

Recommendations for Real-Time Speech MRI

Sajan Goud Lingala, PhD,^{1*} Brad P. Sutton, PhD,²
 Marc E. Miquel, PhD,³ and Krishna S. Nayak, PhD¹

Real-time magnetic resonance imaging (RT-MRI) is being increasingly used for speech and vocal production research studies. Several imaging protocols have emerged based on advances in RT-MRI acquisition, reconstruction, and audio-processing methods. This review summarizes the state-of-the-art, discusses technical considerations, and provides specific guidance for new groups entering this field. We provide recommendations for performing RT-MRI of the upper airway. This is a consensus statement stemming from the ISMRM-endorsed Speech MRI summit held in Los Angeles, February 2014. A major unmet need identified at the summit was the need for consensus on protocols that can be easily adapted by researchers equipped with conventional MRI systems. To this end, we provide a discussion of tradeoffs in RT-MRI in terms of acquisition requirements, a priori assumptions, artifacts, computational load, and performance for different speech tasks. We provide four recommended protocols and identify appropriate acquisition and reconstruction tools. We list pointers to open-source software that facilitate implementation. We conclude by discussing current open challenges in the methodological aspects of RT-MRI of speech.

J. MAGN. RESON. IMAGING 2016;43:28–44.

The upper airway consists of several soft-tissue structures and muscles that are intricately coordinated to perform essential human functions such as speech, swallowing, and breathing. Magnetic resonance imaging (MRI) of the upper airway offers a noninvasive way to visualize the morphological and functional aspects of these upper airway structures.^{1–6} It has several compelling advantages over competing modalities such as x-ray fluoroscopy and ultrasound. MRI provides noninvasive, safe imaging of arbitrary image planes and can visualize deep soft-tissue structures such as the velum, pharyngeal wall, and the larynx.

Real-time MRI (RT-MRI) involves the continuous acquisition of MRI images of a dynamically evolving process. This has emerged as a powerful tool to visualize the complex spatiotemporal coordination of upper airway structures during physiological functions such as speech production. As several definitions of “real-time” imaging may exist, for this work we indicate real-time as producing a series of images capturing the motions during the MRI scan which can be aligned with corecorded audio.⁶ This is in contrast

to gated acquisitions, which produce average motions during scanning, and for which recorded audio cannot be exactly realigned with any specific instance of the motion. Due to a large variability in repetition of even simple speech tasks,⁵ RT-MRI is necessary to understand the relationship between particular articulations and speech sounds.

RT-MRI is continuing to provide new insights into biomechanics of vocal articulators such as the tongue, velum, and pharyngeal wall during speech production.^{6–12} It is emerging as a powerful tool in speech research, which has the potential to address several open questions in the areas of phonetics, phonology, language acquisition, and language disorders.

In addition to applications in speech science, RT-MRI has the potential to impact several clinical applications. Conditions such as velopharyngeal insufficiency (VPI) involve incomplete closure of the velopharyngeal port, which occur either due to structural defects of the velum or pharyngeal walls at the level of the nasopharynx, or due to functional (neurological) conditions that manifest as oral

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.24997

Received Apr 30, 2015, Accepted for publication Jun 23, 2015.

*Address reprint requests to: S.G.L., Ming Hsieh Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089. E-mail: lingala@usc.edu

From the ¹University of Southern California, Los Angeles, California, USA; ²University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois, USA; and ³Barts Health NHS Trust, London, UK

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

motor difficulties. VPI is particularly prevalent in cleft palate patients. Assessing VPI is of significant interest to manage speech, and subsequently inform treatment plans, such as speech therapy and surgical intervention. RT-MRI has been demonstrated to be a promising alternative to current functional tests, which are either invasive (nasoscopy), or involve radiation (x-ray video fluoroscopy), and has been adapted in several clinical VPI studies.^{13–19} The clinical role of RT-MRI for VPI assessment has not yet been fully established, and notable hurdles include the cost and complexity of scanning.^{20,21}

RT-MRI and structural MRI have also been explored for assessment of functional restoration for postsurgical assessment of cleft-palate repair. Although structural MRI has seen increasing use in this population, RT-MRI provides potential for examining the functional implications of the repair by examining movement and VPI.^{22–24} RT-MRI in the fetus has been shown to be effective in early diagnosis of cleft palate compared to structural MRI and ultrasound, providing excellent visualization of midline structures.²⁵

Several studies have demonstrated the feasibility of novel clinical applications of RT-MRI of the upper airway. For instance, three studies have demonstrated the feasibility of RT-MRI in providing sufficient spatiotemporal resolution to characterize swallowing in real time.^{26–28} A few clinical studies of swallowing have been carried out using RT-MRI, for example comparisons to existing invasive modalities on dysphagia patients,²⁹ assessment of swallowing disorders in normal and brainstem lesion patients,³⁰ and evaluation of swallowing in postglossectomy patients have been demonstrated.³¹

RT-MRI has been utilized to analyze and compare speech in normal subjects and postglossectomy patients, where portions of the infected tongue are treated with surgical resection (with or without reconstruction), and radiation therapy.^{32,33} Such analyses may provide patients and their speech therapists with useful information in managing speech posttreatment. Other RT-MRI feasibility studies include characterization of speech in apraxia³⁴ and evaluation of the obstructions in the upper airway during sleep apnea.^{35–38}

In several research and clinical studies, RT-MRI has presented a challenging trade-off between spatial resolution, temporal resolution, signal-to-noise ratio (SNR), and artifact suppression. Several of these factors can be traded differently based on the upper airway task of interest. Over the past decade, a number of technical advancements in acquisition strategies, coil geometry design, reconstruction and artifact correction schemes have emerged, thereby providing a wide range of solutions to RT-MRI tradeoffs.^{1,6,39–47} This article focuses on the technical considerations relevant to implementing an RT-MRI protocol for upper airway imaging.

This article is motivated by an outcome of the Speech MRI Summit meeting (Los Angeles, February 2014,

endorsed by the ISMRM), where the speech MRI research community convened and agreed that there was an unmet need for a recommendations document for RT-MRI of speech. The purpose of this article is to provide a set of recommended protocols that could be easily adapted by new research groups entering the field. We highlight the choices of sequences, sampling, and reconstruction schemes that are appropriate during various speech tasks. We also highlight several open-source software packages that could facilitate implementation of the recommended protocols. We conclude by discussing the current open challenges in the methodological aspects of RT-MRI for speech.

Speech Tasks: Imaging Requirements

The spatiotemporal coordination of upper airway articulators varies in accordance with the speech task at hand. For instance, in a task that involves the production of sustained sounds, the spatial position of the articulators changes on the order of a few seconds (eg, posture of velum in contact with the pharyngeal wall during sustained production of the vowel “\a”). On the other hand, in a task that involves producing consonant clusters movements of the articulators occur on an order of a few milliseconds. Owing to the speech questions being asked, the requirements of spatial and temporal resolution in RT-MRI vary. Based on our experience, evidence from the literature, and a consensus amongst the audience of the 2014 speech MRI summit involving RT-MRI imaging researchers and linguists, we provide RT-MRI spatiotemporal resolution requirements for various speech tasks in Fig. 1, and provide explanations below.

For studying velopharyngeal closure, previous studies have reported time resolutions ranging from 40–200 msec. Since time resolutions >150 msec may carry a risk of missing closure events, we provide a recommendation of 40–150 msec.^{48–52} We further prescribe resolutions less than 100 msec to be a target zone, as velic closure events can be captured over many time frames for normal-paced speakers.⁴⁹ Previous studies reported in-plane spatial resolutions ranging from 1 mm² to 5.2 mm².^{41,50–53} Although there was still a good correlation between closure events visualized by video-fluoroscopy and MRI using a low in-plane resolution (4.8 mm²)⁵⁰, since the velum is a small structure and can vary in size for different subjects (adults vs. children), we recommend spatial resolutions ranging between 1 to 4 mm², and prescribe the mid-sagittal orientation to view velic closure events. Groups particularly interested in quantifying gaps between the velum and the postpharyngeal wall in VPI patients should aim for the higher in-plane resolutions. With the mid-sagittal orientation, slice thicknesses of 6–8 mm are adequate to observe closure events and maintain a high SNR throughout the image series.

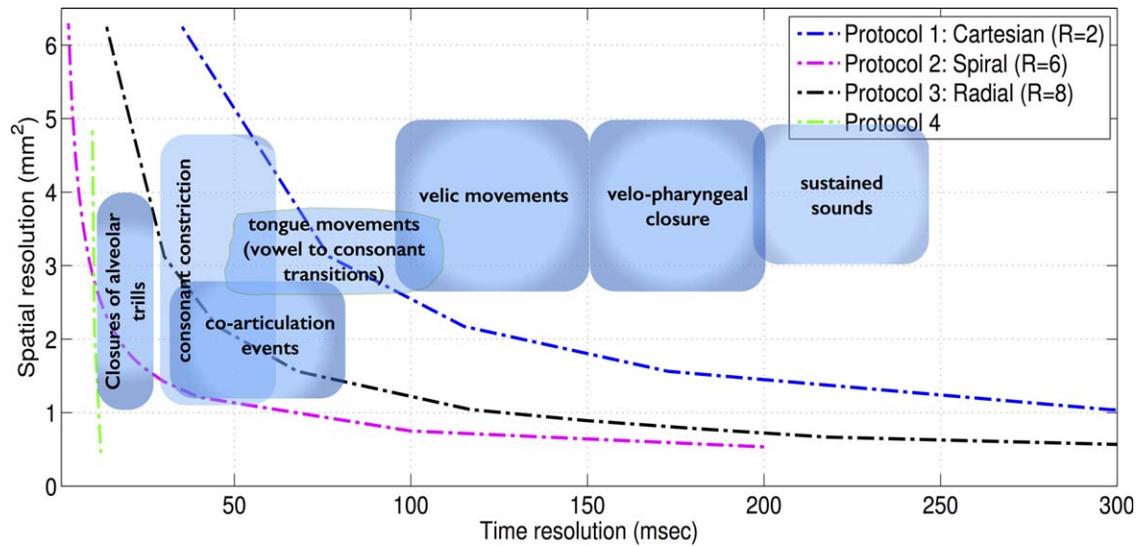


FIGURE 1: Spatiotemporal resolution requirements for various speech tasks. The placement of these “zones” reflects the current consensus opinion among speech imaging researchers in attendance at the 2014 Speech MRI Summit. Boundaries are approximate due to the lack of gold-standard imaging techniques, and are being refined through the more widespread adoption of non-invasive techniques such as RT-MRI. The plots depict the spatiotemporal resolutions that could be realized by the four recommended protocols in Table 1. Note that each of the protocols have tradeoffs beyond spatial and temporal resolutions, which require careful consideration and are discussed in detail in the text. For instance, SNR losses at high spatial resolutions (all protocols), off-resonance artifacts (Protocol 2), long acquisition times (Protocol 4).

