A Fast and Flexible MRI System for the Study of Dynamic Vocal Tract Shaping

Sajan Goud Lingala,[1]* Yinghua Zhu,[1] Yoon-Chul Kim,[2] Asterios Toutios,[1] Shrikanth Narayanan,[1] and Krishna S. Nayak[1]

Purpose: The aim of this work was to develop and evaluate an MRI-based system for study of dynamic vocal tract shaping during speech production, which provides high spatial and temporal resolution.

Methods: The proposed system utilizes (a) custom eight-channel upper airway coils that have high sensitivity to upper airway regions of interest, (b) two-dimensional golden angle spiral gradient echo acquisition, (c) on-the-fly view-sharing reconstruction, and (d) off-line temporal finite difference constrained reconstruction. The system also provides simultaneous noise-cancelled and temporally aligned audio. The system is evaluated in 3 healthy volunteers, and 1 tongue cancer patient, with a broad range of speech tasks.

Results: We report spatiotemporal resolutions of $2.4 \times 2.4\,mm^2$ every 12 ms for single-slice imaging, and $2.4 \times 2.4\,mm^2$ every 36 ms for three-slice imaging, which reflects roughly 7-fold acceleration over Nyquist sampling. This system demonstrates improved temporal fidelity in capturing rapid vocal tract shaping for tasks, such as producing consonant clusters in speech, and beat-boxing sounds. Novel acoustic-articulatory analysis was also demonstrated.

Conclusion: A synergistic combination of custom coils, spiral acquisitions, and constrained reconstruction enables visualization of rapid speech with high spatiotemporal resolution in multiple planes. **Magn Reson Med 77:112–125, 2017.**

Key words: flexible MRI system; rapid vocal tract shaping; custom upper-airway coil; spiral readouts; multi-slice; constrained reconstruction

INTRODUCTION

Speech production involves complex spatiotemporal coordination of several vocal organs in the upper and lower airways. Modalities to study speech production include real-time MRI (RT-MRI), electromagnetic articulography (EMA), electropalatography, ultrasound, and X-ray, or videofluoroscopy (1). In comparison to alternate modalities, RT-MRI provides distinct advantages in terms of (a) noninvasiveness, as opposed to X-rays, videofluoroscopy, and (b) ability to image in arbitrary planes and visualize deep structures (e.g., epiglottis, glottis), which are not possible with ultrasound and EMA. Applications of RT-MRI in speech science and vocal production research are numerous; these include addressing open questions pertaining to understanding the goals of language production, language timing, speech errors, and other topics in phonetics and phonology (1–8) as well as vocal production of song (9,10). It also has the potential to manage and inform treatment plans in several clinical applications, such as clinical assessment of velopharyngeal insufficiency (11–13), cleft palate repair and management (14,15), and surgical planning and post-treatment functional evaluation of speech and swallowing (16,17), in head and neck cancer.

The rates of movements of articulators are highly dependent on the speech task and the subject's speaking style (3,18,19). For example, in the articulation of sustained sounds, such as during singing, the spatial position of the articulators change on the order of seconds, whereas in tasks involving flaps/trills and production of consonant clusters, the motion of articulators occur at a much faster rate on the order of a few milliseconds (also see Figure 1) (3). Whereas modalities such as EMA can operate up to a time resolution of 1 ms, the imaging speed of RT-MRI is restricted by challenges posed as a result of device physics.

Several schemes have been proposed to improve the imaging speed of RT-MRI for speech studies. These can be classified as on-the-fly or off-line schemes (3). On-the-fly schemes are referred to those that allow for immediate visualization of the reconstructed images (with latency less than 500 ms), whereas off-line schemes are referred to those where the reconstructions are implemented off-line. Scott et al (20) utilized on-the-fly imaging with Cartesian trajectories and demonstrated temporal resolutions between 111 and 50 ms at spatial resolution of $1.6–2.7\,mm^2$ for velopharyngeal closure. Other investigators (1,21,22) utilized short spiral readouts to acquire images at a native time resolution of 54–78 ms, and a spatial resolution of $3.0–2.4\,mm^2$, and visualize at 24 frames/sec using view sharing. View sharing was also used with radial imaging in (23,24). Freitas et al (25) compares Cartesian, radial, and spiral trajectories with view sharing and demonstrated spirals to provide the best compromise in terms acquisition efficiency, motion robustness, and signal to noise (SNR).

Iterative constrained reconstruction schemes have shown to enable greater acceleration factors. Utilizing radial trajectories, Niebergall et al (23) proposed

[1]Electrical Engineering, University of Southern California, Los Angeles, CA.
[2]Samsung Medical Center, Seoul, South Korea.

Grant sponsor: National Institutes of Health; Grant number: NIH/NIDCD R01 DC007124.

*Correspondence to: Sajan Goud Lingala, Ph.D., Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, 3740 McClintock Avenue, Los Angeles, CA 90089. E-mail: lingala@usc.edu

Received 24 August 2015; revised 6 November 2015; accepted 24 November 2015
DOI 10.1002/mrm.26090
Published online 17 January 2016 in Wiley Online Library (wileyonlinelibrary.com).

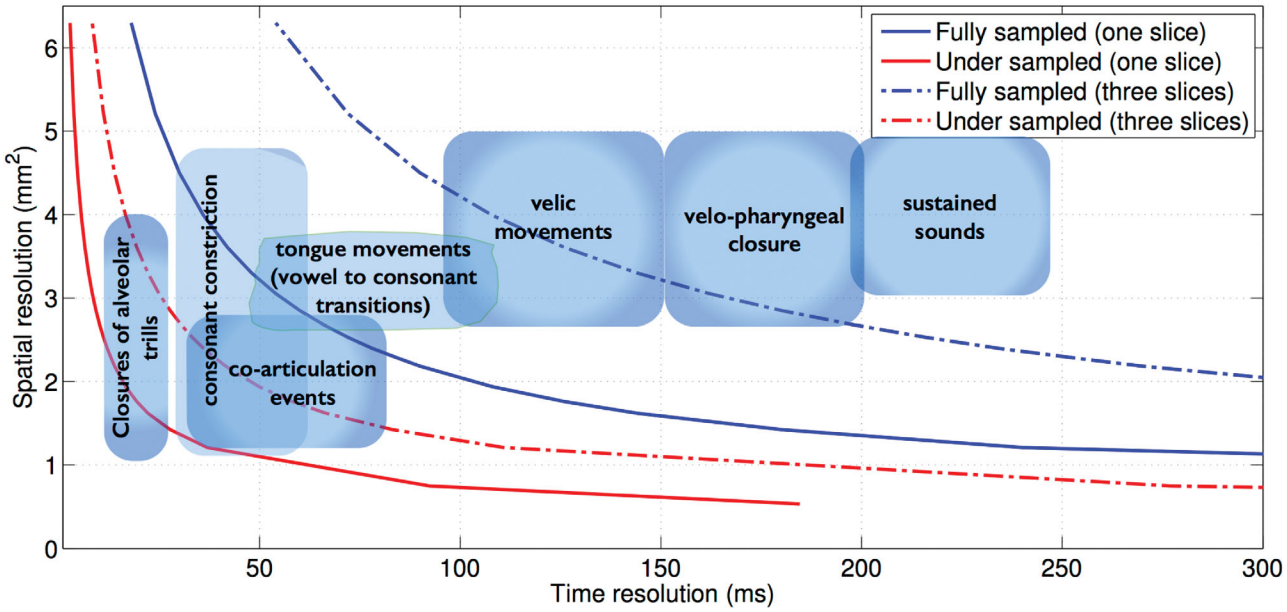

FIG. 1. Spatial versus temporal resolutions trade-offs in RT-MRI using short interleaved spiral trajectories. In comparison to full sampling, sparse sampling reduces spatiotemporal trade-offs and enables improved visualization of several speech tasks both in single- and multiplane imaging. The clouds in the above figure represent a recently reported consensus opinion among speech imaging researchers and linguists (3).

nonlinear temporal regularized reconstruction to enable a temporal resolution of 33 ms, and a spatial resolution of 1.5 mm^2, and studied a variety of tasks (vowel, consonant sounds, and coarticulations events). More recently, Iltis et al (26) demonstrated an on-the-fly implementation of the iterative reconstruction by exploiting parallelization within the reconstruction along with efficient use of graphical processing units. This work used a 64-channel brain coil and demonstrated 100 frames/sec at 1.5 mm^2. Burdumy et al (24) applied off-line spatial total-variation regularization with radial trajectories to provide spatial resolution of 1.8 mm^2, and a native time resolution of 55 ms, and analyzed morphometric measurements of the vocal tract. Fu et al (27) utilized off-line reconstruction with the partial separable (PS) model (28,29), along with spatial-spectral sparsity constraint (30), and demonstrated a frame rate of 102 frames/sec at a spatial resolution of 2.2 mm^2. The low-rank (by the PS model) constraint is essentially a data-driven retrospective binning technique (31) and was fully exploited in Fu et al (27) by utilizing repeated speech utterances. Low-rank constraints are unlikely to apply to short speech utterances with no repetitions, or stimuli involving infrequent distinct movements such as swallowing.

In this work, we developed a MRI-based system for dynamic study of vocal tract shaping during speech production to provide high spatiotemporal resolutions. We propose a system that utilizes (a) a novel eight-channel custom upper airway coil, which has improved sensitivity in upper airway regions of interest (ROIs), (b) a flexible slice selective spiral spoiled gradient echo acquisition with golden angle time interleaving, (c) on-the-fly view-sharing reconstruction, (d) off-line temporal finite difference constrained reconstruction, (e) simultaneous audio acquisition, and off-line temporal alignment of noise-cancelled audio with the reconstructions. The innovation of our system lies in the synergistic combination of the above components. We chose custom upper airway coil design for its ability to provide superior SNR across all articulators of interest. This is advantageous because it can provide an important boost in SNR while operating at 1.5 Tesla (T). Its combination with spirals is complementary, because it enables improved SNR at low fields, low off-resonance–related spiral artifacts, and high sampling efficiency. Multishot short spirals (readout length of 2.4 ms) are used to reduce off-resonance artifacts. Our rationale of choosing spirals is that they have shown to provide improved motion robustness and efficiency over alternate trajectories. Temporal finite difference constraint was chosen because it exploits redundancy based on the prior that the desired information is contained in the moving edges, which directly fits to the end goal assessment of dynamics of air-tissue interfaces. Also, because it exploits similarities among neighboring time frames, it is applicable to a wide variety of speech tasks and does not impose restrictions on the imaging task.

