A Technique to Reduce Motion Artifact for Externally Triggered Cine-MRI(EC-MRI) Based on Detecting the Onset of the Articulated Word with Spectral Analysis

Yasuhiro SHIMADA1*, Hironori NISHIMOTO1,2, Takanori KOCHIYAMA1,3, Ichiro FUJIMOTO1, Hiroaki MANO1,4, Shinobu MASAKI1, and Kenya MURASE5

1Brain Activity Imaging Center, ATR-Promotions Inc.
2–2–2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619–0288; Japan
3ATR Neural Information Analysis Laboratories, Kyoto, Japan
3The Hakubi Project, Primate Research Institute, Kyoto University, Kyoto, Japan
4ATR Cognitive Mechanisms Laboratories, Kyoto, Japan
5Department of Medical Physics and Engineering, Division of Medical Technology and Science, Course of Health Science, Osaka University Graduate School of Medicine, Osaka, Japan

(Received April 13, 2012; Accepted September 6, 2012)

One issue in externally triggered cine-magnetic resonance imaging (EC-MRI) for the dynamic observation of speech organs is motion artifact in the phase-encoding direction caused by unstable repetitions of speech during data acquisition. We propose a technique to reduce such artifact by rearranging the k-space data used to reconstruct MR images based on the analysis of recorded speech sounds.

We recorded the subject’s speech sounds during EC-MRI and used post hoc acoustical processing to reduce scanning noise and detect the onset of each utterance based on analysis of the recorded sounds. We selected each line of k-space from several data acquisition sessions and rearranged them to reconstruct a new series of dynamic MR images according to the analyzed time of utterance onset. Comparative evaluation showed significant reduction in motion artifact signal in the dynamic MR images reconstructed by the proposed method. The quality of the reconstructed images was sufficient to observe the dynamic aspects of speech production mechanisms.

Keywords: acoustic noise reduction, cine-MRI, external trigger, k-space rearrangement, motion artifact

Introduction

Magnetic resonance (MR) imaging has often been used in phonetic and speech science research because it provides excellent image contrast of the soft tissues of the speech organs without negatively affecting the subject. In early applications to study human speech mechanisms, MR imaging was used primarily in the observation of sustained vowels or sustainable consonants.1–4 However, the production of speech sounds by the coordinated movements of several organs, including the tongue, mandible, soft palate, and larynx, has required development of a dynamic imaging technique to visualize those fluid movements with sufficient spatial and temporal resolution.

Dynamic MR imaging techniques developed for speech production studies can be classified into 3 types. The first is a continuous imaging method that employs a single-shot sequence,5–8 the second rearranges the k-space,9–11 and the third utilizes trigger gating.

The first type has included several fast imaging techniques, such as fast low angle shot (FLASH), fast spin echo (SE), parallel imaging, and spiral MR imaging. However, acquisition of spatial resolution of one mm limits temporal resolution to only approximately 10 frames per second (fps). Narayanan’s group achieved dynamic MR imaging with temporal resolution exceeding 20 fps by combining a spiral scan and a sliding window technique at the...
expense of an in-plane spatial resolution of 2.7 mm. To date, this approach has not yielded sufficiently high spatial and temporal resolution for organ imaging.

The second type involves rearrangement of the k-space. The subject repeats the utterance of a word or a behavior in random time during MR imaging, and images are reconstructed based on post hoc rearrangement of the raw data in the k-space using phase information of the moving object or recorded speech sounds. Using this technique, Mohammad and colleagues reported the first dynamic images for speech production studies with temporal resolution of approximately 50 fps for short repetitive utterances of the speech material /pasi/. To obtain sufficient raw data to rearrange the k-space, they randomly repeated MR imaging scans with 2.8 s for slice acquisition during the subject’s repetitions. However, this method requires manual operation to select raw k-space data based on acoustic analysis of the recorded speech during the MR scan, and selection of appropriate raw k-space data requires a sufficiently large data pool to allow for randomly timed speech during the imaging session. Acquisition of sufficient data requires a long MR scan that can fatigue the subject and degrade dynamic image quality. Therefore, this technique might not be best for investigating the mechanisms of articulatory movements during speech production.

