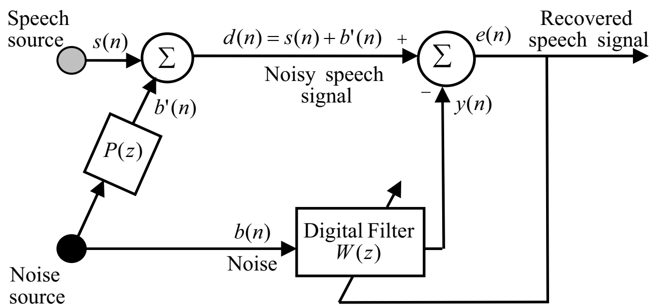

Fig. 2. Two-channel speech enhancement

Fig. 3 shows the block diagram of modified speech enhancement using multi-band filtering. Band-pass filters with non-overlapping frequency responses $\{H_m, m = 1, 2, \ldots, M\}$ partition the noisy speech signal and the reference noise signal into band-limited signals $\{d_m\}$ and $\{b_m\}$ respectively. The adaptive filter $W_m(z)$ tries to model the differential transfer function between d_m and b_m. Thus the filter output, y_m becomes an estimate of the noise present in d_m. The sum of all the e_m is the recovered speech signal which is the estimate of $s(n)$. As the number of samples increases, the LMS filters estimate the differential transfer functions better reducing the background noise and thereby improving the speech quality.

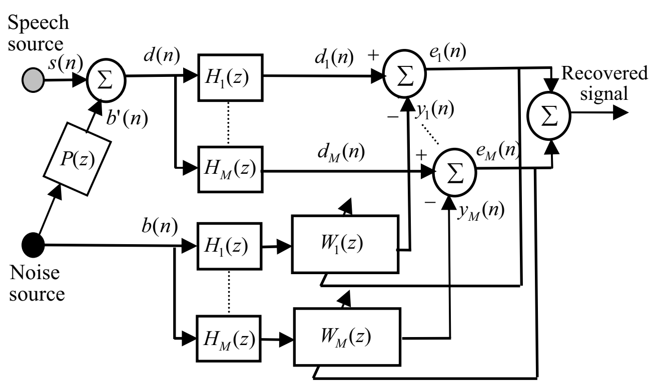

Fig. 3. Two-channel speech enhancement with Multi-banding

Faster convergence is possible in multi-bands because the spectral dynamic range is greatly reduced in each of the frequency band improving the performance for broadband noise signals. Alternatively, the use of multi-band processing for speech enhancement allows diverse processing in each band depending on signal power, noise power and level of coherence between signal and noise in the two channels. The simplest and the most commonly used algorithm for updating filter coefficients is the Least Mean Square (LMS) algorithm proposed by Widrow, et al [7]. A detailed literature study can be found in [8] and [9]. The block diagram of a system that uses LMS algorithm is shown in Fig. 4.

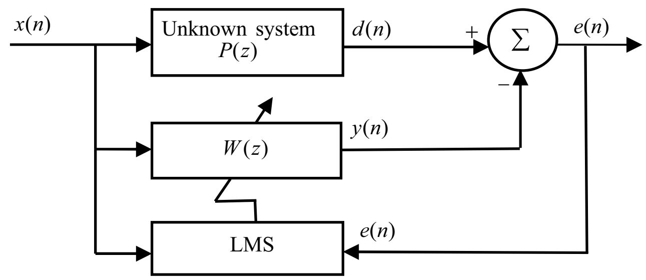

Fig. 4. Block diagram of a system that uses LMS algorithm

In the LMS algorithm, the coefficients of the digital filter $W(z)$ are updated as,

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\, \mathbf{x}(n) e(n) \tag{1}$$

where n is the time index, μ is the step size and $e(n)$ is the error signal and $e(n) = d(n) - [\mathbf{w}^T(n)\mathbf{x}(n)]$.

