



Fast upper airway magnetic resonance imaging for assessment of speech production and sleep apnea

Yoon-Chul Kim

Clinical Research Institute, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Received: July 18, 2018

Revised: September 10, 2018

Accepted: September 14, 2018

Corresponding author:

Yoon-Chul Kim

Clinical Research Institute,
Samsung Medical Center,
Sungkyunkwan University
School of Medicine, 81 Irwon-
ro, Gangnam-gu, Seoul 06351,
Korea

Tel: +82-2-2148-7140

E-mail: yoonckim@skku.edu

ABSTRACT

The human upper airway is involved in various functions, including speech, swallowing, and respiration. Magnetic resonance imaging (MRI) can visualize the motion of the upper airway and has been used in scientific studies to understand the dynamics of vocal tract shaping during speech and for assessment of upper airway abnormalities related to obstructive sleep apnea and swallowing disorders. Acceleration technologies in MRI are crucial in improving spatiotemporal resolution or spatial coverage. Recent trends in technical aspects of upper airway MRI are to develop state-of-the-art image acquisition methods for improved dynamic imaging of the upper airway and develop automatic image analysis methods for efficient and accurate quantification of upper airway parameters of interest. This review covers the fast upper airway magnetic resonance (MR) acquisition and reconstruction, MR experimental issues, image analysis techniques, and applications, mainly with respect to studies of speech production and sleep apnea.

Keywords: Magnetic resonance imaging; Real-time dynamic imaging; Sleep apnea syndromes; Speech; Upper airway

INTRODUCTION

The human upper airway is involved in various functions, including speech, swallowing, and respiration. Speaking involves airflow from the lung and trachea, vibration of the vocal fold, and rapid motion of articulators such as the tongue, soft palate, and lips. Chewing and swallowing activities involve motion of the tongue and soft palate as well as the food itself. The food is transported from the mouth to the esophagus. The epiglottis is moved back toward the esophagus to prevent the food bolus from entering the trachea. Respiration during sleep involves periodic motion of the pharyngeal wall, soft palate, and tongue to maintain the patency of the upper airway during the process of oxygen and carbon dioxide exchange. These functions are dynamic in nature, and the relevant organs are coordinated in a timely manner.

Medical imaging modalities, including X-ray imaging, computed tomography (CT), ultrasound, and magnetic resonance imaging (MRI), are non-invasive in nature and have been used to obtain images of the upper airway. The electromagnetic articulometer is used to track the

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>).

movement of transmitter coils attached to the tongue and lips for linguistic studies [1,2]. However, it is invasive and only provides limited information related to the locations of the coils. Fiberoptic endoscopy is used to visualize the airway lumen during drug-induced sleep in patients with obstructive sleep apnea (OSA) [3]. Optical coherence tomography (OCT) [4] has the potential to provide three-dimensional (3D) images of the upper airway with high spatial resolution [5,6]. Both endoscopy and OCT are invasive and do not provide anatomical information of the surrounding soft tissue.

Unlike X-ray imaging and CT, MRI involves no ionizing radiation, and results in no harmful radiation exposure to the subjects. It can non-invasively visualize the internal structure of the human upper airway and soft tissue, although it is expensive and loud and has a low acquisition speed. The organs involved in speech and swallowing tasks typically move faster than the organs involved in other activities, such as a beating heart or a knee or temporomandibular joint in motion.

One of the goals of upper airway MRI is to obtain high-quality upper airway dynamic image frames with high temporal fidelity. Clinical MRI protocols applied to the brain, heart, abdomen, and knee may not be directly applicable to the upper airway. Thus, development of magnetic resonance (MR) pulse sequences and image reconstruction should be tailored to the upper airway functions that researchers are interested in visualizing. From the perspective of fast dynamic imaging, vocal tract motion during speech production is relatively faster than pharyngeal airway motion during sleep, and generally speech MRI requires higher temporal resolution than sleep MRI. Table 1 lists several fast upper airway imaging protocols used for speech and sleep apnea studies in dynamic 2D, dynamic multi-slice, static 3D, and dynamic 3D imaging.

This review covers recent technologies used in upper airway MR pulse sequence, image reconstruction, and image analysis methods for human speech production and sleep apnea research. Notably, several previous review papers by Lingala et al. [7], Scott et al. [8], and Bresch et al. [9] addressed the technical aspects of upper airway MR acquisition for speech production. Ramanarayanan et al. [10] presented an in-depth review of image analysis techniques on real-time MRIs of vocal tract motion. Nayak and Fleck [11] introduced MRI techniques for assessment of OSA. Compared with the previous review papers, this article relatively puts more emphasis on MR experimental issues that may be of interest to researchers involved in MR experimental studies of the upper airway imaging. Compared with the previous review papers,

Table 1. Comparison of imaging protocols in selected publications

Type	Task	Reference	Acquisition/ reconstruction	No. of coil elements	Spatial resolution (mm ²)	Temporal resolution (ms)	Field of view (mm ³)
Dynamic 2D	Speech	Niebergall et al. (2013) [18]	Radial GRE/PI-tMF	12	1.5×1.5×10	33	192×192×10
	Speech	Lingala et al. (2017) [55]	Spiral GRE/through-time GRAPPA	8	2.4×2.4×6	18	200×200×6
	Sleep	Barrera (2011) [22]	Spiral GRE/gridding	1	2.6×2.6×5	182	200×200×5
	Sleep	Kim et al. (2013) [16]	Cartesian/POCENSE	6	1.6×1.6×5	303	160×160×5
Dynamic multi-slice	Speech	Kim et al. (2012) [38]	Spiral GRE/gridding	4	3.0×3.0×6	163	3 slice, 200×200×6/slice
	Speech	Feng et al. (2018) [39]	Spiral GRE/PI-CS	6	1.2×1.2×10	50	2 slice, 150×150×8/slice
	Sleep	Shin et al. (2013) [40]	Cartesian/FT	2	1.95×1.95×5	2,000	3 slice, 200×200×5/slice
	Airway compliance	Wu et al. (2016) [42]	Radial GA CAIPRINHA/PI-CS	6	1.0×1.0×7	96–128	4 slice, 200×200×7/slice
	Sustained speech	Burdumy et al. (2017) [47]	Stack-of-Stars/PI-CS	64	1.6×1.6×1.3	1,300 ^{a)}	200×200×62
Static 3D	Sustained speech	Kim et al. (2013) [64]	3DFT poisson disk sampling/PI-CS	8	1.25×1.25×1.25	8,000 ^{a)}	200×200×100
Dynamic 3D	Repeated speech	Fu et al. (2017) [51]	Cartesian+cone navigator/PS	12	2.2×2.2×5.0	5.99 ^{b)}	280×280×40
	Sleep	Kim et al. (2014) [49]	GA CAPR/PI-CS	6	1.6×1.6×1.6	602 ^{b)}	160×128×64

2D, two-dimensional; GRE, gradient recalled echo; PI-tMF, parallel imaging temporal median filtering; GRAPPA, generalized autocalibrating partial parallel acquisition; POCENSE, projection on to convex set sensitivity encoding; PI-CS, parallel imaging compressed sensing; FT, fourier transform; GA, golden angle; CAIPRINHA, controlled aliasing in parallel imaging results in higher acceleration; 3D, three-dimensional; DFT, discrete Fourier transform; PS, partial separability; CAPR, Cartesian acquisition with projection-reconstruction-like.

^{a)}Corresponds to scan time rather than temporal resolution; ^{b)}Reflects reconstruction frame rate rather than true temporal resolution.

this article presents a more comprehensive and up-to-date review of fast upper airway MRI, including imaging strategies for both speech production and sleep apnea research.

IMAGE ACQUISITION

Improvement of the acquisition speed in MRI has been an active area of research for more than two decades. Hardware improvements in the design of the gradient and radiofrequency (RF) coils greatly contributed to reducing the image acquisition time and increasing the signal-to-noise ratio (SNR). In addition, parallel imaging [12,13] and compressed sensing [14] revolutionized MRI speeds. MR protocols that use both parallel imaging and compressed sensing have been recently introduced in clinical MR examination protocols under various vendor-specific names, such as the Compressed SENSE by Philips (Best, the Netherlands), Compressed Sensing GRASP-VIBE by Siemens (Erlangen, Germany), and HyperSense by GE (Chicago, IL, USA).

Unlike the beating of the heart, the upper airway and tongue motion during natural speech or swallowing are not necessarily periodic. For example, cardiac electrocardiogram (ECG)-gated imaging, which is typically used in conventional cardiac MRI exams, assumes that the heart motion is periodic. The MRI raw data acquired over multiple heart beats are sorted into appropriate bins of the cardiac cycle based on the ECG

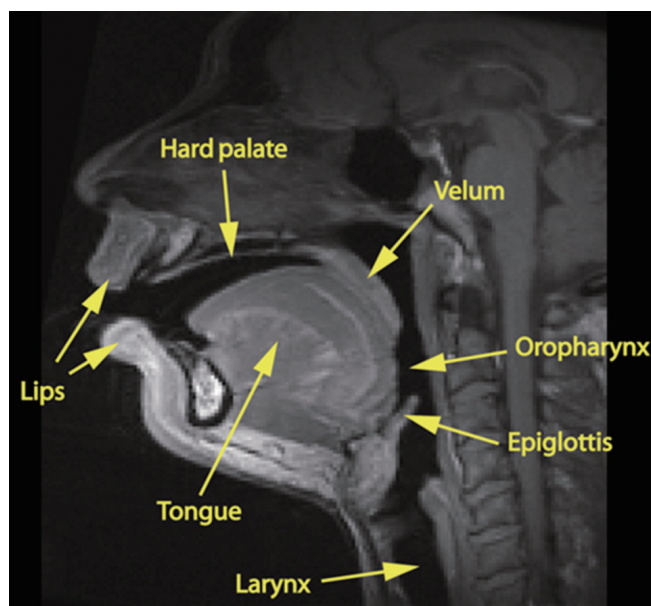


Fig. 1. Midsagittal upper airway magnetic resonance imaging. The midsagittal slice is typically imaged since it displays a relatively full picture of the organs involved in upper airway functions. The primary organ names involved in upper airway functions are shown.

gating information. However, the gated technique is not ideal for dynamic upper airway imaging of fluent speech, during which the motion of the articulators is not periodic in general. Real-time imaging [15] aims to freeze motion by simply reducing the acquisition window and is preferred over gated imaging for visualizing the motion of the articulators as is.