The third type is a trigger gating method. Voldvik and associates used electrocardiogram (ECG) pulse data as a gating trigger to synchronize MR imaging scan and speech production time, and Masaki and colleagues improved this technique by replacing ECG gating with external triggering (EC-MRI, externally triggered cine-MRI). The synchronized trigger generator produces rhythmic sounds to guide the speech time of the subject’s repetitive utterances and triggering signals to indicate the scan time for imaging. Use of this technique could provide sufficient spatial and temporal resolutions in dynamic MR imaging. For instance, it could obtain dynamic images of approximately 200 fps with a spatial resolution of one mm in a single slice and 3-dimensional (3D) dynamic images of 30 fps in multi-slice scanning. EC-MRI could also be combined with tagged MR imaging for quantitative dynamic evaluation of the articulatory movements of speech organs. However, a remaining problem is inconsistency in the subject’s repetitive utterances that leads to motion artifact in the resulting movie of MR images. At present, we can only reduce this motion artifact using classical and passive methods, such as averaging data obtained in multiple EC-MRI sessions, which requires more utterance repetitions that might fatigue subjects.

To overcome this issue, a technique is needed to reduce motion artifacts with a minimum number of utterances. One possible breakthrough technique combines EC-MRI and k-space rearrangement; appropriate raw MR imaging data for k-space is selected using speech timing information for each speech repetition during EC-MRI data acquisition, and k-space data is rearranged for image reconstruction. The key issue with this method is how to select the appropriate raw data for rearranging the k-space data.

We propose a 2-step approach to reduce motion artifacts in EC-MRI. The first is to detect the time of utterance onset from the sound of recorded speech during acquisition of MR imaging data by reducing the acoustic noise of imaging from the sound of recorded speech. The second step is to select and rearrange k-space data based on evaluation of the time of utterance onset. We evaluated the effect of the proposed method on the improvement of dynamic MR imaging quality by comparing the motion artifact signal of each frame of reconstructed MR imaging between the original and rearranged MR images.

Materials and Methods

EC-MRI employs an external trigger for MR imaging and a rhythmic sound to guide the subject in repeating utterances in synchronization with the scan. However, inconsistency of utterance timing lead to the variation of timing of data in the k-space for identical frame resulting in the motion artifact in each frame of the MR image movie.

Figure 1 (a, c) shows examples of typical motion artifacts produced from a session in which the subject repeated speech material at random times. There is significant motion artifact horizontally or in the direction of phase encoding, and it is difficult to detect outlines of the lip, tongue, and soft palate. In contrast, successfully synchronized images (Fig. 1b, d) demonstrate a clear outline of these organs and internal structures without noticeable motion artifact.

We propose a technique to improve the quality of dynamic MR imaging by minimizing temporal variation in raw data for the k-space. This requires selection of appropriate data based on analysis of onset time of each utterance from a data pool obtained during several MR imaging sessions.

Figure 1 (e) is a flowchart of the proposed method. Data processing was divided into 2 stages—
motion artifacts and image processing. In the sound processing stage, we applied an acoustic noise reduction (NR) process to recorded speech sound data and used the noise-reduced speech sound to detect utterance onset. For image processing, we selected appropriate raw data from MR imaging based on the speech time detection process and rearranged k-space data to reconstruct dynamic MR images. To evaluate the proposed method, we compared the motion artifact signal of each frame of reconstructed MR imaging between the original and rearranged MR images.

Data acquisition

The ATR Review Board Ethics Committee and ATR-Promotions Review Board Safety Committee approved the study protocol, and we obtained written informed consent from a volunteer subject.

The experiment was performed using a 3-tesla MR imaging system (MAGNETOM Verio, Siemens, Erlangen, Germany) with a combination of a rear 2-channel neck array coil and a 4-channel small flexible coil. To restrain the translational motion of the subject’s head, we placed sponge foam into the space between the subject’s head and the head holder (Fig. 2a).

We used a FLASH sequence with parameters: field of view (FOV), 256 mm; imaging matrix, 128 × 128 (with zero-fill interpolation); repetition time (TR), 33 ms; echo time (TE), 1.68 ms; flip angle, 20°; slice thickness, 5 mm; trigger interval (T-T), 1000 ms; delay time, 26 ms; acquisition window, 950 ms; and number of frames, 28. Note that the last 4 parameters are used for this particular EC-MRI technique.

