

Published in final edited form as:

*Magn Reson Med.* 2005 March ; 53(3): 739–744. doi:10.1002/mrm.20398.

## Extraction of Overt Verbal Response from the Acoustic Noise in a Functional Magnetic Resonance Imaging Scan by Use of Segmented Active Noise Cancellation

Kwan-Jin Jung<sup>1,2,\*</sup>, Parikshit Prasad<sup>1,3</sup>, Yulin Qin<sup>4</sup>, and John R. Anderson<sup>4</sup>

<sup>1</sup>Brain Imaging Research Center, University of Pittsburgh & Carnegie Mellon University, Pittsburgh, Pennsylvania

<sup>2</sup>Department of Bioengineering, University of Pittsburgh, Pittsburgh, Pennsylvania

<sup>3</sup>Department of Neurobiology, University of Pittsburgh, Pittsburgh, Pennsylvania

<sup>4</sup>Psychology Department, Carnegie Mellon University, Pittsburgh, Pennsylvania

### Abstract

A method to extract the subject's overt verbal response from the obscuring acoustic noise in an fMRI scan is developed by applying active noise cancellation with a conventional MRI microphone. Since the EPI scanning and its accompanying acoustic noise in fMRI are repetitive, the acoustic noise in one time segment was used as a reference noise in suppressing the acoustic noise in subsequent segments. However, the acoustic noise from the scanner was affected by the subject's movements, so the reference noise was adaptively adjusted as the scanner's acoustic properties varied in time. This method was successfully applied to a cognitive fMRI experiment with overt verbal responses.

### Keywords

active noise cancellation; overt verbal response in fMRI; acoustic noise; template matching; dynamic reference noise

---

The fast switching of strong gradient pulses in EPI generates a significant amount of acoustic noise, and this noise becomes greater at higher magnetic fields such as 3 T (1,2). In some fMRI experiments there is a need to obtain a subject's overt verbal response rather than button presses in order to measure the verbal response time (3,4) or to avoid requiring the subject to master complex mappings from nonnumerical responses to finger responses (5). However, the verbal response may be difficult to understand because the background acoustic noise from the scanner obscures the voice signal (6,7). Although headphones can partially shield the acoustic noise to enable the subject to hear an auditory message, the microphone used to receive the subject's verbal response cannot be shielded from acoustic noise generated by the scanner (8). Therefore, we developed a technique to extract the voice signal from the scanner's acoustic noise. While some methods (9–11) use two microphones, one for speech and one just for background noise, we wanted to develop a technique that used only one microphone to avoid problems in matching the background noise between microphones.

EPI scanning and its accompanying acoustic noise are repetitive for each slice and each volume. This property has been noted and suggested as a basis for the cancellation of the acoustic noise from the scanner in fMRI (12,13). However, the acoustic noise from the scanner can be affected by the variation of any acoustic property in the magnet, such as the subject's voluntary or involuntary movement. In this paper the noise cancellation method, which utilizes the repetitive characteristics of the acoustic noise from the scanner, is further developed to adaptively suppress the acoustic noise from the scanner when the scanner's acoustic properties vary in time. This method was applied to a cognitive fMRI experiment in an attempt to extract the subject's verbal responses from the noise.

## THEORY

The recorded sound signal  $s(t)$  is a superposition of voice signal  $v(t)$  and scanner acoustic noise  $n(t)$ , i.e.,

$$s(t) = v(t) + n(t). \quad [1]$$

In fMRI scanning with a single-shot EPI sequence, the acoustic noise from the scanner is repetitive for each slice as well as for each volume. In practice, however, the scanner is designed to maintain a constant repetition time (TR) of the imaging volume, whereas the slice time is calculated within the allowed time resolution of the scanner from a given TR and number of slices. This may result in uneven slice timing during the TR. Therefore, the recorded sound signal can be segmented better into volumes of duration TR as

$$s_m(t) = s((m-1)TR + t), \quad [2]$$

where  $0 \leq t < TR$  and  $s_m(t)$  denotes the sound signal for  $m$ th segment of  $s(t)$ . Similarly, we can have  $v_m(t)$  and  $n_m(t)$  to represent the sound signals of  $m$ th segment of  $v(t)$  and  $n(t)$ , respectively. Then a segment of the signal can be obtained from Eq. [1] as