Dynamic 2D imaging

The midsagittal slice is typically acquired in speech imaging, since it shows a slice of the entire vocal tract from the lips to the glottis (Fig. 1). Dynamic real-time 2D imaging requires image frames at least within 80 ms for speech imaging. Hence, a rapid acquisition technique is necessary to meet the temporal resolution requirement. Gradient echo (also known as gradient recalled echo) sequence (GRE) is usually the choice for rapid imaging of the upper airway. Fig. 2 shows four representative k-space sampling trajectories in 2D imaging.

Cartesian sampling (also known as two-dimensional Fourier transform) is widely adopted in clinical imaging protocols since it is robust to system imperfections, but it may be slow for real-time speech imaging. Fig. 3 shows an example

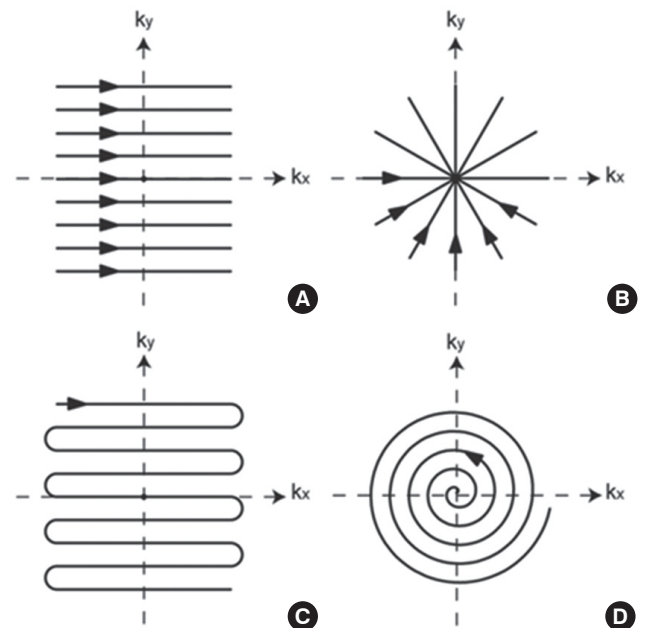


Fig. 2. Basic sampling trajectories. Cartesian sampling in (A) is widely used in clinical magnetic resonance imaging since it is robust to system imperfections, but it has a relatively low acquisition speed. Radial sampling in (B) is less sensitive to motion than Cartesian sampling since it traverses the k-space origin at every repetition time (TR). Echo-planar imaging (EPI) sampling in (C) and spiral sampling in (D) are fast since they can cover the k-space in one TR. However, both EPI and spiral sampling may suffer from artifacts due to system imperfections (e.g., off-resonance, gradient delay, and eddy currents).

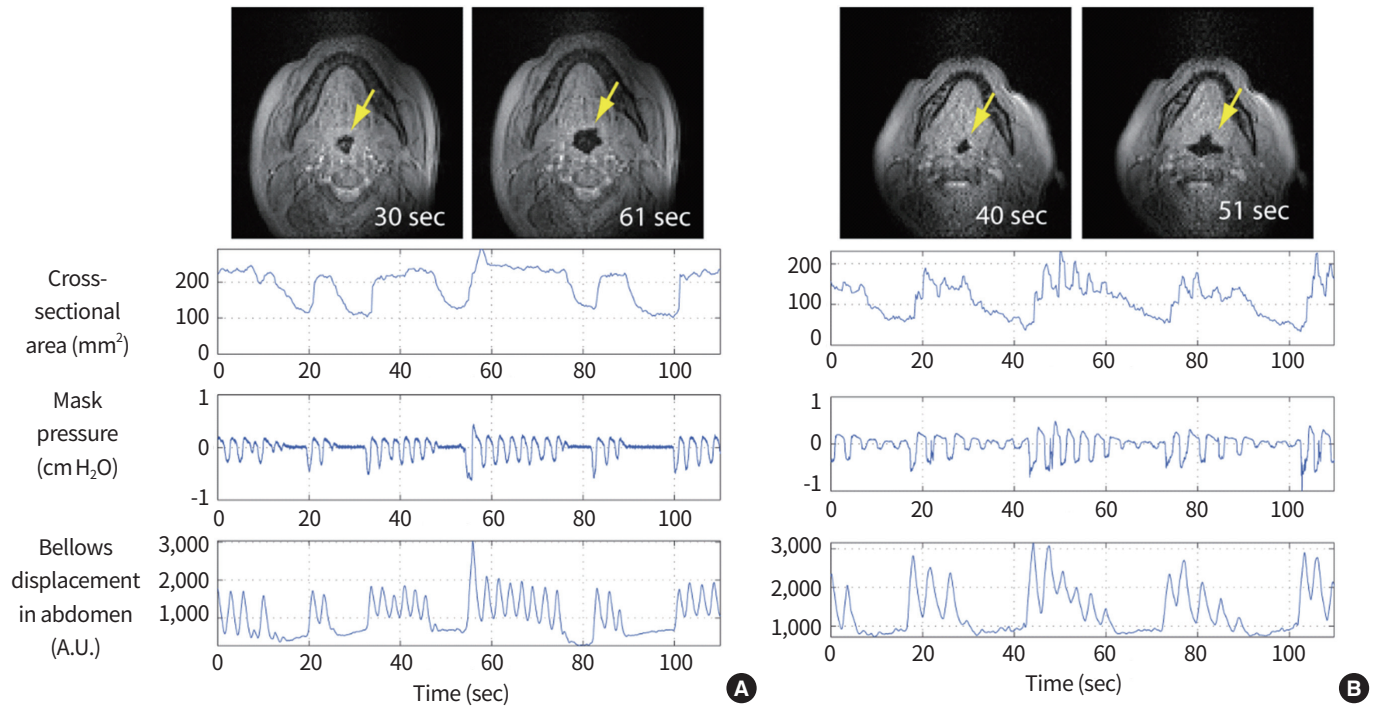


Fig. 3. Selected retroglottal frames and plots of the cross-sectional area, pressure, and respiratory effort in two subjects using real-time two-dimensional magnetic resonance imaging during sleep-disordered breathing. These illustrate (A) central sleep apneas with periodic breathing (14-year-old male) and (B) hypopneas with periodic breathing (17-year-old female). All events and airway dynamics were obtained during spontaneous sleep. Adapted from Kim et al., with permission from International Society for Magnetic Resonance in Medicine [16].

of airway narrowing in the retroglottal slice during sleep-disordered breathing, where a dynamic 2D Cartesian imaging sequence was used for data acquisition [16]. Cartesian imaging with partial k-space undersampling was combined with projection on to convex set sensitivity encoding (POCSENSE) reconstruction [17] in order to recover an un-aliased image and thus speed up real-time dynamic imaging.

Radial sampling is less sensitive to motion than Cartesian sampling and exhibits incoherent spatial aliasing artifacts in undersampling, thus being well-suited to the sparse reconstruction known as compressed sensing. Niebergall et al. [18] demonstrated a radial fast low-angle shot (FLASH) sequence with ultrashort repetition time (i.e., TR) of 2.22 ms to achieve a 33.3 ms acquisition time in speech imaging. More recently, Iltis et al. [19] demonstrated a radial FLASH sequence to achieve a 10.0-ms temporal resolution in capturing rapid tongue motion.

Echo-planar sampling (EPI) is a time-efficient approach that involves acquiring the entire k-space in a single shot or in a few shots, but it may show ghosting and distortion in the presence of motion and off-resonance. EPI is seldom used for speech imaging, mainly because off-resonance from air-tissue boundaries results in a large degree of distortion in images.

Spiral imaging is another time-efficient approach that in-

volves covering the k-space and has been demonstrated in real-time cardiac imaging [20,21], speech imaging [22], and OSA studies [23]. A drawback of spiral imaging is image quality degradation, including spatial blurring due to off-resonance [24]. Given the same readout duration, spatial blurring is more severe at higher magnetic field strengths. In real-time speech imaging, center-frequency adjustment and pre-scan shimming are required to reduce spatial blurring. Off-resonance correction can be performed to de-blur the images, and, in this case, a field map needs to be estimated at each frame either with a pulse sequence with two different echo times [25] or with a focus metric [26,27]. In practice, interleaved spiral trajectories are used to reduce spatial blurring with reduced readout duration. Interleaved spiral trajectories in combination with view-sharing reconstruction increase the frame rate in real-time speech imaging (Fig. 4).