We present several examples with the proposed system for rapid RT-MRI of speech, including visualization of interleaved consonant and vowel sounds, fluent speech sentences that contain consonant clusters, as well as beat-boxing sounds that involve rapid coordination between various vocal articulators, on three healthy volunteers, and one tongue cancer patient referred to glossectomy.

METHODS

Simulation of Spatial Versus Time Resolution Trade-offs With Spiral Sampling

A multishot short spiral readout spoiled gradient echo pulse sequence (flip angle //FA: 15° degrees; slice

thickness: 6 mm; readout time: 2.5 ms, repetition time [TR] = 6.004 ms), which was used in our previous studies, was adapted in this study (32,33). We chose spiral trajectories over alternate trajectories because they have shown to provide a superior trade-off among spatial resolution, time resolution, and robustness to motion artifacts. The spiral trajectories were designed to make maximum use of gradients (40 mT/m maximum gradient amplitude and 150 mT/m/ms slew rate). Simulations were performed to investigate the spatial versus time resolution trade-offs for a field of view (FOV) of 20 cm^2 at Nyquist (full) sampling and rate 6.5-fold undersampling for single-slice and concurrent three-slice imaging (Figure 1). Nyquist sampling was determined by varying the number of spiral interleaves such that the maximum spacing between interleaves equaled the reciprocal of the unaliased FOV. The Nyquist definition was defined based on the assumption of spiral interleaving with a uniform angle distribution. Figure 1 also duplicates the schematic placement of various speech tasks according to their spatial and temporal resolution requirements, as reported in Lingala et al (3).

Single-Slice and Multislice Golden Angle Spiral Time Interleaving

A previously proposed golden angle time interleaved sampling pattern scheme in which successive spiral interleaves are separated by the golden angle $2\pi \times 2/(\sqrt{5}+1)$ was adapted in this work (32). The sampling schedule was repeated after 144 interleaves. Two single-slice sequences with spatial resolutions of 2.4 and 1.76 mm^2 were realized, which respectively corresponded to 13 and 21 spiral interleaves/frame for Nyquist sampling. A flexible multislice time interleaved sequence (33) was also adapted to realize an arbitrary three-slice select sequence at 2.4 mm^2. The golden angle increment for the three-slice sequence occurred every 3 TRs. It should be noted that the unaliased FOV with golden angle sampling slightly differs with that of uniform density sampling used for the simulation in Figure 1 (32). Specifically, the point spread function of golden angle spiral sampling provides more reduced side-lobe energies and provides improved trade-off of the achievable unaliased FOV in comparison to uniform density sampling (32).

Custom Upper Airway Coil

All of our experiments were performed on a GE Signa Excite 1.5T scanner (GE Healthcare, Little Chalfont, UK) with a custom eight-channel upper airway receiver coil that has four elements on either side of the jaw. The elements were designed to be spatially localized and to be in close proximity to the upper airway to provide high SNR over all the important upper airway structures. We chose custom coil design because of its ability to provide superior SNR across all articulators of interest. This is advantageous because it provides an important boost in SNR while operating at 1.5T, and enables efficient combination with spirals is highly complementary, because it together enables improved SNR at low fields, low off-resonance–related spiral artifacts, and high sampling efficiency. The custom coil was developed for two sizes for adult and child arrays, although all the experiments in the current study were performed on adults using the adult-sized array.

Reconstruction

On-the-Fly

Data acquisition was implemented within custom RT-hawk software (34). A view-sharing scheme was used, where data for each frame were combined respectively from 13 and 21 subsequent interleaves, respectively for the 2.4 and 1.76 mm^2 sequences. Images were reconstructed on-the-fly by using a fast implementation of the gridding algorithm within RT-hawk. The minimal latency allowed for instant feedback to the operator and enabled efficient scan plane localizations. In addition, it enabled on-the-fly adjustment of the center frequency to minimize off-resonance blurs. Specifically, the subject being scanned was asked to open their mouth, and in the midsagittal plane, the operator qualitatively adjusted the center frequency such that the air-tissue (majorly: air-tongue, air-velum, air-lip) boundaries were sharp.

Offline

A sparse SENSE temporal finite difference constrained reconstruction scheme was implemented offline. This constraint exploits redundancy based on the fact that the desired information is contained in the moving edges. This directly fits to the application of speech imaging, where the end goal is the assessment of interaction and timing of various articulators, or assessment of the dynamics of the air-tissue interfaces (moving edges). Spatial and temporal finite difference constraints have also been previously used in several studies (e.g., (35–39).

The reconstruction is formulated as as shown by Equation [1]:

$$\min_{f(\mathrm{x},t)} \|A(f) - \mathrm{b}\|_2^2 + \lambda \|D_t(f)\|_1 \qquad [1]$$

where **b** is a concatenated vector containing the spiral noisy *k-t* measurements from each coil and $f(\mathbf{x},t)$ is the dynamic data to be reconstructed at a retrospectively specified time resolution. A models coil sensitivity encoding as well as Fourier encoding on the specified spiral trajectory in each time frame; the coil sensitivities were assumed to be time-invariant, and were estimated by an Eigen decomposition method using time-averaged image data from each coil (40), and the nonuniform Fourier transform (nuFFT) implementation by Fessler and Sutton (41) was used. The l_1-sparsity-based temporal finite difference (D_t) penalty is used to penalize pixels with rapidly varying pixel time profiles. λ is the regularization parameter that controls the balance between the sparsity penalty and the data fidelity. Equation [1] was solved by a nonlinear conjugate gradient (CG) algorithm (42). The algorithm was initialized with the $f=A^H(b)$ estimate and was terminated at 40 iterations, where qualitatively there was no noticeable change in image quality. The reconstructions were implemented within MATLAB (The MathWorks, Inc., Natick, MA) on an Intel Core i7 3.5 GHz machine with 32-GB memory.

In Vivo Experiments and Speech Tasks

Three healthy volunteers (2 male, 1 female; median age: 29) and 1 male tongue cancer patient (62 years) were scanned. The patient was scanned before clinical treatment. All the stimuli were presented in the scanner using a mirror projector setup. A variety of speech tasks were considered. The midsagittal orientation was used for single-slice sequences, whereas orientations for the multislice sequence differed according to the speech task. With the 2.4-mm^2 single-slice sequence, volunteer 1 (male Indian English speaker) was scanned without any speech stimuli on two separate instances: (a) using an eight-channel head coil and (b) using the custom eight-channel upper airway coil. With the custom upper airway coil, the same volunteer was scanned with both the single-slice sequences, using the stimuli: "one-two-three-four-five" at a normal speech rate followed by a rapid speech rate (approximately 4 times faster). A task to produce interleaved consonant and vowel sounds by the repetition of the phrase: "loo-lee-laa-za-na-za" at the normal speech rate was considered on volunteer 1 and imaged using the three-slice sequence (one mid-sagittal, one axial plane at the level of mid-pharyngeal airway, and one coronal plane at the middle of the tongue). Volunteer 2 (male Chinese English speaker) was scanned with the 1.76-mm^2 single-slice sequence, while producing the sentence: "She had your dark suit in greasy wash water," which involves producing sounds that involve rapid articulatory movements (e.g., coarticulation events as a part of running speech). Volunteer 3 (female American English speaker) was a beat boxer and was scanned while producing a variety of beat-boxing sounds. The 2.4-mm^2 single- and concurrent three-slice (one midsagittal slice, one axial slice at the level of velum, and one axial slice at the level of glottis) sequences were considered. The particular axial cuts were chosen to capture the rapid velar and glottis movements during beat boxing. Volunteer 4 (male American English speaker) was a tongue cancer patient and was scanned with the single-slice 2.4-mm^2 sequence. Speech stimuli comprising words, sentences, and a passage were presented, and the ability to produce speech was analyzed. A small subset of these stimuli is presented in this work, which pertain to short words that contain vowels interleaved by consonants: "beat, bit, bait, bet, bat, pot, bought, boat."

Simultaneous Audio Collection

For 3 of 4 volunteers scanned, audio recordings were obtained simultaneously at a sampling frequency of 20 KHz inside the scanner, while the subjects were being imaged, using a commercial fiber optic microphone (Optoacoustics Ltd., Or Yehuda, Israel) (43) and a custom recording setup. Noise cancellation was performed using a data-driven adaptive signal processing algorithm, which is blind to the acoustic properties of noise (44). The final noise-cancelled audio was synchronized with the reconstructed RT-MRI data to facilitate acoustic-articulatory analysis.

Analysis

Comparison of SNR Between Coils

The SNR properties of the custom eight-channel upper airway coil were qualitatively compared with a commercial eight-channel head (brain) coil. The single-slice (2.4 mm^2) sequence was used, where volunteer 1 was scanned with no speech, and 55 interleaves were used to reconstruct $f=A^H(b)$. The ROI SNR in different upper airway regions were quantified as $SNR_{ROI} = \mu\,(S) / \sigma\,(n)$, where S is a vector with image intensities from the ROI containing a specific upper airway structure, and n is a vector with intensities from an ROI in the background capturing only noise. A total of 10 ROIs were defined: 1) upper lip; 2) lower lip; 3) front tongue; 4) middle tongue; 5) back tongue; 6) hard palate; 7) velum; 8) pharyngeal wall; 9) epiglottis; and 10) glottis. A relative measure of SNR between the two coils was evaluated by the factor SNR_{UA}/SNR_{head}.

Choosing Regularization Parameter in Constrained Reconstruction

The regularization parameter in the constrained reconstruction was chosen empirically, with the best trade-off between artifact removal, and temporal stair-casing. With the single-slice 2.4-mm^2 sequence, and a 12-ms time resolution reconstruction, the effect of regularization parameter on the reconstructions was studied. L-curves were obtained for two data sets with different speech stimuli from different volunteers: volunteer 1 with fast speech stimuli of counting numbers and volunteer 3 with beat boxing. For generating the L-curve, the CG iterations were set to a very high number of 120 to ensure no bias in the norm calculations resulting from small numerical errors; however, in practice, a lower number of iterations of around 40 were sufficient for convergence, where, qualitatively, there was no noticeable change in image quality.