$$s_m(t) = v_m(t) + n_m(t). \quad [3]$$

If the acoustic environment in the magnet remains constant during the fMRI run, the initial reference noise  $n_0(t)$  taken from the last shot of dummy scans can be used as the reference noise for the whole run, i.e.,

$$n_{\text{ref}}(t) = n_0(t). \quad [4]$$

Then, the voice signal of the  $m$ th segment can be extracted by subtracting the reference noise from the recorded sound signal as

$$v_m(t) = s_m(t) - n_{\text{ref}}(t). \quad [5]$$

However, the acoustic signal can be affected by the subject's movement during the fMRI run, which will result in a variation of the acoustic noise from the scanner. To achieve noise suppression that adapts to the noise variation, the reference noise is updated continuously through the run by taking the reference noise from the segment that has neither voice signal nor noise transition. The routine to check the presence of voice signal and noise transition in a segment can be divided into two conditions, each depending on their duration relative to the TR. The voice signal and noise transition of a short duration can be detected by

comparing the peak with the mean of  $v_m(t)$  that is processed with the current reference noise, i.e.,

$$\rho_{\text{short}} = \frac{\text{maximum of } |v_m(t)|}{\text{mean of } |v_m(t)|}. \quad [6]$$

On the other hand, the voice signal or noise transition in a long duration can be detected by comparing the mean of  $v_m(t)$  with the mean of  $v_{m-1}(t)$  as

$$\rho_{\text{long}} = \frac{|\text{mean of } |v_m(t)| - \text{mean of } |v_{m-1}(t)|}{\text{mean of } |v_{m-1}(t)|}. \quad [7]$$

If  $\rho_{\text{short}}$  and  $\rho_{\text{long}}$  are lower than thresholds that will be determined experimentally, the  $m$ th segment is considered to have neither voice signal nor noise transition. Under this condition, the reference noise is updated with the signal of the current segment for the processing of subsequent segments as

$$n_{\text{ref}}(t) = s_m(t). \quad [8]$$

We use the term “dynamic reference noise” to refer to the procedure for dynamically updating the reference noise by Eq. [8].

## METHODS

The scanner used was a Siemens Magnetom Allegra head-dedicated 3 T, equipped with a gradient system that has a maximum gradient of 40 mT/m and a rise time of 100  $\mu$ s. The scan parameters of the single-shot gradient echo EPI for fMRI were as follows: TR = 1.5 s, TE = 30 ms, number of slices = 26, image matrix = 64  $\times$  64, pixel bandwidth = 3005 Hz, slice thickness = 3.2 mm, field of view = 200 mm, and number of dummy scans = 3. The audio system used was a Silent Scan Model SS-3100 (Avotec, Inc.) that has an acoustic headphone and microphone. The microphone signal is available at a line output in the console. The microphone was installed at the inner top of the head coil rather than on the subject's chest to avoid interference from cardiac sounds and respiratory motions. The transducer module in the Avotec audio system was initially saturated by the EPI acoustic noise; this was corrected by reducing the gain in the amplifying and filtering stage of the transducer module. The line output of the Avotec console and the scanner synchronization pulse was connected to a stereo line input of a PC (Dell Dimension 8200, Pentium 4, 2.2 GHz, Windows XP) equipped with a sound card (Creative, SB Live). The stereo signal was recorded using Sound Forge 7.0 (Sony Pictures Digital, Inc.), which was triggered by the acoustic noise from the scanner. The scanner synchronization pulse (pulse width = 26 ms) occurred at the beginning of each volume after the dummy scans. The audio sampling rate was set to the maximum available (96 kHz) with a 16-bit width for the stereo channels.

The beginning of each segment must be determined before canceling the noise. The synchronization pulse can be used as an identifier of the beginning of each segment. Alternatively, a template that matches the sound to the reference noise can be used assuming that the acoustic noise from the scanner is not altered significantly. In our case the voice response was brief so template matching did not suffer, even in segments containing a voice signal. Accurate segmentation by template matching was achieved by correlating the template of the reference noise with the sound signal at each segment. The correlation

matching was done by shifting the sample point around TR. The template contained 30 more samples than did TR in order to ensure adequate temporal coverage.

The method was first confirmed by scanning a spherical phantom in order to avoid factors such as the subject's movement that would alter the sound. Next, this method was tested on a human subject in the absence and presence of a controlled motion. For the controlled motion the subject was instructed to rotate the left palm up and down on top of the chest in an alternating manner on each trial. Finally, the method was incorporated into an fMRI study of 15 subjects. All subjects participated after giving their informed consent.