Golden angle sampling [28] emerged as a promising acquisition technique for dynamic imaging [29]. Golden angle sampling, in which the angle of a radial spoke is increased by the golden angle (111.246°) at every TR, bisects the largest azimuthal angle gap at every TR. It is known to provide flexible retrospective selection of temporal resolution and time offset from continuously acquired real-time MR data. It was

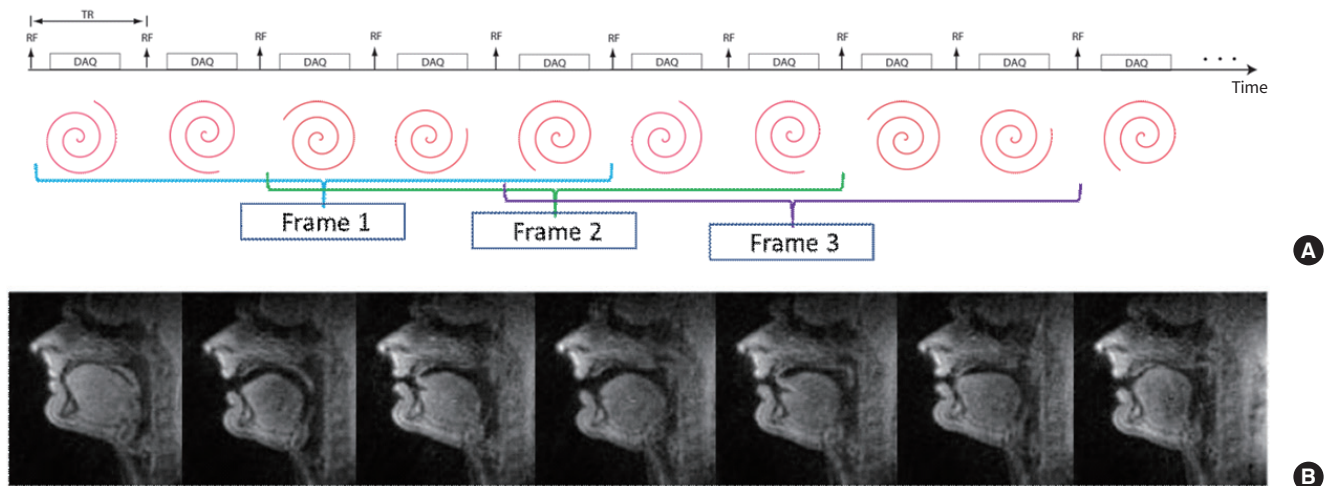


Fig. 4. Schematic of spiral real-time imaging of the vocal tract. (A) View-sharing for dynamic image frame reconstruction. For example, the third and fourth spiral interleaves are shared between frames 1 and 2. View sharing is helpful in increasing the frame rate. (B) An example of the reconstructed image frames. High-speed spiral real-time imaging can capture a variety of vocal tract configurations (e.g., tongue tip constriction, tongue dorsum constriction, velar opening/closing, neutral position) during fluent speech. TR, repetition time; RF, radiofrequency; DAQ, data acquisition.

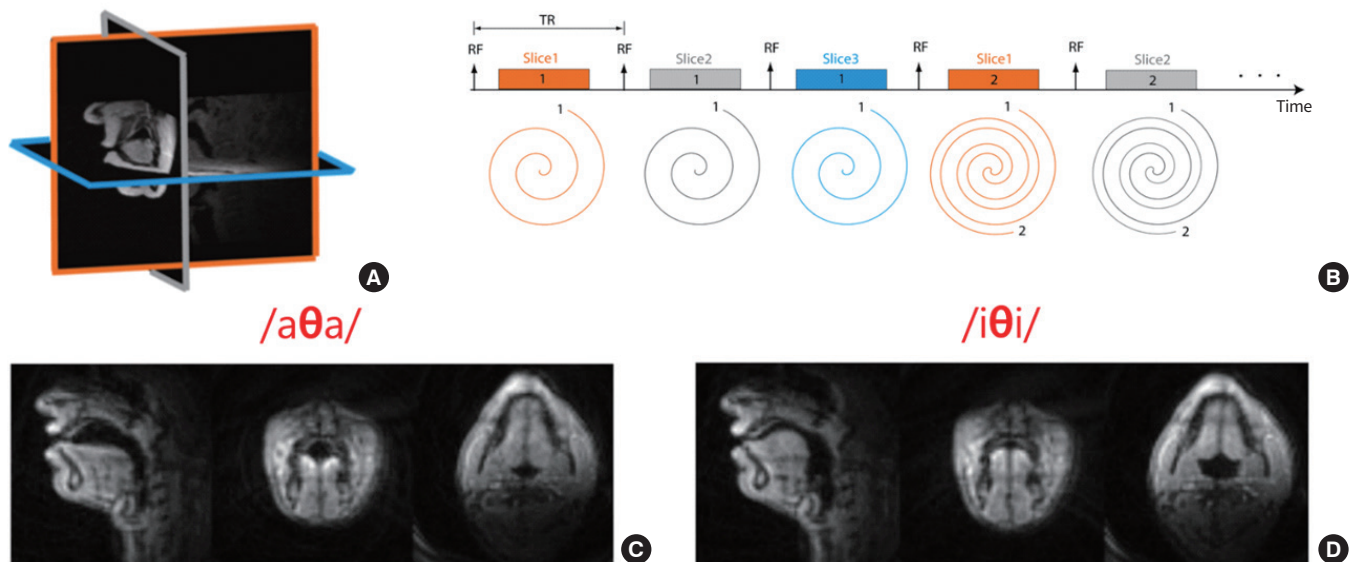


Fig. 5. Multi-planar real-time imaging of the vocal tract. (A) Slice prescription. (B) Slice acquisition sequence along with spiral interleaves. (C, D) Three plane images simultaneously acquired while the subject pronounced /θ/ in the vowel context of /a_θa/ and /i_θi/, respectively. Vocal tract configuration in the three slices of interest can be simultaneously inspected using the multiplanar real-time imaging technique. TR, repetition time; RF, radiofrequency.

demonstrated with radial and spiral trajectories in speech imaging [30–32].

The SNR may be insufficient for high-resolution real-time imaging of the soft palate, which is located far from the surface coil and thus the coil sensitivity is relatively low. Adaptive averaging was demonstrated to improve the SNR from data acquired during repetition of a speech utterance [33].

Tagged cine MRI sequences were used to visualize internal

deformation in the tongue [34,35]. MR tagging is performed using a spatial modulation of magnetization (SPAMM) imaging protocol [36,37]. Many repetitions of the utterance are required to track internal tongue motion on one slice [34]. For every repetition, the subjects should maintain the same articulatory postures and speech rate to avoid image mis-registration. Thus, tagged cine acquisition requires pre-training of the subjects to reduce variability in speech rate and articulation.

Dynamic multi-slice 2D imaging

For speech imaging, the midsagittal slice covers the entire vocal tract from the lips to the glottis, but it does not provide any information about articulation in the parasagittal re-

gions, such as the grooving/doming of the tongue, asymmetries in tongue shape, and lateral shaping of the pharyngeal airway. A multi-slice real-time imaging technique was developed to capture the vocal tract shaping in three orthogonal