Comparison of Constrained Reconstructions at Various Reduction Factors

Golden-angle time interleaving allows for retrospective reconstruction with arbitrary time resolution. Using the speech stimuli of counting numbers at a rapid pace, reconstructions were performed at 5, 3, 2, and 1 TR for the 2.4- and 1.76-mm^2 single-slice sequences. This corresponded to reduction factors (R) between 2.6- and 13-fold for the 2.4-mm^2 sequence and 4.2- to 21-fold for the 1.76-mm^2 sequence. The regularization parameters were chosen empirically for all cases. A $\lambda_{2.4mm2} = 0.1$ and $\lambda_{1.76mm2} = 0.3$ respectively for the 2.4- and 1.76-mm^2 sequence was used for the 5, 3, and 2 TR cases, whereas a higher $\lambda_{2.4mm2} = 0.2$ and $\lambda_{1.76mm2} = 0.4$ was used for the 1 TR case. The trade-off between residual aliasing artifacts, and temporal blurring (resulting from large temporal footprint), and reconstruction time was analyzed in all the cases.

Visualization of Various Stimuli With Different Reconstructions

To qualitatively depict the gains in improved time resolution, and its utility in capturing fast articulatory

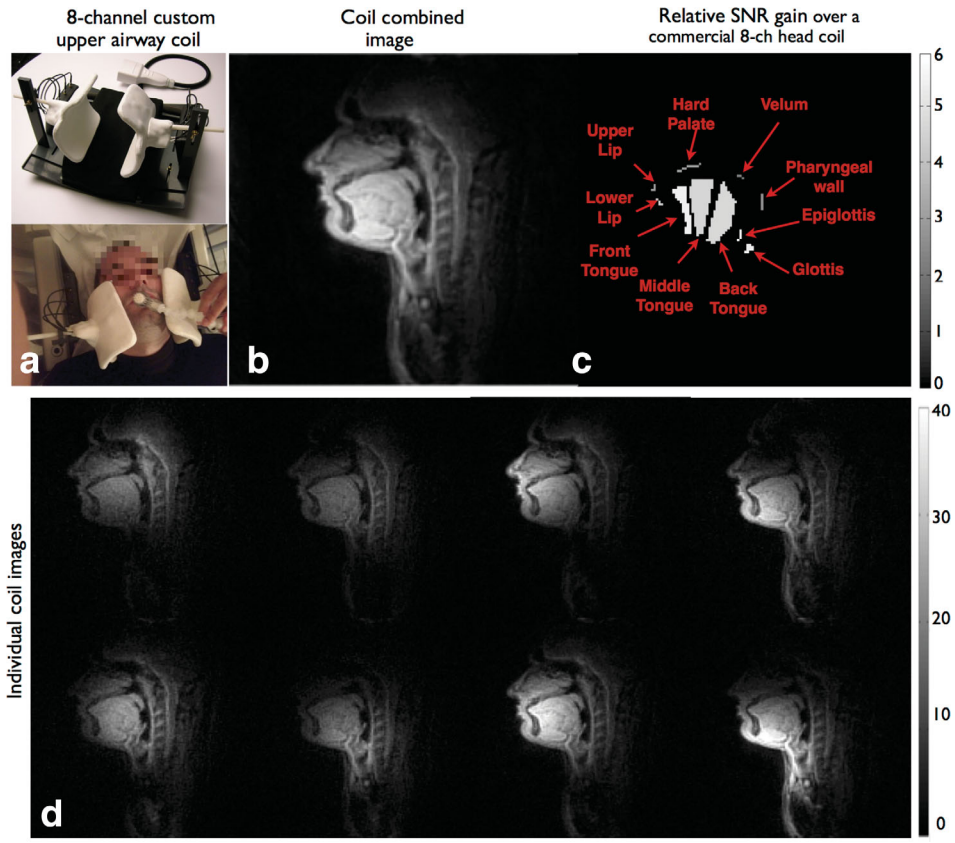

FIG. 2. Custom eight-hannel upper-airway coil and its placement on a subject (**a**). As shown in (**d**), the individual coil images from all the channels depict high signal on all relevant upper airway regions. (**b**) depicts the coil combined image that demonstrates high SNR on all the upper airway regions. (**c**) depicts the relative SNR gain map over a commercial eight-channel head coil, where SNR gains between 2- and 6-fold are observed in all the upper airway ROIs.

movements, constrained reconstructions (from subsampled data) were qualitatively compared to gridding reconstructions obtained at Nyquist (full) sampling using the stimuli of counting numbers (normal pace, followed by rapid pace). A reduction factor of 6.5- to 7.0-fold was used in constrained reconstruction, which corresponded to a temporal footprint of 2 and 3 TR, respectively, for the single-slice 2.4- and 1.76-mm^2 sequences. Constrained reconstructions were also qualitatively compared to the view-sharing reconstruction, which was used in our previous work (21,22). To ensure consistent comparisons, the step size in view sharing was matched to the native time resolution in the constrained reconstruction. Acoustic-articulatory analysis along with articulatory image segmentation was performed on the patient data set. An upper airway image segmentation algorithm (45) was modified to handle image space data and was applied to segment all the important articulators in all the frames; this enabled tracking of the articulators across time.

RESULTS

Simulation of Spatial Versus Time Resolution Trade-offs With Spiral Sampling

Figure 1 shows the improved spatial and temporal resolution trade-off with rate 6.5 undersampled spiral imaging over fully sampled spiral imaging. Fast articulatory movements, such as consonant constrictions, coarticulation events, and flaps, trills are expected to be well depicted at a rate of 6.5-fold single-place spiral imaging over fully sampled spiral imaging. Multiplane imaging compromises the spatial and time resolutions for additional slice coverage. At a rate of 6.5 fold undersampling, three-plane imaging offers spatiotemporal resolutions of up to 2.4 mm^2, and 36 ms/frame, which enables capture of speech tasks, such as consonant constrictions, coarticulation events, and all tongue movements. In contrast, with full sampling, multiplane imaging is notoriously slow for its efficient use in capturing fast articulatory shaping.

Comparison of SNR Between Coils

Figure 2 shows the qualitative comparison of the improvement offered by the custom upper airway eight-channel coil (adult-sized array) over an eight-channel commercial head coil. The individual coil images provide high sensitivity over all relevant upper airway regions. The upper airway coil offered improved SNR for all articulators. The relative SNR gain over the commercial head coil in different upper airway regions were: 2.2, velum; 2.6, pharyngeal wall; 2.9, hard palate; 3, upper lip; 4.3, lower lip; 4.5, midtongue; 4.6, back tongue; 5.2, glottis; 5.4, front tongue; 5.9, epiglottis.