Trigger interval (T-T) refers to the time between triggering pulses to initiate the FLASH sequence and corresponds to the R-R interval for ECG gating. Delay time is the period between triggering and the time of acquisition of raw data for the first frame of EC-MRI. The acquisition window is the time in which the raw data for EC-MRI is acquired and should be shorter than the trigger interval less the delay time. Number of frames refers to the frames of MR imaging to be reproduced in the EC-MRI.

We employed a single microphone to record speech sound from the subject during the MR imaging scan, as proposed by NessAiver’s group, i.e. an optical microphone (FOM-OPTICS 1140; Optoacoustics Ltd., Or-Yehuda, Israel) was mounted on the flexible coil so as to place it in front of the subject’s mouth (Fig. 2a).

Through an air tube-type headset (Kobatel, Kanagawa, Japan), the subject was presented with an amplified rhythmic sound with a T-T interval of 1,000 ms to signal the time for utterance. The rhythmic sounds comprised a tone burst followed by 2 noise bursts with a duple rhythm, each lasting 100 ms (Fig. 2b). The interval between the first tone burst and first noise burst was 500 ms, and that between the first and second noise burst was 250 ms. Simultaneously, a trigger pulse converted into a transistor-transistor logic (TTL) signal of 5 V was provided to drive the scanner system.

Three sessions of EC-MRI comprising 128 repetitions were conducted for speech material /aki/. Each session consisted of 12 subsessions including
Fig. 2. (a) Schematic layout of equipment for magnetic resonance (MR) imaging experiments: receiver coils, headset, and optical microphone. The equipment was arranged as shown. (b-d) Recorded wave data. (b) Rhythmic sound to guide subject to speak. (c) Gradient pulse of slice selection. (d) Speech sound data recorded by optical microphone; the sound overlapped loud acoustic scanner noise. Each of the 40,000 points of speech sound data was extracted from the upstroke time of the slice-selecting gradient pulse (▲).

10 utterance periods and a 13th subsession with only 8 utterance periods to accumulate the total of 128 raw data. In each subsession, the utterance periods were preceded by 2 rehearsal periods, during which MR imaging data was not acquired and the subject was not required to produce the speech material. Each utterance or rehearsal period was identical to the period during the T-T interval (1,000 ms).

We simultaneously recorded the rhythmic sound to guide the utterances (Fig. 2b), the wave form of the slice-selecting gradient current (Fig. 2c), and the speech sound contaminated with acoustic scanner noise (Fig. 2d) into a multi-channel recording system (PowerLab, ADInstruments Pty Ltd., Bella Vista, NSW, Australia) with a sampling frequency of 40 kHz.

Data processing

We processed data (Fig. 1e) primarily using a technical computing language MATLAB R2010a (MathWorks Inc., Natick, MA, USA) and used ImageJ software (http://rsbweb.nih.gov/ij/) to create a masking pattern to restrict the region for quality evaluation of dynamic MR images.

Acoustical processing to reduce scanner noise (NR)

During acquisition of EC-MRI data, the scanner generates acoustic noise at a fixed interval in response to alteration of the gradient magnetic field and determined by the interval of gradient pulse generation for slice selection. Several studies have proposed techniques to reduce such acoustic noise for acquisition of general, dynamic, and functional MR imaging data. In this study, we applied a segmental subtraction approach explained below to reduce the acoustic noise in the recorded speech sound to detect the onset time of utterance of the speech material, /aki/.

We first determined the upstroke of the first gradient pulse for slice selection as the onset of each data acquisition period, which is identical to the T-T interval (1,000 ms) (▲, Fig. 2c). We then extracted speech sound data starting from the onset of each data acquisition period for the following noise reduction process (Fig. 2d), in which we divided an entire session of speech sound data into 128 segments of 40,000-point speech sound data (Vn, where n = 1 to 128) with a data acquisition period of 1,000 ms.

After subdividing each segment of speech sound data (Vni) into 8 parts with 5,000 points of sound data (Vnii, where i = 1 to 8), we then extracted 10,000 points of recorded data containing only scanner noise without speech data from the original 40,000 points of speech sound data (Nn, Fig. 2d). We then calculated the serial cross correlation coefficients between each Vnii and Nn to find the
time lag with the maximum correlation coefficient and finally subtracted the noise component (Vn) from the original speech sound data (Vnsi) at a lag time point to obtain the noise-reduced sound data. Although noise-reduced data obtained in this manner still contain residual noise at both ends of the Vn, these periods do not affect the detection of utterance onset.