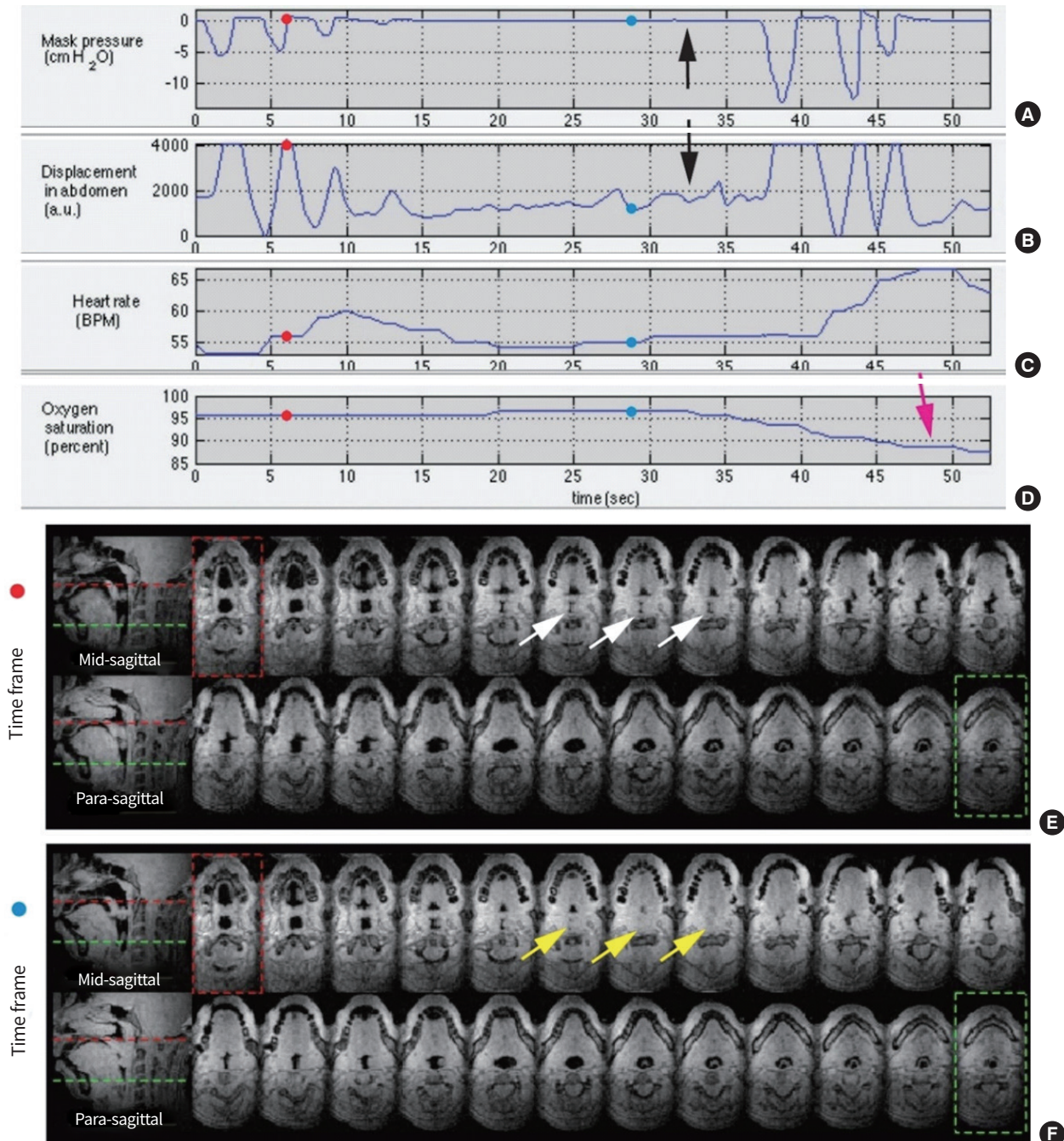


Fig. 6. An example of real-time dynamic three-dimensional imaging during natural sleep. (A-D) Physiological waveforms simultaneously measured during magnetic resonance imaging scans. (E, F) Axial and sagittal slices shown at two frames (E) prior to and (F) during obstructive apnea. During the 25-second obstructive apnea period, as indicated by zero mask pressure and respiratory effort (see black arrows in A, B), total airway collapse is observed in the retropalatal slices (see the yellow arrows in F, in comparison with the narrow airway indicated by the white arrows in E). A reduction in oxygen saturation is observed after the obstructive apnea (see the magenta arrow in D). Adapted from Nayak et al., with permission from IEEE [11].

planes (i.e., midsagittal, axial, coronal) during fluent speech [38]. Fig. 5 illustrates the acquisition and reconstruction of three slice images during the utterance of /θ/ in the vowel contexts of /a_a/ and /i_i/. Notably, multi-slice imaging sacrifices temporal resolution by the number of scan planes. A recent study demonstrated a dual-planar real-time imaging for assessment of velopharyngeal function [39].

For sleep apnea imaging, simultaneous imaging of the midsagittal slice and multiple axial slices would be a preferred approach over a single slice imaging, because additional anatomical information is available from multi-slice imaging. Shin et al. [40] demonstrated imaging of the pharyngeal airway in one midsagittal and two axial planes during natural sleep and noted that the subject's head motion during sleep could have the prescribed midsagittal scan plane not centered in the upper airway. Hence, the midsagittal view, if it is not corrected during imaging, may lead to mis-interpretation of a patent airway as a collapsed airway.

Simultaneous multi-slice (SMS) imaging [41] uses multi-band RF pulses to excite multiple parallel slices simultaneously and resolves the slices with parallel imaging. The benefit of SMS imaging is the reduced geometric factor, which facilitates higher acceleration in multi-slice imaging. SMS real-time MRI was demonstrated in four parallel axial slices using a radial controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) sequence with golden angle view order for upper airway compliance measurement [42].

Static 3D imaging

High-resolution 3D MRI of vocal tract shaping provides insight into the modeling of the vocal tract in association with speech sounds [43-45]. The production of speech sounds is performed within a subject's breath-hold. It is desirable to acquire an entire vocal tract shape in 3D within sustained speech over a period typically ranging from 6 to 8 seconds. This requires highly accelerated imaging to achieve high spatial resolution with

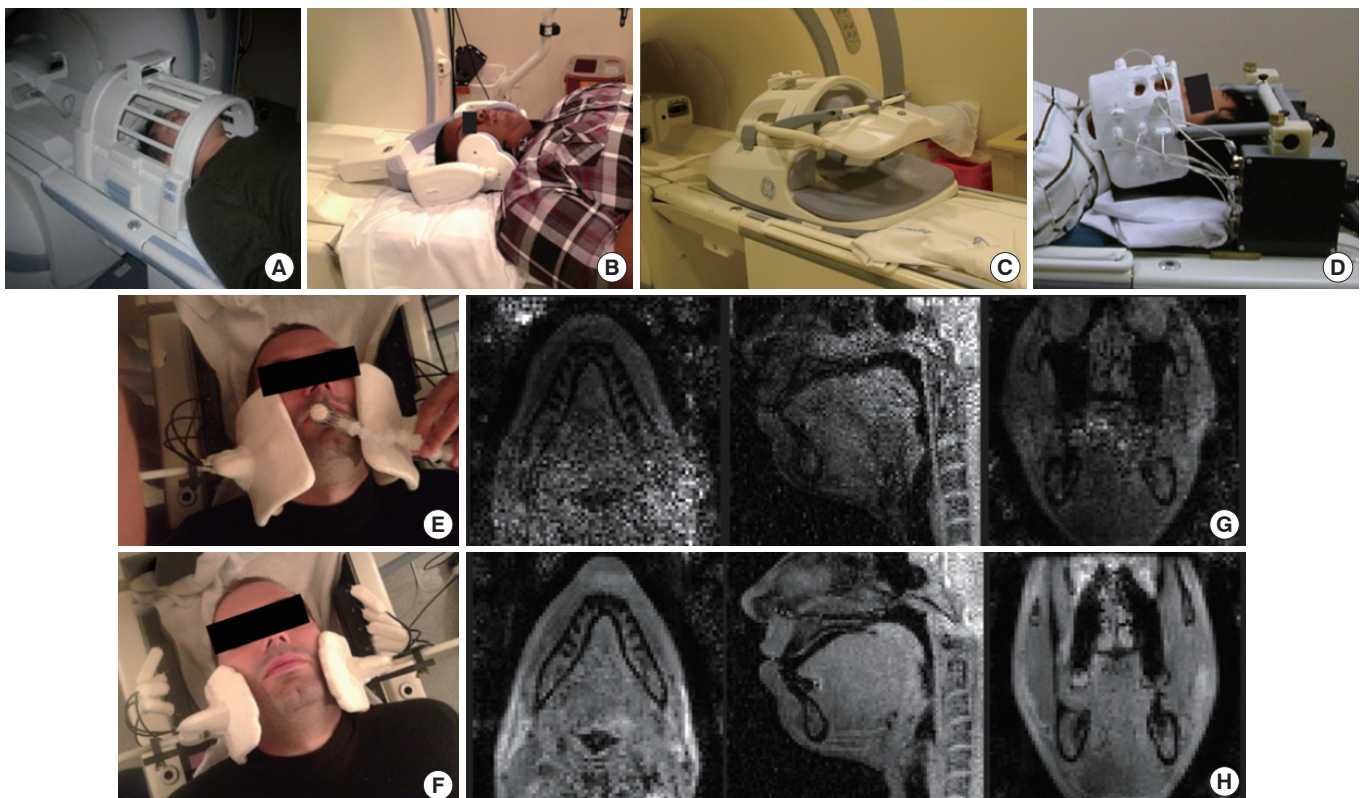


Fig. 7. Radiofrequency coils used for imaging of the upper airway. (A) Commercial single-channel birdcage coil. (B) Commercial 6-channel carotid coil. (C) Commercial 8-channel neurovascular coil. (D) Custom 16-channel coil. (E) Custom adult-sized 8-channel coil. (F) Custom child-sized 8-channel coil. Note that unlike (A, C, D), the coils in (B, E, F) have space for other devices (e.g., a mask attached to the mouth and nose for sleep apnea studies and an optical microphone close to the mouth for speech studies). When a three-dimensional Cartesian gradient echo sequence is used with parallel imaging acceleration rate 6, reconstructed axial/sagittal/coronal images are compared for (G) the 8-channel neurovascular coil and (H) the 16-channel coil. Noise amplification is noticeably higher in (G) than in (H).

complete coverage of the vocal tract. Kim et al. [46] demonstrated the first application of compressed sensing to 3D imaging of the upper airway for speech and achieved a resolution of $1.5 \times 1.5 \times 2.0 \text{ mm}^3$ in 7 seconds of sustained speech production using a single-channel head coil. Recently, Burdumy et al. [47] demonstrated improved full 3D imaging of the vocal tract using a stack-of-stars sequence and compressed sensing to reduce the scan time to 1.3 seconds.

Dynamic 3D imaging

Respiratory gating was used to acquire the 3D dynamics of the upper airway during tidal breathing while awake in patients with OSA [48]. The scan time was proportional to the number of respiratory phases, spatial resolution, and SNR. For successful data acquisition using this technique, subjects' breathing patterns need to be steady without severe movements near the upper airway.

Real-time dynamic 3D acquisition of the upper airway during spontaneous sleep was demonstrated in obese adolescents [49]. The technique showed the potential to provide information of the upper airway collapse site in obstructive apnea events during the subject's natural sleep in MRI. Fig. 6 shows the 2D sagittal and axial slices of the 3D upper airway at a time frame prior to the OSA event and at a time frame during the OSA event, demonstrating the benefit of real-time 3D imaging in evaluating airway obstruction patterns.

