Blind Removal of Gradient Noise From Overt Participant Speech During fMRI

M. D. Kleffner[1], I. C. Atkinson[2], and D. L. Jones[1]

[1]Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [2]Center for MR Research, University of Illinois at Chicago, Chicago, IL, United States

INTRODUCTION

Interpretation of overt participant speech (50-70 dBA) during functional magnetic resonance imaging (fMRI) is made difficult due to significant acoustic noise produced during image acquisition (>100 dBA). Although paradigms involving overt participant speech have the potential to provide insight into rehabilitation strategies for expressive aphasia resulting from stroke, they remain largely unused due to an inability to assess the speech that is masked by gradient noise. Removing this noise while preserving the speech is challenging, in part because mouth movement alters the noise recorded by microphones placed close to the mouth. A recently-developed blind, multi-microphone speech-enhancement algorithm is robust to speaker position, reverberation, and noise.[1] This technique can be applied to overt speech recorded during fMRI to suppress the gradient noise without distorting the desired speech signal.

METHODS

Real-world noise often has lower kurtosis (fourth central statistical moment) than speech. The maximum-kurtosis, distortionless-response (MKDR) algorithm is a frequency-domain beamforming algorithm that recovers speech by finding the direction of maximum kurtosis for each frequency and then applies a minimum-variance, distortionless-response (MVDR) beamformer.[1] An extension of the MKDR (maximum-kurtosis, wiener estimate or MKWE) relies on background-noise estimates to estimate the optimal linear (Wiener) postfilter.[1]

The MKDR and MKWE algorithms are applied to overt participant speech acquired in an fMRI setting. Functional MRI was performed on a healthy native English-speaking male volunteer on a 3T whole-body MR scanner (GE Healthcare) using a standard EPI acquisition (TR=3000 ms, TE=25.3 ms, matrix=64x64, 26 slices, FOV=20x20 cm^2). The volunteer was asked to read English sentences that were presented visually using an MR-compatible synchronization control system (MRIx Technologies, Bannockburn, IL). Participant speech was measured using dual-gradient and single-gradient MR-compatible microphones (Phone-Or, Israel) and recorded at 48 kHz. Each gradient microphone has a bi-directional spatial response along one axis in a three-dimensional coordinate system; the dual-gradient microphone has each gradient microphone oriented such that the spatial responses are 90 degrees apart. Three microphone positions above the volunteer's mouth, sagittal-sagittal (SS), sagittal-paratransverse (ST), and paratransverse-paratransverse (TT), were used in recording spoken sentences from the test subject, where the first plane denotes the plane *containing* the dual-gradient-microphone responses, and the second plane denotes the plane the single-gradient-microphone response is *perpendicular* to.

The audio recordings are downsampled to 8 kHz and four-second segments containing the same spoken sentences are processed offline with the MKDR and MKWE algorithms. The algorithms are applied using all three microphone channels and using only the dual-gradient microphone channels, since the single-gradient microphone has the lowest speech power in all configurations. Since the gradient-noise is periodic (71.5 ms period), the signals are first passed through a series of 10 notch filters designed to remove energy at the 10 frequencies containing the most energy, as defined by summing the power spectra of the signals. The notch-filtered signals are then filtered with a high-pass filter having a cutoff of 350 Hz, since very little speech content is present below the cutoff frequency. Signal SNR is estimated by first squaring the signal and computing a one-gradient-noise-period, sliding average of the squared signal. The maximum value of this signal is assumed to be the sum of the speech and noise power and the minimum value the power of the noise signal, since each recorded sentence has at least 643.5 ms of exclusively noise. The input SNR of the channel resulting in the most overall noise power reduction for the SS, ST, and TT configurations are 1, -1, and 4 dB, respectively. The dual-gradient-only SNRs after notch- and low-pass filtering are, to the nearest 0.5 dB, 4.5, 3.5 and 10 dB. The SNRs become 4.5, 4.0 and 10 dB when all microphones are used.