A speech task involving consonant to vowel sound transitions has the articulators move on the order of a few tens of milliseconds. These transitions involve movements such as bulk tongue movements, subtle tongue tip

movements such as tongue tip hitting the back of teeth, tongue touching the palate, the velum touching the pharyngeal wall, and the airway narrowing. To visualize the tongue movements, we recommend a time resolution of 50–100

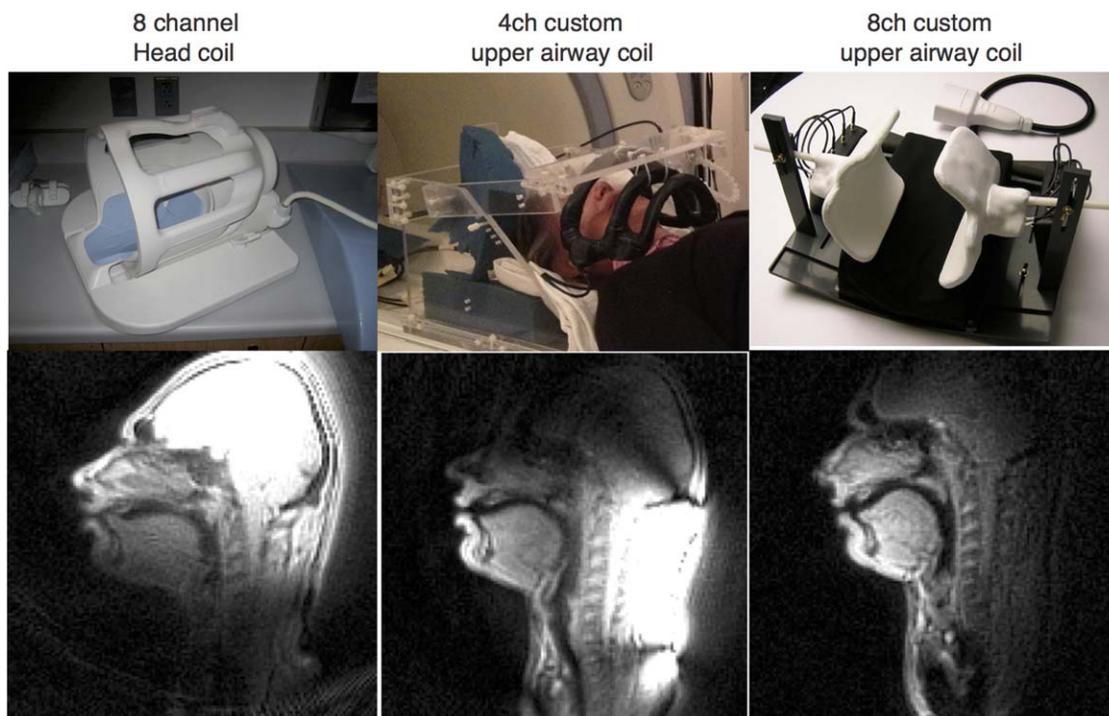


FIGURE 2: Comparison of image quality from different receiver coils in one adult subject. Images were acquired using Protocol 2, spiral gradient echo imaging; $2.4 \times 2.4 \text{ mm}^2$ spatial resolution. Compared to (left) a standard head coil (middle, right) custom airway coils provide improved SNR in all vocal tract articulators including the lips, velum, epiglottis, and tongue. The SNR gain ranges from 50% to 1600%, and is greatest for articulators closest to the coil elements (the lips in this case); also see.⁵⁵ Custom array coils can also provide improved parallel imaging performance (not shown).

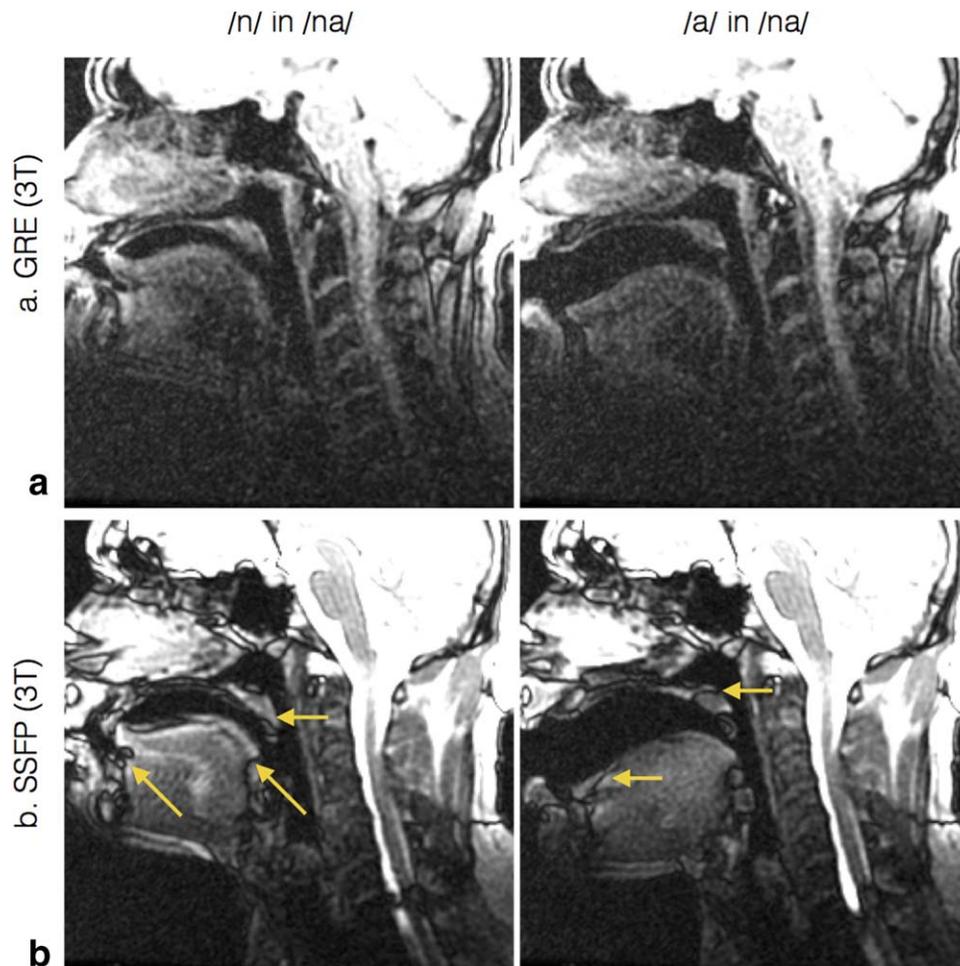


FIGURE 3: Comparison of GRE and SSFP image quality at 3T. Note that the SSFP images have higher SNR but are sensitive to off-resonance and its associated banding artifacts. In this example, the banding artifacts manifested as signal voids on the velum and portions of tongue (see arrows).

msec in the mid-sagittal plane. An in-plane spatial resolution of no more than 3.5 mm^2 is recommended. High time resolutions of below 70 msec are required to study very fast articulatory movements such as those during consonant constrictions and coarticulation events.⁴⁸ Finer spatial resolutions are required for better definition of the border of the tongue, which is important for image segmentation tasks for subsequent quantitative analysis. For movements involving tongue grooving, and airway narrowing, the coronal and axial planes can be used to add complementary information in addition to the mid-sagittal plane.^{39,46,47,54} Multiple planes can be acquired by utilizing time-interleaving sampling of planes or 3D acquisitions.

Acquisition Considerations

Choice of Receiver Coil

Standard clinical 1.5T and 3T scanners are equipped with many receiver coils designed for brain imaging, neurovascular imaging, or combined head and neck imaging. These include single-channel birdcage head coils, multi-channel head coils, multichannel head and neck coils,

and multichannel neurovascular coils. These coils are not optimized for performance over the upper airway vocal tract, but are adequate for some speech experiments. In general, it is favorable for MRI receiver coils 1) to be as close as possible to the tissues of interest (higher signal), and 2) contain highly localized elements (lower noise). These two factors together provide favorable SNR. As depicted in Fig. 2, custom upper-airway coils offer improved SNR in several regions of interest containing the vocal tract articulators (also see⁵⁵). Due to the advantages of high SNR and potential for parallel imaging, we recommend the use of custom coils when available, followed by commercial head and neck coils, and commercial neurovascular coils. Note that superior SNR can be traded off for improved spatial and/or temporal resolution via parallel imaging.