In speech imaging, dynamic 3D visualization of the vocal tract with high temporal resolution was demonstrated [50]. The technique covers the entire vocal tract with 2D real-time MRI while the subject repeats the same utterance for every sagittal slice. The gated technique is capable of yielding dynamic 3D visualization of the vocal tract, but it requires many repetitions of the utterance and substantial post-processing effort for alignment and segmentation. Another method demonstrated the use of low-rank modeling and sparse sampling to substantially accelerate the imaging speed [51].

RF coils

Commercial coils such as the birdcage head coil and the multi-channel head-neck array coil were designed to produce optimal SNR in the brain or neck regions of interest. Custom RF receiver coils were designed and demonstrated to increase image SNR in the tongue, lips, soft palate, and pharyngeal wall [31,52] and to enhance the parallel imaging performance for high acceleration factors (Fig. 7).

IMAGE RECONSTRUCTION

Highly accelerated imaging is realized by image reconstruction from undersampled k-space data. Conventional reconstruction of undersampled k-space data results in spatial aliasing artifacts in images. Parallel imaging reconstruction exploits the spatial information available from multiple channel coils to recover images without spatial aliasing [12]. Compressed sensing reconstruction exploits transform sparsity and incoherent aliasing from a pseudo-random undersampling scheme to recover images [14]. This typically involves minimization of the sum of the data consistency L2 norm and the sparsity-promoting L1 norm weighted by a regularization parameter.

Low-latency reconstruction

The image reconstruction period is a particularly important factor for real-time interactive imaging, where an ideal time interval between image acquisition and display is within 100 ms for fast interactions and scan parameter changes by the MRI operator. Spiral imaging is inherently fast and was an option for real-time interactive imaging [20,53]. Gridding reconstruction is performed in radial or spiral imaging to map non-Cartesian data to a Cartesian grid followed by fast Fourier transform (FFT) to reconstruct an image [54]. The gridding reconstruction process is sufficiently fast to guarantee acceptable latency in real-time interactive imaging, but it does not use any image acceleration framework such as parallel imaging.

Recently, Lingala et al. [55] demonstrated the feasibility of through-time spiral generalized autocalibrating partial parallel acquisition (GRAPPA) [21] for low-latency reconstruction in real-time speech MRI. The through-time spiral GRAPPA achieved four-fold acceleration and reconstructed image frames at a rate of 18 ms/frame with eight processors [55].

Iterative reconstruction

Constrained reconstruction (also known as compressed sensing parallel imaging) is used to obtain accurate image estimates iteratively from undersampled k-space data. Since it updates the image estimate at every iteration, the reconstruction time is proportional to the iteration number. The iterative reconstruction is based on the minimization of a cost function that is typically the sum of data consistency L2 norm and the sparsity-promoting L1 norm weighted by a regularization parameter. It is well known that iterative reconstruction is often performed off-line and takes more time than

conventional FFT-based reconstructions. Parallelization over multiple graphic processing units is an advanced method for acceleration of the reconstruction period [56,57].

Deep learning-based reconstruction

Recent trends in MRI reconstruction investigate the effectiveness of machine learning in optimizing parameters related to image reconstruction algorithms [58,59]. The parameter learning involves learning of the regularization parameter and filter coefficients concerned with image unaliasing. Once the parameters are trained by learning algorithms, they are theoretically not necessary to tune after data acquisition. Another advantage of the learning-based reconstruction, compared to compressed sensing iterative reconstruction, is efficient reconstruction time [58]. It is highly expected that the learning-based reconstruction framework will be applied to upper airway MRI.

IMAGE ANALYSIS

Vocal tract analysis

Dynamic real-time speech MRI typically acquires thousands of image frames over 20 seconds of fluent speech. Articulatory information is obtained from a single midsagittal slice image; thus, manual segmentation and analysis of all image frames is time-consuming and laborious. Hence, custom automatic or semi-automatic analysis techniques were developed by speech scientists who were interested in articulatory analysis of MRI data.

A variety of analysis methods have been developed by several research groups. A statistical shape model was built from training images and was used to automatically segment unseen vocal tract images [60]. Automatic segmentation and labeling of the vocal tract articulators in midsagittal dynamic images was demonstrated in another study [61]. Another report demonstrated an approach based on a graphical user interface for vocal tract segmentation in midsagittal dynamic

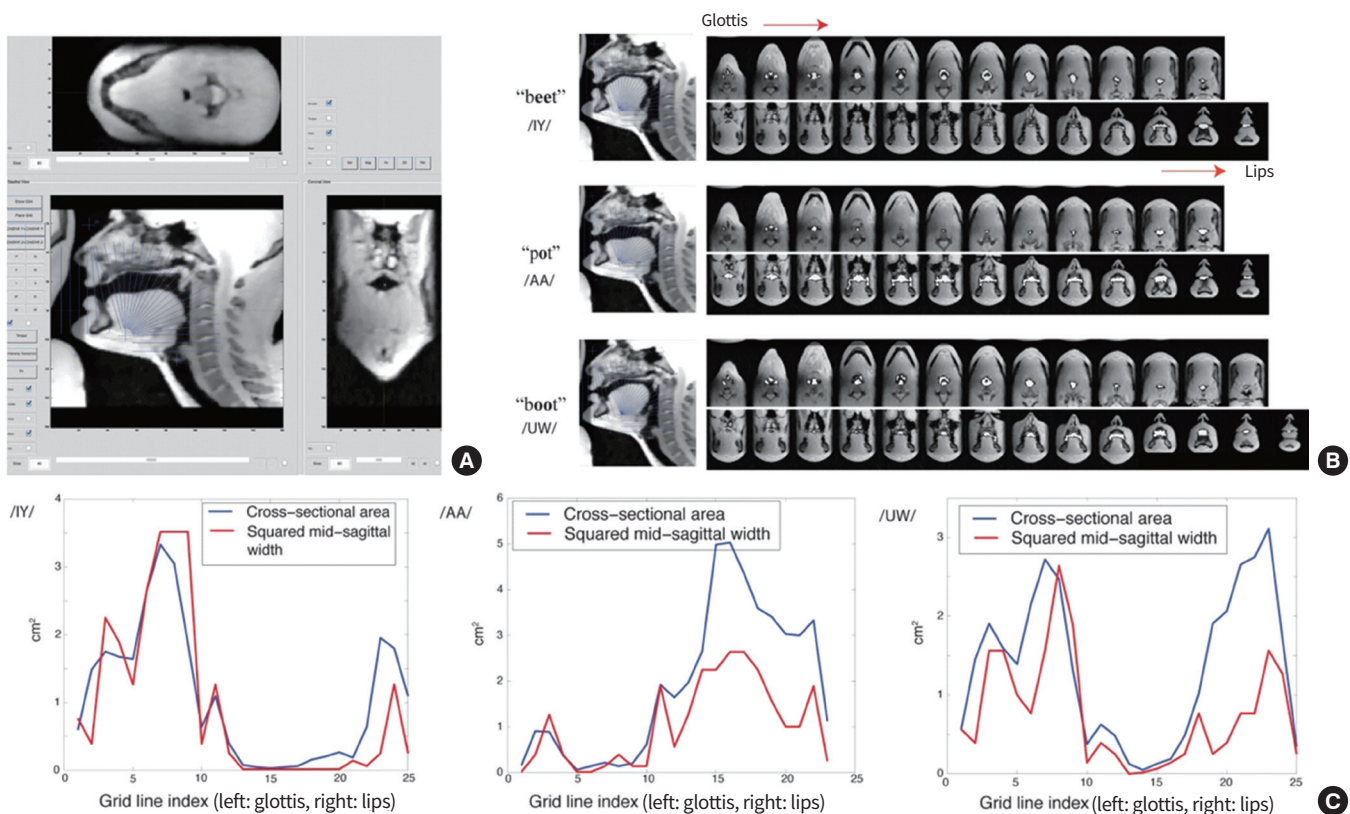


Fig. 8. An example of semi-automatic vocal tract area function estimation from three-dimensional (3D) vocal tract data. (A) Axial, midsagittal, and coronal views of the 3D vocal tract during sustained speech. Slice prescriptions (indicated by blue lines) from the glottis to the lips are performed semi-automatically in the midsagittal slice. (B) Segmentations of cross-sectional airways are automatically performed in the prescribed slices from the glottis to the lips. (C) Cross-sectional airway area and squared midsagittal width are plotted as a function of the distance from the glottis.

images [62]. Machine learning has also been used to automatically segment individual articulators in real-time mid-sagittal images [63]. Semi-automatic estimation of vocal tract area function was demonstrated using a graphical user interface in accelerated 3D MRI data of sustained speech [64]. It involves the user's annotation of anatomical landmarks, centerline extraction, cross-sectional slicing of the airway, and automatic segmentation of the cross-sectional airway. Fig. 8 illustrates the semi-automatic procedures for estimating vocal tract area functions.

Internal tissue deformation analysis

Internal tissue motion tracking is often performed on images acquired with preparation pulses of SPAMM. Tagged images in the tongue can be analyzed using harmonic phase (HARP) MRI [34]. Measurements involve displacement and velocity of tissue points and strain of specific muscles. The trajectories of tissue points can be visualized as path lines.

Airway narrowing/collapse analysis

A custom graphical user interface was developed to synchronously visualize real-time 3D MRI movies and measured physiological signals (Fig. 9). This allowed the user to rapidly inspect sleep apnea events and associated upper airway im-

ages from approximately 20- to 30-minute data. Frame-by-frame semi-automatic quantification of pharyngeal airway volume was demonstrated by using 3D region growing segmentation in real-time 3D MRI data [65]. The technique also enabled automatic detection of an airway collapse event from 4D airway data.