**Detection of utterance onset**

Figure 3 shows speech sound wave forms with their spectrograms for (a) pre- and (b) post-NR processing. To detect the utterance onset of the initial vowel /a/ in speech material /aki/ from the noise-reduced speech sound data, we first calculated the linear predictive coding (LPC) spectra of post-NR data to extract formant frequency components from the spoken material. The gray and black lines in Figure 3 show the LPC spectra calculated from the periods of (c) preutterance, (d) production of vowel /a/, and (e) /i/ in /aki/ for pre- and post-NR sounds. These figures show the effect of NR processing on extraction of the clear LPC spectra. In Figure 3d, the inverted triangles (▼) indicate the 4 formants for /a/, and the first formant for /a/ is seen to have maximum power. Therefore, we selected the rising onset of the power for this formant as the index of utterance onset.

In accordance with the previous decision to select the first formant as the cue of utterance onset, we then calculated the sum of power (SOP) for the frequency range from 300 to 900 Hz, which includes the first formant frequency for /a/. For the calculation, we used a window width of 50 ms and window shift interval of one ms.

We then normalized the obtained time courses of the SOP by converting them into values relative to the maximum of SOP in each period of repetitive production of /aki/ (Fig. 3f-h). Finally, we determined utterance onset by the time when the relative SOP value reached more than 1% of the maximum SOP for noise-reduced speech sound data.

**Image reconstruction using the selected and rearranged k-space data**

Figure 4 shows: (a-c) the utterance onset time of /a/ in each repetitive production of /aki/ (horizontal order) for each of 3 MR imaging experiment sessions (vertical order); (e-g) the histograms of onset time variation of /a/ for each of the 3 sessions; (d) selected time data nearest to the average onset time of utterance onset, which was calculated from these data shown in (a-c); and (h) selected time data from the 3 sessions.

In the reconstruction process to obtain a set of clear dynamic MR images, phase-encoding data in the k-space should be selected from all phase-encoding data to minimize variation in utterance onset. To satisfy this restriction, we chose 128 raw data for rearrangement of the k-space among the 3-session data (Fig. 5a) corresponding to the selected onset time data (Fig. 4d), reconstructed MR imaging for each frame (Fig. 5b) from the rearranged raw data (Fig. 5a), and presented a set of MR images sequentially as a movie of MR images (Fig. 5c).

**Image evaluation**

For quantitative evaluation of the proposed method on motion artifact reduction, we assessed reduction in motion artifact signal, adopting standard deviation (SD) as an index for the motion artifact. We considered 2 issues—normalization of the signal intensity of the MR imaging movie and the area of each image in which SD was calculated.

Differences in tuning and scaling factors between sessions even on the same scanner influence signal intensities. To exclude such influence on the evaluation of motion artifact, we normalized the signal intensities of MR images for 3 sessions and for a newly rearranged version to achieve an identical average value (50) and SD (10) for each series of dynamic MR images as follows:

\[ NI(i,x,y) = 50 + 10(SI(i,x,y) - \mu) / \sigma_S, \]

where \( NI(i,x,y) \) is the normalized intensity of the x, y coordinate in i-th frame; \( SI(i,x,y) \) is the signal intensity of the x, y coordinate in i-th frame; and \( \mu \) and \( \sigma_S \) are the mean and standard deviation of image intensity for the series.

Determining the area in which to evaluate the motion artifact signal is crucial. Motion artifacts usually overlap the target objects (in this study, speech organs) because they appear horizontally at the same level as the moving object. Thus, the SD value is influenced not only by the motion artifact signal but also by the changing images of the moving objects, and it is difficult to evaluate only the motion artifact signal in the area that includes the target object. To exclude the influence of moving objects on the evaluation of motion artifacts, we calculated the motion artifact signal only in the white area of the “masking pattern” shown in Figure 6. We selected this area to include only the outside of the body at the same horizontal level as the moving speech organs and to exclude the inside of the body, which includes the moving objects.

We calculated the SD of the normalized signal intensity within the background area defined above for each image frame and compared the resulting...
Results

As described, Figure 3 illustrates the improved detection of utterance onset from recorded sound using the NR process. Utterance onset was obscured in pre-NR data by MR scanning noise that masked the sound of the subject’s speech (Fig. 3a). Use of the NR process greatly reduced scanner noise and recognition of formant frequencies (▼, Fig. 3d). Though we applied a rather simple subtraction approach for the NR process, spectra for the preutterance period (Fig. 3c) indicated a 20- to 40-dB reduction in sound pressure.