The fMRI paradigm and the placement of the microphone were identical for both phantom and human subjects. For each subject there were 12 runs of the cognitive task following the structural imaging. Each run consisted of 12 trials for a total of 232 volumes during 5 min 53 s and was programmed in E-Prime (Psychology Software Tools, Inc.) (5). On each trial there was a sequence of three audio communications as illustrated in Fig. 1, i.e., a prime to the subject, a response from the subject, and a feedback to the subject. The prime was a sequence of "Dick," "Fred," and "Tom" in a random order for each of the 12 trials. The task was to reorder the three names according to a subsequent instruction given as a visual message and to speak overtly the three names in their new order. It took fewer than 2 s to speak the three names.

## RESULTS

The scanner's acoustic noise, recorded with the phantom as the subject, is shown in Fig. 2a. The segmentation using the synchronization pulse and template matching was compared by processing with a constant reference noise as shown in Fig. 2b and c, respectively. The segmentation with the synchronization pulse suffered from an intermodulation effect between the base frequency of the scanner synchronization pulse and the audio sampling frequency. The segmentation using template matching required information about the TR and the sampling frequency of the recorded sound. Since there was a potential mismatch between the scanner and the audio sampling frequency, the number of samples in a segment of TR was first obtained by matching the segment for the initial reference noise with the next segment. In this experiment, the number of samples in a TR segment was about 18–19 samples larger than the theoretical number, i.e., 144000, which can be attributed to a mismatch between the clock frequency in the sound card and the scanner frequency. The gradual increase of the remnant noise might be due to instability within the sound card. The increase was eliminated by use of the dynamic reference noise as demonstrated in Fig. 2d. We used thresholds of 9 for  $\rho_{\text{short}}$  and 2.1 for  $\rho_{\text{long}}$ . The dynamic reference noise achieved a noise suppression of 25 dB and an improvement of 8 dB at the end of the run with respect to the constant reference noise.

The results with a volunteer human subject are shown in Figs. 3 and 4 in the absence and presence of the controlled motion, respectively. The variation of the scanner's acoustic noise caused by the subject's motion is clearly visible after processing. The dynamic reference noise took one or two segments to adapt to the abrupt change of the noise. The thresholds for  $\rho_{\text{short}}$  and  $\rho_{\text{long}}$  were 5.5 and 2.1, respectively. The sound signal contained crosstalk from the prime and feedback signals and therefore the threshold for  $\rho_{\text{short}}$  was set low enough to eliminate the segment with crosstalk from being considered reference noise. The detection of the segment with the sound or noise transition was very reliable for  $\rho_{\text{short}}$  and  $\rho_{\text{long}}$  in the range of 5.5–10 and 2.0–4.5, respectively. The noise suppression by the dynamic reference noise was about 25 dB at the dominant frequency of 1.58 kHz and around 20 dB for frequencies up to 3 kHz, as shown in Fig. 5.

The signals recorded from 15 subjects in the fMRI study were processed by use of both the constant and the dynamic reference noise. For the dynamic reference noise,  $\rho_{\text{short}}$  and  $\rho_{\text{long}}$  were constant at 9.0 and 2.1, respectively, for all runs and subjects. One subject with severe motion artifacts in the EPI images was excluded and 4 runs from the remaining subjects were not considered due to the poor quality of the recorded audio signal, leaving 164 runs for the voice extraction. To quantitatively analyze the improvement of the dynamic reference noise over the constant reference noise, we calculated the root mean square (RMS) value of the remnant noises under the two conditions. For a segment without voice or other abrupt transients between the 11th and 12th verbal responses in a run, a RMS ratio was calculated as RMS of remnant noise processed with dynamic reference noise over RMS of remnant noise processed with constant reference noise. The distribution of RMS ratios is shown as a histogram in Fig. 6. The ratio was always less than unity, which indicates that the dynamic reference noise improved the result in all runs. The average RMS ratio was 0.46, corresponding to an improvement of more than twofold. When the ratio was less than 0.46 it was often difficult to interpret the verbal responses after processing with constant reference noise. The appearance of a secondary peak where the RMS ratio was less than 0.05 corresponds to runs with large subject movements, where the constant reference noise failed to adequately cancel the scanner noise.

