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

Real Time Speech Enhancement for the Noisy MRI Environment

Nishank Pathak, Issa Panahi, *Senior Member, IEEE*, P Devineni and Richard Briggs, *Member, IEEE*

***Abstract*— Performance of two Adaptive (nLMS and Normalized Sign-error LMS) and a single channel (LogMMSE) speech enhancement algorithms are tested on a floating point DSP to reveal their effectiveness in enhancing speech corrupted in noisy MRI environment with very low SNR. The purpose of experiments is to reduce the fatigue of the listener by eliminating the strong MRI noise. The experiments use actual data set collected from a 3-Tesla MRI machine. Results of the experiments and performance of the speech enhancement system are presented in this paper. The speech enhancement system is automated. Our experiments reveal that after enhancement of the speech signal using Sign-Error LMS, the residual noise shows characteristics of white noise in contrast to the residual noise of the other algorithms which is more structured. It is also shown that the Sign-Error LMS offers fast convergence in comparison to the other two methods.**

I. INTRODUCTION

INTERACTION between the patient and the medical staff and recording and filing of the speech data for analysis are very important part of MRI (Magnetic Resonance Imaging) experiments. Modern fMRI (functional MRI) using very high magnetic fields (3Tesla) that interact with rapidly switched magnetic gradients create vibration in the scanner coil structure. This produces up to 130 dB acoustic noise [1]. High noise levels corrupt the speech of the patient and make the interaction between the doctor and the patient difficult [1]. Besides corrupting the speech of the patient, the noise may mask the non verbal cues of the speech.

MRI noise has a structure shown in Fig.5. The MRI noise used in our experiments were recorded by a scanning performed at 40EPI (Echo Planar Imaging) sequences per 2 seconds. The noise has a frame structure of 2 seconds. The whole noise sequence consists of such frames repeated one after another. In between any two frames there is a dead band. The effect of the dead band is discussed in the later sections.

A real-time implementation of a speech enhancement system could be used to enhance the speech of the patient in the MRI control room. Enhanced speech signal can be stored for processing and used for communication between the doctor and the patient. This paper presents and investigates the performance of two adaptive algorithms and a single channel speech enhancement algorithm by designing a real-time DSP implementation of speech enhancement system. The experiments were performed on speech data collected from an actual scanning experiment performed on a subject (Fig.1). The implementation and in-depth analysis of the real time system make this work different from the previous work published in [2][3]. Section II gives a brief description of the algorithms used to update the adaptive filter and the single channel speech enhancement algorithm. Section III gives a description of the hardware setup. Section IV describes the software used. Section V discusses the result.

Fig. 1. Speech Enhancement setup on an MRI machine. The picture shows two microphones. The primary microphone is held by the subject and the reference is connected to the head matrix. The microphones are kept from 15 to 20 cm apart.

II. SPEECH ENHANCEMENT ALGORITHMS

The equations for nLMS(Normalized LMS), nSLMS (Normalized Sign-error LMS, which are used to update the adaptive filter coefficients, and LogMMSE are given below. Fig.2 shows the basic adaptive speech enhancement algorithm setup

A. Normalized Least Mean Square (nLMS)

The adaptive filter coefficients $\hat{\mathbf{w}}(n)$ are calculated and the estimated noise $\hat{d}(n)$ is removed from the noisy speech signal. The error signal which also contains the residual speech is fed back to the adaptive filter.

$$\hat{d}(n) = \hat{\mathbf{w}}(n)\mathbf{u}(n) \tag{1}$$

This study was supported by the 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 U.S. government, and no official endorsement should be inferred

Nishank Pathak is a Graduate Teaching Assistant at the University of Texas at Dallas, Richardson, TX 75080 USA (phone: 214-830-0799; fax: 972-883-2710) nishank.pathak@student.utdallas.edu

I. M. S. Panahi is an Assistant Professor with the Department of Electrical Engineering, University of Texas, Dallas, TX 75080 USA : issa.panahi@ieee.org

R. Briggs works with the Department of Radiology, University of Texas Southwestern Medical Center, Dallas, TX 75390 : richard.briggs@utsouthwestern.edu