Figure 3 (f-h) shows the time course of the normalized SOP used in utterance onset detection, as described above. Onset of speech material production varied by 100 ms at most, which is comparable to 3 frames of dynamic MR imaging under our experimental conditions.

Figure 4 shows (a-c) distributions of utterance onset for 3 sessions and (e-g) histograms of their distribution. The averages and SDs of the utterance onset times were $467.8 \pm 25.7$ ms for the first session (Fig. 4a, e), $474.3 \pm 24.5$ ms for the second session (Fig. 4b, f), and $478.0 \pm 23.0$ ms for the third session (Fig. 4c, g). The overall average and SD for the 3 sessions were $473.4 \pm 24.7$ ms. Fig. 4d shows the selected onset time for each phase from the 3 sessions and (4h) the distribution histogram. We selected phase data using the time of onset nearest the overall average time of onset. As a result, the average time for the selected data became 476.1 ms, and the SD range of utterance onset was reduced to nearly half, from 24.7 to 12.9 ms.

Figure 6 shows the time course of the motion artifact signal calculated as described above. We excluded the first and final frames from this motion artifact evaluation because the k-spaces for the frames at both ends were not filled in the new dynamic series. The graph shows that motion artifact signal starts at a higher value and decreases until about the time of utterance onset between frames 2 and 15. This might be caused by the flashing effect that usually emerges in the early frames of dynamic MR imaging. The abrupt increase in SD value observed from #17 to #20 could be due to the rapid and inconsistent movements of the tongue in producing the consonant /k/ in /aki/.

Figure 7 shows examples of the images of dynamic MR imaging. Frame #12 shows the preutterance stage; #15, onset of production of initial vowel...
Fig. 4. (a–c) Time of utterance onset of initial word sound /a/ for the (a) first, (b) second, and (c) third sessions. (d) Selected time data nearest to the average onset time for all 3 sessions; gray circle, selected data from first session; open triangle, from second session; and gray square, third session. Their distribution histograms are shown in the right panel (e–h).

Fig. 5. New k-space data (a) were selected and rearranged from the original k-space data for 3 sessions based on analyses of utterance onset times. Then, new image (b) and dynamic series (c) were reconstructed from rearranged k-space data (a).

Fig. 6. Time course of standard deviation (SD) values as an index of motion artifact for the white part of the masking pattern in the figures. The graph shows higher SD values in the first part of the time course because of the flashing effect in the early stage of the magnetic resonance (MR) imaging scan. The motion artifact signal for each frame is lower for all frames of dynamic MR imaging reconstructed by the proposed technique compared to those based on the original 3 sessions.
Fig. 7. Comparison of reconstructed dynamic magnetic resonance (MR) imaging between the original 3 sessions and the rearranged version for frames of #12 (a-d), #15 (e-h), and #18 (i-l). Motion artifacts observed in the front of the mouth are obviously reduced by the proposed method compared to the images reconstructed from the original sessions.

/a/; and #18, the mid-point of transition from /a/ to the consonant /k/. Figure 7d, h, and l show the reconstructed images based on the proposed method. Comparison of these images with those reconstructed using the original data indicates a fair reduction in motion artifacts in the front of the mouth.

Discussion

EC-MRI is useful for observing articulatory movements because its temporal and spatial resolution are sufficient to visualize the rapid movement and fine structure of human speech organs. However, motion artifacts result from inconsistencies among the many utterance repetitions this technique requires. In particular, image quality is seriously damaged by contamination of MR imaging data from irregular motion during acquisition that affects the central region of the k-space. In this study, we proposed a new method to improve the quality of the resulting images of EC-MRI by reducing acoustic noise in the subject's speech sound data, detecting exact onset time of each utterance, and rearranging the k-space based on the detected onset time.

The rearrangement of k-space data is a promising technique in speech research using MR imaging. Nishimoto and his associates used this technique to create a series of moving MR images that demonstrated tongue tip movements during production of the alveolar trill with extremely high temporal resolution at 1,290 fps. It is impossible to synchronize speech production with MR scanning procedures to visualize such a high-speed object during speech production, but a moving target vibrating at a constant frequency can be visualized using a rearrangement technique. Such application requires detection of the exact time of some event or the phase of the moving object to reduce motion artifacts.