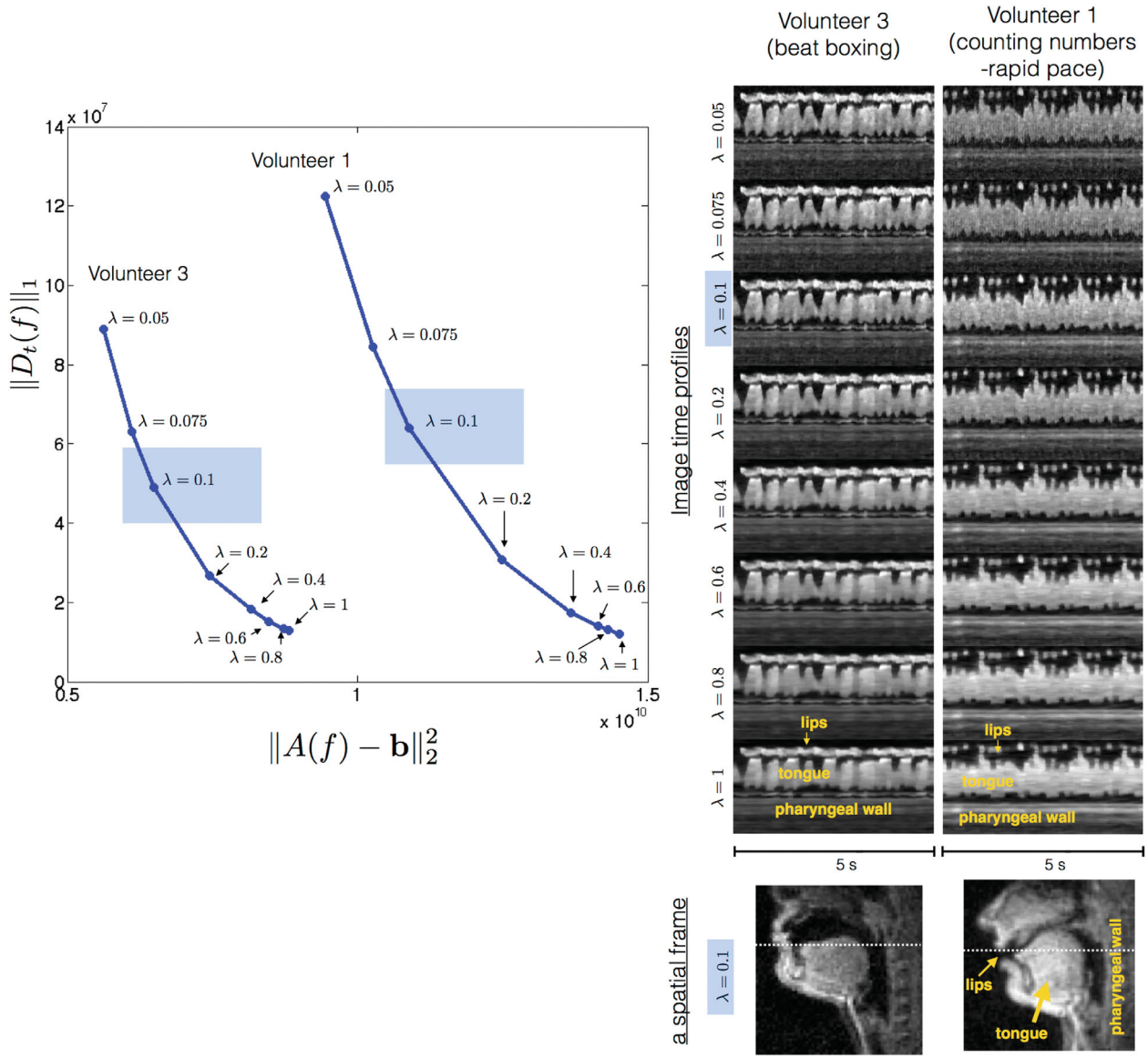

FIG. 3. Determining the regularization parameter λ: The L-curve determines the choice of λ that best controls the bias between the sparsity constraint and data consistency on a global perspective. Our choice, however, was motivated by the local spatiotemporal fidelity of the images. To find the best λ, we relied on an empirical based approach where the temporal fidelity of fast articulators was qualitatively analyzed. The choices of λ greater than 0.5 resulted in greater denoizing within the tissue (e.g., interior of tongue), but occurred at the expense of temporal blurring of the fast articulators (e.g., lips, tongue tip as shown from the image time profiles above). $\lambda = 0.1$ was chosen, which resolved all aliases with maintained temporal fidelity of the articulators and slight residual noise in the interiors of the tissue. This choice correlated with a bias toward data-consistency in the L-curve and a slightly higher sparsity norm in comparison to the inflection point of the L-curve. The behavior was consistent across data sets from different volunteers, with different speech tasks.

Choosing Regularization Parameter in Constrained Reconstruction

Figure 3 shows consistent behavior of L-curve for two subjects with two stimuli that were different: (a) counting numbers at a rapid pace (from volunteer 1) and (b) producing beat boxing sounds (from volunteer 3). The reconstructions produced expected image quality and artifacts with varying choice of λ. The choices of $\lambda > 0.4$ resulted in greater denoizing within the tissue (e.g., interior of tongue), but this occurred at the expense of temporal blurring of the fast articulators (e.g., lips tongue tip). The choice of extremely low λs (<0.05) resulted in residual noise and aliasing. The choice of $\lambda = 0.1$ was empirically chosen—which resolved all aliasing while maintaining temporal fidelity of fast-moving articulators and slight residual noise in the interiors of the tissue. This choice correlated with a bias toward data consistency in the L-curve and a slightly higher sparsity norm in comparison to the inflection point on the L-curve.

Comparison of Constrained Reconstructions at Various Reduction Factors

Figure 4 demonstrates constrained reconstructions at 1, 2, 3, and 5 TR time resolutions with the 2.4- and 1.76-

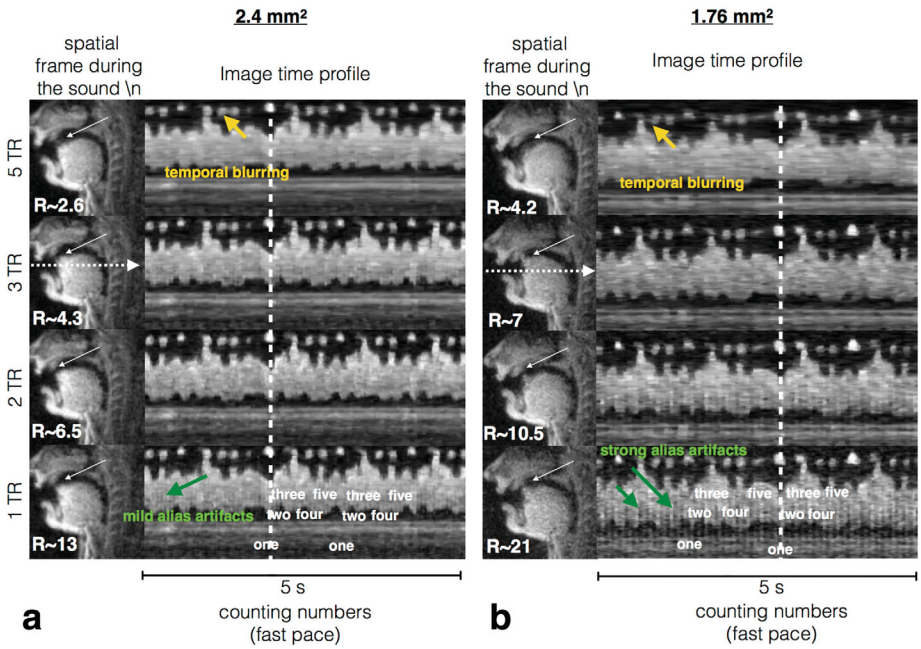

FIG. 4. Evaluation of constrained reconstruction at different subsampling factors using a 5-second speech stimuli of counting numbers at a rapid pace. Constrained reconstructions at 5, 3, 2, and 1 TR time resolutions are compared with the (**a**) 2.4- and (**b**) 1.76-mm^2 single-slice sequences. The spatial frame depicts the event of producing the sound /n/ in the word "one," where the tongue hits the hard palate. The image time profile corresponds to the time evolution of the cut depicted by the horizontal dotted arrows in (**a**) and (**b**). As depicted by the white arrows in the spatial frames, the 5 TR reconstructions produce temporal blurring in capturing the formation of the /n. sound, whereas the 3, 2, 1 TR reconstructions depict this event well. The temporal blurring in the 5 TR reconstructions is also evident in the image time profiles (see yellow arrows). As the acceleration factor is increased, alias artifacts are predominant in the images as seen with R∼21 in the 1 TR reconstruction at 1.76 mm^2 spatial resolution (green arrows on the image time profiles). With R∼13, mild alias artifacts were observed in the 1 TR reconstruction at 2.4-mm^2 spatial resolution. Based on these observations, an acceleration factor of 6.5- to 7.0-fold is chosen for rest of the experiments in this work.

mm^2 single-slice sequences. Whereas the reconstructions with 5 TR demonstrate blurring of the articulatory movements, such as the tongue tip touching the hard palate during the formation of the sound /n/, the reconstructions with time resolutions of at least 3 TR qualitatively demonstrates capturing these events well. This is depicted as increased sharpness along the image time profiles for the 3, 2, and 1 TR reconstructions versus the 5 TR reconstructions in Figure 4. The unresolved aliasing artifacts were qualitatively substantial only at very high subsampling factors, which corresponds to 1 TR reconstruction with the 1.76-mm^2 sequence (i.e., 21-fold acceleration level). Mild aliasing artifacts were present with the 2.4-mm^2 1 TR and 1.76-mm^2 2TR reconstructions. Reconstruction times respectively for reconstructing a 5-second sample with 2.4/1.76 mm^2 sequences were: (a) 5TR, 12/17 minutes); (b) 3 TR, 19/26 minutes; (c) 2 TR, 24/36 minutes; and (d) 1 TR, 45/76 minutes). Based on the trade-offs between temporal resolution (because of large temporal footprint), residual alias artifact energy, and reconstruction times, a conservative acceleration factor of 6.5- to 7.0-fold was empirically chosen in this study.

Comparison of Undersampled Constrained Reconstruction With Fully Sampled Gridding Reconstructions Without and With View Sharing

Figure 5 compares fully sampled (Nyquist) acquisitions using gridding reconstructions against undersampled acquisitions with constrained reconstruction. As depicted in the image time profiles, the proposed reconstruction enables marked improvement in temporal fidelity and enables robust visualization of the fast articulatory shaping, such as tongue tip movements and opening and closure of lips.

Figure 6 demonstrates comparisons of constrained reconstruction against view sharing using the 1.76-mm^2 single-slice sequence. Whereas view-sharing reconstructions qualitatively demonstrated minimal temporal blurring with normal speech rate stimuli, it showed temporal blurring with fast speech rate (approximately 4 times faster than normal pace). For instance, the events of opening of lips and raising of tongue tip toward the hard palate during the production of initial vowel sound in "one" and the lip movement during the sound /f/ in "five," were considerably blurred with view sharing, whereas these events were robustly depicted with constrained reconstruction (Fig 6).

Figure 7 demonstrates one more example from volunteer 2 with free running speech. Similar performances are shown in this figure, where constrained reconstruction showed improved temporal fidelity over view sharing—in particular, in capturing the production of consonant clusters, such as [kr], which involves rapid transitions of articulatory positions, such as the tongue base touching the velar region followed by tongue tip touching hard palate.