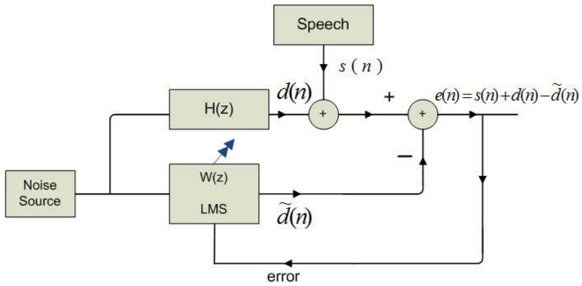


Fig. 2. 2-Channel Speech Enhancement.

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{\mu \mathbf{u}(n) e^*(n)}{||\mathbf{u}(n)||^2 + \varepsilon} \tag{2}$$

Where $\mathbf{u}(n)$ is the input signal from the reference microphone, μ is the step size and ε is the bias factor.

B. Normalized Sign-Error LMS (nSLMS)

NSLMS is a variation of LMS algorithm [4]. Retaining the name of all the parameters in nLMS, the update equation for nSLMS is given by

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{\mu \mathbf{u}(n) sign(e(n))}{||\mathbf{u}(n)||^2 + \varepsilon} \tag{3}$$

C. Log Minimum Mean Square Estimate (LogMMSE)

LogMMSE, which is a single channel speech enhancement method [5], is known to perform good enhancement of speech corrupted by Gaussian white noise. This estimator minimizes the mean-square error of the log-magnitude spectra $E\{(\log(x_k) - \log(\hat{x}_k))^2\}$. The optimal LogMMSE estimator can be obtained by evaluating the conditional mean of $\log(\hat{x}_k)$ and assuming a Gaussian model of the noise, i.e.

$$\hat{x}_k = \frac{\xi_k}{\xi_k + 1} \exp\{\frac{1}{2} \int_{v_k}^{\infty} \frac{e^{-t}}{t} dt\} y_k \tag{4}$$

$\hat{x}_k$ is the estimated magnitude of the clean speech. x_k is the actual magnitude of the clean speech. y_k is the noisy speech. ξ_k is the apriori SNR (Signal to Noise Ratio).

III. HARDWARE SETUP

The test-bed hardware used for real-time realization of speech enhancement includes a high-speed signal processor TMS320C6713 DSK from Texas Instruments (TI), two 4942 Brüel & Kjær (BK) microphones, CT 4200 crown amplifier and two RC 65i Polkaudio loudspeakers. To automate the data I/O (Input /Output), NI6733 digital to analog converter (DAC) and NI4472 analog to digital converter (ADC) from National Instruments (NI) are used. The noisy speech and the reference noise are captured in an actual fMRI scanning experiment using the BK microphones as shown in Fig.1 at 64 KHz sampling rate and 32 and 40 slices per two seconds Echo Planar Imaging (EPI) sequences. The results shown here are for the higher EPI sequence. The speech enhancement setup on the test-bed in the laboratory is shown in Fig.3. The audio signals from the computer are generated through a buffer to the NI 6733 and given as input to the DSK. The three adaptive speech enhancement algorithms are separately coded in on a TMS320C6713 DSK using TI Code Composer Studio. The board has a 2 channel ADC and a 2 channel DAC with a maximum sampling rate of 96 KHz. The processor has a clock rate of 225MHz, on chip memory of 64KB and 192 KB of SRAM memory. The data collected from the actual fMRI experiment was decimated to 8 KHz by setting the sampling rate of the ADC to 8 KHz. The DSK has a built in low pass filter. The filtered speech signal from the DSK is recorded on the computer using NI 4472. A laboratory simulation of the fMRI noise generation in an MRI bore model was also performed in the laboratory. This was required to objectively measure the speech enhancement quality. Objective measures of speech enhancement require clean speech for comparison with the enhanced speech signal. Since it is not possible to record both clean and noisy speech at the same time in an fMRI scan, we performed this simulation in laboratory. The fMRI bore is simulated by an acrylic half cylinder on a wooden base. The half cylinder is 1.52m long and has a 0.76m diameter and 1 inch thickness. The noise generating loudspeaker is placed at one end of the bore and the clean speech generating loudspeaker is kept horizontal inside the bore. Two microphones were placed close to two loudspeakers and their outputs were fed to the DSK. The enhanced speech from DSK was given to NI4472 DAC to be stored in the computer.

