A Noise Cancellation Scheme for fMRI Involving Participant Speech

I. Atkinson[1], T. Claiborne[1], M. P. Flannery[1], K. R. Thulborn[1]

[1]Center for Magnetic Resonance Research, University of Illinois at Chicago, Chicago, IL, United States

INTRODUCTION

Echo-planar imaging (EPI) sequences that are often used for functional magnetic resonance imaging (fMRI) produce substantial acoustic noise due to rapid gradient switching. Noise levels during an EPI scan at 3.0T can reach 115 dBA or more while normal human speech is only 50-70 dBA. To limit head motion caused by speech production, functional paradigms involving participant speech must use unforced speech (e.g. conversational volume). As speech is overpowered by the gradient noise, its correct interpretation when recorded during fMRI studies is challenging. One approach to this problem is to add an additional delay between certain acquisitions to allow time for the subject to speak without interference from gradient switching noise. Careful synchronization of the speech and stimulus with the imaging sequence is required. This requires a cooperative participant, compromises temporal resolution, imposes an increased acquisition time, and requires correction for effects of variable T1. We have adopted an approach that avoids these issues. The speech is recorded as corrupted by the gradient noise and then this noise contamination is suppressed by applying an adaptive noise cancellation technique to yield readily comprehensible speech.

METHODS

An exploitable feature of the acoustic noise produced during an EPI scan is that the gradient switching, and therefore the noise it creates, is periodic. This allows the true gradient noise to be estimated from the source data at times when no speech is present and subtracted from the noisy signal to suppress the gradient noise. We estimate the magnitude spectrum of the corrupting gradient noise using a bank of adaptive filters and employ spectral subtraction to suppress the noise by subtracting the estimated noise magnitude spectrum from the magnitude spectrum of the observed data.

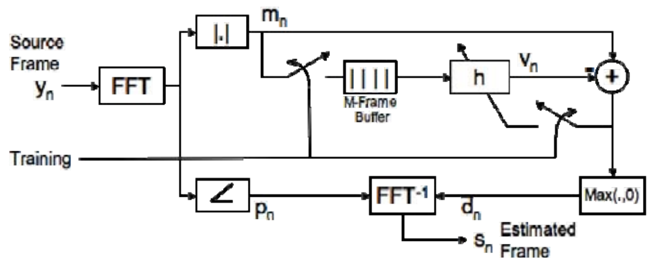

Figure 1: Gradient Noise Cancellation Scheme

The observed source data set is divided into frames of length $N+N_O$ where N is the period of the gradient noise and $N_O \geq 0$ is the overlap of adjacent blocks. Frames that are known not to contain any speech (due to the paradigm design) are termed *training* frames and are used to estimate the gradient noise. For a given frame of source data (y_n), the spectrum is computed using the fast Fourier transform (FFT) and separated into its magnitude and phase components, which we denote m_n and p_n, respectively. Training frames are marked for insertion into a length-M buffer, which by design contains the magnitude spectrum of the M most recent training blocks. This buffer is used to form an estimate of the current noise magnitude spectrum by filtering its contents with an adaptive filter (h) to produce an estimate of the noise magnitude spectrum (v_n). We subtract this estimated noise spectrum from y_n to produce d_n, which is an estimate of the magnitude spectrum of the true signal for frame n. For training frames, d_n is be a measure of how well the magnitude of each frequency was estimated and is used to update the coefficients of h so that future estimates are more accurate. Since the gradient noise varies little from frame to frame, the least mean-squares (LMS) update rule can be used to adjust the filter coefficients. Updates to the filter coefficients are not made for non-training frames since d_n contains speech information and would cause the filter to converge toward, and therefore filter out, the speech. The final step in the noise cancellation process is to reconstruct the estimated frame in the time domain, which is done by combining the original phase of the source with the estimated magnitude and computing the inverse FFT. If a non-zero frame overlap is used, the estimated frames are windowed in the time-domain and combined using the overlap-add method. A block diagram of the noise cancellation scheme (without block recombining) is shown in Figure 1.