To address the concern of motion artifacts in the fMRI analysis from the overt speech, a run of images from a single subject was analyzed for a linear correlation with the overt verbal response time measured from the noise-cancelled voice signal. There were positive activations on the frontal cortex surfaces and the lateral ventricle, while there were negative activations on the superior regions of paranasal sinuses (14). However, there were no artifactual activations in the cortex areas targeted in this study (5).

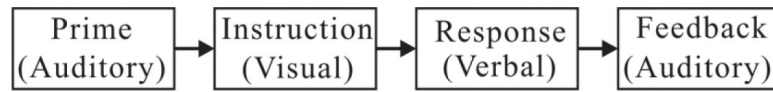
## CONCLUSIONS

Voice signals can be extracted from EPI scanning noise in fMRI by means of a segmented active noise cancellation technique that uses a conventional headphone with one microphone. However, the acoustic properties in the magnet vary during the fMRI run due to movements of the subject. Because the acoustic noise changed over time it was essential to update the reference noise dynamically in order to maintain a high level of noise suppression during the fMRI run. The algorithm for such a dynamic update of the reference noise was very effective in our fMRI study and is expected to be generally applicable to variations of the acoustic noise from the scanner during an fMRI run. Our method can be used with or without the help of the synchronization pulse. The simplicity of our method and the calculation interval of TR for the update of the reference noise can allow real-time processing for both the extraction of the voice signal and the suppression of scanner's acoustic noise experienced by the subject.

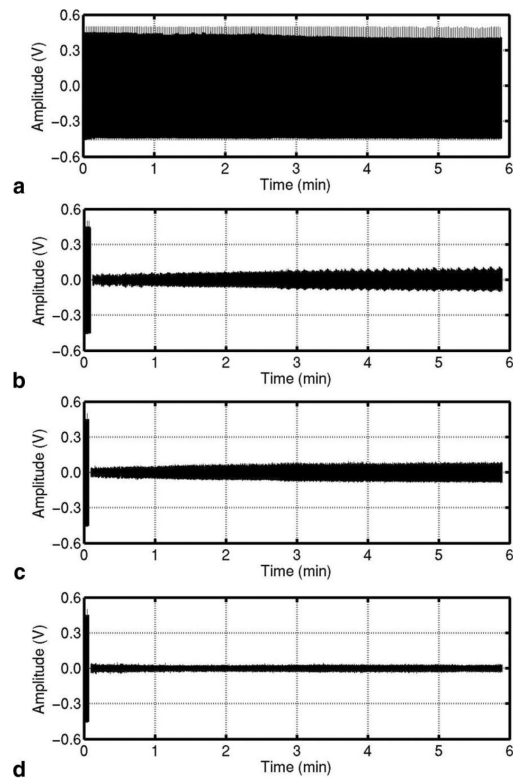
## REFERENCES

1. Schmitt, F.; Stehling, MK.; Turner, R. Echo-planar imaging theory, technique, and application. Springer; New York: 1998.
2. Cho ZH, Park SH, Kim JH, Chung SC, Chung ST, Chung JY, Moon CW, Yi JH, Sin CH, Wong EK. Analysis of acoustic noise in MRI. *Magn Reson Imaging*. 1997; 15:815–822. [PubMed: 9309612]
3. Phelps EA, Hyder F, Blamire AM, Shulman RG. FMRI of the prefrontal cortex during overt verbal fluency. *Neuroreport*. 1997; 8:561–565. [PubMed: 9080448]
4. Schlaggar BL, Brown TT, Lugar HM, Visscher KM, Miezin FM, Petersen SE. Functional neuroanatomical differences between adults and school-age children in the processing of single words. *Science*. 2002; 296:1476–1479. [PubMed: 12029136]