IV. SOFTWARE AND AUTOMATION

The IDE (Integrated Development Environment) used in the code development is TI Code Composer Studio (CCS) v2.2. The coding is done in C by using the TI DSP library (dsplib) functions and Fast Run- Time Support library (fastrts.lib). The DAQmx (Data Acquisition) module of LabVIEW 8.5 from NI is used for audio signal generation and capture. Also, the graphical user interface is created in LabVIEW 8.5. The LabVIEW script generates the reference noise and the noisy speech using the DAC as inputs for the DSP, runs the CCS and captures the clean speech using the ADC. This process is repeated for 21 seconds giving us enough data to evaluate the performance of the speech enhancement algorithms.

V. ANALYSIS

Since we are working with speech we set our sampling rate to 8 KHz [6]. Limiting the sampling frequency to 8 KHz allows us to use a larger filter length. For the adaptive algorithms the filter length was varied and it was reconfirmed that the enhancement is better for larger filter lengths. We observed that there was a tradeoff between bandwidth and enhancement quality. When we increased the sampling frequency, we had to reduce the filter length because of the limited processing power of the DSK. Ideally, when the sampling rate is increased we should use a larger filter length to obtain high level of enhancement.

As discussed in [7] the adaptation of the filter need not be stopped when we are performing speech enhancement. This

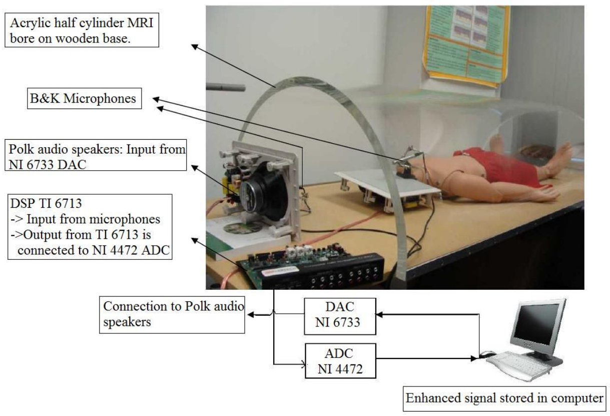


Fig. 3. Hardware setup in the laboratory.

is due to the fact that speech and noise are uncorrelated with each other.

For the LogMMSE algorithm the entire processing was done in frequency domain by taking the Fourier transform of the signals using optimized FFT function. This made the filtering operation easier to perform.

For all the algorithms the processor is programmed to take a block of data at a time. The frame size used for the adaptive algorithms was 80 and for the single channel algorithm was 128. The frame size was chosen as 128 to perform a 256 point overlap and add frame processing. The frames were overlapped by 50%. Two frames of data were concatenated and processed for each input frame. Out of the two frames one was always the old frame which was retained and shifted left and the second was the new input frame concatenated to the right. This way only 50% of the old data was retained. Similarly, the output frame was constructed by overlapping the new output with the old output frame by half.

Table.I shows the performance of all the 3 speech enhancement algorithms. We used PESQ and Segmental SNR [6] as the enhancement quality measure. Our experiments confirm that the original file, which is the noisy audio file recorded in the experiment, has the best speech quality as proved in [6]. The results in Fig.4 show the spectrogram of the noisy and recovered speech signals usig the 3 different algorithms. Segmental SNR is highest for nLMS, but it does not take into account the fact that the noise is mainly MRI noise. nSLMS on the other hand has a worse segmental SNR when compared to the nLMS, but the noise is more similar to white noise in this case (Fig.4).

Silence segment is the segment of data in which there is no speech, but only noise present. Noise suppression was measured as $20\log_{10}(||x_e||/||x_n||)$ where $||x_e||$ is the second norm of 100ms of the silence segment of the enhanced signal, and $||x_n||$ is the second norm of the 100ms of the silence segment of the noisy signal.