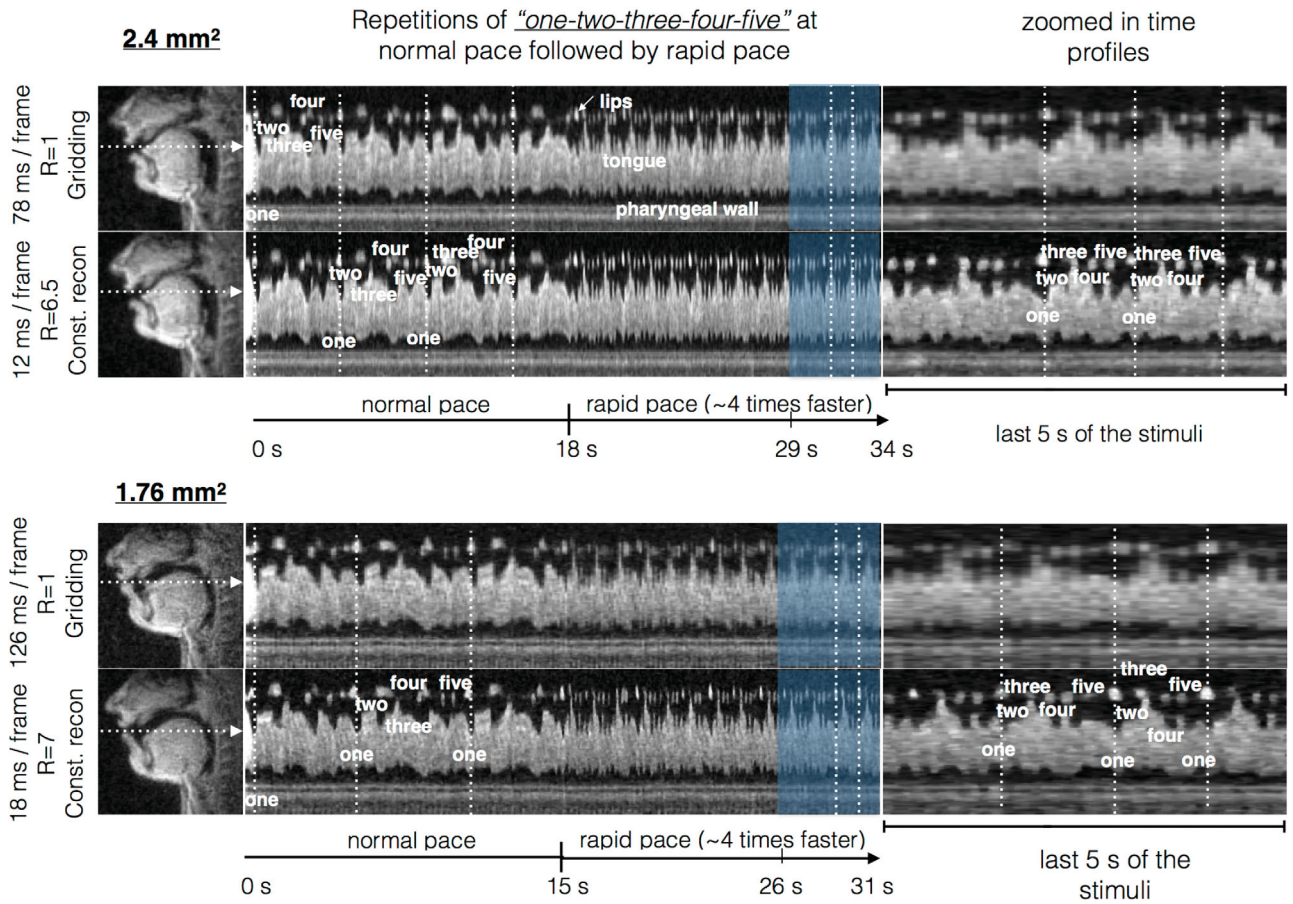

FIG. 5. Demonstration of improved temporal resolution using sparse sampling and constrained reconstruction: A speech task of repeated counting of numbers "one-two-three-four-five" at a normal pace followed by a rapid pace (~4 times faster) was performed. The top and bottom rows show the comparisons against Nyquist spiral sampling and gridding reconstruction with the 2.4- and 1.76-mm^2 sequences respectively, where the time profiles corresponds to the cut marked by the white arrow in the spatial frames (first column). Note the marked improvement in the temporal fidelity (in terms of crispness along time profiles) of the accelerated reconstructions in comparison with the fully sampled reconstructions. As demonstrated in the last column, the gains in time resolution ensure the capture of fast articulatory movements.

Multiplane Imaging of Interleaved Consonant and Vowel Sounds

Figure 8 demonstrates an example of producing the sequence "loo-lee-la-za-na-za" simultaneously in midsagittal, coronal, and axial planes at 36 ms/frame with constrained reconstruction. Concurrent coronal and axial imaging enabled capturing of articulatory shaping that is complementary to midsagittal orientation, such as tongue grooving in the coronal plane, and airway opening and closures in two dimensions at the lower pharyngeal airway level.

Axial Imaging of the Glottis and Velum

Figure 9 shows examples of axial imaging in the regions of the glottis and the velum, concurrently with the midsagittal plane during recording of a short beat-boxing session. Figure 9a shows the changing area of the velopharyngeal port (also see arrows). During the production of the beat-boxing sound, the velum rises to close against the posterior pharyngeal wall, and the lateral pharyngeal walls move to close against the soft palate, which results in a sphincter-type closure that obstructs the flow of air toward the nasal cavity. The third row of Figure 9b gives insight to the dynamics of the glottal cross-sectional area. This area has been modeled as the sum of two components: a slow-varying component (speed comparable to the other speech articulators) and a fast-varying component (vibrating at typically >100 Hz during voiced sounds) (46). The slow-varying component distinguishes between voiced (small cross-sectional area) and unvoiced (large area) sounds and is expected to spread temporally with coarticulation. The time course of the slow-varying glottal component is typically only inferred (47,48). Though the fast-varying glottal component (which determines pitch) is too fast to be observed by MRI, we posit that the slow-varying component can be observed.

Articulatory-Acoustic Analysis of a Preoperative Tongue Cancer Patient

Figure 10 shows the result from an articulatory segmentation algorithm, which segments the nose, upper lip, lower lip, hard palate, tongue, epiglottis, glottis, velum, and pharyngeal wall. This allows for identification and

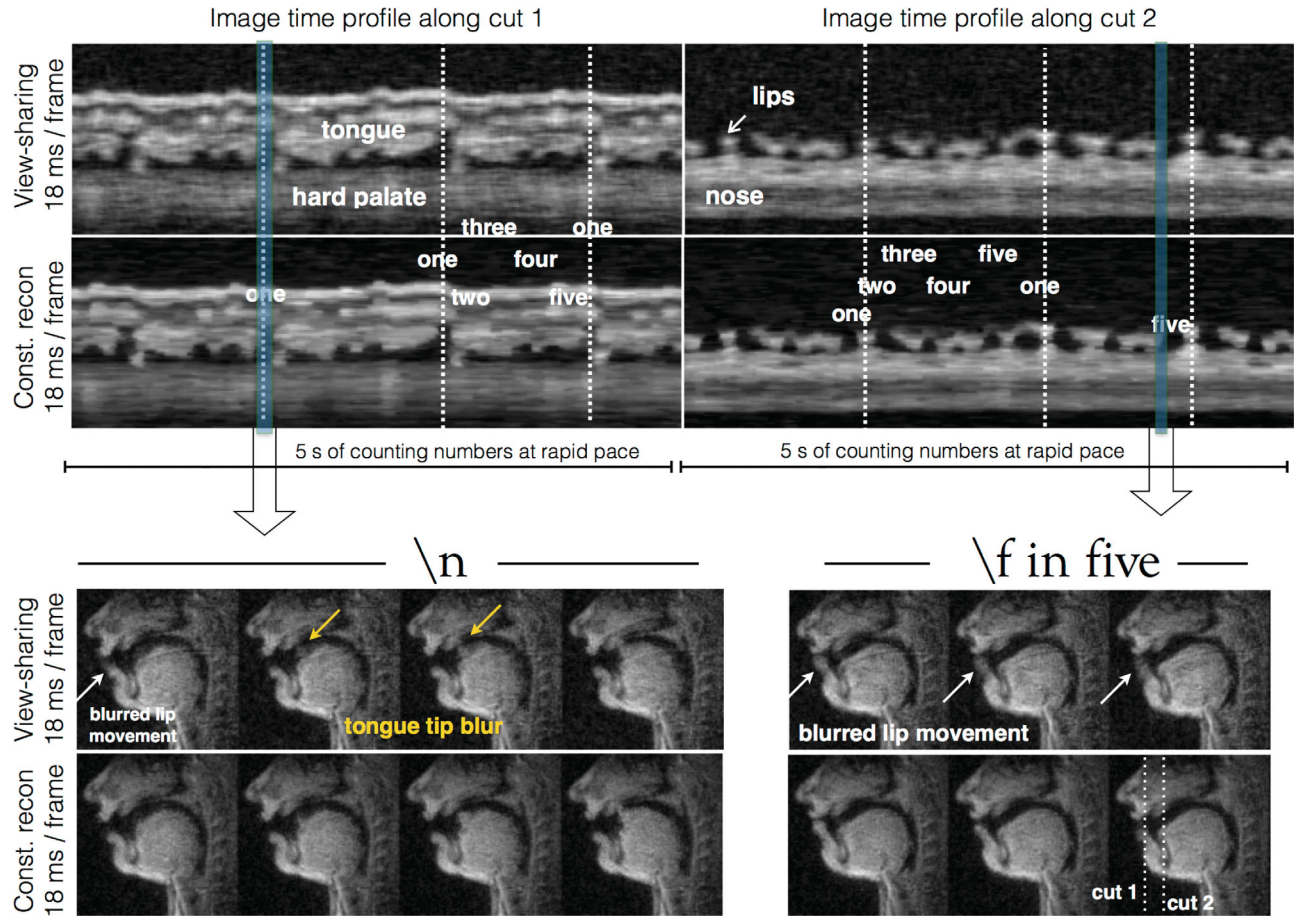

FIG. 6. Demonstration of improved spatiotemporal fidelity with constrained reconstruction, over view-sharing reconstruction. The top row shows two image time profiles corresponding to the cuts marked in the spatial frame of the bottom row. View sharing depicted good image quality in articulators that moved slowly, such as velar motion, but, however, showed motion artifacts of faster articulators, such as those involving lips and the tongue tip. In comparison, constrained reconstruction showed improved overall temporal fidelity (see crispness along time profiles in the top row). The bottom row shows examples of motion blurs in view sharing resulting from fast lip and tongue movements (see white and yellow arrows).

tracking of subtle articulatory motion patterns, such as the tongue contour shape changing, and opening of lips during the production of the word [bit]. ROI analysis is also shown, which gives a more compact view of the articulatory kinematics along certain parts of the vocal tract. Three averaged ROI time profiles are aligned with the noise-cancelled spectrogram of simultaneously acquired audio. In all the words, the first "red" segment in the spectrogram for a healthy subject should correspond to the vowel and the second red segment to the release of the plosive consonant. Plosive consonants, such as /p/,/k/,/t/,/b/,/g/,/d/, have (a) a closure phase; which corresponds to a silence phase in the spectrogram, and (b) a release phase resulting from the sudden opening of the airway.

This example was drawn from the recording of a preoperative cancer patient, who was able to produce normal speech while producing the short words listed in Figure 10. The patient's articulatory trajectories followed normal patterns as depicted by the closure phase of /b/ being marked by a local maximum in the lip trajectory, the closure phase of the /t/ sound being marked by a local maximum in the ROI 2 (tongue tip) trajectory, and its release by the sudden drop in the same ROI 2 (tongue tip) trajectory.