Choice of Field Strength and Pulse Sequences

Imaging at higher field strengths has the benefit of providing high SNR. However, upper airway imaging presents challenges due to the magnetic susceptibility differences

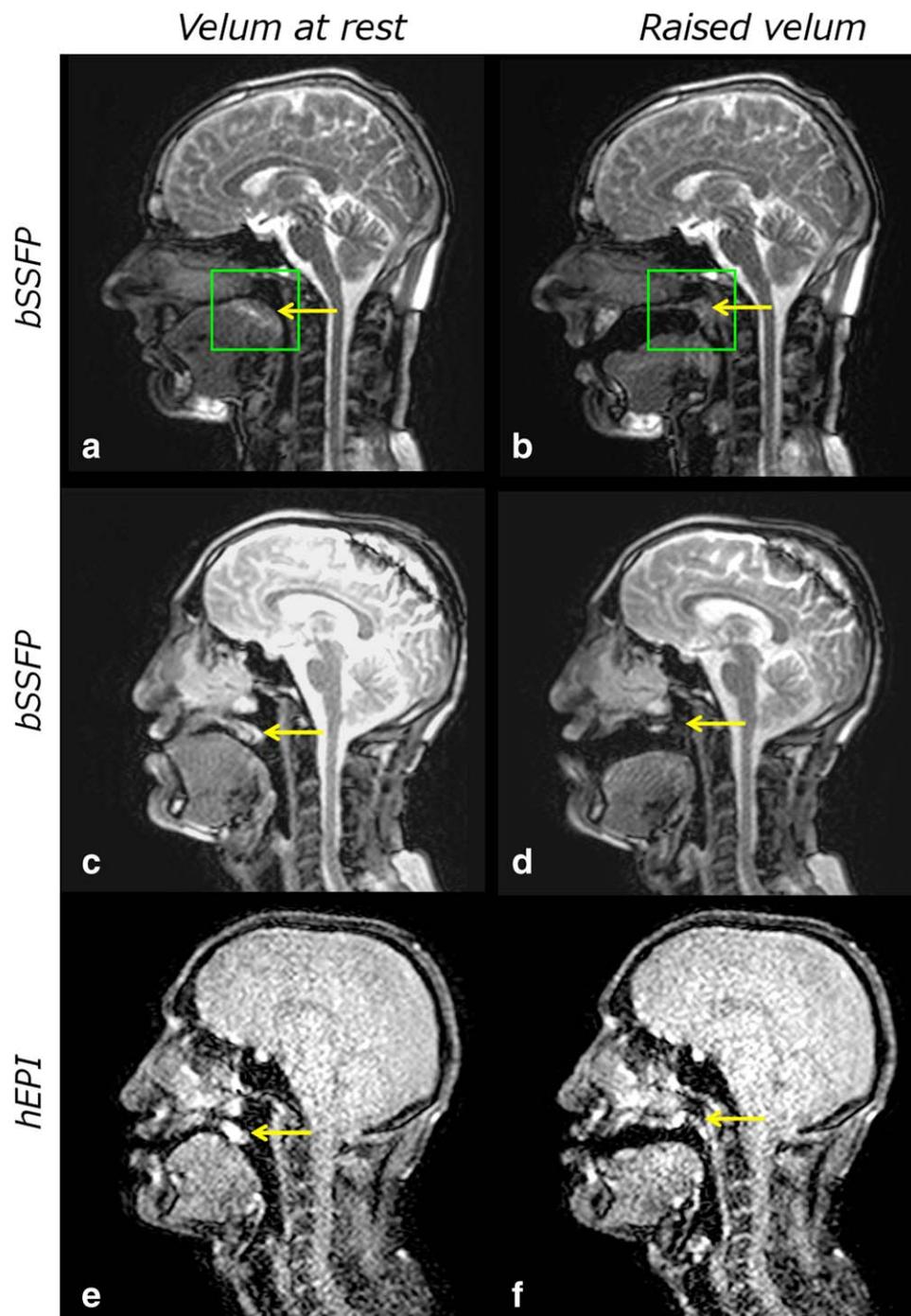


FIGURE 4: Example images of velopharyngeal closure obtained at 1.5T using protocol 1. (a,b) The placement of the shim volume (green box) centered around the velum (yellow arrow) for a 39-year-old male volunteer. Despite careful shimming images can degrade substantially in some subjects during the speech sample acquisition as demonstrated by images (c,d) in a 28-year-old patient with repaired cleft lip and palate. The velum (yellow arrow) almost completely disappears in the elevated position (d). Although h-EPI sequences have a lower SNR (e,f), it is possible to track the velum (yellow arrow) throughout the speech sample.

between air and tissue (roughly 9.41 ppm^{56}). At higher field strengths, this results in a large degree of off-resonance around air-tissue interfaces. Consider the main imaging targets of RT-MRI: the tongue, velum, and palate. These structures are in contact at rest and they separate during speech, thus creating large areas of air-tissue interfaces that are prone to magnetic susceptibility artifacts.

Both gradient echo and spin echo sequences have been used for upper airway imaging, with the latter being used for static imaging. Rapid gradient echo sequences are widely used in RT-MRI applications. These sequences can be split into two categories based on how the remaining transverse magnetization is dealt with at the end of each repetition (TR). Radiofrequency (RF) spoiled gradient echo sequences

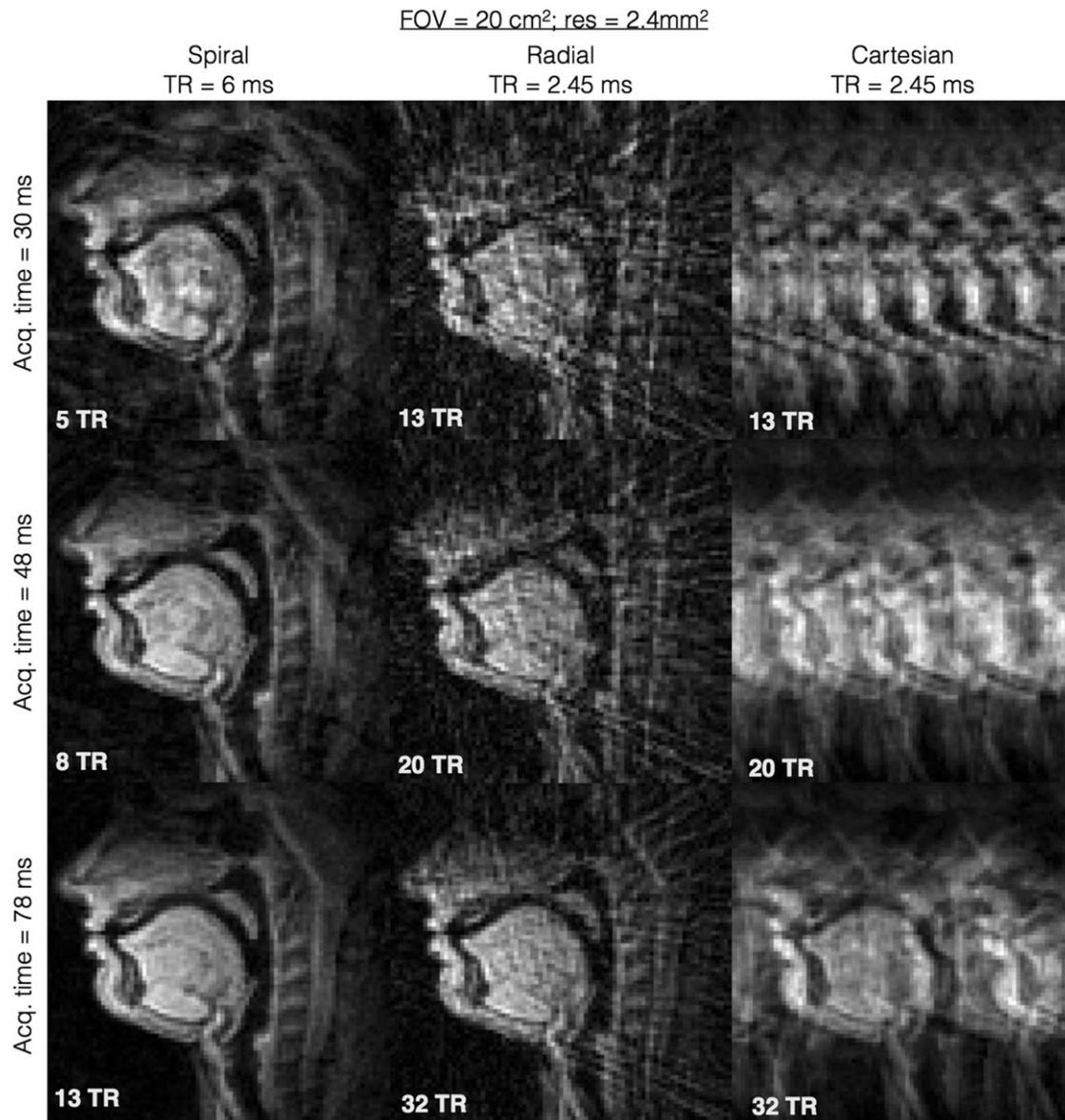


FIGURE 5: Simulated aliasing artifacts for spiral, radial, and Cartesian acquisitions. Simulations are based on $2.4 \times 2.4\text{ mm}^2$ resolution with $20 \times 20\text{ cm}^2$ FOV of a mid-sagittal slice, and are realized by retrospective subsampling of a reconstructed image formed from a 34 interleave spiral dataset, using different trajectories. All the trajectories were simulated with realistic TRs that were obtained from a 1.5T GE scanner with modern gradient specifications (40 mT/m amplitude and 150 mT/m/ms slew rate). The left, middle, and right columns respectively correspond to Protocols 2, 3, 1; all based on gradient echo imaging. In comparison to Cartesian aliases, spiral, and radial aliasing artifacts are less coherent and less visually detrimental to visualization of the vocal tract air space and articulators.

eliminate the residual transverse magnetization while maintaining the T_1 weighting by incrementing the phase of the RF pulse in combination with gradient spoiling, while steady-state free precession (SSFP) sequences utilize the residual magnetization to improve the SNR leading to a T_2/T_1 contrast.⁵⁷ SSFP sequences are particularly sensitive to magnetic field inhomogeneities, and can suffer from large areas of signal nulling or banding artifacts (eg, Fig. 3). It is possible to minimize off-resonance artifacts through careful shimming over the regions of interest, and tuning the choice of center frequency. However, in practice it is difficult to reliably obtain good image quality at 3T with SSFP

sequences. At 3T, we would recommend the use of spoiled gradient echo sequences, as they have the advantage of reduced sensitivity to off-resonance, and are simpler to implement. At 1.5T, SSFP sequences are easier to shim and, due to their increased SNR, are the preferred option for RT-MRI.