Several studies of acoustic noise reduction and control for MR imaging could reduce sound pressure approximately 20 to 40 dB using a subtraction algorithm with a cross-correlation or least-mean-square algorithm with several adaptive filters. Most applied the methods to sound data recorded using a multiple microphone system. In contrast, we used a cross-correlation-based segmented subtraction technique on sound data recorded using a single microphone system and were able to reduce noise from the MR scanner that contaminated speech sound to almost the same level as in the previous reports.

Acoustic NR has often been tried for MR imaging but rarely applied to detect utterance onset.
Remarkably, NessAiver and colleagues\(^\text{16}\) applied the NR technique to sound recorded during acquisition of MR imaging data to improve the quality of the reconstructed dynamic MR images. To reconstruct tagged dynamic MR images, they used complementary spatial modulation of magnetization (CSPAMM), which requires multiplying together the subtraction of two pairs of images. However, images obtained using this method are significantly degraded by small misalignments. The authors attempted to avoid reconstruction of degraded images by adopting a temporal alignment of k-space data using the recorded speech sound data as a reference, but the quality of reconstructed images was disappointing. They attributed this to temporal variability, such as that of speech onset relative to external triggering of MRI scan and that from inconsistency in sound duration.\(^\text{16}\) Perhaps consistent speech time would have been achieved if they presented the subject with a rhythmic sound to signal the onset of the utterance and the onset of the following syllables, thereby improving image quality of their tagged MR imaging movie.

In the current study, rearrangement of raw data on reconstructed dynamic MR images yielded consistently lower motion artifact signal in all frames of processed images than those of the 3 original sessions (Fig. 6), but the observed SD value increased abruptly in response to inconsistent movement of the tongue in producing the consonant /k/ in /aki/. These results suggest that a parallel temporal shift of raw data is insufficient to reduce motion artifact in the reconstructed MR imaging movie, especially for consonant production, which usually requires fast articulatory movements and is often inconsistent. Application of such innovative technique as dynamic time warping\(^\text{21}\) might aid adjustment of temporal coordination based on evaluation of similar acoustical characteristics of speech sounds.

At present, it is difficult to evaluate inconsistency of utterance times during EC-MRI data acquisition; we must finish the session to confirm image quality. Real-time evaluation of speech inconsistency and its effect on image quality would help reduce the number of sessions and consequent subject fatigue. As well, fewer utterance repetitions could improve consistency of speech production and image quality. Further application studies of the proposed techniques to real-time monitoring system will contribute significantly to research in phonetics and speech science.

Our study is limited because we present findings from only one subject for the utterance of one particular speech material, /aki/. Application of the proposed method to speech research requires some considerations such as subject selection to acquire high quality MR imaging movie, detection of onset of various speech materials, and image uniformity for comparing movies between subjects.

The quality of the MR imaging movie depends on the subject’s capacity to repeat speech sounds and maintain regular speech timing for reproduction. Our subject’s maintenance of relatively consistent speech timing during acquisition of raw data may explain the success of our proposed method in improving the movie images reconstructed from k-space data in the 3 sessions of raw data acquisition. A method for quantitative evaluation of the relationship between subject performance and quality of MR movie images should be established for selecting subjects who can repeat speech production consistently.

Because the speech material of our study, /aki/, starts with a vowel, utterance onset could be effectively determined using linear predictive coding (LPC) analysis. However, the LPC technique is not suitable for detecting utterance onset when the initial sound is a consonant, especially a word-initial voiceless fricative consonant with little sound power. Further study is required to develop a method of detecting onset time for a variety of speech materials.

For studies that compare individual speech performance among multiple subjects, consistency of subject position and quality of reconstructed MR movie images are required between sessions and/or between subjects. In order to realize such consistency, suitable technique for alignment of the subject, optimum receiver coil, and imaging sequences should be selected. Moreover, since parallel imaging become a primary imaging technique in the future because of its short scanning time, some issues will likely arise relating to the technique to keep the uniformity of images reconstructed from raw data detected by multiple coils.

**Conclusion**

We propose a technique for EC-MRI that reduces motion artifact by rearranging the k-space. We detected the onset of each utterance from noise-reduced speech sound recorded during an EC-MRI scan, selecting phase-encoding data in the k-space acquired from a database of several MR imaging sessions and rearranging the data according to time of utterance onset. Using the proposed method, we found a significant decrease in motion artifacts by quantitative evaluation of motion artifact signal for each frame of reconstructed dynamic MR imag-
ing as well as qualitative evaluation by visual inspection. This technique will be useful for investigating the dynamics and mechanisms of speech production.

