Magnetic Resonance in Medicine

Dynamic off-resonance correction for spiral real-time MRI of speech

Yongwan Lim | Sajan Goud Lingala | Shrikanth S. Narayanan | Krishna S. Nayak

Ming Hsieh Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, California

Correspondence
Yongwan Lim, 3740 McClintock Ave, EEB 400, University of Southern California, Los Angeles, CA 90089-2564.
Email: yongwanl@usc.edu

Purpose: To improve the depiction and tracking of vocal tract articulators in spiral real-time MRI (RT-MRI) of speech production by estimating and correcting for dynamic changes in off-resonance.

Methods: The proposed method computes a dynamic field map from the phase of single-TE dynamic images after a coil phase compensation where complex coil sensitivity maps are estimated from the single-TE dynamic scan itself. This method is tested using simulations and in vivo data. The depiction of air–tissue boundaries is evaluated quantitatively using a sharpness metric and visual inspection.

Results: Simulations demonstrate that the proposed method provides robust off-resonance correction for spiral readout durations up to 5 ms at 1.5T. In -vivo experiments during human speech production demonstrate that image sharpness is improved in a majority of data sets at air–tissue boundaries including the upper lip, hard palate, soft palate, and tongue boundaries, whereas the lower lip shows little improvement in the edge sharpness after correction.

Conclusion: Dynamic off-resonance correction is feasible from single-TE spiral RT-MRI data, and provides a practical performance improvement in articulator sharpness when applied to speech production imaging.

KEYWORDS
off-resonance correction, real-time MRI, speech production, spiral

1 | INTRODUCTION

Real-time MRI (RT-MRI) has become a valuable tool for speech production research[1-3] and is now a preferred tool in speech science to alternative imaging modalities including x-ray microbeam,[4] electromagnetic articulography,[5] and ultrasound.[6] RT-MRI provides a non-invasive capture of the dynamics of deep articulatory structures (e.g., pharynx, glottis, and epiglottis) during speech production and allows for arbitrary imaging planes. In this context, spiral RT-MRI scanning is desirable because it allows for a time-efficient acquisition, given that spirals can provide higher spatio-temporal resolution than alternative schemes.[1]

A key drawback of spiral MRI is signal loss and/or blurring artifacts that result from field inhomogeneity, also called "off-resonance".[7] This can be significant at air–tissue interfaces because of their magnetic susceptibility difference ($\Delta\chi = 9.41$ parts per million).[8] Furthermore, these artifacts near the air–tissue boundaries[9] are more pronounced with long spiral readout or at high field strength MRI scanners. To mitigate this artifact, current RT-MRI studies for speech production are most often conducted using short duration readouts ($\sim$2.5 ms) and at lower field strength (1.5T) MRI scanners.[10-12]

Off-resonance artifacts have significant potential impact on the analysis of articulator dynamics, which is of prime

A preliminary version of this work was presented at the 25th Annual Meeting of the ISMRM Scientific Sessions, Honolulu, HI, 2017. Abstract #4017.

interest in speech science. The articulators of interest include the surfaces of the lips, tongue, hard palate, soft palate (velum), and structures along the pharyngeal airway. These are located at air–tissue interfaces and therefore are vulnerable to the artifacts. Previously used speech RT-MRI biomarkers, such as average pixel intensity[13,14] in regions of interest (ROI), are prone to error because of artefactual airway area perturbation. Any temporally varying blur of soft tissues can result in changes in the detected patent airway and will disrupt the estimation of constriction kinematics, such as timing in consonant production.[13] Air–tissue boundary segmentation[15-17] is required as a pre-processing step in acquiring vocal tract area functions[18] and suffers in the presence of ambiguous boundaries with poor contrast. Velopharyngeal insufficiency[19-23] is caused by incomplete closure between the soft palate and the posterior and lateral pharyngeal walls, and its assessment can be hampered by signal loss near the soft palate.

Several deblurring methods in spiral scanning have been proposed in the literature,[24-30] most of which require a measurement of a frequency offset image, also called a "field map."[24-26] A previous study applied this approach to spiral RT-MRI of vocal tract[31] where spirals with 2 different TEs were obtained in an interleaved fashion, and a dynamic field map was estimated using each pair of consecutive images. This field map-based method showed improvement of image quality in the tongue and soft palate. The reconstructed images, however, could suffer from flickering artifact between consecutive images reconstructed with different TEs. This scheme also requires a compromise in temporal and/or spatial resolution[31] and is not applicable to previously collected single-echo-time data.

An alternative approach is to estimate the field map directly from the data set itself, known as "auto-focus."[27-30] Auto-focus methods use an image-domain focus metric that provides local information about the presence of residual off-resonance artifacts based on the off-resonance point spread function (PSF). A widely used metric is the absolute value of the imaginary component of the image (after correcting for a coil phase) at an image location.[28] It assumes that the imaginary component should be zero when the local effects of off-resonance have been corrected. These methods have shown comparable results to the methods that acquire the field map. However, these are computationally demanding and performance depends on the focus metric used and can be sensitive to experimental factors, such as MRI sequence parameters, SNR, and the accuracy of coil sensitivity maps (especially their phase). Additionally, spurious minima of the focus metric can occur as the range of off-resonance at air–tissue interfaces ($\sim$600 Hz at 1.5T) is large enough to produce more than 1 cycle of phase accrual ($>2\pi$) even during a short spiral readout ($\sim$2.5 ms).[27,32,33]

In this work, we present a simple dynamic off-resonance estimation method for spiral imaging where a dynamic field map is directly estimated from the phase of single-TE dynamic images after a coil phase compensation. We estimate complex coil sensitivity map from the single-TE scan itself. Our approach does not require a dynamic two-echo measurement of a field map nor the use of a focus metric. Therefore, it can be performed on conventional real-time spiral data without the need for additional scanning and is not computationally intensive. We evaluate this method using simulations and on an existing multi-speaker data set of speech RT-MRI. We demonstrate improvements in the depiction of air–tissue boundaries quantitatively using an image sharpness metric and visual inspection, and the practical use of this method on the boundary segmentation and distance metric as a use case example.

2 | THEORY

2.1 | Spiral imaging in the presence of the field inhomogeneity

In spiral MRI, ignoring relaxation and noise, the signal equation of an object with a transverse magnetization $m_0(\boldsymbol{r})$ is given by

$$s(\tau) = \int_{\boldsymbol{r}} m(\boldsymbol{r}) e^{-j2\pi f(\boldsymbol{r})\tau} e^{-j2\pi \boldsymbol{k}(\tau)\cdot\boldsymbol{r}} d\boldsymbol{r}, \qquad (1)$$

where $\tau \in [0, T_{read}]$ is time variable defining $\tau = 0$ as the start of the readout; T_{read} is the readout duration. $\boldsymbol{r}$ and $\boldsymbol{k}(\tau)$ are the spatial coordinate and the k-space trajectory, respectively. In $m(\boldsymbol{r}) = m_0(\boldsymbol{r})C(\boldsymbol{r})e^{-j2\pi f(\boldsymbol{r})TE}$, $f(\boldsymbol{r})$ is the off-resonance frequency presented at $\boldsymbol{r}$, and $C(\boldsymbol{r})$ is the complex coil sensitivity map.