Potential users should remain aware of the difficulty in obtaining artifact-free images across large regions. Consequently, users have to focus their efforts on shimming the part of the vocal tract that is specifically of interest. Figure 4a,b illustrates the region shimmed for a velopharyngeal closure study. This will increase image quality in the

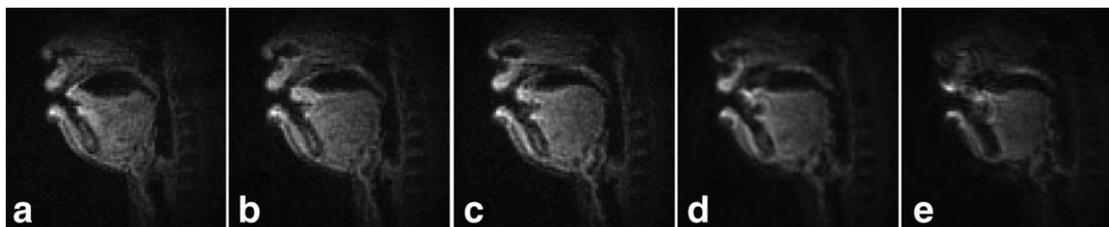


FIGURE 6: Effect of spiral readout duration on image quality. Each image represents a single frame captured when the subject sustained the nasal sound /n/. The spiral readout durations were (a) 2.520 msec, (b) 3.584 msec, (c) 4.576 msec, (d) 6.368 msec, and (e) 10.560 msec. Identical shim values were used for (a–e). Blurring artifacts are most prominent near air-tissue interfaces of the tongue, lips, and velum. (Figure courtesy: Y-C. Kim, Samsung Research, Seoul, South Korea.)

region of interest but will increase artifacts in other areas, for example the brain, as shown in Fig. 4c,d. We recommend that users test their shim on short speech samples (eg, counting from 1 to 5) as the shim might not be adequate for the range of movement, causing images to degrade substantially during speech. Adjusting the position and size, both in-plane and through-plane, of the shim volume will resolve the issue in most cases, although artifacts caused by motion are likely to remain. Degradation of image quality is particularly prevalent in the case of the velum, as illustrated in Fig. 4c,d.

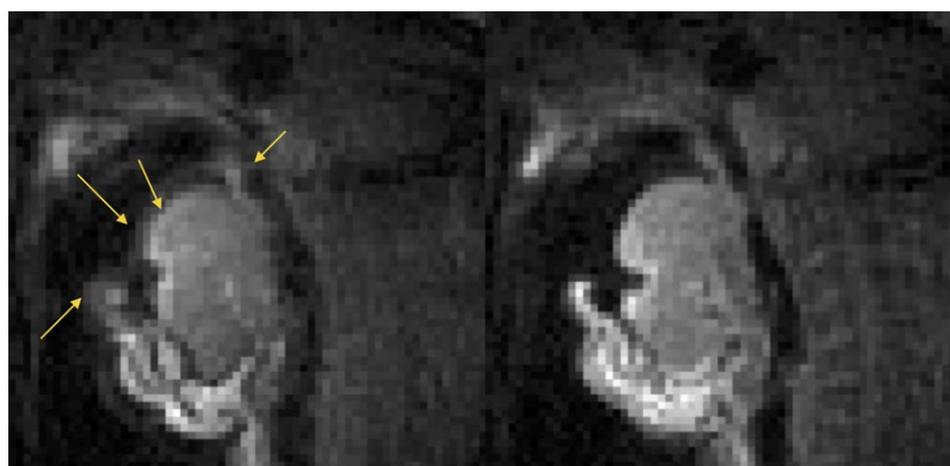
Image quality is subject-dependent and in some cases it can be difficult to maintain image quality throughout the speech sample. In those cases, it is recommended to use spoiled sequences; for example, hybrid echo planar imaging (EPI) can provide a viable alternative (Fig. 4e,f) with fewer artifacts in the velum at the expense of a lower SNR.⁵⁸

It is hard to predict which subjects will have images of lower quality; even the disruptions associated with dental work do not consistently degrade image quality. Groups planning to study patient populations who are undergoing orthodontics treatment should be aware that orthodontics devices can drastically alter shim and image quality and in the most extreme cases render RT-MRI impossible.⁵⁹

Choice of k -space Sampling Strategy

Cartesian sampling is widely used in MRI due to simplicity of reconstruction and the fact that it is robust to many artifacts (except motion). 2D Cartesian sampling (eg, 2D Fourier Transform [2DFT]) is used by real-time interactive imaging sequences available on several commercial platforms. It can be accelerated by 2–3-fold using half Fourier techniques in combination with parallel imaging, using widely available coil arrays. As shown in Fig. 1, such a setup can be used to study a wide range of speech tasks.

Non-Cartesian trajectories are employed when higher spatial and/or temporal resolution is desired.^{6,42–46,60} Spiral sampling provides higher efficiency compared to Cartesian, as a large fraction of k -space is acquired during each TR. Spiral and radial sampling provide reduced motion artifact because they both naturally oversample the center of k -space.⁶⁰ Undersampled non-Cartesian imaging results in relatively incoherent aliasing (eg, radial streaking, spiral ringing). This artifact can often be read through, and is less detrimental to image quality compared to coherent aliasing in undersampled 2DFT. While radial sampling is $\pi/2$ less efficient than Cartesian sampling, it is common to under-sample by at least that much because the aliasing artifacts are mild in appearance (Fig. 5). Spiral sampling typically



a In-correct center frequency

b Adjusted center frequency

FIGURE 7: Off-resonance blurring in spiral MRI can be mitigated by real-time adjustment of the center frequency. Note that air-tissue interfaces (yellow arrows) are the locations of largest variation in resonant frequency.

requires longer readout times, which increases the sensitivity to off-resonance and manifests as spatial blurring (Fig. 6). We recommend use of ≤ 2.5 msec readouts at 1.5T in combination with careful shimming.^{2,6} This can be done by acquiring field maps, or by visual inspection during real-time adjustment of the center frequency (Fig. 7). When operating at higher fields (3T), reconstruction should incorporate magnetic field inhomogeneity correction; these schemes require acquisition of dynamic field maps.⁴⁰

Simultaneous Audio Acquisition

Simultaneous acquisition of an audio signal along with the RT-MRI data is essential for subsequent analysis and modeling of acoustic-articulatory relationships. MR-compatible fiberoptic microphones are popular, commercially available, and have been used successfully in several MRI studies for audio acquisition (eg, by Opto-acoustics^{5,61}). In addition to recording the audio, there is a need to cancel the acoustic noise caused by the MRI gradients during imaging, which is in the audible frequency range and often over 100 dB.^{62,63}

The Opto-acoustics commercial audio acquisition setup utilizes two microphones that are mounted at different angles. The main transducer is positioned towards the mouth of the subject, in a unidirectional manner to capture the subject's audio. The second transducer is positioned at a 90° angle from the first, and is positioned in an omnidirectional orientation to capture only the gradient noise components. Integrated software based on adaptive noise cancellation offers real-time noise cancellation based on the signals recorded from these two microphones.

Custom noise cancellation algorithms can be utilized to further improve the SNR in the recorded audio. These algorithms operate on the raw signal from the microphones, and are usually implemented offline. One such algorithm leverages the knowledge of periodicity of MRI acoustic noise (eg, spiral/radial pulse sequences where the spiral interleaves/radial spokes are repeated with a predetermined period) to create an artificial noise signal, which is then used as a reference noise signal in standard adaptive noise cancellation algorithms.⁶⁴ Other techniques based on correlation subtraction have also been proposed; unlike adaptive filtering, these require data from only one microphone.⁵ Two recordings of the noise from the same microphone are taken with and without speech. These recordings are aligned to obtain a maximum discrete cross-correlation, and then the signal without speech is subtracted from the signal with speech to filter the periodic acoustic noise. This method has been recently improved by combining it with spectral domain filtering, and is available as an open-source package.^{62,65} A data-driven algorithm that is blind to the acoustic properties of noise has also been recently proposed, and is also available as open-source software.^{66,67}

Reconstruction Considerations

On-the-Fly Reconstruction

We refer in “on-the-fly” reconstruction to schemes that are capable of providing reconstructed images with latency reconstruction times of less than 500 msec. On-the-fly reconstruction is desirable in RT-MRI due to the flexibility of changing scan planes on-the-fly along with visualization and immediate compensation of artifacts. This is particularly beneficial during scan localization, and should be used whenever feasible. For instance, obtaining an exact mid-sagittal slice location often requires traversing through several cuts along the articulators in different planes. Also, localizations may be performed during short speech tasks to visualize the articulatory movements and determine a plane that can best capture the task. Clinical RT-MRI applications such as VPI, cleft palate, and postglossectomy imaging all the more require the frequent updates of scan planes during localization, as the geometry of the articulators are considerably different from normal subjects. For instance, patient-specific oblique planes (eg, angular cuts along the velopharyngeal port in VPI) in addition to the mid-sagittal plane may be required.

Cartesian parallel imaging reconstructions such as sensitivity encoding (SENSE) and generalized autocalibrating partially parallel acquisitions (GRAPPA) have minimal latency reconstruction times, and are widely available on modern-day scanners.¹ By combining the parallel imaging capabilities with the head and neck coil geometries, and Cartesian trajectories, studies have reported high-quality reconstructions up to an acceleration factor of 2–3-fold.¹

Parallel imaging acceleration leads to a spatially varying reduction of SNR, and the noise distribution in the resulting images depends on the coil geometry and scan prescription. Custom upper airway coils offer improved SNR, and are favorable to parallel imaging reconstructions. Reconstruction schemes from non-Cartesian trajectories are not commonly available on clinical scanners; however, many research sites have successfully developed on-the-fly reconstructions for non-Cartesian radial and spiral trajectories.^{6,42,45,47,60,68,69} As described in the previous section, non-Cartesian sampling has advantages in terms of efficient time sampling compared to Cartesian trajectories, but may have additional artifacts due to magnetic field inhomogeneity.

Gridding in combination with a sliding window reconstruction has been implemented.⁶ At higher undersampling factors, the incoherent artifacts in the gridded images can often be read-through. Although the final image series may be reconstructed offline for analysis, it is still desirable to have on-the-fly reconstruction options to ensure the correct imaging prescription and data acquisition.