Computational fluid dynamics (CFD) simulations were performed to predict flow pressure measurements in human airways from 3D anatomical MRI or CT data [66-68]. The study demonstrated the potential of CFD modeling to elucidate the mechanism of OSA. Finite element analysis modeling was developed to predict the airway closing pressure and airway collapse site under different surgical treatment options [69].

Deep learning

As deep learning [70], more specifically deep convolutional neural network (CNN), shows incredible performance compared with conventional machine learning methods in computer vision [71], its applicability has been investigated in other domains, including medical image analysis. Since the development platforms (e.g., Keras, Tensorflow, PyTorch, Caffe) provide open source software for deep learning, researchers easily access and choose deep learning libraries on

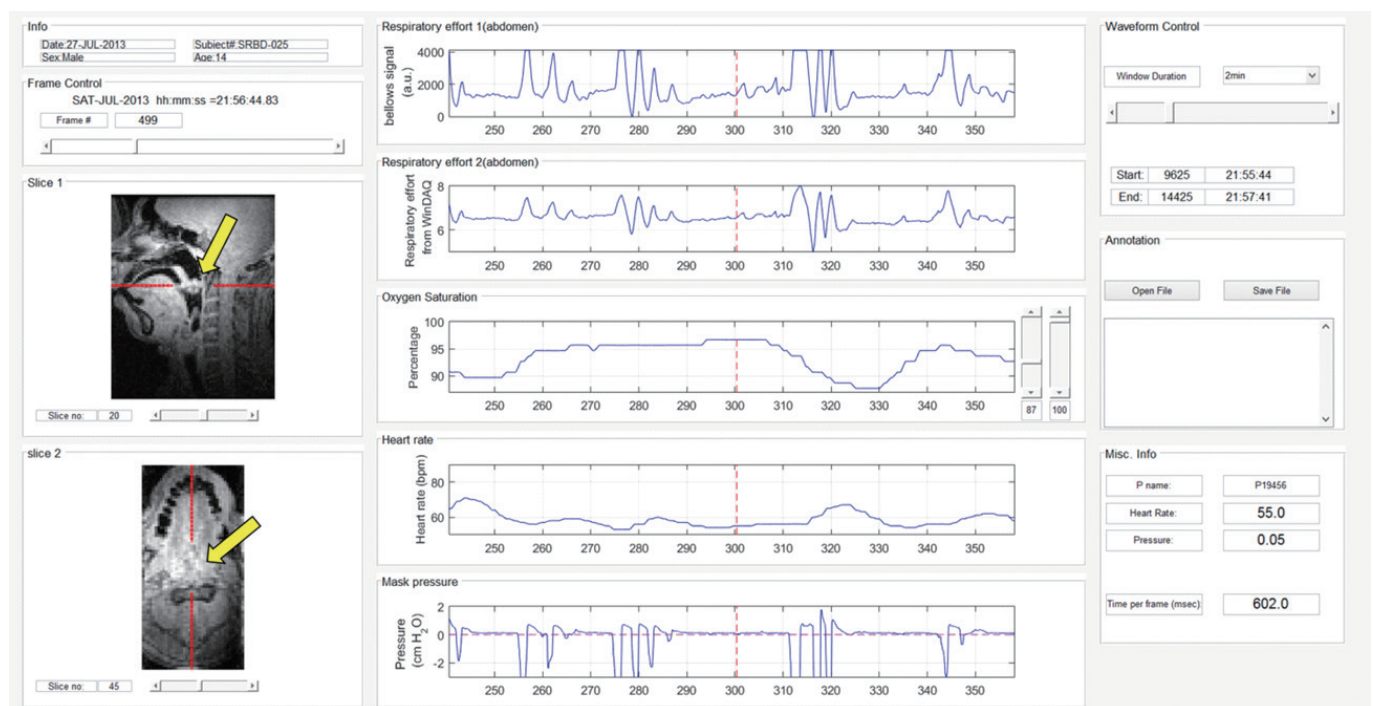


Fig. 9. A custom MATLAB graphical user interface for synchronous inspection of dynamic real-time three-dimensional image frames and simultaneously measured physiological signals. This illustrates airway obstruction (see yellow arrows) at the retropalatal slice at the time (indicated by red dashed lines) when an obstructive sleep apnea event occurs.

their own purposes and modify their programming scripts to implement deep learning algorithms. In MRI, a variety of deep CNN methods have been demonstrated in a variety of applications: for example, brain tissue segmentation [72], cerebral infarct segmentation [73], and cerebral microbleeds detection [74].

Deep learning-based image analysis was recently demonstrated in vocal tract shape analysis. An encoder-decoder CNN was demonstrated to automatically extract the vocal tract air-tissue boundaries [75,76].

MR EXPERIMENTAL ISSUES

Acoustic noise

MRI scans produce loud sounds, which are caused by the vibrations of the three pairs of gradient coils and have been reported to exceed 130 dBA in extreme cases on commercial 3 T scanners [77]. The noise is more severe at higher magnetic field strengths [78]. This may cause patients to experience difficulties in falling asleep in sleep MRI. Sequence parameters such as TR can be adjusted to potentially reduce the sound pressure level. Continuous scanning with a lengthened TR is empirically known to produce a low tone and a steady “humming” sound.

Mirror projector setup

In speech MRI, a mirror projector setup wherein the subject lies on the scanner bed and reads the sentences through the mirror is useful. The experimenter, who is positioned outside the MRI scanner room, operates a laptop and plays the slides containing sentences to read. A projector is used to display the content of the slides to the subject. If the subject is near-sighted and the subject's glasses contain metallic objects, MR-compatible glasses may be provided prior to the scan in order to help the subject read the stimuli.

Synchronized acquisition

Obstructive apneic events occur during a subject's sleep. Electroencephalograms are recorded synchronously with MRI to objectively determine sleep and wakefulness [79]. Respiration, heart rate, and oxygen saturation signals are simultaneously recorded to infer central or obstructive apneic events [49]. A pressure transducer can be used to monitor mask pressure. A continuous positive airway pressure device is used in an MRI environment to control airway pressure [80].

Speech MRI synchronously records the audio signals from the subject. The optical microphone is an MR-compatible de-

vice, so it can be used inside the magnet room. The raw speech signal acquired during MRI scans is corrupted with the MR noise. Recovery of uncorrupted speech sounds requires signal-processing techniques, including adaptive signal processing [81], dictionary learning [82], and combined operations in the time and frequency domains [83]. Audio denoising is often applied off-line after the MRI and audio acquisitions are completed.

Supine position

The supine position in MRI examinations is not an ideal position for speech and swallowing tasks. An open-type MRI scanner can be used to acquire images in the upright position, but its availability is typically more limited than conventional MRI scanners. One study compared vocal tract configurations between the supine and upright positions [84].

Real-time interactive imaging

Real-time interactive imaging software [85] has proved useful for a variety of applications, including high-intensity focused ultrasound ablation of the liver [86], OSA [23,40], and speech imaging [53]. Real-time interactive imaging for speech involves (1) the operator's rapid prescription of a midsagittal scan plane based on anatomical landmarks (e.g., nose tip, pharyngeal airway) in the upper airway, (2) center-frequency adjustment and linear shimming for rapid and interactive de-blurring in spiral imaging, and (3) the user-interactive gradient delay correction. With real-time interactive imaging, the subject's inappropriate behaviors, such as non-speaking or non-responsiveness, can be monitored during scanning. In particular, RTHawk (HeartVista Inc., Los Altos, CA, USA), a commercial real-time interactive imaging software package, has been mainly developed for comprehensive cardiac MRI exams and also has been used for more than a decade to establish a database of real-time speech MRI [22].

Patient comfort

The MRI examination is a loud and claustrophobic process, and thus is not a friendly environment for the subject's natural sleep. Instructions to properly use the ear plugs are helpful in reducing the noise perceived by the subjects. Bedding with memory foam can also help improve patient comfort.

Repeated real-time speech MRI scans without sufficient pauses between the scans can lead to excessive heating in the gradient amplifiers. The gradient amplifier heating issue sometimes results in inadvertent interruptions in the scan-

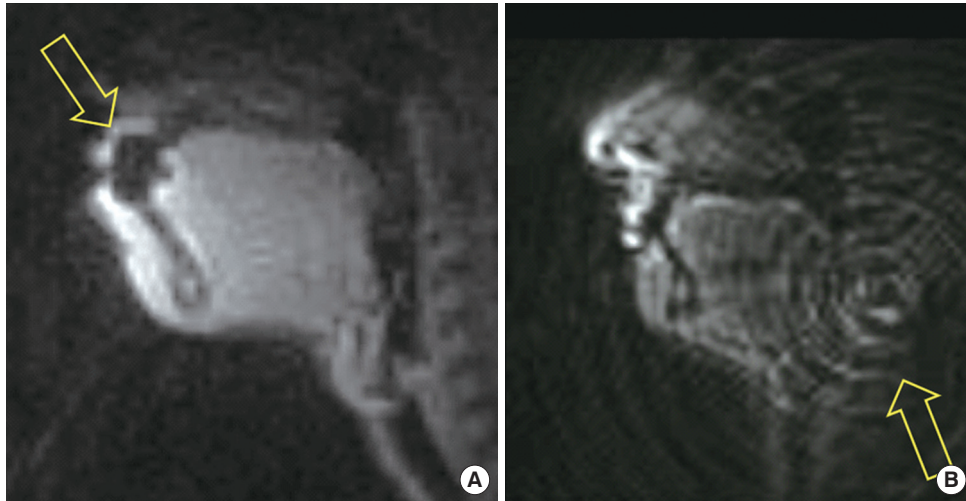


Fig. 10. Examples of image artifacts encountered during real-time speech magnetic resonance imaging scans. (A) The signal void (indicated by the arrow) results from intravoxel dephasing caused by large magnetic susceptibility differences in the presence of metallic dental work. (B) The ring-like artifact (indicated by the arrow) results from radiofrequency (RF) interference when using spiral imaging. The RF leakage can happen when the two scanners are operating at sites close to each other and the magnet room door is slightly open for the audio recording.

ner's operation and causes long waiting times for the subjects until the scanner resumes working. This can be avoided by adequate pauses between the scans.