Consider the image signal ($\tilde{m}(\boldsymbol{r})$) reconstructed from $s(\tau)$ without off-resonance correction as follows:

$$\tilde{m}(\boldsymbol{r}) = \int_{\boldsymbol{r}'} m(\boldsymbol{r}') \; PSF\left(\boldsymbol{r}', \boldsymbol{r}; f(\boldsymbol{r}')\right) d\boldsymbol{r}', \qquad (2)$$

where $PSF\left(\boldsymbol{r}', \boldsymbol{r}; f(\boldsymbol{r}')\right) = \int_0^{T_{read}} W(\tau) e^{-j2\pi\{f(\boldsymbol{r}')\tau + \boldsymbol{k}(\tau)\cdot(\boldsymbol{r}'-\boldsymbol{r})\}} d\tau$ is a PSF of an imaging system using a particular k-space trajectory in the presence of $f(\boldsymbol{r})$; $W(\tau)$ denotes the pre-density compensation function for the trajectory. When $f(\boldsymbol{r}) \cdot T_{read} \approx 0$, we can ignore a phase accrual because of off-resonance during the readout. Then, the PSF in Equation 2 is a sharp impulse at $\boldsymbol{r}$ so that the image signal in Equation 2 can be approximated by $\tilde{m}(\boldsymbol{r}) \approx m(\boldsymbol{r}) = m_0(\boldsymbol{r})C(\boldsymbol{r})e^{-j2\pi f(\boldsymbol{r})TE}$.

2.2 | Field map estimation in spiral imaging

Consider spiral RT-MRI, where the image time series ($m_i(\boldsymbol{r}, t)$) for i-th coil is:

$$m_i(\mathbf{r}, t) \approx m_0(\mathbf{r}, t)C_i(\mathbf{r})e^{-j2\pi f(\mathbf{r},t)TE}, \quad (3)$$

where t represents time frame, $f(\mathbf{r}, t)$ is dynamic off-resonance, and $C_i(\mathbf{r})$ is the complex coil sensitivity map that is spatially smooth and independent of time. Phase accrual during the spiral readout is ignored. Assuming that $m_0(\mathbf{r}, t)$ is real, we can compute an estimate of the dynamic field map, $\hat{f}(\mathbf{r}, t)$, as follows:

$$\hat{f}(\mathbf{r}, t) = \angle \hat{m}_0(\mathbf{r}, t)/(-2\pi TE). \quad (4)$$

where $\hat{m}_0(\mathbf{r}, t)$ denotes a coil-composite image using the optimal B1 combination,[34] which is given by

$$\hat{m}_0(\mathbf{r}, t) = \sum_{i=1}^{Nc} m_i(\mathbf{r}, t)\hat{C}_i^*(\mathbf{r}), \quad (5)$$

and $\hat{C}_i(\mathbf{r})$ is an estimate of the sensitivity maps, N_c is the number of coil components, and $\hat{C}_i^*(\mathbf{r})$ is the complex conjugate of $\hat{C}_i(\mathbf{r})$.

3 | METHODS

3.1 | Implementation of field map estimation for speech RT-MRI

Figure 1 illustrates the proposed field map estimation process. The individual coil image frames $m_i(\mathbf{r}, t)$ are first reconstructed from raw k-space $s_i(\mathbf{k}, t)$ using sliding window view-sharing with the non-uniform fast Fourier transform (NUFFT).[35] For sliding window view-sharing, reconstructions were performed every 4 spirals using a temporal window of 13 spirals (fully sampled k-space). Note that this number matches to a frame rate of dynamic images to be reconstructed in off-resonance correction, which will be described more in the "Off-resonance correction" section. The multi-coil images are then merged into composite image frames $\hat{m}_0(\mathbf{r}, t)$ based on Equation 5 using complex coil sensitivity maps, whose estimation will be discussed later. $\hat{m}_0(\mathbf{r}, t)$ is then smoothed by convolution with a 3D Hanning window ($\mathbf{r}$-t) with size $3 \times 3 \times 3$ to reduce noise, and masked by either of 0 or 1 based on a threshold (2% of maximum of the absolute squared value of the smoothed image) to control uninitialized values in air spaces that result from a lack of image signal. Consequently, a dynamic field map is estimated from the smoothed and masked images of $\hat{m}_0(\mathbf{r}, t)$ based on Equation 4.

Complex coil sensitivity maps $\hat{C}_i(\mathbf{r})$ (the 'i' subscript indicates the i-th coil element) are estimated from a temporally averaged and spatially low-pass filtered image. The individual coil image frames $m_i(\mathbf{r}, t)$ (shown in Figure 1) are averaged over time and low-pass filtered by a 2D Hanning window with size 15×15 (FWHM $\approx$ 8 pixels). Note that this low-pass filter is different from the smoothing applied to $\hat{m}_0(\mathbf{r}, t)$ and is comparable to a low-pass filter that takes
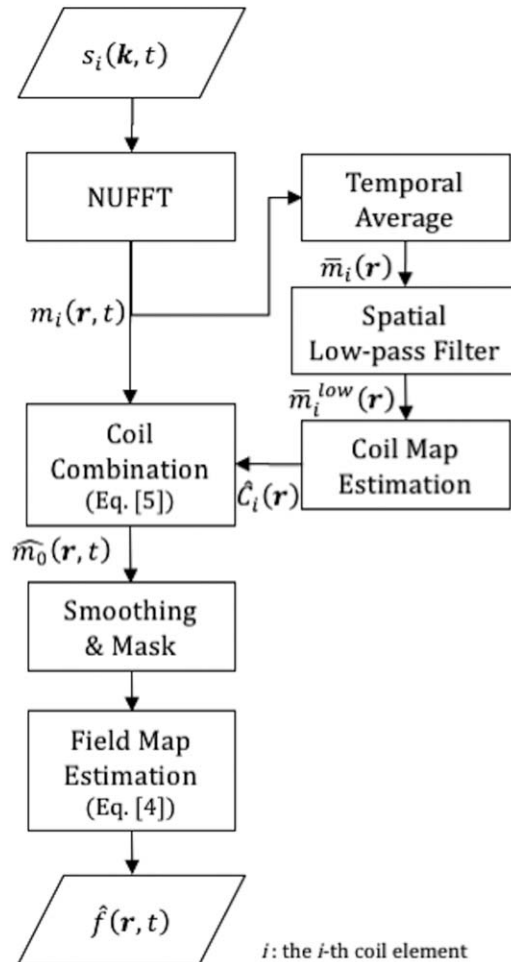

FIGURE 1 Flow chart illustrating the proposed field map estimation method. The raw image frames from individual coils are first reconstructed from the raw k-space data using view-sharing with NUFFT. The coil sensitivity maps are estimated from the multi-coil image frames after temporal average and spatial low-pass filter. The multi-coil image frames are then merged into composite image frames using the complex coil maps by Equation 5. The composite images are smoothed and masked, and a dynamic field map is estimated from the phase of the resulting image frames by Equation 4

12.5% of the central part of the k-space. These settings were chosen empirically. Then, the resultant image $\bar{m}_i^{low}(\mathbf{r})$ is used to estimate the coil map by $\hat{C}_i(\mathbf{r}) = \bar{m}_i^{low}(\mathbf{r})/\sqrt{\sum_i |\bar{m}_i^{low}(\mathbf{r})|^2}$. A drawback of this approach is that the spatially smooth portion of the time-averaged field map will be spuriously included in the coil sensitivity map and will not be corrected, which will be extensively discussed in the "Discussion" section.