Offline Reconstruction

In order to improve the tradeoff between spatial resolution and frame rate, recent interest has been on methods that

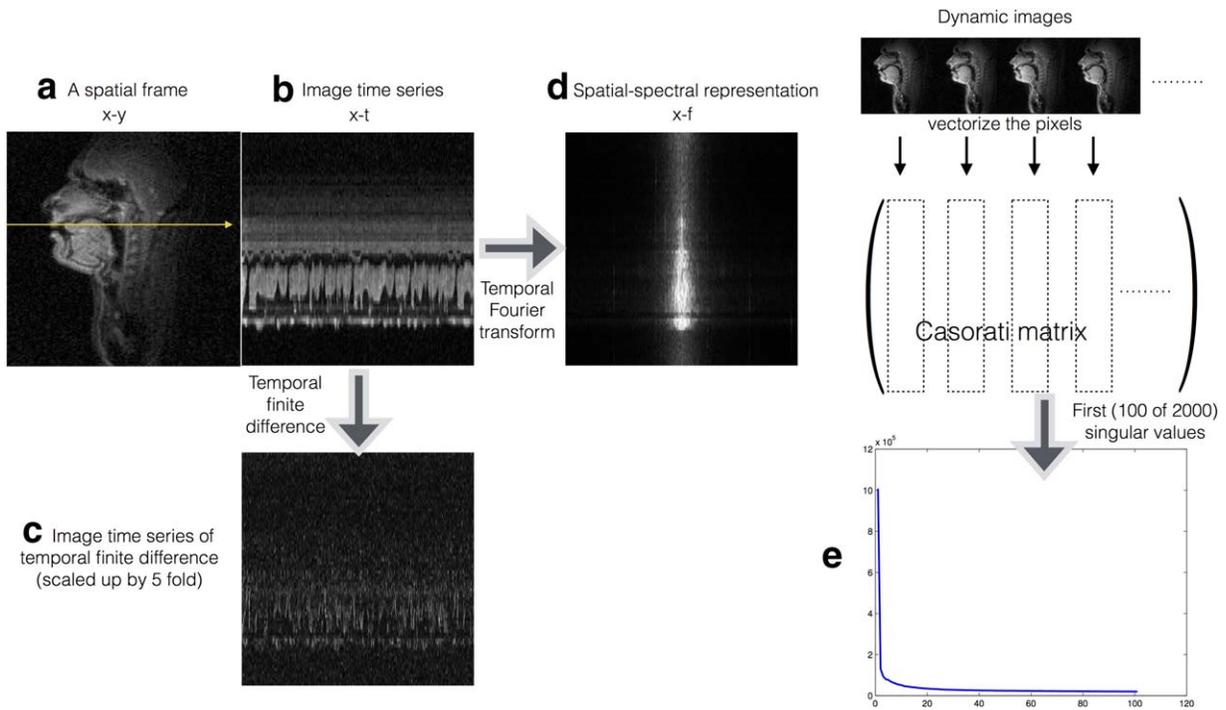


FIGURE 8: Redundancy in dynamic speech data: A reconstructed time series with spatial resolution of 2.4 mm^2 , and temporal resolution of 12 msec/frame , obtained from an ~ 6.5 -fold accelerated reconstruction is considered for demonstration. (a) A representative spatial frame of the time series, and (b) the time profile along the arrow specified in (a). (c) The temporal finite difference of the time series, which is obtained by taking the pixel-wise finite difference of the dynamic data along time. The sparse representation of the temporal finite difference transform was exploited to realize the accelerated reconstruction in (a,b). The dynamic data also have sparse representations in the spatial-spectral domain (d); this representation is obtained by evaluating the Fourier transform of the data along the temporal dimension. The dynamic data can be rearranged as a Casorati matrix (vectorized and represented column-wise). A singular value decomposition of this matrix reveals the redundancies amongst the pixel time profiles, ie, fast decay of the singular values as depicted in the (e), which is exploited in the low rank and partially separable models.

exploit redundancies in the dynamic MRI data. Figure 8 demonstrates an illustration of the redundancy of dynamic data in RT-speech MRI. Imaging schemes based on sparse sampling and regularized/constrained reconstruction have evolved that exploit these redundancies, and incorporate them appropriately into the reconstruction process.

Models such as UNFOLD/ k -t broad-use linear acquisition speed-up technique (BLAST) exploit the banded spatial-spectral support of the dynamic image time series, and can be used to accelerate the acquisition. These algorithms require modifications in the sampling patterns. They use close-to-fully-sampled low-resolution images to learn/estimate the spatial spectral support. In the second stage they use customized sampling patterns that minimize alias foldover artifacts onto the spatial-spectral support.^{70–72} The reconstruction in these schemes is linear and can be interpreted as a filtering problem.

Constrained reconstruction and compressed sensing schemes that exploit sparse representations of the dynamic image time series in an appropriate transform domain have gained considerable interest in recent years. Methods that exploit sparsity in the spatial-spectral domain or temporal finite difference transform domain have been popularly used in several cardiac MRI applications. These schemes have

advantages over support-based schemes in that they do not rely on fully sampled training data. They instead involve acquiring the samples in an incoherent manner resulting in a highly incoherent residual aliasing. The reconstruction is posed as a nonlinear optimization problem, which results in a time-consuming iterative algorithm. These algorithms are usually computed offline and have been used by several research groups.^{42,46,73,74} However, more recent advances in graphic processing units have potential to enable fast reconstruction pipelines.^{68,69,75,76} The choice of an appropriate constraint in speech imaging is an open area. Several research groups have relied on different types of constraints based on the speech task at hand. For instance, a median filter in combination with a temporal total variation constraint was used in⁴² to study production of vowel, consonant sounds, and co-articulations events.

Another approach to speech imaging is based on the partially separable (PS) model.³⁹ The PS model exploits the similarity of the pixel time profiles, and utilizes a set of fewer orthogonal temporal bases functions to represent the temporal variations present in the data to be reconstructed. This method leverages the imaging data matrix being low rank, enabling reconstruction of the full dynamic image series by estimating a set of temporal and spatial basis

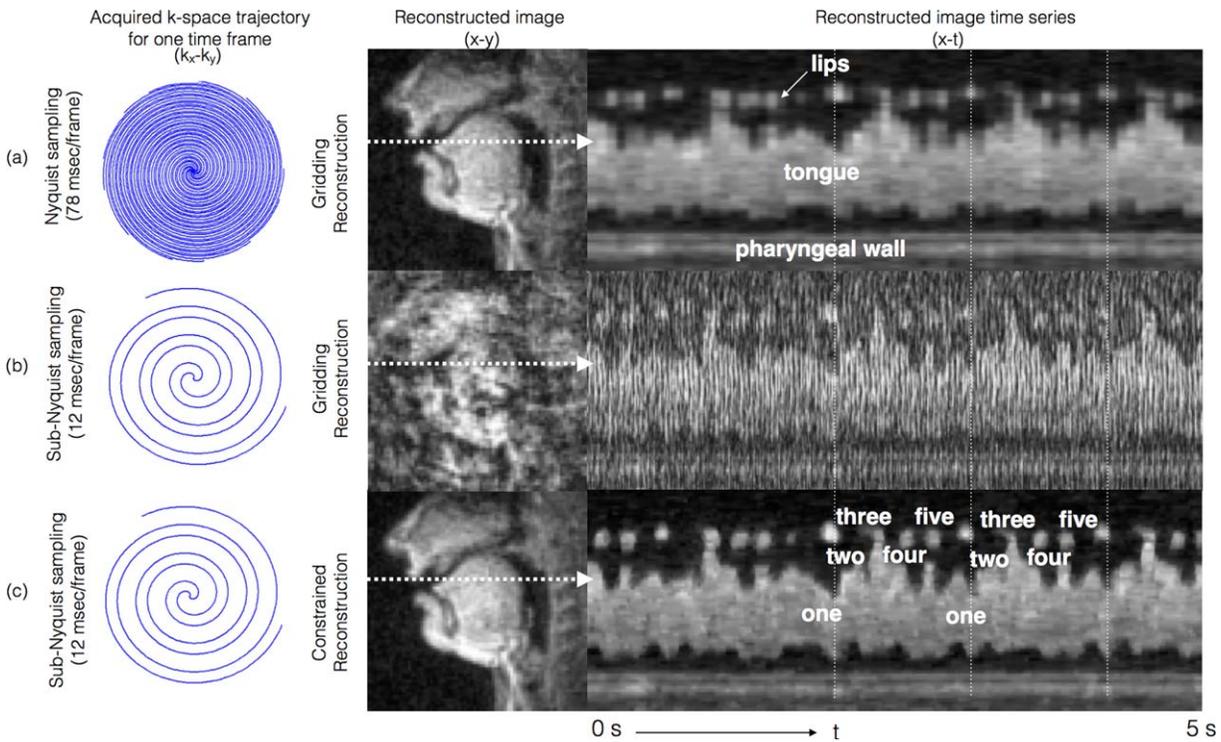


FIGURE 9: Subsampling in combination with efficient constrained reconstruction allows for improved time resolution in comparison to Nyquist sampling; a spiral-based acquisition as depicted in Protocol 2 is used. Images reconstructed using (a) Nyquist sampling and online gridding reconstruction result in time resolution of 78 msec/frame, while (b) sub-Nyquist sampling and online gridding reconstruction allows for significantly improved time resolution, at the expense of aliasing artifacts. Offline constrained reconstruction addresses this tradeoff by resolving the aliasing at the native time resolution of 12 msec/frame. Note the apparent advantage of the increase in time resolution of (c) vs. (a) in terms of crispness along the time axis. The task was to repeatedly count numbers “one-two-three-four-five” at a rapid pace (the subject spoke ~ 4 times faster than his normal speech pace).

functions. This is achieved in a two-step strategy, where the first step is to acquire data with low spatial encoding but with fine temporal resolution to estimate the temporal basis functions. Interleaved with the temporal sampling is a slower acquisition of high-quality, high-resolution image data, as part of the second step. These image data are acquired one line of k -space at a time for the 3D volume, but assumed to be acquired in time with the adjacent navigator data. Several full frames of imaging data are acquired and then fit to the PS model and temporal basis functions to estimate the high temporal and spatial resolution dynamic images. Often, the second step also exploits other properties of dynamic speech data, such as sparse spatial-spectral support, and sparse spatiotemporal image gradients.

Due to the separation of the temporal sampling and the image data sampling with the PS model, both acquisitions can be optimized for their required purpose. Previously,³⁹ a spiral acquisition was used for temporal sampling to sample a large extent of k -space efficiently. However, Cartesian data are used for the imaging data to minimize sensitivity to magnetic field inhomogeneity effects. A high-quality image results, as seen in Fig. 11. A potential disadvantage of the PS speech imaging approach is that for high-resolution or 3D datasets, the total acquisition time

can be on the order of 10 minutes to acquire sufficient imaging data to fit the PS model. However, a large number of images are obtained during this time at a rate of >100 frames per second for up to 8 slices, with the subject potentially performing multiple speech tasks.³⁹

Other algorithms have been proposed in the dynamic MRI literature, which reinterpret the PS model as a low-rank matrix recovery problem.^{77,78} These do not require training data, but directly estimate the dynamic images at hand by explicitly imposing the low rank, and sparsity constraints in the form of regularizers.