Teeth

Visualizing the contours of the teeth is important in measuring the degree of constriction in the production of fricative sounds. However, dynamic real-time MRI cannot visualize tooth contours because teeth, which are bony structures, lack hydrogen. MRI data were acquired while the subject held blueberry juice as a contrast medium for MRI [87]. Trays containing ferric ammonium citrate gels, which are T1-shortening contrast media, were used to visualize the teeth [88]. A relatively simple approach involves acquiring MRI data while the subject wraps his or her tongue to the upper teeth followed by the lower teeth. These methods enable extraction of the contours of the teeth, which are superimposed over the dynamic real-time images.

Image artifact issues

Dental work often causes image artifacts near the mouth in the subject's upper airway. MRI screening procedures prior to MRI scanning should check if the subject is wearing dental braces or has undergone metallic dental work. Fig. 10A illustrates an example of the signal voids caused by the presence of dental work.

RF interference can cause image artifacts during real-time MRI. In spiral imaging, the artifact appears as a ring-like pat-

tern (Fig. 10B) instead of the zipper artifact pattern observable in Cartesian imaging.

APPLICATIONS

Upper airway MRI has a wide range of applications in research on sleep apnea and speech production as well as in research on other areas such as swallowing and singing. Related works are briefly described for each application.

Obstructive sleep apnea

OSA is a disease characterized by repetitive episodes of upper airway collapse during sleep [89-91]. Dynamic upper airway MRI was used to identify airway narrowing or obstruction sites in OSA patients with sedation [92,93]. It was also used to evaluate airway narrowing during wakefulness and natural sleep [94].

Deformation of airway tissue can be measured using tagged MRI sequences. Analysis of displacements of the tag lines indicated that the genioglossus moved anteriorly during inspiration in healthy awake subjects [95]. Different patterns of tongue motion were observed during awake breathing in OSA patients [96].

Static 3D imaging methods with conventional MR pulse sequences have been used to quantify anatomical regions of interest for clinical OSA research. A T1-weighted sequence was used to measure the tongue volume and lateral pharyngeal wall volume [97]. A three-point Dixon water-fat separa-

tion sequence was used to measure tongue fat volume from fat images [98].

Speech and singing

Dynamic real-time MRI of speech can provide unique and valuable insights into understanding the spatial and temporal aspects in vocal tract shaping. This technique has been adopted by linguists or speech scientists to investigate speech production and articulatory gestures in a variety of speech tasks or language settings. For example, real-time speech MRI was exploited to investigate the temporal dynamics of vocal tract articulators of interest in nasal sounds [99], English diphthongs [100], Tamil retroflex consonants [101], etc.

Speech pathology has been investigated using dynamic real-time MRI. Real-time MRI of a patient with speech apraxia could capture vocal tract shaping in silent initiation gestures at speech onset and that during covert articulation of words [102].

Vocal tract shaping from real-time MRI was investigated in resonance tuning in soprano singing [103] and in tenors' *passaggio* [104].

Velopharyngeal function

Three-dimensional high-resolution anatomical MRI can be used to visualize the velopharynx including the levator muscle. Early work demonstrated the use of real-time MRI to evaluate velopharyngeal closure in patients with velopharyngeal insufficiency [105]. High frame-rate dynamic speech imaging was used to assess the velopharyngeal anatomy in a midsagittal and an oblique coronal scan plane [106]. The technique enabled visualization of the movements of the soft palate and pharyngeal wall during speech production containing nasal sounds. The imaging is known to be useful in assessing velopharyngeal function in subjects with cleft lip and palate.

Swallowing disorder

An early study investigated three different pulse sequences to evaluate image quality during swallowing [107]. The potential of real-time spiral MRI at 1.5 T was demonstrated in evaluating swallowing function in patients receiving tongue cancer treatment [108]. Real-time MRI was performed at 3 T using GRE sequences [109,110] and a radial FLASH with undersampling [111].

CONCLUSION

Upper airway imaging techniques for speech production and sleep apnea research have been covered in this review. Image acquisition and reconstruction methods for high spatial and temporal resolution as well as full 3D coverage of the upper airway have been investigated by numerous MRI research groups. These techniques involve the use of compressed sensing, parallel imaging, and custom or commercial RF coils, which are sensitive to the upper airway regions of interest. When these research protocols on pulse sequence and reconstruction are not available, the use of commercial pulse sequences is possible for dynamic upper airway MRI [112]. The use of real-time interactive imaging software is beneficial especially for efficient and robust data acquisition in non-Cartesian imaging. In addition, speech and sleep MRI experimental research studies require other MR-compatible measurement devices (e.g., fiberoptic microphone, facial mask), which are not often available for routine clinical exams. High-field MRI (e.g., 7-Tesla MR) is gaining popularity in MR research, but, for dynamic real-time imaging of speech and sleep apnea, lower field strength magnets (e.g., 1.5-Tesla or below) may be advantageous because of the moderate sound pressure level and lower resonance offset from the large degree of magnetic susceptibility between the air and tissue.

Deep learning has recently gained popularity in computer vision [113] and medical image analysis [114]. It is also emerging in the areas of MRI reconstruction and image segmentation. Current trends in deep learning-based image reconstruction suggest that real-time upper airway imaging may benefit from faster reconstruction times using a reconstruction method with a learned model than with conventional iterative reconstructions (e.g., compressed sensing parallel imaging). Automatic post-processing of upper airway image data may benefit from deep learning-based techniques in the areas of image segmentation and landmark detection.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

This work was supported by Basic Science Research Program

through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (Grant Number: NRF-2015 R1C1A1A02036340, NRF-2018 R1D1A1B07042692). The author acknowledges the Speech Production and Articulation kNowledge group at the University of Southern California and thanks anonymous reviewers for valuable comments.

REFERENCES

1. Stone M. Laboratory techniques for investigating speech articulation. In: Hardcastle WJ, Laver J, Gibbon F, editors. *The handbook of phonetic sciences*. 2nd ed. Chichester (UK): Wiley-Blackwell; 2010. p. 9-38.
2. Hiiemae KM, Palmer JB. Tongue movements in feeding and speech. *Crit Rev Oral Biol Med* 2003;14:413-29.
3. Kezirian EJ, Hohenhorst W, de Vries N. Drug-induced sleep endoscopy: the VOTE classification. *Eur Arch Otorhinolaryngol* 2011;268:1233-6.
4. Zysk AM, Nguyen FT, Oldenburg AL, Marks DL, Boppart SA. Optical coherence tomography: a review of clinical development from bench to bedside. *J Biomed Opt* 2007;12:051403.
5. Jing J, Zhang J, Loy AC, Wong BJ, Chen Z. High-speed upper-airway imaging using full-range optical coherence tomography. *J Biomed Opt* 2012;17:110507.
6. Armstrong JJ, Leigh MS, Sampson DD, Walsh JH, Hillman DR, Eastwood PR. Quantitative upper airway imaging with anatomic optical coherence tomography. *Am J Respir Crit Care Med* 2006;173:226-33.
7. Lingala SG, Sutton BP, Miquel ME, Nayak KS. Recommendations for real-time speech MRI. *J Magn Reson Imaging* 2016;43:28-44.
8. Scott AD, Wylezinska M, Birch MJ, Miquel ME. Speech MRI: morphology and function. *Phys Med* 2014;30:604-18.
9. Bresch E, Kim YC, Nayak K, Byrd D, Narayanan S. Seeing speech: capturing vocal tract shaping using real-time magnetic resonance imaging [Exploratory DSP]. *IEEE Signal Process Mag* 2008;25:123-32.
10. Ramanarayanan V, Tilsen S, Proctor M, Toger J, Goldstein L, Nayak KS, et al. Analysis of speech production real-time MRI. *Comput Speech Lang* 2018;52:1-22.
11. Nayak KS, Fleck RJ. Seeing sleep: dynamic imaging of upper airway collapse and collapsibility in children. *IEEE Pulse* 2014;5:40-4.
12. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med* 1999;42:952-62.
13. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med* 2002;47:1202-10.
14. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med* 2007;58:1182-95.
15. Uecker M, Zhang S, Voit D, Karaus A, Merboldt KD, Frahm J. Real-time MRI at a resolution of 20 ms. *NMR Biomed* 2010;23:986-94.
16. Kim YC, Loloyan S, Wu Z, Tran W, Kato R, Ward SLD, et al. Real-time MRI can differentiate sleep-related breathing disorders in children. In: 21st Annual ISMRM Scientific Meeting and Exhibition 2013; 2013 Apr 20-26; Salt Lake City, UT. p. 251.
17. Samsonov AA, Kholmovski EG, Parker DL, Johnson CR. POCSense: POCS-based reconstruction for sensitivity encoded magnetic resonance imaging. *Magn Reson Med* 2004;52:1397-406.
18. Niebergall A, Zhang S, Kunay E, Keydana G, Job M, Uecker M, et al. Real-time MRI of speaking at a resolution of 33 ms: undersampled radial FLASH with nonlinear inverse reconstruction. *Magn Reson Med* 2013;69:477-85.
19. Itlis 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;5:374-81.
20. Nayak KS, Cunningham CH, Santos JM, Pauly JM. Real-time cardiac MRI at 3 tesla. *Magn Reson Med* 2004;51:655-60.
21. Seiberlich N, Lee G, Ehse P, Duerk JL, Gilkeson R, Griswold M. Improved temporal resolution in cardiac imaging using through-time spiral GRAPPA. *Magn Reson Med* 2011;66:1682-8.
22. Narayanan S, Toutios A, Ramanarayanan V, Lammert A, Kim J, Lee S, et al. Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC). *J Acoust Soc Am* 2014;136:1307.
23. Barrera JE. Sleep magnetic resonance imaging: dynamic characteristics of the airway during sleep in obstructive sleep apnea syndrome. *Laryngoscope* 2011;121:1327-35.
24. Block KT, Frahm J. Spiral imaging: a critical appraisal. *J Magn Reson Imaging* 2005;21:657-68.