DISCUSSION

The proposed MRI-based system combining custom upper airway coil acquisition, flexible multiplane interleaved fast spiral readouts, constrained reconstruction, simultaneous audio recording, and noise cancellation demonstrated improved visualization of rapid vocal tract dynamics with time resolutions of up to 12 and 36 ms respectively for single- and three-slice imaging. In comparison to fully sampled spiral imaging with gridding reconstruction, undersampled spiral imaging with temporal finite difference constrained reconstruction demonstrated marked improvement in visualization of articulators during fast speech (e.g., consonant constrictions and coarticulation events). View-sharing reconstructions were adapted as a part of on-the-fly reconstruction during data acquisition. Comparisons of the offline constrained reconstruction against view-sharing demonstrated improvements in preserving true temporal resolution with constrained reconstruction, by

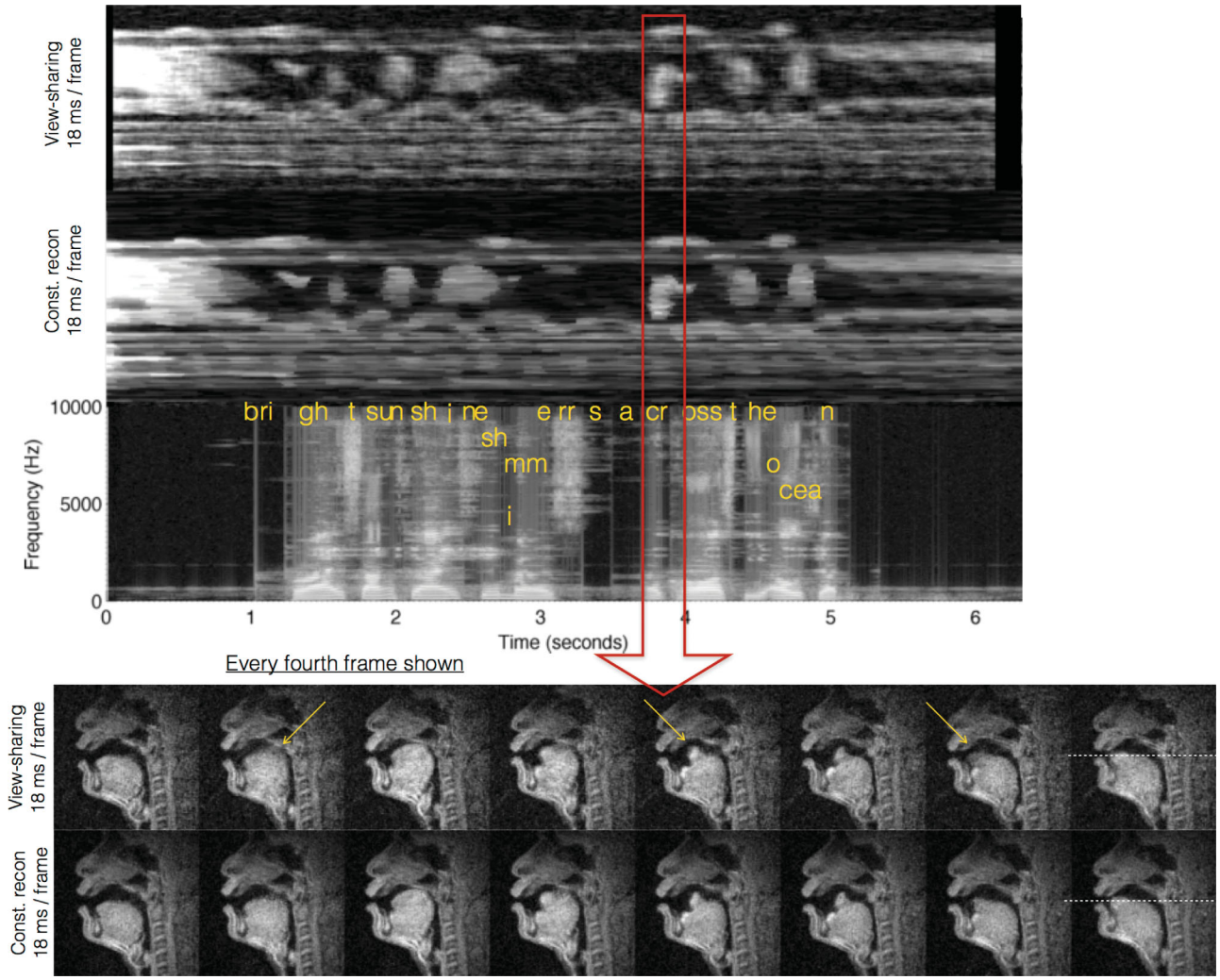

FIG. 7. Single-slice RT-MRI acquisition aligned with a simultaneous acquired audio of speech after acoustic gradient noise cancellation: The speech task was the phrase "bright sunshine shimmers across the ocean" at the subject's normal pace. This phrase contains several consonant clusters and fricatives, which involve fast articulatory movements. For instance, the time window highlighted by the red arrow depicts the formation of the consonant cluster [kr], where the tongue retracts back, and then touches the hard palate. The bottom row shows every fourth frame (for compactness) of the reconstructions to depict the articulatory movement during this sound. Note the constrained reconstruction-based images show less motion blur and loss of temporal fidelity in comparison to view sharing (see yellow arrows).

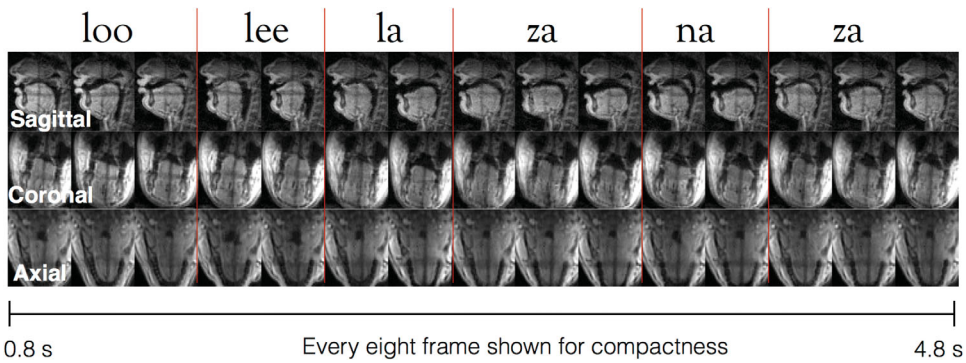

FIG. 8. Simultaneous visualization of sagittal, coronal, and axial planes during production of loo-lee-la-za-na-za sounds using constrained reconstruction at a time resolution of 36 ms and a spatial resolution of 2.4 mm^2. Note movements such as tongue grooving, and pharyngeal airway shaping in two dimensions can be best visualized in the coronal and axial planes, and add complementary information to the sagittal view.

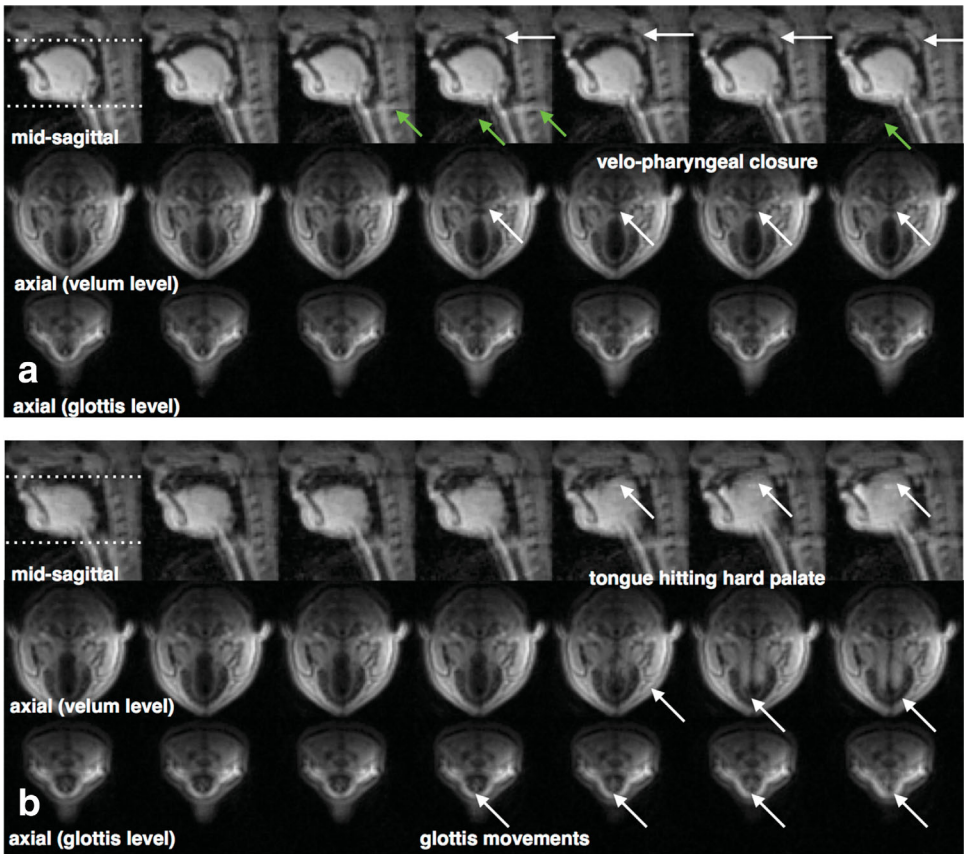

FIG. 9. Demonstration of axial imaging at the areas of the glottis and the velum. The two example sequences are drawn from a short beat-boxing section and correspond to imitations of percussion instruments. Note the potential of imaging the dynamics of the area of the velo-pharyngeal port (which is critical for the characterizing nasal vs. oral speech segments) and the dynamics of the glottal aperture (critical for characterizing voiced vs. unvoiced segments).

reducing the temporal footprint of the number of interleaves required to form a single frame. These differences were visually evident with single-slice 1.76-mm^2 sequence and the multislice 2.4-mm^2 sequences.