Due to the nonlinear recovery in several of the constrained reconstruction schemes, there exists a major challenge in quantifying the properties (such as interpreting the true SNR, true resolution, and spatiotemporal varying artifacts) of the reconstructions. While this is an ongoing research area, it is highly recommended to be aware of such effects introduced with the reconstructions while interpreting the dynamic images. There also exists an open challenge to automatically reconstruct datasets with minimal human intervention in choosing the reconstruction parameters (such as the regularization, model parameters). The reconstruction algorithm run time is beginning to be addressed by modern hardware support, and also new software-based optimization tools.

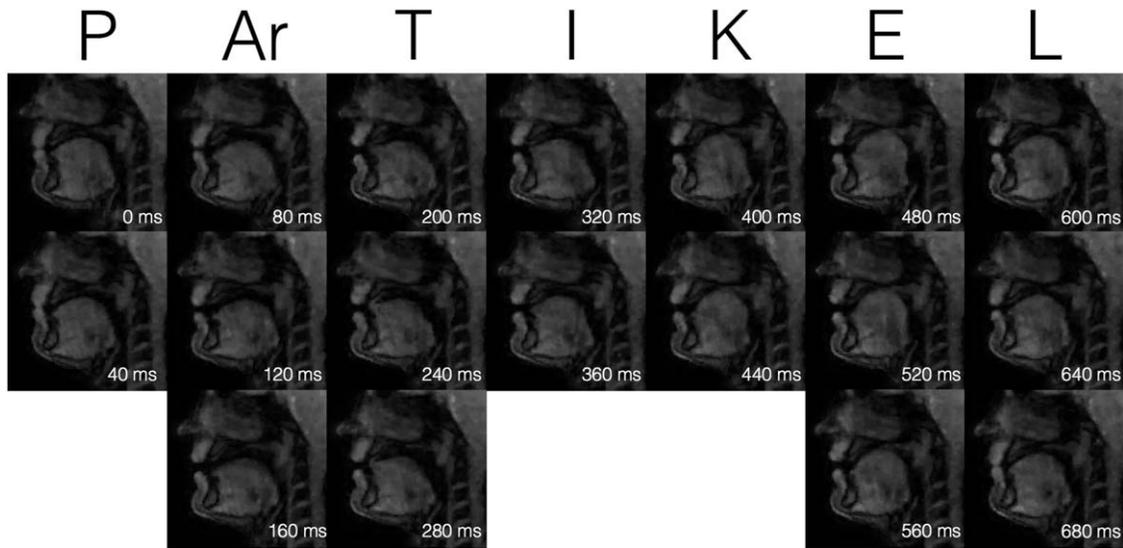


FIGURE 10: Example of acquisition with Protocol 3: The raw data were acquired with a radial gradient echo sequence with golden angle rotation. Images were reconstructed using a conjugate gradient-SENSE method that applies regularization using a spatial total variation (TV) operator. Twenty-five subsequent echoes were used to calculate low resolution sensitivity maps for each coil and a complete image with a spatial resolution of $1.8 \times 1.8 \times 10 \text{ mm}^3$, leading to a native temporal resolution of 55 msec that was further accelerated to 40 msec by applying a sliding window. (Figure courtesy: M. Burdumy, University Medical Center Freiburg, Germany.)

Open-Source Software

We recommend the following open-source reconstruction packages that are applicable for speech imaging: 1) Partially separable function model with spatial-spectral sparsity;^{39,79} 2) K-t SLR: Low rank and spatiotemporal total variation sparsity regularized reconstruction;^{77,80} 3) GRASP: golden angle radial acquisition with temporal total variation sparsity regularization;^{74,81} 4) K-t FOCUSS: spatial-spectral sparsity based regularization;^{73,82} 5) gadgetron: GPU based reconstruction tools;^{69,83} 6) IMPATIENT: GPU-based reconstruction tools.^{76,84}

Recommended RT-MRI Speech Protocols

We recommend four RT-MRI protocols for visualizing the upper airway during speech, summarized in Table 1. Protocol 1 is based on Cartesian trajectories with the reconstruction done on-the-fly. Protocol 2 is based on spiral

trajectories with the reconstruction done on-the-fly and/or offline. Protocol 3 is based on radial trajectories with the reconstruction done on-the-fly and/or offline. Protocol 4 is based on combination of Cartesian and non-Cartesian trajectories with the reconstruction done offline only.

These protocols are based on different choices of acquisition and reconstruction tools, and are classified based on their ease of implementation, and their abilities to perform on-the-fly reconstruction versus offline-only reconstruction. Significant differences in the properties of the resulting time series of the images result from choices of the different protocols and even between on-the-fly and offline reconstruction of the same protocol.

Protocol 1 is based on sequences that are likely available on all commercial scanners. It is based on Cartesian trajectories with the on-the-fly reconstruction options such as partial Fourier and/or parallel imaging. As discussed

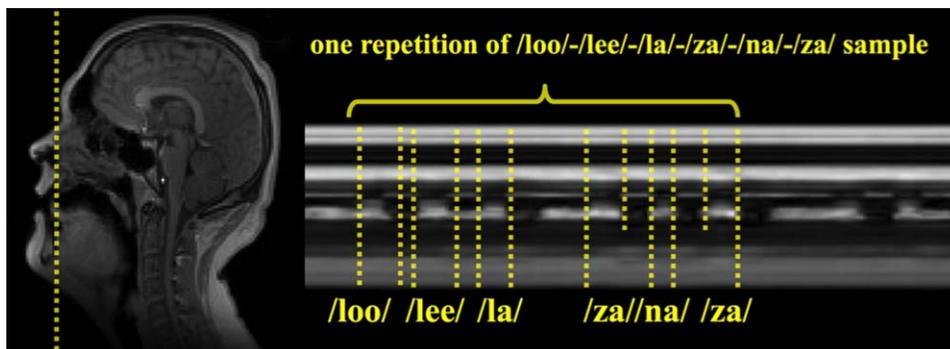


FIGURE 11: Strip plot through a single slice at the tip of the tongue (indicated by the dotted line) during an acquisition with Protocol 4, using spatial-spectral regularization and the PS model for reconstruction. The acquisition acquired eight simultaneous sagittal slices at 102 frames per second while the subject uttered a simple speech sample.

TABLE 1. Recommended RT-MRI Protocols for Speech Imaging

Recommended protocols	<u>Protocol 1</u>	<u>Protocol 2</u>	<u>Protocol 3</u>	<u>Protocol 4</u>
Accessibility	Widely available on most commercial scanners	Custom implementation	Custom implementation	Custom implementation
Key advantages	Compatible with linear reconstruction schemes, which enables on-the-fly reconstruction, a feature beneficial for frequently updating scan planes, and diagnosis, & correction for artifacts on the fly	Motion robust spiral readouts Compatible with iterative offline reconstruction schemes Enables high spatiotemporal resolutions (eg, up to 1.5–2.5mm ² ; 10–50 ms/frame)	Motion robust radial readouts	High temporal resolution
Speech tasks: On-line visualization (also see Fig 1 for spatiotemporal resolutions requirements)	Sustained sounds Velo-pharyngeal closure Bulk tongue movements		Sustained sounds Velopharyngeal closure Bulk tongue and few rapid tongue movements (eg, consonant to vowel) Parts of consonant constriction sounds	
Speech tasks: Off-line visualization (also see Fig 1 for spatiotemporal resolutions requirements)			Sustained sounds Velopharyngeal closure Most rapid tongue movements Consonant constriction sounds Coarticulation events	
Receiver coil		In order of preference (Also see Fig. 2): Custom upper airway coil Head and neck coil Neurovascular coil Head coil		
FOV		20 × 20 to 30 × 30 cm (dependent on the coil geometry)		
Orientation		Dependent on speech task; Mid-sagittal for most tasks		
Slice thickness		5–10 mm		
Field strength	1.5 T or 3T	1.5 T	1.5 T (preferred) 3 T	1.5 T 3 T (preferred)

TABLE 1: Continued

Recommended protocols	Protocol 1	Protocol 2	Protocol 3	Protocol 4
Sampling	Cartesian	Spiral	Radial	Specialized (Spiral + Cartesian)
Sequence	GRE or SSFP	GRE	GRE	GRE
Example sequence parameters	SSFP 3T FA: 15°, TE/TR: 1.1/2.3 msec, GRE 3T FA: 30°, TE/TR: 1.0/2.2 msec	TR = 6 msec, FA = 15°, spiral readout length = 2.4 msec	1.5 T; TE/TR: 1.44/2.2 ms, FA = 5°	At 3T: Spiral FLASH TE/TR = 0.85/9.8 ms Cartesian FLASH TE/TR = 2.3/9.8 ms
Specifications of the sampling pattern	Cartesian subsampling (2–3 fold)	Incoherent repetition of short spiral interleaves along time	Incoherent repetition of radial readouts along time	Temporal navigation by short spiral readouts. Randomized Cartesian sub-sampling for imaging data
On-the-fly reconstruction	✓	✓	✓	(x)
Parallel Imaging, Partial Fourier				
Gridding combined with sliding window				
Offline reconstruction	(x)	✓	✓	✓
Iterative constrained reconstruction				
Eg. spatial-spectral sparsity, spatiotemporal finite difference sparsity, low rank constraints				
Figures demonstrating quality and expected artifacts	Figs. 3–5	Figs. 5–9	Figs. (5 and 10)	Fig. 11

The various speech tasks that can be captured by the recommended four protocols are classified based on the latency in visualization (on-the-fly versus off-line). Protocol 1 can be enabled by Cartesian sequences, which are widely available on commercial scanners. Protocols 2, 3 are based on non-Cartesian sampling trajectories, and provide options for both on-the-fly and off-line reconstructions. Protocol 4 relies on special sampling schemes, and is based on offline reconstruction. On-the-fly reconstructions allow for real-time visualization of the speech events, and provide flexibility in prescription of planes, diagnosis of artifacts. In comparison to on-the-fly reconstructions, off-line iterative constrained reconstructions offer capabilities of imaging at high spatiotemporal resolutions. As discussed in the text, careful prescriptions of these protocols are required to avoid some common artifacts.

earlier, on-the-fly reconstructions allow for interactive imaging, a feature that is beneficial for frequently updating the scan planes, and for identifying and correcting artifacts.