5. Anderson JR, Qin Y, Stenger VA, Carter CS. The relationship of three cortical regions to an information-processing model. *Cogn Neurosci*. 2004; 16:637–653.
6. Moelker A, Pattynama PMT. Acoustic noise concerns in functional magnetic resonance imaging. *Hum Brain Mapp*. 2003; 20:123–141. [PubMed: 14601139]
7. Amaro E Jr, Williams SC, Shergill SS, Fu CH, MacSweeney M, Picchioni MM, Brammer MJ, McGuire PK. Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. *J Magn Reson Imaging*. 2002; 16:497–510. [PubMed: 12412026]
8. Ravicz ME, Melcher JR. Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: examination of noise conduction through the ear canal, head, and body. *J Acoust Soc Am*. 2001; 109:216–231. [PubMed: 11206150]
9. Chen CK, Chiueh TD, Chen JH. Active cancellation system of acoustic noise in MR imaging. *IEEE Trans Biomed Eng*. 1999; 46:186–191. [PubMed: 9932340]
10. McJury M, Stewart RW, Crawford D, Toma E. The use of active noise control (ANC) to reduce acoustic noise generated during MRI scanning: some initial results. *Magn Reson Imaging*. 1997; 15:319–322. [PubMed: 9201679]
11. Chambers J, Akeroyd MA, Summerfield AQ, Palmer AR. Active control of the volume acquisition noise in functional magnetic resonance imaging: method and psychoacoustical evaluation. *J Acoust Soc Am*. 2001; 110:3041–3054. [PubMed: 11785805]
12. Nelles JL, Lugar HM, Coalson RS, Miezin FM, Petersen SE, Schlaggar BL. Automated method for extracting response latencies of subject vocalizations in event-related fMRI experiments. *NeuroImage*. 2003; 20:1865–1871. [PubMed: 14642496]
13. Ravicz ME, Melcher JR, Kiang NY. Acoustic noise during functional magnetic resonance imaging. *J Acoust Soc Am*. 2000; 108:1683–1696. [PubMed: 11051496]
14. Barch DM, Sabb FW, Carter CS, Braver TS, Noll DC, Cohen JD. Overt verbal responding during fMRI scanning: empirical investigations of problems and potential solutions. *NeuroImage*. 1999; 10:642–657. [PubMed: 10600410]