A direct comparison of the proposed system against state-of-the-art methods (e.g., (26,27) has not been performed because of practical challenges in implementation at our site (e.g., implementation at 3T, custom k-t sampling requirements (27), acquisition with a 64-channel head coil, and on-the-fly reconstruction with graphics processing units (GPUs) (26)). Comparisons of some of the individual components used in the proposed work against alternate components have been performed in the literature. For instance, the SNR benefits, with a custom coil have been utilized toward improved parallel imaging in Kim et al (49). The trade-off among spatial resolution, time resolution, and artifacts, with different trajectories, has been reported in previous studies (3,25).

With a single temporal finite difference constraint, high acceleration factors (>10-fold) demonstrated only mild artifacts (Figure 4). We attribute this to the improved SNR offered by the custom airway coil and SNR-efficient spiral acquisitions. Performance at these high accelerations may be further improved by enforcing additional spatial sparsity constraints and/or suitable postprocessing methods, such as median filtering (26). Note that additional constraints also introduce additional tuning of regularization parameters. Reconstruction times in this work may also be improved by the use of coil compression (e.g., Buehrer et al (50)), efficient optimization algorithms, such as alternating direction of method of multipliers (51), advanced parallelization of reconstruction (26), use of GPUs (26,52–54)(26, 52–54). Automatic tuning of regularization parameters, as demonstrated in Ramani et al (55), may also be feasible. On our systems, we have found gradient errors to be insignificant with very short spiral readouts (2.4 ms), as used in this work, because there is insufficient time for these to accumulate into significant k-space trajectory errors. However, we acknowledge that a controlled experiment should be performed to confirm this, and that performance may vary among MRI vendors or with gradient subsystems. The coil map estimation in this work assumed the coil maps to be time invariant and were estimated using the eigen decomposition method before reconstruction. Joint estimation of time-varying coil maps along with the reconstruction may further improve the well posedness of the problem and may improve the image quality (e.g., Ying and Shen 56). All the above extensions require detailed investigations and are the scope of future work.

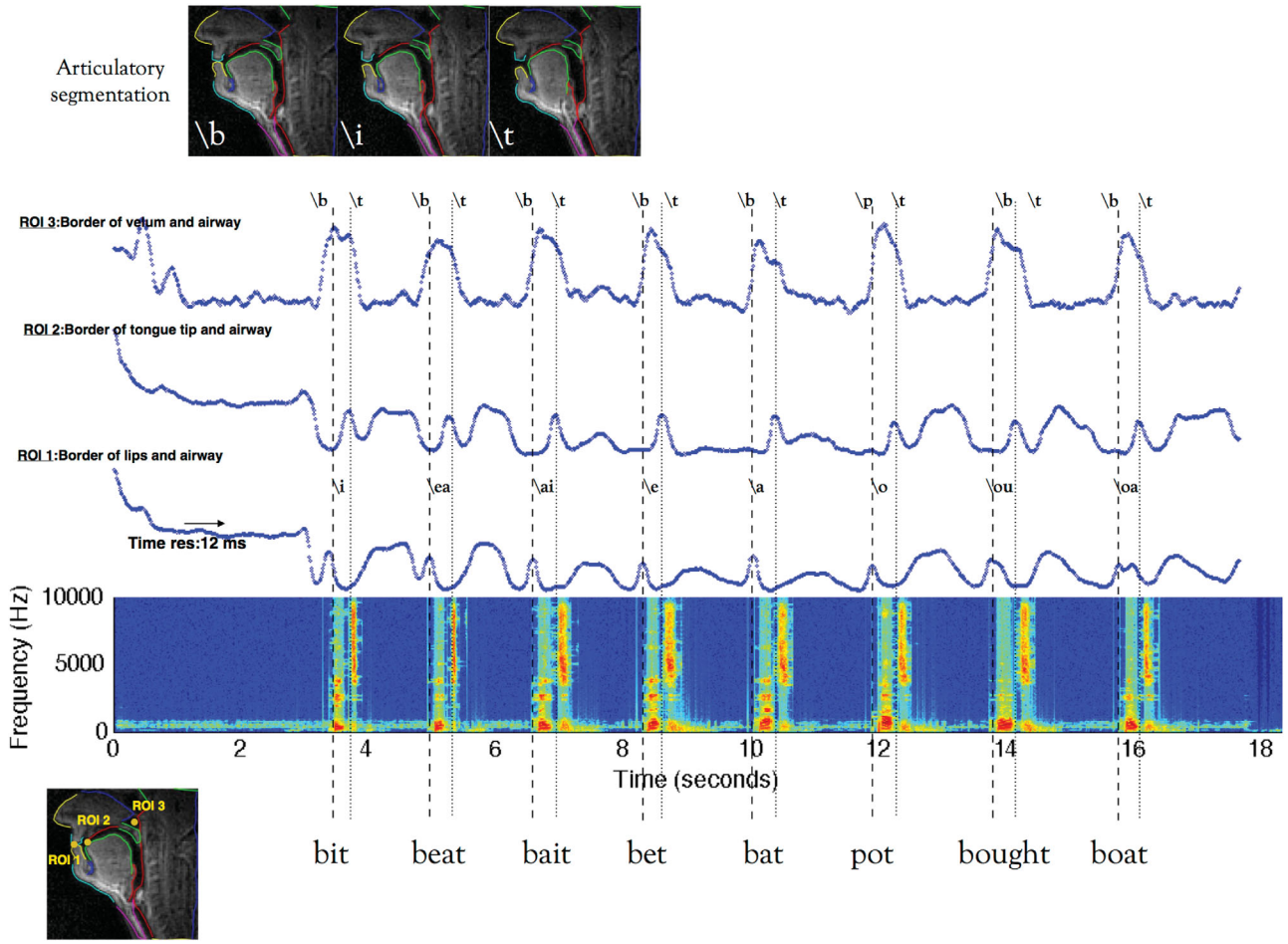

FIG. 10. Single-slice RT-imaging of a preoperative tongue cancer patient at 12 ms/frame: The patient's ability to produce consonant and vowel sounds was studied. Speech stimuli of several short words with the consonant sounds interleaved with vowel sounds as consonant-vowel-consonant was used. ROI-averaged time-course profiles at the airway between lips (ROI 1), tongue tip (ROI 2), and airway between back of velum and pharyngeal wall (ROI 3) are shown at the native time resolution of 12 ms. Note that the high time resolution depicts the formation of all the sounds with excellent fidelity and correlates well with the simultaneously acquired audio signal.

In 2 of the 4 subjects, we observed a cardiac pulsation artifact (Figure 10). Although not shown in this article, we have observed this in several previous studies and have confirmed that the artifact relates to the precise tilt of the subjects head. This artifact occurs when the midsagittal airway plane also intersects with the heart and/or large vessels far outside the imaging FOV. This artifact can typically be corrected on-the-fly by adjusting the scan plane. Offline constrained reconstruction was observed to partially suppress this artifact relative to view-sharing reconstruction .

In this work, image quality from constrained reconstruction was assessed qualitatively because quantitative assessment is challenging in several ways because of the lack of ground truth and the nonlinear nature of the reconstruction. Ground-truth images obtained from mechanically rotating phantoms with known velocities have been previously used to assess temporal fidelity of temporal constraints (57). Rotating phantoms, however, do not fully describe articulatory motion patterns during speech production. It may be possible to extract a ground truth from higher temporal resolution modalities, such as EMA, or high-speed camera imaging for plainly visible articulators, such as the lips and tongue. Comparison of the articulatory motion profiles obtained from constrained reconstruction with RT-MRI against these modalities may be appropriate for investigation of speech tasks over a broad range of speech rate.

Example data from three normal subjects, and 1 patient are shown in this work. The supporting videos (S1-S6) contain RT-MRI movies from parts of our ongoing speech science studies. Additional video material is hosted at http://sail.usc.edu/span/fastmri.html. These include example reconstructions from linguistically driven Puerto Rican Spanish study and a comprehensive study of acquiring all sounds in the entire International Phonetic Alphabet (IPA) chart from four experienced phoneticians.

REFERENCES

1. Bresch E, Kim YC, Nayak K, Byrd D, Narayanan S. Seeing speech: capturing vocal tract shaping using real-time magnetic resonance imaging. IEEE Signal Processing Magazine 2008;25:123–132.
2. Scott AD, Wylezinska M, Birch MJ, Miquel ME. Speech MRI: morphology and function. Physica Medica 2014;30:604–618.