The choice of the step size, μ determines the rate of convergence and the stability of the filter $W(z)$.A variation of the LMS algorithm is the NLMS algorithm proposed by Bershad [7]. In NLMS algorithm, the step size is normalized by the energy of the data vector. Convergence in NLMS is faster than LMS at very little extra cost. The weight updating case of the NLMS algorithm is given as,

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu\, e(n)\mathbf{x}(n)}{\sum_{i=1}^{N} \mathbf{x}^2(n-i+1)} \tag{2}$$

where N is the length of the adaptive filter.

4. SIMULATIONS

Five sentences uttered by male speakers and five sentences uttered by female speakers taken from the NOIZEUS[1] (noisy speech corpus) database were used to evaluate the proposed method. The original wideband signals sampled at 25 kHz were used in the evaluations, after downsampling them to 16 kHz. T[1]he fMRI noise (recorded from the 3T Siemens Trio) modified by the transfer function $P(z)$ is added to the speech. The frequency spectrum of the fMRI noise, sampled at 16 kHz, is shown in Fig. 5.

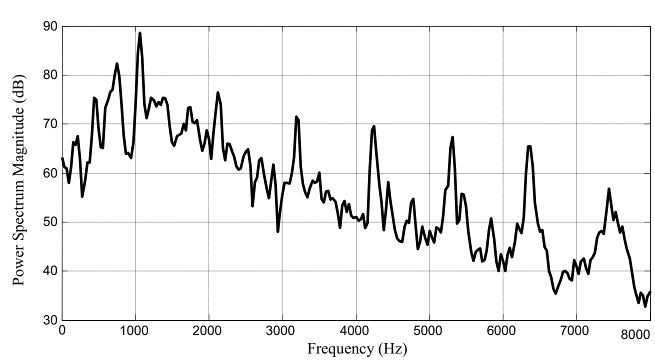

Fig. 5. Frequency spectrum of fMRI noise

1 Available at : http://www.utdallas.edu/~loizou/speech/noizeus

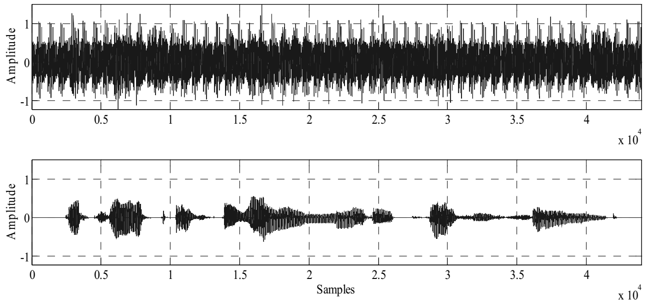

Fig. 6. From the top, noisy speech signal of -10 dB SNR and clean speech signal

The Signal to Noise Ratio (SNR) is calculated as follows,

$$SNR = 10\ log(S/E) \text{ dB} \quad (3)$$

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

The SNR of the noisy speech signal is chosen as -10 dB. In the calculation of the SNR using (3), E refers to the fMRI noise power. The time waveforms of the noisy and clean speech signals are shown in Fig. 6. We have used the 25-pole IIR filter given in [5] for the acoustic path model *P(z)*. The frequency response of *P(z)* is shown in Fig. 7.

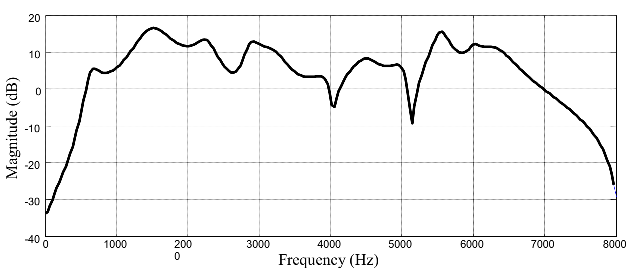

Fig. 7. Frequency response of the primary path, *P(z)*

When there is a transition from low energy (silence or unvoiced) to high energy (voiced) or vice versa in the speech signal, the step size (μ) changes drastically in the case of the conventional LMS algorithm. This can cause the filter to become unstable. So we used an adaptive filter that uses the NLMS algorithm for speech enhancement. In NLMS algorithm, the step size is normalized by the energy of the input speech signal. This makes the filter immune to the fluctuations in the signal power. In the simulation, an FIR filter of length 70 was used to model the transfer function *P(z).* The step size used for updating the filter coefficients was 0.01. The recovered speech signal obtained using the Conventional NLMS (CNLMS) algorithm is shown in Fig. 8.