A. nLMS

The maximum filter length that could be implemented on a TMS 320C6713 DSP (the DSK) for nLMS algorithm was 900. The step size (μ) was set to 0.01. nLMS resulted in 18.4677 db of noise suppression in the output (the enhanced speech signal) when compared to the input (noisy speech signal) signal. The residual MRI noise is high in contrast to the output of nSLMS as described in the next section.

B. sNLMS

As can be seen in section II, nSLMS and nLMS are of the same complexity. Hence the number of taps for both the cases was same. Also the μ was set to 0.01. Both by simulations in Matlab and by analyzing the output of the DSK we found that nSLMS whitens the MRI noise better than nLMS. We found that nSLMS is very aggressive in suppressing noise. The problem we faced with nSLMS was that the small presence of speech in the reference microphone resulted in some distortion of speech in the output signal. The noise suppression in the output is 15.5740 db.

C. LogMMSE

The performance of this single channel method was poor. This could be due to the fact that this algorithm models the initial noise segment as white Gaussian [5], which is not correct in the case of MRI noise. The noise suppression in the output is 13.1529 db.

VI. CONCLUSION

1) Although we successfully enhanced the speech by three different algorithms, no single algorithm could be chosen as the best performer. There was a tradeoff between noise level, noise types and additional disturbances.
2) MRI noise is not white Gaussian in statistics and hence cannot be cancelled well by many conventional single channel speech enhancement methods.
3) As we experienced, the 2 adaptive algorithms show a whitening effect on the noise component of the noisy speech signal, but they also introduce disturbance at the end of the input frames. This is caused due to the dead band in between the two frames. Removal of this periodic impulse noise is a pending problem.

REFERENCES

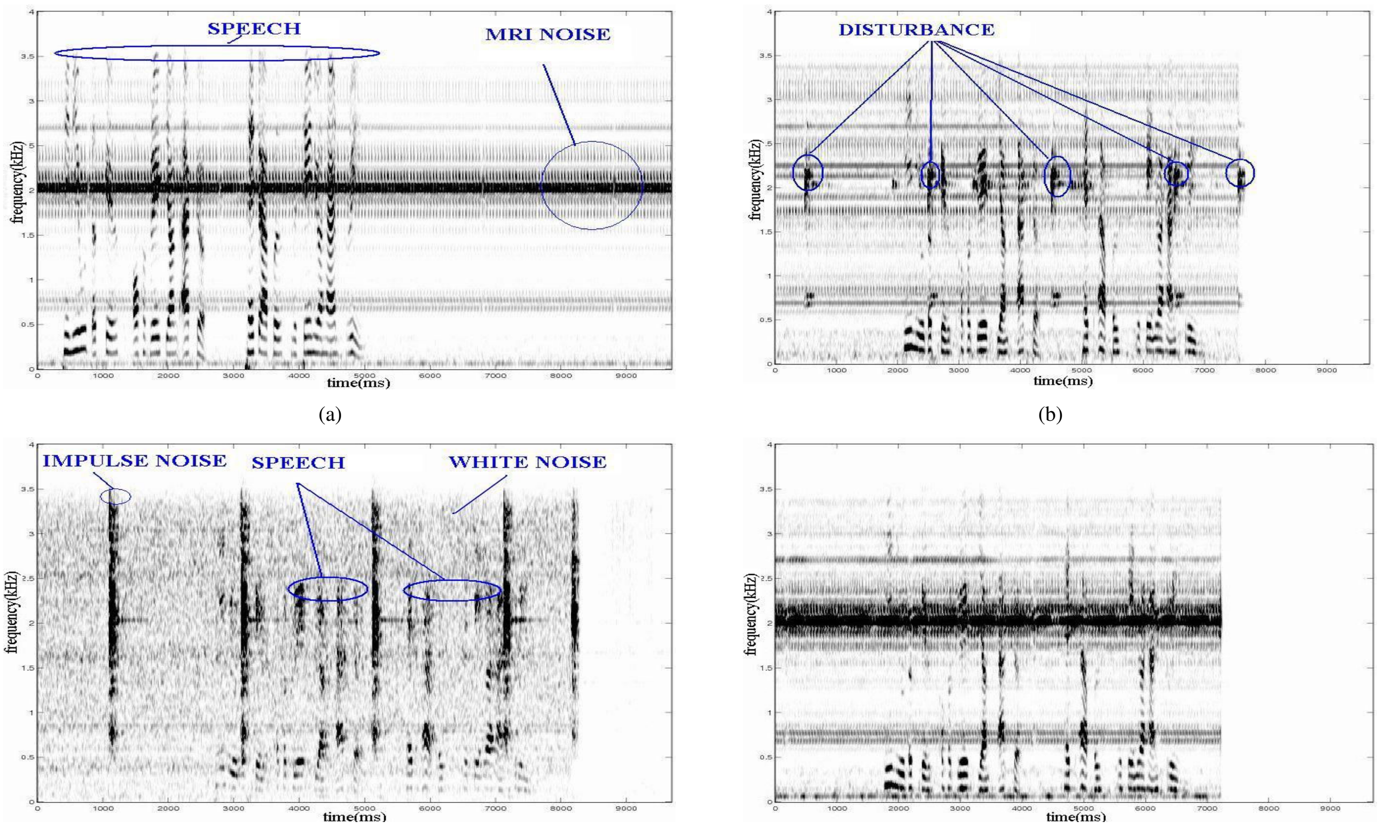
[1] M. Ravicz, J. Melcher, and N. Kiang, “Acoustic noise during functional magnetic resonance imaging,” *Journal of Acoustic Society of America*, vol. 108, no. 4, October 2000.