3.2 | Simulation

To assess the accuracy of the proposed field map estimation, a simulation was performed with various spiral readout
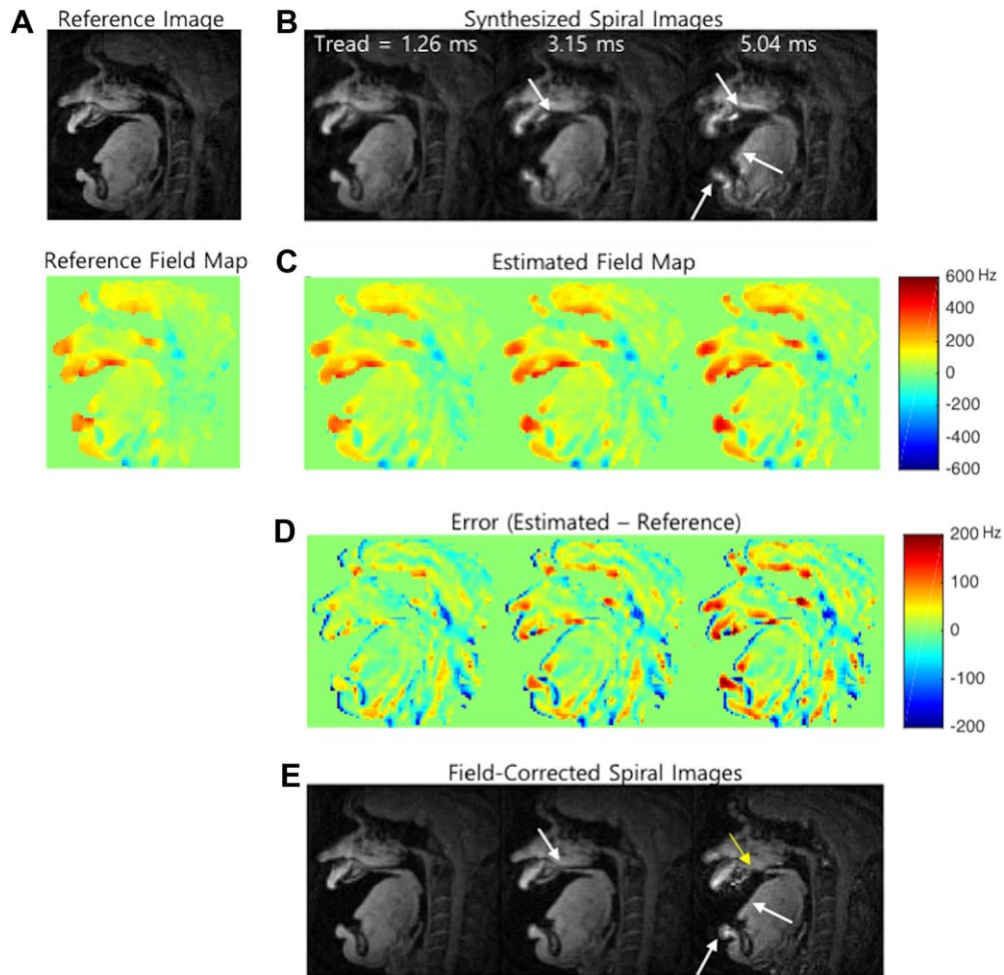
FIGURE 2 Representative simulation results. (A) A magnitude image and reference field map acquired from Cartesian dual-TE acquisition. (B) Synthesized spiral images using the magnitude image and reference field map with different readout durations (1.26, 3.15, and 5.04 ms). Off-resonance blurring is most apparent near the lips, hard palate, and tongue boundary and becomes worse with the longer readouts. (C) Field maps (unit: Hz) estimated from phase of the spiral complex images shown in (B). (D) Estimation errors in the field map (error maps amplified by a factor of 3 for better visualization). (E) Spiral images after correction for off-resonance based on the estimated field map represented in (C)

durations as follows: Cartesian images with 2 TEs ($\Delta TE = 1$ ms) were acquired from a healthy subject at 5 postures including mouth open at varying angles such as mouth fully open and mouth half open, mouth closed, and tongue tip raised to the front of the palate. For each of the postures, a reference field map was obtained from the phase difference between the images acquired at 2 TEs divided by ΔTE shown in Figure 2A. For a given spiral trajectory, spiral k-space data were synthesized from the magnitude of the Cartesian image from the first TE based on Equation 1. The reference field map was used to simulate off-resonance effects on the synthesized spiral k-space data. Those data simulations were performed with different readout durations varying from 0 ms to 6.3 ms with 0.63-ms increment. Finally, we estimated a field map from the simulated data and attempted to correct for off-resonance based on the estimated field map.

3.3 | Application to existing speech RT-MRI data

Experiments were performed on a speech RT-MRI dataset collected at our institution using a standardized vocal-tract protocol.[36] It currently contains more than 20 healthy subjects' data on a wide variety of speech tasks to capture salient, static and dynamic, articulatory characteristics of speech production as well as morphological aspects of the vocal tract.[36] Notice that the degree of blurring artifacts in their images varies depending on the subjects and speech tasks. We selected 20 subjects ($N = 20$, 10 F/10 M; age 19–31 y) with several speech tasks from the data set.

Imaging was performed using a real-time interactive imaging platform (RT-Hawk, Heart Vista, Los Altos, CA) [37] on a commercial 1.5T scanner (Signa Excite, GE Healthcare, Waukesha, WI). The body coil was used for RF transmission, and a custom 8-channel upper airway coil[12] was used

for signal reception. A 13-interleaf spiral spoiled gradient echo pulse sequence was used. Imaging was performed in the mid-sagittal plane. Imaging parameters used were: $T_{read} = 2.52$ ms, spatial resolution $= 2.4 \times 2.4$ mm^2, slice thickness $= 6$ mm, FOV $= 200 \times 200$ mm^2, TR $= 6.004$ ms, TE $= 0.8$ ms, receiver bandwidth $= \pm 125$ kHz, and flip angle $= 15°$. In addition to the automatic shimming provided by the prescan calibration from the scanner, we performed a manual adjustment of the center frequency as described in Lingala et al.[1,12] Specifically, we on-the-fly adjusted the center frequency in a way that air–tongue boundary is sharp in the mid-sagittal plane while the subject being scanned is in a neutral open-mouth position.

3.4 | Off-resonance correction

We use an iterative approach[38,39] where the off-resonance exponential term is approximated by a set of bases to improve computational speed and to reconstruct a deblurred image. We integrate this approach into a recent sparse-SENSE reconstruction method[12] that uses temporal finite difference constraint to improve time resolution in the time-series of spiral images of speech. Specifically, off-resonance exponential term shown in Equation 1 is approximated by non-exponential bases at each time frame, by using histogram principal components ($K = 40$ bins) and singular value decomposition analysis ($L = 6$) described in Equations 19 and 20 from Fessler et al.[39] The approximated bases are incorporated into the imaging model used in the sparse-SENSE reconstruction.[12] Raw k-space data and an estimated coil map are then fed into the reconstruction algorithm as inputs. In turn, it generates a corrected time-series of images. For evaluating the effectiveness of off-resonance correction, the original time-series of images were also reconstructed using the sparse-SENSE reconstruction without the modification. All the images were reconstructed with a temporal resolution of 24 ms/frame (41.66 frames/s, 4 spiral interleaves/frame, and with reduction factor $R = 3.25$). For implementation, a nonlinear conjugate gradient (CG) algorithm with NUFFT was coded using MATLAB (The MathWorks, Natick, MA) on using 8 cores on a 16-core Intel(R) Xeon(R) CPU E5-2698 v3; 2.30 GHz with 40 MB of L3 cache. The computation time was ~60 s to estimate the coil sensitivity maps and the field maps for 400 time frames from raw k-space data (~10 s long dynamic images) and 30 and 180 mins to reconstruct images without and with off-resonance correction, respectively.

3.5 | Sharpness score

We introduce an image sharpness measure to investigate the impact of the proposed method on articulator air–tissue boundaries. We quantitatively compare the metric scores between the images with and without correction. We hypothesize that the proposed method would improve the image depiction at air–tissue articulator boundaries in 2 ways —the blurred-edge width be narrowed and/or the contrast at the edge be enhanced. We define an edge-slope metric for sharpness as follows.