The time-frequency parameters used in the MKDR and MKWE algorithms are as follows. The speech is divided into 128 ms (1024-sample) segments, with 96 ms (75 percent) overlap, and fast-Fourier transforms, zero-padded to 2048 samples, are computed. A segment length roughly twice the gradient period was chosen so that the beamforming filters (having length equal to the segment length) are capable of canceling the gradient noise via subtracting a delayed version of the signal from itself. However, more reverberation is introduced with longer beamforming filters. For the MKWE algorithm, the lowest 20^{th} percentile of signal powers in each frequency bin is used to estimate the noise, and the biases in these noise estimates are removed via scale factors computed from a four-second, noise-only portion of the recording. The enhanced speech was evaluated by informal listening tests and estimated output SNR.

RESULTS

The top waveform in Figure 1 is the noisy speech signal as recorded from the channels of the TT-configuration, dual-gradient microphone and the bottom waveform is the output from the MKDR algorithm applied exclusively to the dual-gradient microphone. The bar graph in Figure 1 shows the SNRs produced by the MKDR and MKWE algorithms, applied to the notch-filtered signals, for the aforementioned configurations using only the dual-gradient microphone (2 channels) and both microphones (3 channels). In terms of SNR, the MKDR and MKWE algorithms perform well in all cases, with the MKWE having an advantage over MKDR when the dual-gradient is used exclusively. Informal listening tests confirm such an advantage. Informal listening tests also indicate that using all three microphone channels reduces the "peakiness" of the noise more than using only the dual-gradient microphone, but the resulting speech is more reverberant. It is presently unknown whether this reverberance can be reduced by using a microphone containing three gradient microphones in a fixed, co-located position.

Although the output SNRs are similar, informal listening tests indicate that the TT configuration has the smallest amount of reverberation introduced by the enhancement process. This correlates to the TT configuration having the best input SNR, the best gain from the notch- and low-pass filtering, and the lowest gain from the MKDR and MKWE algorithms.

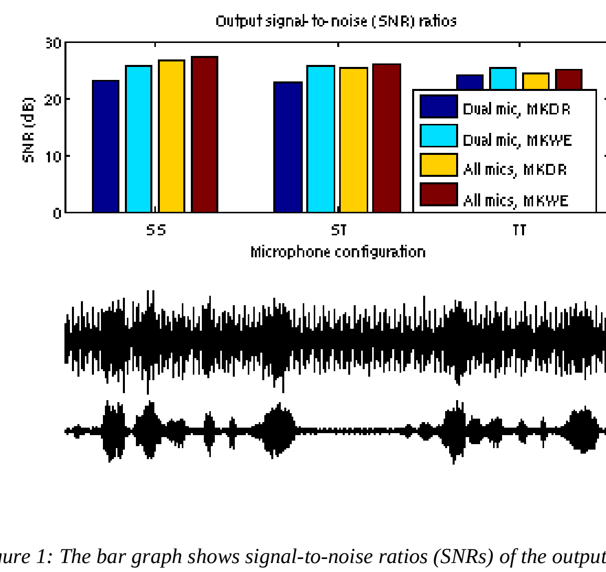

Figure 1: The bar graph shows signal-to-noise ratios (SNRs) of the output signals produced by the gradient-noise-removal process. The top waveform is one of the channels of the dual-gradient microphone in the TT configuration, and the bottom is the output of the MKDR algorithm applied exclusively to both channels of the dual-gradient microphone.

CONCLUSION

The MKDR and MKWE algorithms, when combined with pre-filtering, increase the SNR of overt participant speech by 20-27 dB to 23-27 dB for multiple microphone configurations. Using only the dual-gradient microphone results in less reverberation than using the dual-gradient microphone and a single-gradient microphone, but the noise "peakiness" is reduced by a smaller amount.

REFERENCES

1. M.D. Kleffner and D.L. Jones, "Blind Recovery of a Speech Source in Noisy, Reverberant Environments", submitted to the *Journal of the Acoustical Society of America*.

2. I. Atkinson, T. Claiborne, M. P. Flannery, and K. R. Thulborn, "A noise cancellation scheme for fMRI involving participant speech", *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM)*, 2006, Abstract No. 5304.