In calculating (3) for the speech enhancement, E represents power of the difference between the clean and the recovered speech signals. The initial convergence period of 2 seconds (32000 samples) was discarded when calculating the SNR. SNR improvement (in dB) refers to the difference between the SNR of the recovered speech signal and SNR of the original noisy speech signal.

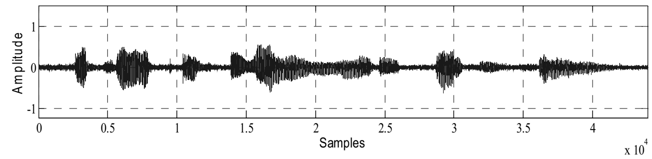

Fig. 8. Recovered speech signal (CNLMS)

The recovered speech signal has an SNR of 10.3 dB i.e., an SNR improvement of 20.3 dB is observed when single band NLMS algorithm is used.

We partitioned the frequency spectrum (ranging from 0 to 8000 Hz) into M bands linearly. In our case, we have chosen M = 2, 4, 7 and 8 bands. The noisy speech signal and the reference noise signal are filtered into M different frequency bands using linear phase FIR filters. All the filters were chosen to have the same order so as to have equal delay in all the frequency bands. Noise reduction was done independently in each of the bands using an FIR adaptive filter of order 70. The recovered speech signals for 2, 4, 7 and 8 bands are shown in Fig. 9. The SNR improvement was calculated for all the 10 sentences. Average SNR improvement obtained for varying number of bands is tabulated in Table 1.

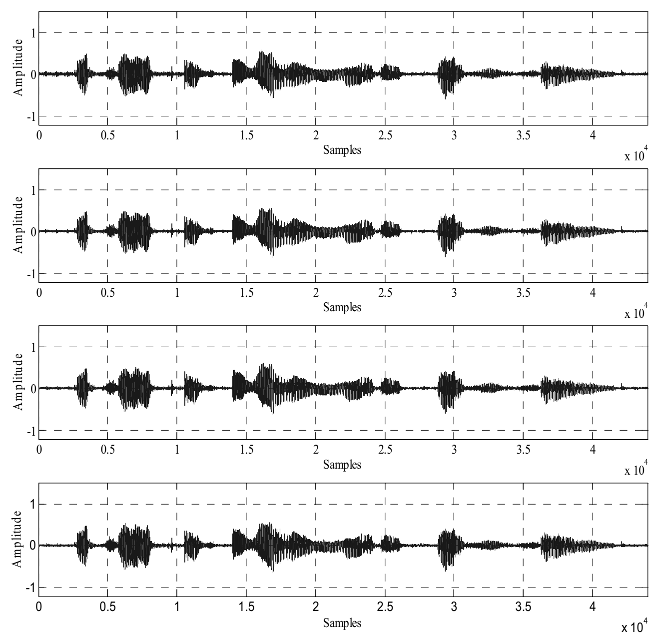

Fig. 9. Recovered speech signals
(From the top, NLMS with 2, 4, 7 and 8-band filters)

Number of bands	Average SNR improvement (dB)
1	21.13
2	23.93
4	25.66
7	22.61
8	22.14

Table 1. Multi-band filtering – Performance comparison

We observe from Table 1 that the SNR does not improve for a higher number of bands. This is because of the distortion introduced by the frequency-partitioning filters. Error noise, E consists of two components; the residual noise and distortion. As the number of filters increases, the distortion increases. This causes the reduction in the SNR.