Using a semi-automatic boundary extraction method,[16] we extract the superior–posterior (upper) boundary and the inferior–anterior (lower) boundary as shown in Figure 3A. Intensity profiles (grid lines) perpendicular to the upper and lower boundary (Figure 3B) of the patent airway are chosen and extracted from a reconstructed image series with a normalized intensity between 0 and 1 and linearly interpolated to generate 10 times greater spatial resolution. Finally, the sharpness score (S) is calculated (Figure 3C) as follows:

$$S = \alpha \frac{CNR}{d}, \tag{6}$$

where α is a scaling factor associated with the intensity normalization, $d = |p_{80} - p_{20}|$, and $CNR = (I(p_{80}) - I(p_{20}))/\sigma$; p_{80} and p_{20} are points (nearby the extracted boundary pixel location) at 80% and 20% of the maximum intensity value in grid lines, respectively; $I(p)$ is an intensity value at point p; σ is the SD of an ROI outside the object where there is no signal. The sharpness score was calculated over valid time frames in which a distance between upper and lower boundary pixel locations is >5 pixels. The sharpness score was compared using paired t-tests for statistical analysis, assuming that the samples collected along the grid lines are uncorrelated. A P-value of < 0.001 was used to determine statistical significance.

3.6 | Practical use of the off-resonance correction

Finally, to determine the practical use of the off-resonance correction on an end use case, we measure vocal tract distance, which is a desired metric that is often used in the speech RT-MRI analysis to obtain constriction degree[40-42] or vocal tract area function.[43-45] The distance metric is defined as the physical distance between the upper and lower boundaries shown in Figure 3A. The boundaries are extracted using the aforementioned method[16] with the same initialization in both sets of images, without and with off-resonance correction. Distances were measured from both images.

4 | RESULTS

4.1 | Simulation

Figure 2 shows a representative example (static posture with mouth fully opened) of simulation results with
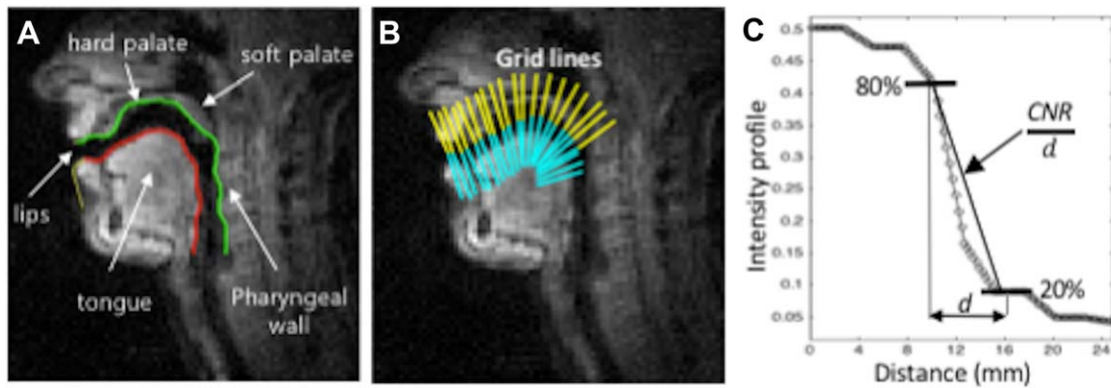
FIGURE 3 Illustration of articulator boundary identification and sharpness score evaluation. (A) Airway boundary segmentation with the upper (superior–posterior) boundary (green, color online) and the lower (inferior–anterior) boundary (red, color online). (B) Gridlines of the upper (yellow) and lower boundaries (cyan) at several locations along the airway are chosen to obtain intensity profiles. (C) Intensity profile of the gridline is plotted where a sharpness metric is measured as a slope between the points of 80% and 20% of the maximum intensity values (*CNR/d*)

different spiral readout durations. Off-resonance blurring is seen most clearly at the lips, hard palate, and tongue boundary and becomes more severe with the longer readouts as shown in Figure 2B. As the duration of the readout is longer, the estimated field maps (Figure 2C) tend to be blurred and amplified in some areas such as the tongue surface and lips surface. Accordingly, high spatial frequency error can be seen in those area (Figure 2D). The estimated field map fails to correct for the simulated off-resonance for the longer readout duration (>5 ms), and the blurred anatomic structures remain unresolved.

4.2 | Existing speech RT-MRI data

Figure 4 contains representative mid-sagittal image frames and the corresponding field map estimated for 4 subjects, which, on visual assessment, presented the most significant blurring artifacts among the 20 subjects. Note that subject numbers of 4, 6, 9, and 13 shown in Figure 4 correspond to those shown in Figure 5. For every image reconstructed with off-resonance correction, the soft palate, hard palate, and medial surface of the tongue become more intense and sharper compared to the blurred images (see yellow arrows). For all the 4 subjects, posterior to the alveolar ridge, the hard palate appears sharper up to the soft palate in the deblurred images. Correspondingly, in the estimated field maps, the regions that have shown blurred anatomical structures represent high off-resonance frequency values of >200 Hz.

Figure 6 shows the profiles that are extracted at the solid lines in the sample image frames from the 3 subjects. For subject 9, the intensity profile from the deblurred image provides a clear delineation of the soft palate movements. For subjects 6 and 13, the intensity in the hard palate in the deblurred image sequence is more constant along time than the intensity value in the blurred image sequence. This result

agrees with the fact that the hard palate, which is a bony structure covered by a thin layer of tissue, does not change its shape during speech production.[17] Furthermore, the intensity profile from the deblurred image exhibits sharper boundary between tongue and air.

Figure 7 illustrates 1 more example of correction result from subject 4, especially showing the estimated field map over time. As depicted in the off-resonance frequency value versus time profile, the proposed method enables capturing of the dynamic change in off-resonance at the tissue boundaries. Whereas the estimated field map shows high off-resonance frequency values at the hard palate and tongue boundaries over time, it shows a low frequency value at those boundaries during the event of the tongue touching the hard palate because there is no air between the tongue and hard palate (see white arrows).