[2] V.Ramachandran, I. Panahi, Y. Hu, P. Loizou, R. Briggs, and S. McCaslin, “Multi-band speech enhancement for functional mri,” in *Acoustics, Speech and Signal Processing, 2006 IEEE International Conference on*, vol. 3, May 2006, p. III.

[3] V. Ramachandran, G. Kannan, A. Milani, and I. Panahi, “Speech enhancement in functional mri environment using adaptive sub-band algorithms,” in *Acoustics, Speech and Signal Processing, 2007 IEEE International Conference on*, vol. 2, April 2007, pp. II–341–II–344.

[4] A. Sayed, *Fundamentals of Adaptive FIltering*. Wiley,New York, 2003.

[5] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 33, pp. 443–445, April 1985.



(a) (b) (c) (d)

Fig. 4. Spectrogram plot for (a) noisy speech (b) speech after enhancement by nLMS. Periodic components of the original noise are still present (c) speech after enhancement by nSLMS. Notice the absence of peiodic component and the presence of white noise. In addition there is an impulse noise. (d) speech after enhancement by LogMMSE. All the plots are for 40 EPI sequences per second.

TABLE I
PERFORMANCE ANALYSIS

	PESQ	Segmental SNR
Original noisy speech	2.8671	-6.5894
Speech enhanced by nLMS	2.3555	-1.9757
Speech enhanced by nSLMS	1.9283	-5.7008
Speech enhanced by LogMMSE	2.7311	-5.8747

[6] P. C. Loizou, *Speech Enhancement Theory and Practice*. CRC Press, Florida, 2007.
[7] B. Widrow, J. R. Glover, J. M. Mccool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, R. C. Goodlin, and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692–1716, 1975.

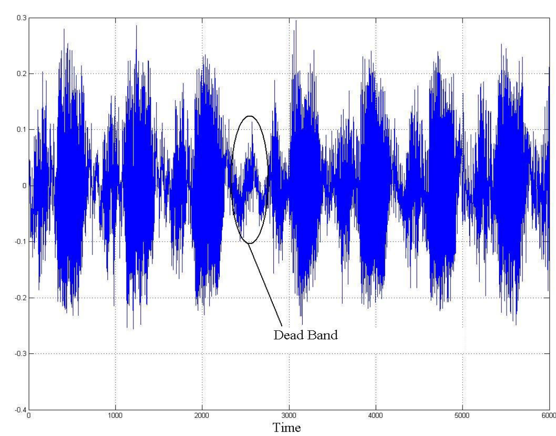


Fig. 5. Time domain plot of a section of MRI noise sampled at 8Khz with dead band in highlighted.