3. Lingala SG, Sutton BP, Miquel ME, Nayak KS. Recommendations for real-time speech MRI. J Magn Reson Imaging 2016;43:28–44.
4. Demolin D, Hassid S, Metens T, Soquet A. Real-time MRI and articulatory coordination in speech. Comptes Rendus Biol 2002;325:547–556.
5. Byrd D, Tobin S, Bresch E, Narayanan S. Timing effects of syllable structure and stress on nasals: a real-time MRI examination. J Phonet 2009;37:91–110.
6. Iltis PW, Schoonderwaldt E, Zhang S, Frahm J, Altenmüller E. Real-time MRI comparisons of brass players: a methodological pilot study. Hum Mov Sci 2015;42:132–145.
7. Perry J, Sutton BP, Kuehn DP, Gamage JK. Using MRI for assessing velopharyngeal structures and function. Cleft Palate Craniofac J 2014; 51:476–485.
8. Ramanarayanan V, Byrd D, Goldstein L, Narayanan SS. Investigating articulatory setting—pauses, ready position, and rest—using real-time MRI. Proceedings of Interspeech 2010 (pp. 1994–1997), Makuhari, Japan, 2010.
9. Proctor M, Bresch E, Byrd D, Nayak K, Narayanan S. Paralinguistic mechanisms of production in human 'beatboxing:' a real-time magnetic resonance imaging study. J Acoust Soc Am 2013;133:1043–1054.
10. Bresch E, Narayanan S. Real-time MRI investigation of resonance tuning in soprano singing. J Acoust Soc Am Express Lett 2010;128: EL335–EL341.
11. 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.
12. Maturo S, Silver A, Nimkin K, Sagar P, Ashland J, van der Kouwe AJ, Hartnick C. MRI with synchronized audio to evaluate velopharyngeal insufficiency. Cleft Palate Craniofac J 2012;49:761–763.
13. 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.
14. Tian W, Li Y, Yin H, Zhao SF, Li S, Wang Y, Shi B. Magnetic resonance imaging assessment of velopharyngeal motion in Chinese children after primary palatal repair. J Craniofac Surg 2010;21:578–587.
15. Kazan-Tannus JF, Levine D, McKenzie C, Lim KH, Cohen B, Farrar N, Busse RF, Mulliken JB. Real-time magnetic resonance imaging aids prenatal diagnosis of isolated cleft palate. J Ultrasound Med 2005;24:1533–1540.
16. Hagedorn C, Lammert A, Bassily M, Zu Y, Sinha U, Goldstein L, Narayanan SS. Characterizing post-glossectomy speech using real-time MRI. In Proceedings of the 10th International Seminar on Speech Production (ISSP), Cologne, Germany, 2014.
17. Zu Y, Narayanan S, Kim YC, Nayak K, Bronson-Lowe C, Villegas B, Ouyoung M, Sinha U. Evaluation of swallow function post tongue cancer treatment using real-time MRI: a pilot study. JAMA Otolaryngol Head Neck Surg 2013;139:1312-1319.
18. Adams SG, Weismer G, Kent RD. Speaking rate and speech movement velocity profiles. J Speech Hear Res 1993;36:41–54.
19. Tasko SM, McClean MD. Variations in articulatory movement with changes in speech task. J Speech Lang Hear Res 2004;47:85–100.
20. 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 3 T. Br J Radiol 2012;85:e1083–e1092.
21. 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.
22. Narayanan S, Toutios A, Ramanarayanan V, Lammert A, Kim J, Lee S, Nayak K, Kim YC, Zhu Y, Goldstein L, Byrd D, Bresch E, Ghosh P, Katsamanis A, Proctor M. Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research. J Acoust Soc Am 2014;136:1307–1311.
23. Niebergall A, Zhang S, Kunay E, Keydana G, Job M, Uecker M, Frahm J. Real-time MRI of speaking at a resolution of 33 ms: Undersampled radial FLASH with nonlinear inverse reconstruction. Magn Reson Med 2013;69:477–485.
24. Burdumy M, Traser L, Richter B, Echternach M, Korvink JG, Hennig J, Zaitsev M, Acceleration of MRI of the vocal tract provides additional insight into articulator modifications. J Magn Reson Imaging 2015;42:925–935.
25. Freitas AC, Wylezinska M, Birch M, Petersen SE, Miquel ME. Real time speech MRI: a comparison of Cartesian and non-Cartesian sequences. In Proceedings of ISMRM 23rd Scientific Sessions (p. 655), Toronto, Canada, 2015.
26. Iltis PW, Frahm J, Voit D, Joseph AA, Schoonderwaldt E, Altenmüller E. High-speed real-time magnetic resonance imaging of fast tongue movements in elite horn players. Quant Imaging Med Surg 2015;5:374–381.
27. Fu M, Zhao B, Carignan C, Shosted RK, Perry JL, Kuehn DP, Liang ZP, #Sutton BP. High resolution dynamic speech imaging with joint low-rank and sparsity constraints. Magn Reson Med 2015;73:1820–1832.
28. Gupta AS, Liang ZP. Dynamic imaging by temporal modeling with principal component analysis. In Proceedings of the 9th Annual Meeting of ISMRM, Glasgow, Scotland, UK, 2001.
29. Liang ZP. Spatiotemporal imaging with partially separable functions. In Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging, 2007. NFSI-ICFBI 2007. Joint Meeting of the 6th International Symposium on (pp. 181–182). New York: IEEE.
30. Zhao B, Haldar JP, Christodoulou AG, Liang ZP. Image reconstruction from highly undersampled-space data with joint partial separability and sparsity constraints. IEEE Transact Med Imaging 2012;31:1809–1820.
31. Hu Y, Lingala SG, Jacob M. High resolution structural free breathing cardiac MRI using k-t SLR, In: Proceedings of the ISMRM 19th Scientific Sessions (p. 4382), Montreal, Canada, May 7–13, 2015.
32. 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.
33. Kim YC, Proctor MI, Narayanan SS, Nayak KS. Improved imaging of lingual articulation using real-time multislice MRI. J Magn Reson Imaging 2012;35:943–948.
34. Santos JM, Wrigh G, Pauly JM. Flexible real-time magnetic resonance imaging framework. In Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE (Vol. 1, pp. 1048–1051). New York: IEEE.
35. Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magnc Reson Med 2007;57:1086–1098.
36. Liu B, King K, Steckner M, Xie J, Sheng J, Ying L. Regularized sensitivity encoding (SENSE) reconstruction using bregman iterations. Magn Reson Med 2009;61:145–152.
37. Todd N, Adluru G, Payne A, DiBella EV, Parker D. Temporally constrained reconstruction applied to MRI temperature data. Magn Reson Med 2009;62:406–419.
38. Adluru G, McGann C, Speier P, Kholmovski EG, Shaaban A, Dibella EV. Acquisition and reconstruction of undersampled radial data for myocardial perfusion magnetic resonance imaging. J Magn Reson Imaging 2009;29:466–473.
39. Feng L, Grimm R, Tobias Block K, Chandarana H, Kim S, Xu J, Axel L, Sodickson DK, Otazo R. Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. Magn Reson Med 2013;717:707–717.
40. Walsh DO, Gmitro AF, Marcellin MW. Adaptive reconstruction of phased array MR imagery. Magn Reson Med 2000;43:682–690.
41. Fessler J, Sutton BP. Nonuniform fast Fourier transforms using min-max interpolation. IEEE Transact Signal Processing 2003;51:560–574.
42. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn Reson Med 2007;58: 1182–1195.
43. Opto-acoustics: MR-compatible fiber optic microphone. http://www.optoacoustics.com/medical/fomri-iii/features; last accessed on 11 January 2016.
44. Vaz C, Ramanarayanan V, Narayanan S. A two-step technique for MRI audio enhancement using dictionary learning and wavelet packet analysis. In Proceedings of InterSpeech (pp. 1312–1315), Lyon, France, August 25–29, 2013.
45. Bresch E, Narayanan S. Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images. IEEE Transact Med Imaging 2009;28:323–338.
46. Maeda, S. Phonemes as concatenable units: VCV synthesis using a vocal-tract synthesizer. In Sound Patterns of Connected Speech: Description, Models and Explanation. Proceedings of the symposium held at Kiel University, Arbeitsberichte des Institut für Phonetik und

digitale Spachverarbeitung der Universitaet Kiel:31, Simpson AP, Pötzold, M, eds., June 1996, pp. 145–164.
47. Toutios A, Maeda S. Articulatory VCV synthesis from EMA data. In Proceedings of InterSpeech, Portland, Oregon, USA, 2012.
48. Laprie Y, Loosvelt M, Maeda S, Sock R, Hirsch F et al. Articulatory copy synthesis from cine X-ray films. In Proceedings of the Inter-Speech 14th Annual Conference of the International Speech Communication Association, Lyon France, 2013.
49. Kim YC, Hayes CE, Narayanan SS, Nayak KS. Novel 16-channel receive coil array for accelerated upper airway MRI at 3 Tesla. Magn Reson Med 2011;65:1711–1717.
50. Buehrer M, Pruessmann KP, Boesiger P, Kozerke S. Array compression for MRI with large coil arrays. Magn Reson Med 2007;57:1131–1139.
51. Ramani S, Fessler J. Regularized parallel MRI reconstruction using an alternating direction method of multipliers. In Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on (pp. 385–388). New York: IEEE.
52. Hansen MS, Sørensen TS. Gadgetron: an open source framework for medical image reconstruction. Magn Reson Med 2011;69:1768–1776.
53. Pryor G, Lucey B, Maddipatla S, McClanahan C, Melonakos J, Venugopalakrishnan V, Patel K, Yalamanchili P, Malcolm J. High-level GPU computing with Jacket for MATLAB and C/C++. In SPIE Defense, Security, and Sensing (pp. 806005–806005). International Society for Optics and Photonics, Orlando, Florida, USA, April 25–29, 2011.
54. Kong J, Dimitrov M, Yang Y, Liyanage J, Cao L, Staples J, Mantor M, Zhou H. Accelerating MATLAB image processing toolbox functions on GPUs. In Proceedings of the 3rd Workshop on General-Purpose Computation on Graphics Processing Units (pp. 75–85). New York: ACM; 2010.
55. Ramani S, Liu Z, Rosen J, Nielsen J, Fessler JA. Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods. IEEE Trans Image Process 2012;21:3659–3672.
56. Ying L, Sheng J. Joint image reconstruction and sensitivity estimation in SENSE (JSENSE). Magn Reson Med 2007;57:1196–1202.
57. Frahm J, Schätz S, Untenberger M, Zhang S, Voit D, Merboldt KD, Sohns JM, Lotz J, Uecker M. On the temporal fidelity of nonlinear inverse reconstructions for real-time MRI—the motion challenge. Open Med Imaging J 2014;8:1–7.

SUPPORTING INFORMATION

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

Supporting Video S1. Single-slice RT-imaging using the 2.4-mm^2 sequence at 12 ms time resolution. The speech task involved counting numbers at a normal pace followed by a rapid pace (approximately 4 times faster). The video also shows the result from an articulatory segmentation algorithm, which segments the nose, upper lip, lower lip, hard palate, tongue, epiglottis, glottis, velum, and pharyngeal wall. Segmentation of articulators allow for tracking

Supporting Video S2. Rapid beat-boxing sounds captured with 2.4-mm^2 spatial resolution at 12 ms time resolution.

Supporting Video S3. Vowel sounds produced by an expert phonetician and captured at 12 ms time resolution. These sounds are a part of a study, which involved imaging all the sounds in the entire International Phonetic Alphabet (IPA) chart.

Supporting Video S4. Consonant sounds produced by an expert phonetician and captured at 12 ms time resolution. These sounds are a part of a study, which involved imaging all the sounds in the entire International Phonetic Alphabet (IPA) chart.

Supporting Video S5. Fluent speech (rainbow passage) produced by an expert phonetician and captured at 12 ms time resolution.

Supporting Video S6. Beat-boxing sounds imaged concurrently in midsagittal, and two axial planes at the level of velum, and glottis at 36 ms/frame.