4.3 | Sharpness score

Figure 5 illustrates the sharpness scores and summary table. Sharpness scores without and with correction were measured at upper airway boundaries (upper lip, hard palate, and soft palate) and lower boundaries (lower lip, anterior-, medial-, and posterior-tongue) described in Figure 3 and averaged over time. The boundary extraction method used failed to segment the image from 1 subject because of low image quality, which was excluded in this sharpness analysis. Overall, the sharpness scores show a statistically significant difference in mean values (correction > no correction, $P > 0.001$) for the subjects tested at a majority of the boundaries. The lower lip shows negligible sharpness improvement in 10 subjects and worse sharpness score in 3 subjects when correction was applied. The hard palate exhibits worse sharpness score in 3 subjects after correction compared to no correction, whereas 15 subjects show improvement in sharpness score after correction.
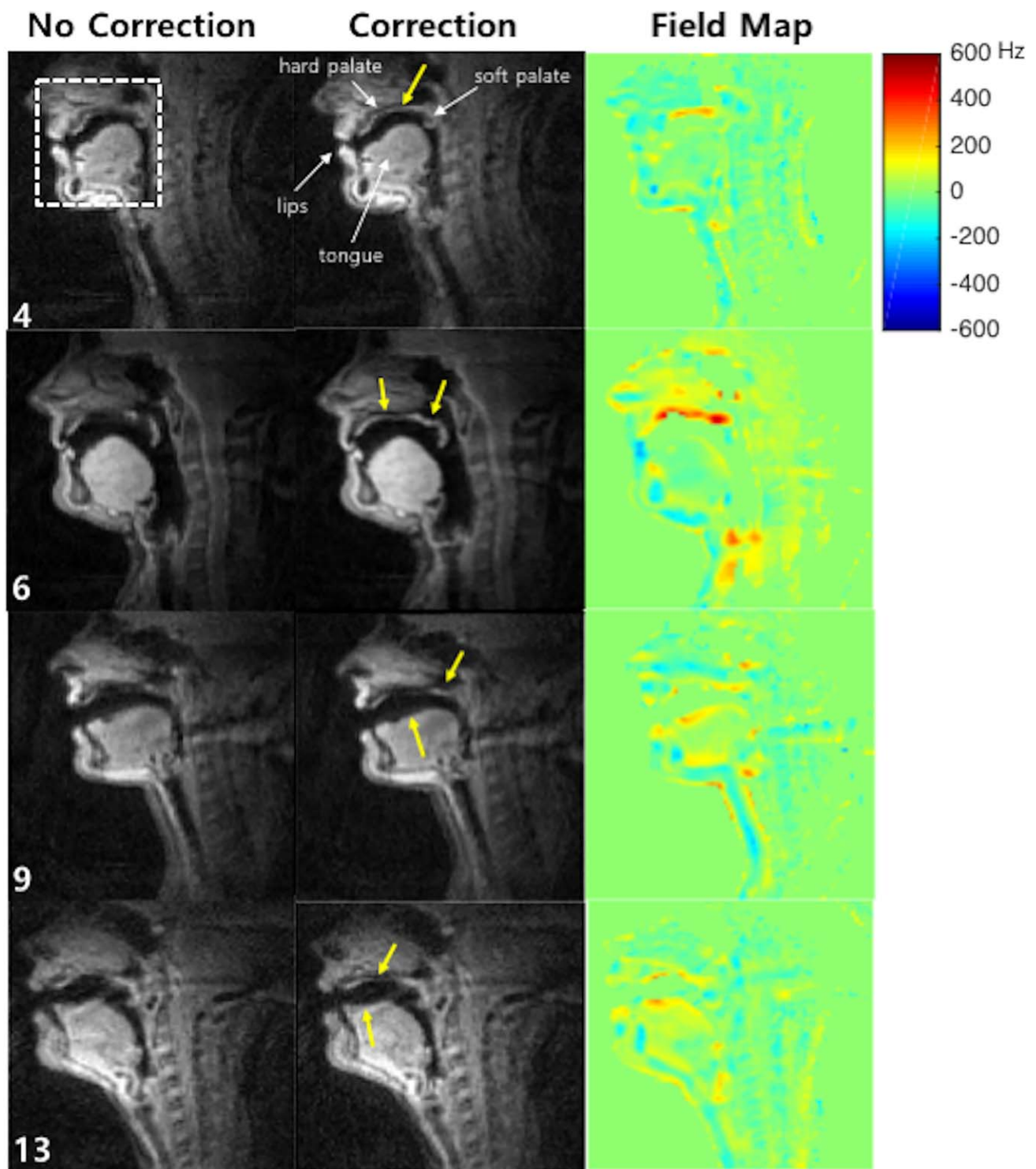
FIGURE 4 Representative mid-sagittal image frames of vocal tracts for 4 subjects, which, on visual assessment, presented the most significant blurring artifacts and were selected among the 20 subjects. The first and the second columns show images reconstructed with no correction and with correction, respectively. The last column shows the estimated field maps corresponding to those image time frames. Yellow arrows point out the regions that are most affected by off-resonance blurring, and corrected by the proposed method. See also Supporting Information Video S1

4.4 | Practical use of the off-resonance correction

Figure 8 illustrates airway boundary segmentation result based on which the corresponding vocal tract distance measured from images without and with correction from subject 6 shown in Figure 4. The uncorrected image exhibits noticeable errors in the segmentation because of off-resonance-induced blurring around the hard palate and soft palate, as indicated with arrows in Figure 8A and erroneous
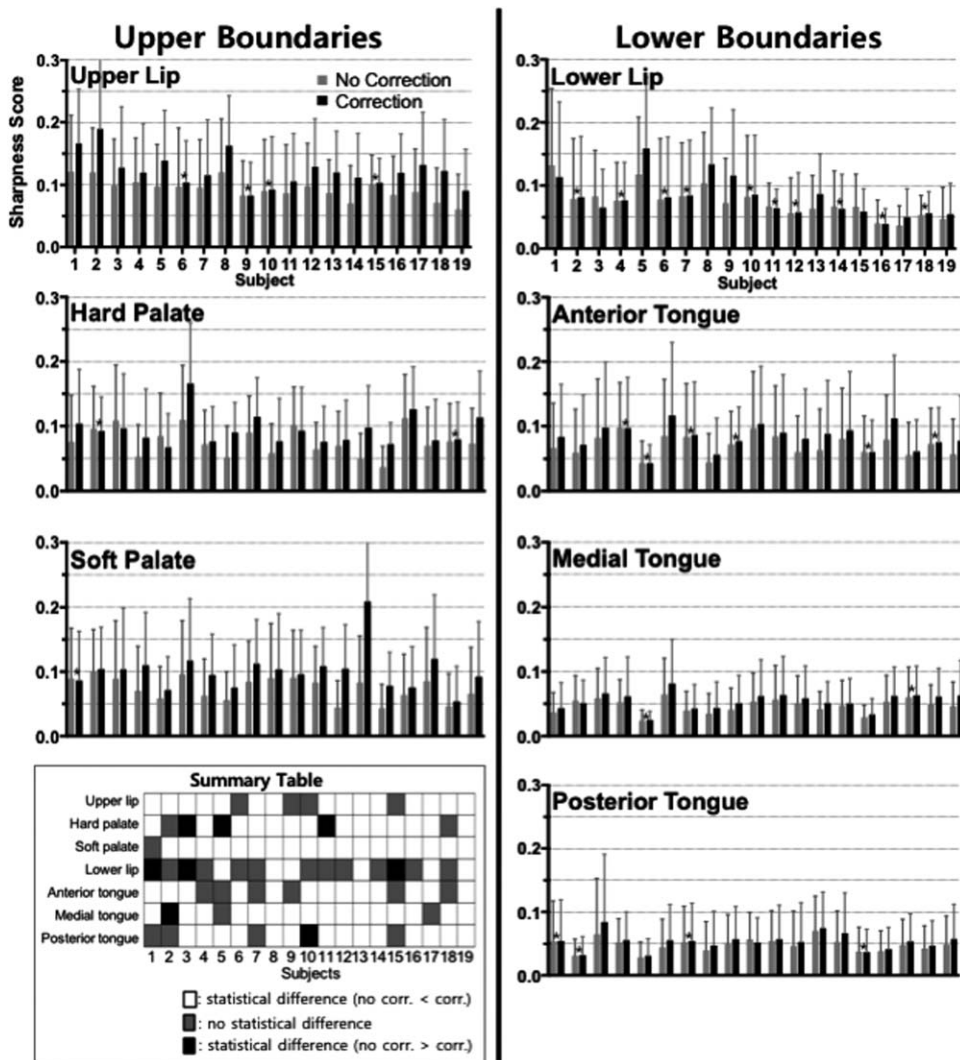
FIGURE 5 Sharpness without and with correction at different articulator boundary locations. Sharpness scores are measured at the upper boundaries (upper lip, hard palate, and soft palate) and lower boundaries (lower lip, anterior-, medial-, and posterior-tongue) along time. The mean and the SD of the sharpness scores over time are shown here where the 19 subjects are presented in descending order of average uncorrected sharpness score. A paired t-test was performed at each articulator boundary for each individual subject to test for the significance of the sharpness difference. The sharpness scores marked with an asterisk (*) were not found to be statistically different. All remaining scores were found to have significant mean differences ($P < 0.001$). Summary table in the bottom left panel summarizes the significance of mean sharpness score difference between no correction and correction in three different categories: (white) no correction < correction, (gray) no significant difference between no correction and correction, and (black) no correction > correction

results on the corresponding vocal tract distance in those area as shown in Figure 8B.