Acknowledgments

This study was partially supported by KAKENHI (#22700195).

References

1. Rokkaku M, Hashimoto K, Imaizumi S, Niimi S, Kiritani S. Measurement of the three dimensional shape of the vocal tract based on the magnetic resonance imaging technique. Ann Bull RILP 1986; 20:47–54.
2. Baer T, Gore JC, Boyce S, Nye PW. Application of MRI to the analysis of speech production. Magn Reson Imaging 1987; 5:1–7.
3. Narayanan SS, Alwan AA, Haker K. An articulatory study of fricative consonants using magnetic resonance imaging. J Acoust Soc Am 1995; 98: 1325–1347.
4. Masaki S, Akahane-Yamada R, Tiede MK, Shimada Y, Fujimoto I. An MRI-based analysis of the English /r/ and /l/ articulations. Proc 4th Int Conf Spoken Language Processing 1996; Volume 3:1581–1584.
5. Crary MA, Kotzur IM, Gauger J, Gorham M, Burton S. Dynamic magnetic resonance imaging in the study of vocal tract configuration. J Voice 1996; 10:378–388.
6. Demolin D, George M, Lecuit V, Metens T, Soquet A, Raeymaekers H. Coarticulation and articulatory compensations studied by dynamic MRI. Proc 5th Eurospeech 97 Conference 1997; Rhodes, Greece, 31–34.
7. Mady K, Sader R, Zimmermann A, et al. Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction. Proc 4th International Speech Motor Conference 2001; Nijmegen, Netherlands, 142–145.
8. Narayanan S, Nayak K, Lee S, Sethy A, Byrd D. An approach to real-time magnetic resonance imaging for speech production. J Acoust Soc Am 2004; 115:1771–1776.
9. Mohammad M, Moore E, Carter JN, Shadle CH, Gunn SR. Using MRI to image the moving vocal tract during speech. Proc Eurospeech 97 1997; 14: 2027–2030.
10. Nishimoto H, Shimada Y, Fujimoto I, Masaki S. [Super high temporal resolution MRI movie for observing dynamic aspect of speech organs]. IEICE Tech Rep 2009; 109:SP2009–26, 17–22. [in Japanese]
11. Nishimoto H, Shimada Y, Fujimoto I, Masaki S, Murase K. Observation of the tongue movement producing alveolar trill by high frame rate MRI-Movie. Proc 9th Int Sem Speech Production (ISSP), Montreal, Canada, June 20–23, 2011: 33–40.
12. Foldvik AK, Husby O, Kvaerness J, Nordli IC, Rinck PA. MRI (Magnetic Resonance Imaging) film of articulatory movements. Proc Int Conf Spoken Language Processing 1990; 1:421–422.
13. Masaki S, Tiede M, Honda K, et al. MRI-based speech production study using a synchronized sampling method. J Acoust Soc Jpn (E) 1999; 20: 375–379.
14. Takemoto H, Honda K, Masaki S, Shimada Y, Fujimoto I. Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. J Acoust Soc Am 2006; 119:1037–1049.
15. Stone M, Davis EP, Douglas AS, et al. Modeling the motion of the internal tongue from tagged cine-MRI images. J Acoust Soc Am 2001; 109:2974–2982.
16. NessAiver MS, Stone M, Parthasarathy V, Kahana Y, Paritsky A. Recording high quality speech during tagged cine-MRI studies using a fiber optic microphone. J Magn Reson Imaging 2006; 23:92–97.
17. Chen CK, Chueh TD, Chen JH. Active cancellation system of acoustic noise in MR imaging. IEEE Trans Biomed Eng 1999; 46:186–191.
18. Bresch E, Nielsen J, Nayak K, Narayanan S. Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans (L). J Acoust Soc Am 2006; 120:1791–1794.
19. Ramachandran VR, Panahi IM, Milani AA. Objective and subjective evaluation of adaptive speech enhancement methods for functional MRI. J Magn Reson Imaging 2010; 31:46–55.
20. Atal BS, Schroeder MH. Predictive coding of speech signals. IEEE Proc Conf Commun Process 1967:360–361.
21. Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans ASSP 1978; 26:43–49.