The adaptive filters were trained using 32000 samples of fMRI noise. Let R denote the variance of the last 1500 samples of the residual noise. To find the distortion variance, the clean speech signal $s(n)$ was filtered using the multi-band filters and the filter outputs were added to get $s'(n)$. The distortion variance, D is calculated as,

$$D = \frac{1}{L} \| s(n) - s'(n) \|^2 \qquad (4)$$

where $\|.\|$ denotes the norm of the signal and L is the number of samples.

Average noise variance, R_{avg} was calculated for 10 segments of fMRI noise and average distortion variance, D_{avg} was calculated for the 10 sentences. Table 2 shows the R_{avg} and the D_{avg} for varying number of bands.

Number of bands	Average noise variance, R_{avg}(X 10^{-4})	Average distortion variance, D_{avg}
1	7.78	0
2	3.62	0.32
4	1.22	0.81
7	1.30	8.90
8	1.11	13.5

Table 2. Residual noise and distortion comparison

From Table 2, we see that the best performance is obtained for M = 4, i.e., both the noise variance and the distortion variance are low compared to the other bands.

5. CONCLUSION

The multi-band adaptive speech enhancement shows a definite improvement over the conventional NLMS method. However, the increase in distortion decreases the overall SNR for a larger number of bands. We find that four linearly-spaced frequency bands are adequate in obtaining good speech quality when it is corrupted by the 3T fMRI-scanner noise.

6. 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 Materiel 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.

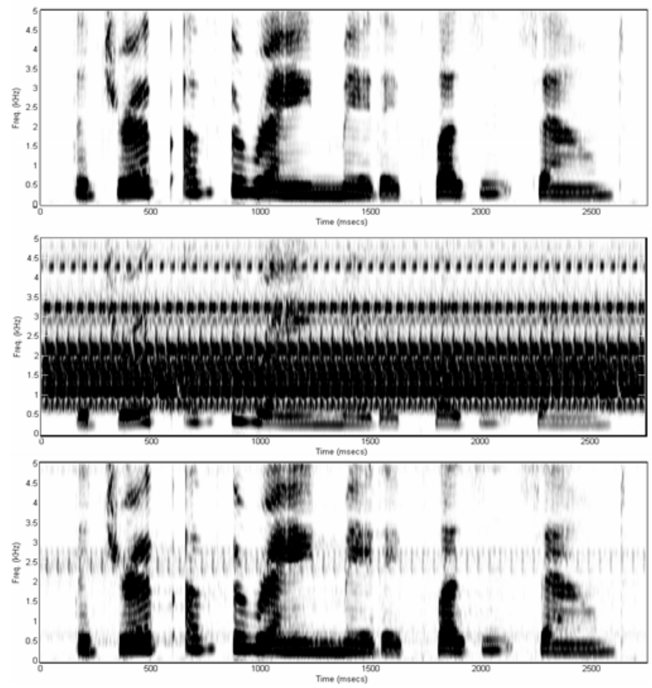

Fig 10. From the top, spectrograms of clean speech, noisy speech and recovered speech (using 4-band NLMS) signals

7. 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] D. J. Darlington and D. R. Campbell, "Sub-Band, Dual-Channel Adaptive Noise Cancellation Using Normalized LMS", IEEE Digital Signal Processing Workshop Proceedings, September 1996.

[3] S. M. Kuo and D. R. Morgan, "Active Noise Control: A Tutorial Review", Proceedings of the IEEE, Vol. 87, No. 6, June 1999.

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

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

[6] S. D. Kamath and P. C. Loizou "A Multi-band Spectral Subtraction Method for Enhancing Speech Corrupted by Colored Noise", Proceedings of ICASSP-2002, Orlando, FL, May 2002.

[7] N. J. Bershad, "Analysis of the Normalized with LMS Algorithm Gaussian Inputs", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 34, No. 4, August 1986.

[8] B. Widrow and S. D. Stearns, "Adaptive Signal Processing", Prentice-Hall, Inc., 1985.

[9] S. Haykin, "Adaptive Filter Theory", Prentice Hall; 4th edition, 2001.