5 | DISCUSSION

We have developed a dynamic field map estimation method for spiral RT-MRI where a dynamic field map is directly estimated from the phase of single-TE dynamic images after a coil phase compensation. We estimated complex coil sensitivities from single-echo data itself—temporally averaged and spatially low-pass filtered image. The proposed method could provide partial off-resonance correction for previously

collected spiral RT-MRI data sets because it does not require the additional acquisition of the coil sensitivity map. The proposed method is simple, computationally less demanding, and when combined with the iterative image reconstruction, improves sharpness of the vocal tract articulator boundaries including the upper lip, hard palate, soft palate, and tongue boundaries (except for the lower lip) in a majority of the 19 subjects tested. This has the potential to improve the downstream analysis of the dynamics of articulators during speech.

The signal equation in Equation 3 ignores phase accrual during the spiral readout. This assumption is not strictly true and becomes less valid for long spiral readout duration and/
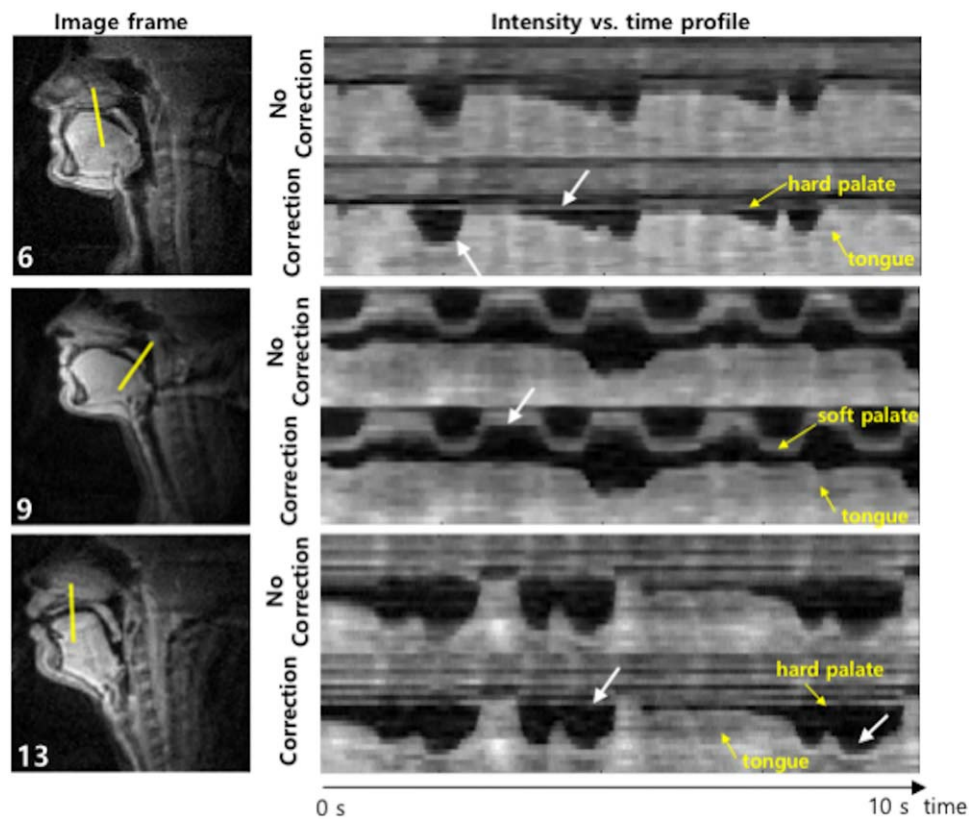
FIGURE 6 Illustration of improved sharpness of articulator boundaries. The first column shows an example frame for 3 different subjects and the second column shows intensity versus time profiles marked by the solid lines in the first column images where each of the solid lines corresponds to one of the gridlines shown in Figure 3. For all subjects, the intensity time profiles from image sequences reconstructed with correction exhibit sharper boundary between tongue and air than that from image sequences with no correction. For subject 9, the intensity profile from the correction provides a clear delineation of the soft palate movements. For subjects 6 and 13, the correction method provides more constant intensity in the hard palate along time than image sequence with no correction
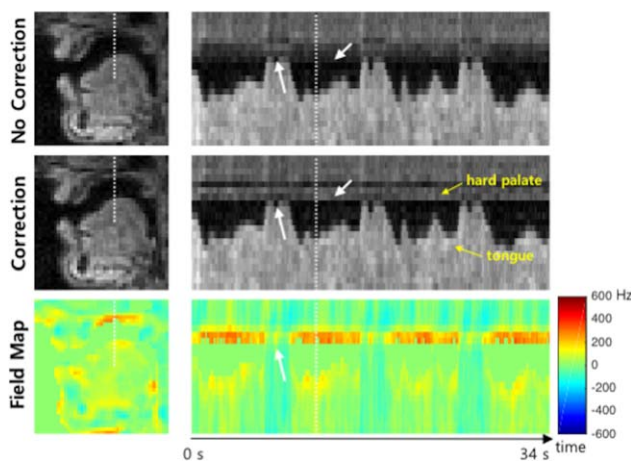
FIGURE 7 Illustration of the estimated field map over time. The first column shows example frames of reconstructed images and field map corresponding to the white dot box shown in Figure 4. The second column shows intensity versus time profiles marked by the dot lines in the first column images. In the estimated field map, high off-resonance frequency values are shown at the hard palate (400 Hz) and tongue (200 Hz) boundaries over time except when the tongue contacts the hard palate. This is because when the tongue touches the hard palate, there is neither air and susceptibility difference between them. See also Supporting Information Video S2

or large resonant frequency offsets. In most cases, the PSF in Equation 2 is no longer sharp impulse nor pure real at the origin, which distorts the complex images used for the field map estimation. This PSF distortion is the basis of auto-focus methods. As readout duration is increased, phase, and therefore the estimated field map, tend to be erroneously blurred and amplified as can be seen in the simulation result (Figure 2C). These are practical limitations to the proposed method. Our findings suggest that for speech RT-MRI at 1.5T, the proposed method will fail to work reliably for read-out durations >5 ms. An area of future work is investigating and predicting phase error caused by the non-ideal impulse with longer spiral readout.

An important issue in the field map estimation relates to the accuracy of the coil sensitivity maps. We low-pass-filtered the time-averaged image to estimate the coil map. This stems from an assumption that the coil maps contain only low spatial-frequency information and are stationary. Although the deblurred result demonstrated improvement in the sharpness at the boundaries compared to the original uncorrected images, the correction based on this coil map estimation depends on whether the anatomic structure and its

A   No Correction   Correction

B   Vocal Tract Distance
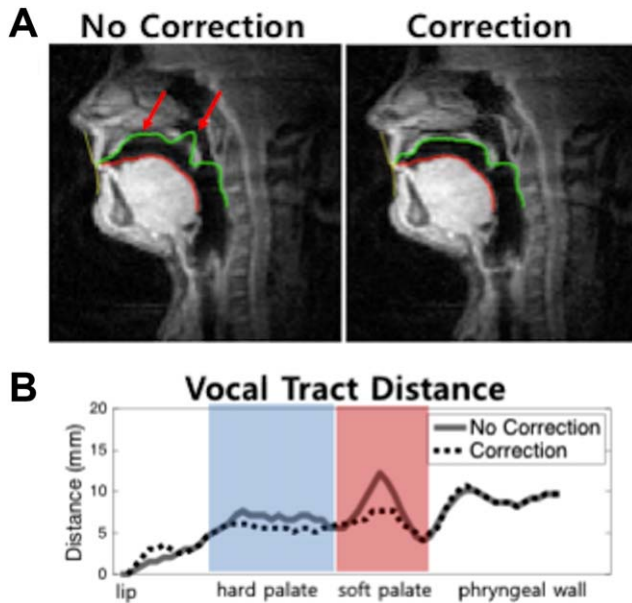
FIGURE 8 Representative illustration of airway boundary segmentation results on images without and with correction from subject 6. (A) Airway boundary segmentation with a same initialization was performed on images without and with correction, to extract the upper and lower boundaries (green and red contours). As indicated by red arrows, the uncorrected image shows segmentation errors at the hard palate and soft palate because of off-resonance-induced blurring. (B) Vocal tract distance, defined as the distance between the upper and lower boundaries, is plotted. Discernible errors are observed around the hard palate and soft palate in the uncorrected data.