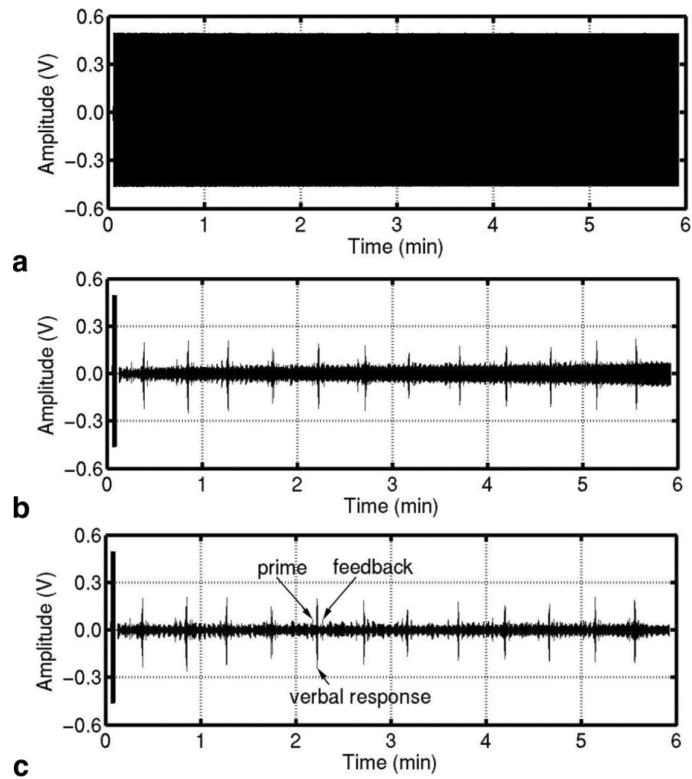


**FIG. 1.**  
A block diagram of the fMRI paradigm.

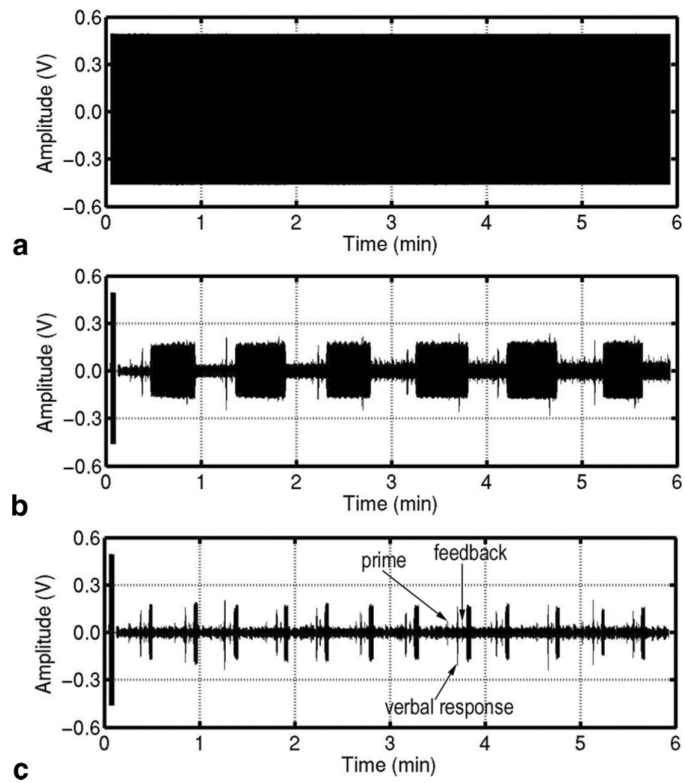
**FIG. 2.**

An experiment using a phantom subject. **(a)** Acoustic noise from the scanner. **(b)** Processed waveform using the synchronization pulse for the segmentation. **(c)** and **(d)** Processed waveforms using template matching with a constant and a dynamic reference noise, respectively. The segment corresponding to the reference noise is nulled and the scanner's acoustic noise before the reference noise is left un-processed.

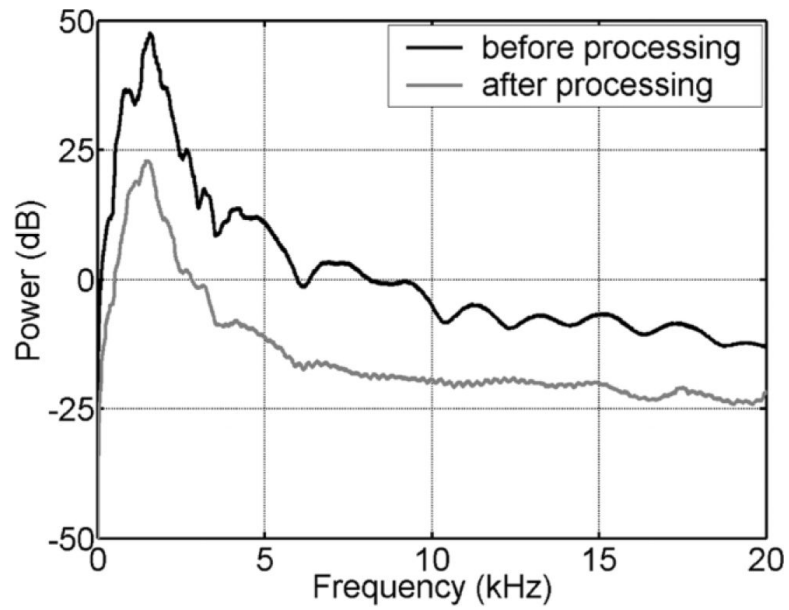




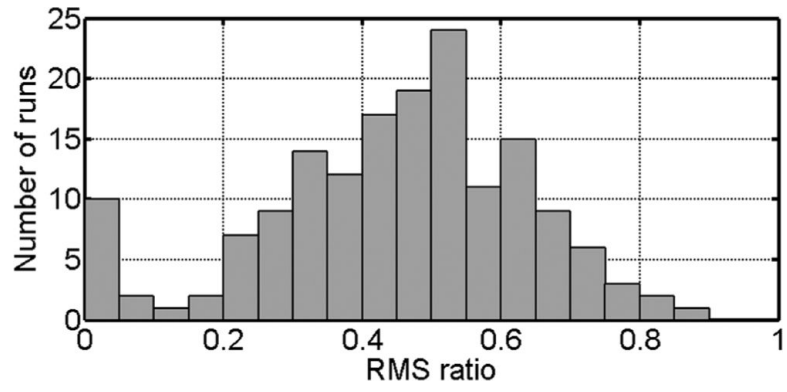
**FIG. 3.** An experiment using a human subject without voluntary movements. **(a)** Waveform before processing. **(b)** and **(c)** Waveforms after processing with a constant and a dynamic reference noise, respectively.



**FIG. 4.** An experiment using a human subject with voluntary movements. (a) Waveform before processing. (b) and (c) Waveforms after processing with a constant and a dynamic reference noise, respectively.



**FIG. 5.** Spectral power of the acoustic noise from the scanner before and after processing with the dynamic reference noise for one segment in Fig. 3.



**FIG. 6.** Histogram of the RMS ratios of remnant noise processed with dynamic reference noise to that processed with constant reference noise for a total of 164 runs from 14 subjects.