Protocols 2 and 3 are respectively based on non-Cartesian spiral and radial trajectories. In comparison to Cartesian trajectories, these provide better resilience to motion artifacts, less detrimental undersampling artifacts, and offer improved spatial and temporal resolution capabilities. The reconstructions in Protocols 2 and 3 can be performed on-the-fly with linear reconstruction strategies such as sliding window in combination with fast gridding reconstruction. In addition, offline reconstructions could be performed on the same data acquired with Protocols 2 and 3 to enable higher spatial/temporal resolution, as depicted in Figures 9,10. These offline reconstructions involve solving iterative constrained reconstruction algorithms, several of which can be adapted by open-source software packages.

Protocol 4 is an offline technique based on combination of both Cartesian and non-Cartesian trajectories offering even higher spatiotemporal resolution capabilities. We contrast these protocols in Table 1, followed by detailed descriptions of the acquisition, reconstruction, and artifact reduction strategies associated with each of these protocols.

Discussion

Recent developments in RT-MRI provide unprecedented looks into the oropharyngeal dynamics of speech. Advances in hardware, acquisition, and reconstruction methods have enabled spatial resolutions capable of distinguishing contact of the velum and temporal resolutions capable of seeing tongue tip movements. The new RT-MRI speech imager is faced with many choices. We have outlined four strategies of varying complexity in acquisition and reconstruction that can achieve adequate spatiotemporal resolution to visualize soft-tissue dynamics with MRI.

In this article we have only provided qualitative demonstrations of the reconstructed image quality, and artifacts with all of the four recommended protocols. Performing quantitative comparisons among the protocols requires a controlled experimentation setting. For instance, evaluation of the different sampling trajectories for off-resonance artifacts, alias energies, and motion artifacts could be performed with retrospective undersampling of a ground truth phantom. In such experiments, it is desirable to evaluate quality using several choices of quantitative metrics that characterize different features in the reconstructed images. Metrics such as region-based root mean square error, structural similarity, point spread function specifications (main lobe width, side lobe energies) have been previously adapted.

Characterizations of parallel MRI linear reconstruction schemes are feasible using g-factor measures as described previously.⁵⁵ However, the quantitative assessments of

images obtained from nonlinear iterative constrained reconstruction schemes are nontrivial. Methods that derive local point spread functions for compressed sensing reconstruction characterization have been proposed, but are not fully established.⁸⁵ Few groups have proposed regional metrics that characterize SNR in upper airway regions of interest at several undersampling factors.^{41,60,86} A controlled task-specific based RT-MRI experimentation was proposed to evaluate the temporal finite difference constraint along with median filtering;⁸⁷ the spatiotemporal fidelity of the reconstructions were evaluated on a mechanical motion phantom that consisted of water filled tubes rotating at predefined angular velocities. However, such a rotating phantom may not be optimal for other regularizers such as those involving low rank constraints. Akin to advances in using numerical, or mechanistic phantoms in other MRI applications such as cardiovascular first-pass perfusion, cine MRI,^{88,89} realistic phantoms simulating fluent speech, repeated speech utterances with flexibility of varying the speech rates need to be developed for a thorough evaluation of several constrained reconstruction methods.

Apart from Protocol 1, all the other recommended protocols are based on methodologies that are currently nonroutine, such as those involving non-Cartesian sampling, and iterative constrained reconstruction. These methods are currently associated with shortcomings of lack of a systematic evaluation for reproducibility across sites. However, guidelines in this article could pave the way for future reproducibility studies, which could potentially refine the protocols into a set of routine methods in terms of best practices.

RT-MRI of speech production or swallowing is currently done in a supine position, which is not a natural posture for speech or for swallowing, which are almost exclusively performed in the upright position. Upright scanners can potentially overcome these problems, and there have been a few studies that demonstrate their utility in RT-MRI of speech research.⁹⁰⁻⁹² The use of iterative constrained reconstruction algorithms have been shown to improve the SNR from low field simulated measurements, and these hold great potential in future use of upright low field scanners for RT-MRI of the upper airway.⁹³

The clinical role of RT-MRI in the assessment of abnormal vocal tract and upper airway dynamics has not yet been determined, and remains an exciting area for investigation. VPI, swallowing-aspiration, and obstructive apnea are among the most compelling initial applications due to the prevalence of these conditions, and the lack of safe nonionizing alternatives. For each of these applications, RT-MRI has provided unique insights into the disease process, but has been hampered by cost, complexity, and availability of RT-MRI testing.

Acknowledgments

We thank the Ming Hsieh Institute at the University of Southern California for generous funding support towards the ISMRM endorsed Speech MRI Summit held in Los Angeles in February 2014. We also thank all attendees of the summit for their feedback.

References

1. Scott AD, Wylezinska M, Birch MJ, Miquel ME. Speech MRI: morphology and function. *Phys Med* 2014;30:604–618.
2. Bresch E, Kim Y-C, Nayak K, Byrd D, Narayanan S. Seeing speech: capturing vocal tract shaping using real-time magnetic resonance imaging. *IEEE Signal Proc Mag* 2008;25:123–132.
3. Demolin D, Hassid S, Metens T, Soquet A. Real-time MRI and articulatory coordination in speech. *Comptes Rendus Biol* 2002;325:547–556.
4. Honda K, Takemoto H, Kitamura T, Fujita S, Takano S. Exploring human speech production mechanisms by MRI. *IEICE Trans Inform Syst* 2004;87:1050–1058.
5. NessAiver MS, Stone M, Parthasarathy V, Kahana Y, Paritsky A. Recording high quality speech during tagged cine-MRI studies using a fiber optic microphone. *J Magn Reson Imaging* 2006;23:92–97.
6. 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.
7. 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;37:97–110.
8. Ramanarayanan V, Byrd D, Goldstein L, et al. Investigating articulatory setting—pauses, ready position, and rest—using real-time MRI. *Inter-speech* 2010;1994–1997.
9. Echternach M, Sundberg J, Arndt S, Markl M, Schumacher M, Richter B. Vocal tract in female registers — a dynamic real-time MRI study. *J Voice* 2010;24:133–139.
10. Ventura SR, Freitas DR, Tavares JMRS. Application of MRI and biomedical engineering in speech production study. *Comput Methods Biomech Biomed Eng* 2009;12:671–681.
11. Teixeira A, Martins P, Oliveira C, Ferreira C, Silva A, Shosted R. Real-time MRI for Portuguese. In: *Computational Processing of the Portuguese Language*. Berlin: Springer; 2012. p 306–317.
12. Iltis PW, Schoonderwaldt E, Zhang S, Frahm J, Altenmüller E. Real-time MRI comparisons of brass players: a methodological pilot study. *Hum Move Sci* 2015;42:132–145.
13. 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.
14. Maturo S, Silver A, Nimkin K, et al. MRI with synchronized audio to evaluate velopharyngeal insufficiency. *Cleft Palate Craniofac J* 2012;49:761–763.
15. 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.
16. Silver AL, Nimkin K, Ashland JE, et al. Cine magnetic resonance imaging with simultaneous audio to evaluate pediatric velopharyngeal insufficiency. *Arch Otolaryngol Head Neck Surg* 2011;137:258–263.
17. Sagar P, Nimkin K. Feasibility study to assess clinical applications of 3-T cine MRI coupled with synchronous audio recording during speech in evaluation of velopharyngeal insufficiency in children. *Pediatr Radiol* 2015;45:217–227.
18. Miquel ME, Wylezinska-Arridge M, Pinkstone M, Theobald C, Birch M, Scott A. Assessment of velopharyngeal closure and soft palate anatomy using MRI in cleft palate patients. *Med Phys Int* 2013;1:459.
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. Shprintzen RJ, Marrinan E. Velopharyngeal insufficiency: diagnosis and management. *Curr Opin Otolaryngol Head Neck Surg* 2009;17:302–307.
21. Raol N, Sagar P, Nimkin K, Hartnick CJ. New technology: use of cine MRI for velopharyngeal insufficiency. *Surg Pediatr Velopharyngeal Insufficiency* 2015;76:27–32.
22. Perry JL, Sutton BP, Kuehn DP, Gamage JK. Using MRI for assessing velopharyngeal structures and function. *Cleft Palate Craniofac J* 2014;51:476–485.
23. Kuehn DP, Ettema SL, Goldwasser MS, Barkmeier JC. Magnetic resonance imaging of the levator veli palatini muscle before and after primary palatoplasty. *Cleft Palate Craniofac J* 2004;41:584–592.
24. Tian W, Li Y, Yin H, et al. Magnetic resonance imaging assessment of velopharyngeal motion in Chinese children after primary palatal repair. *J Craniofac Surg* 2010;21:578–587.
25. Kazan-Tannus JF, Levine D, McKenzie C, et al. Real-time magnetic resonance imaging aids prenatal diagnosis of isolated cleft palate. *J Ultrasound Med* 2005;24:1533–1540.
26. Zhang S, Olthoff A, Frahm J. Real-time magnetic resonance imaging of normal swallowing. *J Magn Reson Imaging* 2012;35:1372–1379.
27. Olthoff A, Zhang S, Schweizer R, Frahm J. On the physiology of normal swallowing as revealed by magnetic resonance imaging in real time. *Gastroenterol Res Pract* 2014.
28. Sutton BP, Conway C, Bae Y, Brinegar C, Liang Z-P, Kuehn DP. Dynamic imaging of speech and swallowing with MRI. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE 2009*;6651–6654.
29. Carstens PO, Zhang S, Olthoff A, Bremen E, Lotz J, Frahm J, Schmidt J. Real-time MRI for evaluation of dysphagia in inclusion body myositis (IBM) (P2. 015). *Neurology* 2015;84.14(Suppl):P2-015.
30. Vijay Kumar KV, Shankar V, Santosham R. Assessment of swallowing and its disorders—a dynamic MRI study. *Eur J Radiol* 2012;82:215–219.
31. Zu Y, Narayanan S, Kim Y, et al. Evaluation of swallow function after tongue cancer treatment using real-time magnetic resonance imaging: a pilot study. *JAMA Otolaryngol Head Neck Surg* 2013;139:1312–1319.
32. Mády K, Sader R, Zimmermann A, et al. Assessment of consonant articulation in glossectomy speech by dynamic MRI. *Interspeech* 2002.
33. Hagedorn C, Lammert A, Bassily M, et al. Characterizing post-glossectomy speech using real-time MRI. *International Seminar on Speech Production*, Cologne, Germany 2014.
34. Hagedorn C, Proctor MI, Goldstein L, Gorno-Tempini ML, Narayanan S. Characterizing covert articulation in apraxic speech using real-time MRI. *Interspeech* 2012.
35. Nayak KS, Fleck RJ. Seeing sleep: dynamic imaging of upper airway collapse and collapsibility in children. *IEEE Pulse* 2014;5:40–44.
36. Schaaf WE, Wootten CT, Donnelly LF, Ying J, Shott SR. Findings on MR sleep studies as biomarkers to predict outcome of genioglossus advancement in the treatment of obstructive sleep apnea in children and young adults. *AJR Am J Roentgenol* 2010;194:1204–1209.
37. Schwab R, Pasirstein M, Pierson R, et al. Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging. *Am J Respir Crit Care Med* 2003;168:522–530.
38. Kim Y-C, Lebel RM, Wu Z, et al. Real-time 3D magnetic resonance imaging of the pharyngeal airway in sleep apnea. *Magn Reson Med* 2014;71:1501–1510.