field map are passed by its filtering process and show up in the sensitivity map or not. The proposed method corrects field inhomogeneity that is not low-pass filtered. Note that low-pass filtered (and time-averaged) phase is assigned as coil phase. The window size of the low pass filter needs to be chosen as large as possible not to capture abruptly varying phase because of off-resonance at articulator boundaries while it also needs to be kept adequate to obtain the spatially smoothly varying coil phase. However, it would be hard for one to optimize the choice of the size without knowing the object and the coil configuration in detail. In addition, as we described earlier, a precise shimming is required because the zero- and first-order field inhomogeneity is highly likely to be included in the estimated coil map and could be a main source of the error in the estimated field map. An alternative solution to these limitations of the coil sensitivity map estimation would be to use an additional 2-echo, static scan to estimate coil sensitivity maps that are free of phase because of off-resonance and B_0 field inhomogeneity.[46] This solution is a work in progress in terms of comprehensive data collection and validation.

Another consideration for the field map estimation is to maintain an acceptable SNR level for the complex image. This is because error in phase is closely related to the SNR

of the magnitude image (i.e., $\sigma_\theta = 1/SNR$),[47] as is the field map error (i.e., $\sigma_f = \sigma_\theta/(2\pi \times TE) = 1/(2\pi \times TE \times SNR)$). For example, if SNR = 10 and TE = 0.8 ms, the field map standard deviation is σ_f = 19.9 Hz. At readout duration of 2.5 and 5.0 ms, this error causes phase accrual error during spiral readout at the edge of the k-space of 18° and 36°, respectively. Therefore, it is important to have sufficient SNR with respect to the given TE and readout duration so that the accuracy of the estimation is less affected by noise. We chose a $3 \times 3 \times 3$ Hanning window (in **r**-t) to maintain an adequate SNR > 60 in the ROI so that $\sigma_f < 3.3$ Hz theoretically. Note that SNR is approximately increased by $1/\sqrt{\sum (w_i)^2}$ where w_i is the weight of the Hanning window. However, the use of large window could also result in smoothing out high frequency features.

Field map was estimated from images reconstructed using view-sharing with a temporal window of 78 ms (fully sampled k-space, 13 spirals). It is possible that articulator movement within the temporal window ($<<$78 ms) could result in temporal blurring of the field map or residual spiral artifact. Temporal blurring could give rise to errors in the artifact-corrected image as there is a discrepancy in the temporal windows between the estimated field maps and the corrected images. For example, if the tongue tip moves so rapidly that temporal blurring around the tongue tip appears in the field map but not in the image to be reconstructed, there could be unresolved blurring by off-resonance around the tongue tip. Residual spiral artifact that affects the phase of the complex image could also lead to erroneous field map. This is 1 of the limitations of the view-sharing scheme used in this work for field map estimation.

We excluded noise-only area in the estimated field map using a mask. The mask was calculated from the distorted complex images where signal loss often manifests at some boundaries such as the hard palate and soft palate. Therefore, locations containing a high frequency feature could erroneously be masked out as zero. A more sophisticated method for generating field map masks should be investigated to mitigate this type of error.

We measured the sharpness score in several specific air–tissue boundary locations along the vocal tract to quantitatively evaluate the effectiveness of the proposed method. However, no metric is perfect, and the sharpness score was found to be sensitive to several factors. The boundary sharpness score is highly dependent on the location pre-identified as the true boundary. In the presence of signal loss because of off-resonance effect, the semi-automatic boundary segmentation method may fail. Specifically, the boundary location can be incorrectly identified. We often found this case in the original uncorrected image. For example, the boundary at the hard palate and soft palate is ambiguous and segmented erroneously as shown in Figure 8A. In this case, it is hard to

fairly compare the scores between the uncorrected and corrected images. To address this problem in this work, we used a boundary location extracted from the corrected image to measure the score in both the uncorrected and corrected images.

Ultimately, it is important to evaluate the impact of the off-resonance correction on RT-MRI analysis in speech science. For example, in Figure 8, we have conducted segmentation of the vocal tract and shown observable improvement in the segmentation and measurement of the vocal tract distance after correction is applied as a use case example in RT-MRI analysis. Nevertheless, because in many cases the improvement would be not so much noticeable by visual inspection as shown in Figure 8, a better way to evaluate improvement in the segmentation result would be to compare the segmentation results with manual segmentation results. However, because of the very large number of frames in the RT-MRI data sets, performing a manual segmentation is not practical. Hence, in ongoing work, we are investigating a methodology to evaluate the segmentation results without manual reference.

6 | CONCLUSIONS

We have developed and demonstrated a simple method for estimating a dynamic field map from spiral RT-MRI data of speech and incorporating the correction of the off-resonance into the constrained image reconstruction. We use the base image phase from single-echo data, after some initial processing, to estimate the field map directly by assigning the smoothly varying time-averaged phase to be used as coil phase and the residual high-frequency phase variations to the dynamic field map. We have demonstrated improvements in depiction of the vocal tract articulators at several air–tissue boundaries both visually and through a sharpness metric and the practical use of this method on the boundary segmentation and distance metric as a use case example.

REFERENCES

[1] Lingala SG, Sutton BP, Miquel ME, Nayak KS. Recommendations for real-time speech MRI. *J Magn Reson Imaging.* 2016; 43:28-44.

[2] Bresch E, Kim YC, Nayak KS, Byrd D, Narayanan SS. Seeing speech: capturing vocal tract shaping using real-time magnetic resonance imaging. *IEEE Signal Process Mag.* 2008;25:123-129.

[3] Scott AD, Wylezinska M, Birch MJ, Miquel ME. Speech MRI: morphology and function. *Phys Med.* 2014;30:604-618.

[4] Westbury JR. The significance and measurement of head position during speech production experiments using the x-ray microbeam system. *J Acoust Soc Am.* 1991;89:1782-1791.

[5] Perkell JS, Cohen MH, Svirsky MA, Matthies ML, Garabieta I, Jackson MT. Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *J Acoust Soc Am.* 1992;92:3078-3096.

[6] Denby B, Stone M. Speech synthesis from real time ultrasound images of the tongue. In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, Canada, 2004. pp. 685-688.

[7] Block KT, Frahm J. Spiral imaging: a critical appraisal. *J Magn Reson Imaging.* 2005;21:657-668.

[8] Schenck JF. The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds. *Med Phys.* 1996;23:815-850.

[9] Meyer CH, Hu BS, Nishimura DG, Macovski A. Fast spiral coronary artery imaging. *Magn Reson Med.* 1992;28:202-213.

[10] Narayanan SS, Nayak KS, 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.

[11] Kim YC, Narayanan SS, Nayak KS. Flexible retrospective selection of temporal resolution in real-time speech MRI using a golden-ratio spiral view order. *Magn Reson Med.* 2011;65:1365-1371.

[12] Lingala SG, Zhu Y, Kim Y, Toutios A, Narayanan SS, Nayak KS. A fast and flexible MRI system for the study of dynamic vocal tract shaping. *Magn Reson Med.* 2017;77:112-125.