Functional MRI was performed at 3.0T on a whole-body long-bore scanner (3T94 HDMR Signa, GE Healthcare, Milwaukee, WI) using a standard, single-channel, quadrature head coil and EPI acquisition (TR = 3000 ms, TE = 25.3 ms, matrix = 64x64, 26 slices, FOV=20 cm^2). Audio signals were measured using an MR-compatible microphone (Phone-Or Ltd., Israel), which was positioned in the scanner bore and digitally sampled at 44.1 kHz by a standard computer sound card. Recorded audio signals were post-processed offline (MATLAB, MathWorks, Natick, MA) with the described adaptive spectral subtraction noise cancellation algorithm.

The algorithm was first tested across the frequency range of human voice using a known source consisting of 15 seconds of silence followed by 13 cycles of a linear frequency sweep from 0 Hz to 5000 Hz at a sweep rate of 1000 Hz/second and 5 second blocks of silence. This signal was injected into the scanner bore near the microphone through plastic tubing. Digital recordings were made without (Reference) and with (Noisy Source) the scanner operating. The recording of the Noisy Source was processed offline using the described algorithm with the times of known silence marked as training blocks (Processed).

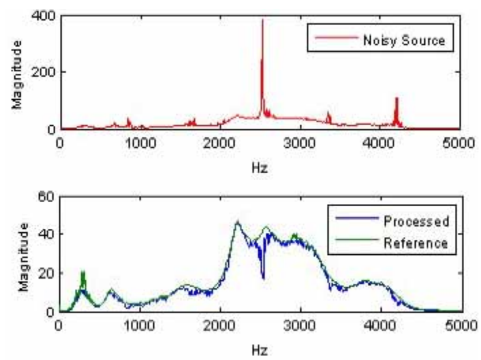

Figure 2: Magnitude Spectra of Noisy Source, Reference, and Processed data

Functionality was then tested with two native English-speaking volunteers, one male and one female. These volunteers were asked to read individual English words that were presented visually using an MR-compatible synchronization control system (MRIx Technologies, Bannockburn, IL) during the fMRI experiment. The participant speech was recorded as contaminated by scanner operation and processed offline with the described noise cancellation scheme. Rigid body motion analysis performed (SPM) on both datasets determined that speech production caused no significant head motion (<1 mm of translation and <1° rotation in any direction). The unprocessed and noise cancelled speech recordings were then scored by three independent, naïve listeners by writing down words that could be understood. The words were then scored for number correctly identified and the total time required by the listener to complete the identification. Overall performance was measured as the percentage change in number of words correctly identified and the percentage change in total identification time.

RESULTS

The average magnitude spectrum of the linear frequency sweep portions of the Reference, Noisy Source, and Processed data are shown in Figure 2. Substantial noise is introduced by the switching gradients at approximately 2500 Hz and 4200 Hz, which greatly contaminates the reference signal. Processing the recorded audio with the proposed noise cancellation algorithm effectively removes the gradient noise and yields a magnitude spectrum that is closely matched with the reference spectrum.

Processing human speech contaminated by gradient noise with the described noise cancellation algorithm improved the ability of listeners to correctly identify individual English words by an average of 21.6%. In addition, the time required identify those words was decreased by an average of 17.5%. The results of the word scoring experiment are summarized in Table 1.

Subject	Listener	% Improvement	
		Accuracy	Time
Female	1	38.3	39.5
	2	23.0	22.1
	3	20.0	21.4
Female Average		**27.1**	**27.7**
Male	1	34.5	12.5
	2	7.8	-2.5
	3	6.0	11.8
Male Average		**16.1**	**7.3**
Overall Average		**21.6**	**17.5**

Table 1: Word scoring results

CONCLUSION

The proposed noise cancellation scheme applied to human speech captured during fMRI improved speed (17.5%) and accuracy (21.6%) of intelligibility.