39. Fu M, Zhao B, Carignan C, et al. High-resolution dynamic speech imaging with joint low-rank and sparsity constraints. *Magn Reson Med* 2015;73:1820–1832.
40. 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.
41. 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.
42. Niebergall A, Zhang S, Kunay E, et al. 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.
43. Zhang S, Block KT, Frahm J. Magnetic resonance imaging in real time: Advances using radial FLASH. *J Magn Reson Imaging* 2010;31:101–109.
44. Uecker M, Zhang S, Voit D, Karaus A, Merboldt KD, Frahm J. Real-time MRI at a resolution of 20 ms. *NMR Biomed* 2010;23:986–994.
45. Burdumy M, Traser L, Richter B, et al. Acceleration of MRI of the vocal tract provides additional insight into articulator modifications, *J Magn Reson Imaging* 2015.
46. Lingala SG, Zhu, Y, Kim Y-C, et al. High spatio-temporal resolution multi-slice real time MRI of speech using golden angle spiral imaging, constrained reconstruction, and a novel upper airway coil. In: *Proc ISMRM 23rd Scientific Sessions 2015*;689.
47. Feng X, Inouye J, Blemker S, et al. Assessment of velopharyngeal function with multi-planar high-resolution real-time spiral dynamic MRI. In: *Proc ISMRM 21st Scientific Sessions 2013*;1228.
48. Kuehn DP. A cineradiographic investigation of velar movement variables in two normals. *Cleft Palate J* 1976;13:88–103.
49. Miquel ME, Freitas AC, Wylezinska M. Evaluating velopharyngeal closure with real-time MRI. *Pediatr Radiol* 2014;45:941–942.
50. Beer AJ, Hellerhoff P, Zimmermann A, et al. Dynamic near-real-time magnetic resonance imaging for analyzing the velopharyngeal closure in comparison with videofluoroscopy. *J Magn Reson Imaging* 2004;20:791–797.
51. Bae Y, Kuehn DP, Conway C, Sutton B. Real-time magnetic resonance imaging of velopharyngeal activities with simultaneous speech recordings. *Cleft Palate Craniofac J* 2011;4:695–707.
52. Teixeira A, Martins P, Oliveira C, et al. Real-Time MRI for Portuguese. Database, methods and applications. *Computational Processing of the Portuguese Language Coimbra* 2012;306–317.
53. Martins P, Oliveira C, Silva S, Teixeira A. Velar movement in European Portuguese nasal vowels. In: *Proc IberSpeech 2012 VII Jornadas en Tecnología del Habla and Iberian SLTech Workshop 2012*;pp:231–240.
54. 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.
55. Kim Y-C, 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.
56. Schenck J. The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds. *Med Phys* 1996;23:815–850.
57. Hargreaves B. Rapid gradient echo imaging. *JMRI J Magn Reson Imaging* 2012;36:1300–1313.
58. Scott AD, Boubertakh R, Birch MJ, Miquel ME. Adaptive averaging applied to dynamic imaging of the soft palate. *Magn Reson Med* 2013;70:865–874.
59. Wylezinska M, Pinkstone M, Hay N, et al. Impact of orthodontic appliances on the quality of craniofacial anatomical magnetic resonance imaging and real-time speech imaging. *Eur J Orthodont* 2015;DOI: 10.1093/ejo/cju103.
60. Freitas AC, Wylezinska M, Birch M, Petersen SE, Miquel ME. Real time speech MRI: a comparison of Cartesian and Non-Cartesian sequences. In: *Proc ISMRM 23rd Scientific Sessions 2015*;655.
61. Opto-acoustics: MR-compatible fiber optic microphone. <http://www.optoacoustics.com/medical/fomri-iii/features>
62. Inouye JM, Blemker SS, Inouye DI. Towards undistorted and noise-free speech in an MRI scanner: correlation subtraction followed by spectral noise gating. *J Acoust Soc Am* 2014;135:1019–1022.
63. Counter SA, Olofsson A, Grahn H, Borg E. MRI acoustic noise: sound pressure and frequency analysis. *J Magn Reson Imaging* 1997;7:606–611.
64. Bresch E, Nielsen J, Nayak K, Narayanan S. Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans. *J Acoust Soc Am* 2006;120:1791–1794
65. Open source package of Audio denoising by correlation subtraction followed by spectral noise gating. <http://bme.virginia.edu/muscle/cleftpalate/noisecancellation.html>
66. Vaz C, Ramanarayanan V, Narayanan S, A two-step technique for MRI audio enhancement using dictionary learning and wavelet packet analysis. In: *Proc Inter-Speech, Lyon, France; 2013*. p 1312–1315.
67. Open source package of Audio denoising by data-driven denoising algorithm using dictionary learning and wavelet packet analysis. <http://sail.usc.edu/~cvaz/denoisingtoolbox.html>
68. Sorensen TS, Atkinson D, Schaeffter T, Hansen MS. Real-time reconstruction of sensitivity encoded radial magnetic resonance imaging using a graphics-processing unit. *IEEE Trans Med Imaging* 2009;28:1974–1985.
69. Hansen MS, Sørensen TS. Gadgetron: an open source framework for medical image reconstruction. *Magn Reson Med* 2013;69:1768–1776
70. Madore B, Glover GH, Pelc NJ. Unaliasing by Fourier-encoding the overlaps using the temporal dimension :applied to cardiac imaging and fMRI. *Magn Reson Med* 1999;42:813–828.
71. Hansen MS, Muthurangu V, Baltes C, et al. Real-time imaging of speech production using radial k-t SENSE. In: *Proc 14th Annual Meeting ISMRM, Seattle; 2006*:3187.
72. Sharif B, Derbyshire JA, Faranesh AZ, Bresler Y. Patient-adaptive reconstruction and acquisition in dynamic imaging with sensitivity encoding (PARADISE). *Magn Reson Med* 2010;64:501–513.
73. Jung H, Sung K, Nayak KS, Kim EY, Ye JC. k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI. *Magn Reson Med* 2009;61:103–116.
74. Feng L, Grimm R, Block KT, et al.. 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;72:707–717.
75. Gai J, Obeid N, Holtrop JL, et al. More IMPATIENT: a gridding-accelerated Toeplitz-based strategy for non-Cartesian high-resolution 3D MRI on GPUs. *J Parallel Distribut Comput* 2013;73:686–697.
76. Wu X-L, Gai J, Lam F, et al. Impatient MRI: Illinois massively parallel acceleration toolkit for image reconstruction with enhanced throughput in MRI. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro 2011*;69–72.
77. Lingala SG, Hu Y, Dibella E, Jacob M. Accelerated dynamic MRI exploiting sparsity and low-rank structure: K-t SLR. *IEEE Trans Med Imaging* 2011;30:1042–1054.
78. Haldar JP, Liang ZP. Spatiotemporal imaging with partially separable functions: a matrix recovery approach. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro 2010*;716–719.
79. Partially separable function model with spatial-spectral sparsity: Matlab code package: http://mri.beckman.uiuc.edu/for_download/PSSsparse_recon_tool_v0.1.zip
80. K-t SLR :Matlab code package: <https://research.engineering.uiowa.edu/cbig/content/matlab-codes-k-t-slr>.

81. GRASP Matlab code package: <http://cai2r.net/resources/software/grasp-matlab-code>.
82. K-t FOCUSS code package: <http://bispl.weebly.com/k-t-focuss.html>.
83. Gadgetron: <http://sourceforge.net/p/gadgetron/home>.
84. IMPATIENT: <http://impact.crhc.illinois.edu/mri.aspx>
85. Wech T, Staub D, Budich JC, et al. Resolution evaluation of MR images reconstructed by iterative thresholding algorithms for compressed sensing. *Med Phys* 2012.
86. Wylezinska M, Freitas A, Birch M, Miquel M. K-t BLAST/k-t FOCUSS in Real time imaging of the soft palate during speech. In: *Proc ISMRM 23rd Scientific Sessions 2015*:2302.
87. Frahm J, Schätz S, Untenberger M, et al. On the temporal fidelity of nonlinear inverse reconstructions for real-time MRI-the motion challenge. *Open Med Imaging J* 2014;8:1–7.
88. Chiribiri A, Schuster A, Ishida M, et al. Perfusion phantom: an efficient and reproducible method to simulate myocardial first-pass perfusion measurements with cardiovascular magnetic resonance. *Magn Reson Med* 2013;69:698–707.
89. Wissmann L, Santelli C, Segars WP, Kozerke S. MRXCAT: realistic numerical phantoms for cardiovascular magnetic resonance. *J Cardiovasc Magn Reson* 2014;16:63.
90. Santos J, Butts Pauly K, Popelka G, Pauly J. Real-Time MRI of swallowing in upright position. In: *Proc 16th Scientific Meeting ISMRM 2008*:2002.
91. Perry JL. Variations in velopharyngeal structures between upright and supine positions using upright magnetic resonance imaging. *Cleft Palate Craniofac J* 2011;48:123–133.
92. Traser L, Burdumy M, Richter B, Vicari M, Echtermach M. Weight-bearing MR imaging as an option in the study of gravitational effects on the vocal tract of untrained subjects in singing phonation *PLoS One* 2014;e112405.
93. Wu Z, Chen W, Nayak KS. Minimum field strength requirements for proton density weighted MRI. University of Southern California, Signal and Image Processing Institute, Technical report 2015;417.