[13] Proctor MI, Lammert A, Katsamanis A, Goldstein L, Hagedorn C, Narayanan SS. Direct estimation of articulatory kinematics from real-time magnetic resonance image sequences. In Proceedings of the Annual Conference of INTERSPEECH, Florence, Italy, August 27-31, 2011. pp. 281-284.

[14] Lammert A, Ramanarayanan V, Proctor MI, Narayanan SS. Vocal tract cross-distance estimation from real-time MRI using region-of-interest analysis. In Proceedings of the Annual Conference of INTERSPEECH, Lyon, France, August 25-29, 2013. pp. 959-962.

[15] Proctor MI, Bone D, Katsamanis N, Narayanan SS. Rapid semi-automatic segmentation of real-time magnetic resonance images for parametric vocal tract analysis. In Proceedings of the Annual Conference of INTERSPEECH, Makuhari, Japan, September 26-30, 2010. pp. 1576-1579.

[16] Kim J, Kumar N, Lee S, Narayanan SS. Enhanced airway-tissue boundary segmentation for real-time magnetic resonance imaging data. In Proceedings of the 10th International Seminar on Speech Production (ISSP), Cologne, Germany, May 5-8, 2014. pp. 222-225.

[17] Bresch E, Narayanan SS. Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images. *IEEE Trans Med Imaging.* 2009;28:323-338.

[18] Browman C, Goldstein LM. Towards an articulatory phonology. *Phonology Yearbook.* 1986;3:219-252.

[19] Atik B, Bekerecioglu M, Tan O, Etlik O, Davran R, Arslan H. Evaluation of dynamic magnetic resonance imaging in assessing velopharyngeal insufficiency during phonation. *J Craniofac Surg.* 2008;19:566-572.

[20] Drissi C, Mitrofanoff M, Talandier C, Falip C, Le Couls V, Adamsbaum C. Feasibility of dynamic MRI for evaluating velopharyngeal insufficiency in children. *Eur Radiol.* 2011;21:1462-1469.

[21] Scott AD, Boubertakh R, Birch MJ, Miquel ME. Towards clinical assessment of velopharyngeal closure using MRI: Evaluation of real-time MRI sequences at 1.5 and 3T. *Br J Radiol.* 2012; 85:1083-1092.

[22] Freitas AC, Wylezinska M, Birch MJ, Petersen SE, Miquel ME. Comparison of Cartesian and non-Cartesian real-time MRI sequences at 1.5T to assess velar motion and velopharyngeal closure during speech. *PLoS ONE.* 2016;11:1-16.

[23] Bae Y, Kuehn DP, Conway CA, Sutton BP. Real-time magnetic resonance imaging of velopharyngeal activities with simultaneous speech recordings. *Cleft Palate Craniofac J.* 2011;48:695-707.

[24] Noll DC, Meyer CH, Pauly JM, Nishimura DG, Macovski A. A homogeneity correction method for magnetic resonance imaging with time-varying gradients. *IEEE Trans Med Imaging.* 1991;10: 629-637.

[25] Man LC, Pauly JM, Macovski A. Multifrequency interpolation for fast off-resonance correction. *Magn Reson Med.* 1997;37: 785-792.

[26] Nayak KS, Tsai CM, Meyer CH, Nishimura DG. Efficient off-resonance correction for spiral imaging. *Magn Reson Med.* 2001; 45:521-524.

[27] Man LC, Pauly JM, Macovski A. Improved automatic off-resonance correction without a field map in spiral imaging. *Magn Reson Med.* 1997;37:906-913.

[28] Noll DC, Pauly JM, Meyer CH, Nishimura DG, Macovski A. Deblurring for non-2D Fourier transform magnetic resonance imaging. *Magn Reson Med.* 1992;25:319-333.

[29] Chen W, Meyer CH. Fast automatic linear off-resonance correction method for spiral imaging. *Magn Reson Med.* 2006;56:457-462.

[30] Smith TB, Nayak KS. Automatic off-resonance correction in spiral imaging with piecewise linear autofocus. *Magn Reson Med.* 2013;69:82-90.

[31] Sutton BP, Conway CA, Bae Y, Seethamraju R, Kuehn DP. Faster dynamic imaging of speech with field inhomogeneity corrected spiral fast low angle shot (FLASH) at 3 T. *J Magn Reson Imaging.* 2010;32:1228-1237.

[32] Chen W, Meyer CH. Semiautomatic off-resonance correction in spiral imaging. *Magn Reson Med.* 2008;59:1212-1219.

[33] Lee D, Nayak KS, Pauly J. Reducing spurious minima in automatic off-resonance correction for spiral imaging. In Proceedings of the 12th Annual Meeting of ISMRM, Kyoto, Japan, 2004. p. 2678.

[34] Roemer PB, Edelstein WA, Hayes CE, Souza SP, Mueller OM. The NMR phased array. *Magn Reson Med.* 1990;16:192-225.

[35] Fessler JA, Sutton BP. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Trans Signal Proc.* 2003;51:560-574.

[36] Lingala SG, Toutios A, Toger J, et al. State-of-the-art MRI protocol for comprehensive assessment of vocal tract structure and function. In Proceedings of the Annual Conference of INTER-SPEECH, San Francisco, CA, September 8-12, 2016. pp. 475-479.

[37] Santos JM, Wright GA, Pauly JM. Flexible real-time magnetic resonance imaging framework. *Conf Proc IEEE Eng Med Biol Soc.* 2004;2:1048-1051.

[38] Sutton BP, Noll DC, Fessler JA. Fast, iterative, field-corrected image reconstruction for MRI. *IEEE Trans Med Imaging.* 2003; 22:178-188.

[39] Fessler JA, Lee S, Olafsson VT, Shi HR, Noll DC. Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity. *IEEE Trans Signal Proc.* 2005; 53:3393-3402.

[40] Vaz C, Toutios A, Narayanan SS. Convex hull convolutive non-negative matrix factorization for uncovering temporal patterns in multivariate time-series data. In Proceedings of the Annual Conference of INTERSPEECH, San Francisco, CA, September 8-12, 2016. pp. 963-967.

[41] Kim J, Lammert AC, Kumar Ghosh P, Narayanan SS. Co-registration of speech production datasets from electromagnetic articulography and real-time magnetic resonance imaging. *J Acoust Soc Am.* 2014;135:EL115-EL121.

[42] Töger J, Sorensen T, Somandepalli K, et al. Test–retest repeatability of human speech biomarkers from static and real-time dynamic magnetic resonance imaging. *J Acoust Soc Am.* 2017; 141:3323-3336.

[43] Story BH, Titze IR, Hoffman EA. Vocal tract area functions from magnetic resonance imaging. *J Acoust Soc Am.* 1996;100: 537-554.

[44] Kim YC, Kim J, Proctor MI, et al. Toward automatic vocal tract area function estimation from accelerated three-dimensional magnetic resonance imaging. In Proceedings of the ISCA Workshop on Speech Production in Automatic Speech Recognition, Lyon, France, August 30, 2013. pp. 2-5.

[45] Skordilis ZI, Toutios A, Toger J, Narayanan SS. Estimation of vocal tract area function from volumetric magnetic resonance imaging. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, LA, March 5-9, 2017. pp. 924-928.

[46] Robinson S, Grabner G, Witoszynskyj S, Trattnig S. Combining phase images from multi-channel RF coils using 3D phase offset maps derived from a dual-echo scan. *Magn Reson Med.* 2011; 65:1638-1648.

[47] Brown RW, Cheng YCN, Haacke EM, Thompson MR, Venkatesan R. *Magnetic resonance imaging: physical principles and sequence design.* Second Edition. Hoboken: John Wiley & Sons; 2014. 1008 p.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

VIDEO S1 Comparison of results with no correction and correction for 4 different subjects. This supporting information video corresponds to Figure 4

VIDEO S2 Illustration of the estimated field map over time. This supporting information video corresponds to Figure 7