25. 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;32:1228-37.
26. Lim Y, Lingala SG, Toutios A, Narayanan S, Nayak KS. Improved depiction of tissue boundaries in vocal tract real-time MRI using automatic off-resonance correction. *Interspeech* 2016;2016:1765-9.
27. Lim Y, Lingala SG, Narayanan SS, Nayak KS. Dynamic off-resonance correction for spiral real-time MRI of speech. *Magn Reson Med* 2019;81:234-46.
28. Winkelmann S, Schaeffter T, Koehler T, Eggers H, Doesel O. An optimal radial profile order based on the Golden Ratio for time-resolved MRI. *IEEE Trans Med Imaging* 2007;26:68-76.
29. Feng L, Grimm R, Block KT, Chandarana H, Kim S, Xu J, 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* 2014;72:707-17.
30. Burdumy M, Traser L, Richter B, Echternach M, Korvink JG, Hennig J, et al. Acceleration of MRI of the vocal tract provides additional insight into articulator modifications. *J Magn Reson Imaging* 2015;42:925-35.
31. Lingala SG, Zhu Y, Kim YC, Toutios A, Narayanan S, Nayak KS. A fast and flexible MRI system for the study of dynamic vocal tract shaping. *Magn Reson Med* 2017;77:112-25.
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;65:1365-71.
33. Scott AD, Boubertakh R, Birch MJ, Miquel ME. Adaptive averaging applied to dynamic imaging of the soft palate. *Magn Reson Med* 2013;70:865-74.
34. Parthasarathy V, Prince JL, Stone M, Murano EZ, Nesaiver M. Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing. *J Acoust Soc Am* 2007;121:491-504.
35. Stone M, Davis EP, Douglas AS, Nesaiver M, Gullapalli R, Levine WS, et al. Modeling the motion of the internal tongue from tagged cine-MRI images. *J Acoust Soc Am* 2001;109:2974-82.
36. Fischer SE, McKinnon GC, Maier SE, Boesiger P. Improved myocardial tagging contrast. *Magn Reson Med* 1993;30:191-200.
37. Axel L, Dougherty L. Heart wall motion: improved method of spatial modulation of magnetization for MR imaging. *Radiology* 1989;172:349-50.
38. Kim YC, Proctor MI, Narayanan SS, Nayak KS. Improved imaging of lingual articulation using real-time multislice MRI. *J Magn Reson Imaging* 2012;35:943-8.
39. Feng X, Blemker SS, Inouye J, Pelland CM, Zhao L, Meyer CH. Assessment of velopharyngeal function with dual-planar high-resolution real-time spiral dynamic MRI. *Magn Reson Med* 2018;80:1467-74.
40. Shin LK, Holbrook AB, Capasso R, Kushida CA, Powell NB, Fischbein NJ, et al. Improved sleep MRI at 3 tesla in patients with obstructive sleep apnea. *J Magn Reson Imaging* 2013;38:1261-6.
41. Breuer FA, Blaimer M, Heidemann RM, Mueller MF, Griswold MA, Jakob PM. Controlled aliasing in parallel imaging results in higher acceleration (CAIPRINHA) for multi-slice imaging. *Magn Reson Med* 2005;53:684-91.
42. Wu Z, Chen W, Khoo MC, Davidson Ward SL, Nayak KS. Evaluation of upper airway collapsibility using real-time MRI. *J Magn Reson Imaging* 2016;44:158-67.
43. Story BH, Titze IR, Hoffman EA. Vocal tract area functions from magnetic resonance imaging. *J Acoust Soc Am* 1996;100:537-54.
44. Narayanan SS, Alwan AA, Haker K. An articulatory study of fricative consonants using magnetic resonance imaging. *J Acoust Soc Am* 1995;98:1325-47.
45. Zhou X, Espy-Wilson CY, Boyce S, Tiede M, Holland C, Choe A. A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *J Acoust Soc Am* 2008;123:4466-81.
46. Kim YC, Narayanan SS, Nayak KS. Accelerated three-dimensional upper airway MRI using compressed sensing. *Magn Reson Med* 2009;61:1434-40.
47. Burdumy M, Traser L, Burk F, Richter B, Echternach M, Korvink JG, et al. One-second MRI of a three-dimensional vocal tract to measure dynamic articulator modifications. *J Magn Reson Imaging* 2017;46:94-101.
48. Wagshul ME, Sin S, Lipton ML, Shifteh K, Arens R. Novel retrospective, respiratory-gating method enables 3D, high resolution, dynamic imaging of the upper airway during tidal breathing. *Magn Reson Med* 2013;70:1580-90.
49. Kim YC, Lebel RM, Wu Z, Ward SL, Khoo MC, Nayak KS. Real-time 3D magnetic resonance imaging of the pharyngeal airway in sleep apnea. *Magn Reson Med* 2014;71:1501-10.
50. Zhu Y, Kim YC, Proctor MI, Narayanan SS, Nayak KS. Dy-

- namic 3-D visualization of vocal tract shaping during speech. *IEEE Trans Med Imaging* 2013;32:838-48.
51. Fu M, Barlaz MS, Holtrop JL, Perry JL, Kuehn DP, Shosted RK, et al. High-frame-rate full-vocal-tract 3D dynamic speech imaging. *Magn Reson Med* 2017;77:1619-29.
52. 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;65:1711-7.
53. 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;115:1771-6.
54. Jackson JI, Meyer CH, Nishimura DG, Macovski A. Selection of a convolution function for Fourier inversion using gridding [computerised tomography application]. *IEEE Trans Med Imaging* 1991;10:473-8.
55. Lingala SG, Zhu Y, Lim Y, Toutios A, Ji Y, Lo WC, et al. Feasibility of through-time spiral generalized autocalibrating partial parallel acquisition for low latency accelerated real-time MRI of speech. *Magn Reson Med* 2017;78:2275-82.
56. Stone SS, Haldar JP, Tsao SC, Hwu WM, Sutton BP, Liang ZP. Accelerating advanced MRI reconstructions on GPUs. *J Parallel Distrib Comput* 2008;68:1307-18.
57. Murphy M, Alley M, Demmel J, Keutzer K, Vasanawala S, Lustig M. Fast l(1)-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime. *IEEE Trans Med Imaging* 2012;31:1250-62.
58. Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med* 2018;79:3055-71.
59. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature* 2018;555:487-92.
60. Vasconcelos MJ, Ventura SM, Freitas DR, Tavares JM. Towards the automatic study of the vocal tract from magnetic resonance images. *J Voice* 2011;25:732-42.
61. Bresch E, Narayanan S. Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images. *IEEE Trans Med Imaging* 2009;28:323-38.
62. Proctor MI, Bone D, Katsamanis A, Narayanan SS. Rapid semi-automatic segmentation of real-time magnetic resonance images for parametric vocal tract analysis. In: 11th Annual Conference of the International Speech Communication Association 2010; 2010 Sep 26-30; Chiba, JP. Red Hook (NY): Curran Associates, Inc.; 2010. p. 1576-9.
63. Labrunie M, Badin P, Voit D, Joseph AA, Frahm J, Lamalle L, et al. Automatic segmentation of speech articulators from real-time midsagittal MRI based on supervised learning. *Speech Commun* 2018;99:27-46.
64. Kim YC, Kim J, Proctor M, Toutios A, Nayak K, Lee S, et al. Toward automatic vocal tract area function estimation from accelerated three-dimensional magnetic resonance imaging. In: Workshop on Speech Production in Automatic Speech Recognition; 2013 Aug 30; Lyon, FR.
65. Javed A, Kim YC, Khoo MC, Ward SL, Nayak KS. Dynamic 3-D MR visualization and detection of upper airway obstruction during sleep using region-growing segmentation. *IEEE Trans Biomed Eng* 2016;63:431-7.
66. Xu C, Sin S, McDonough JM, Udupa JK, Guez A, Arens R, et al. Computational fluid dynamics modeling of the upper airway of children with obstructive sleep apnea syndrome in steady flow. *J Biomech* 2006;39:2043-54.
67. Mylavarapu G, Murugappan S, Mihaescu M, Kalra M, Khosla S, Gutmark E. Validation of computational fluid dynamics methodology used for human upper airway flow simulations. *J Biomech* 2009;42:1553-9.
68. Lucey AD, King AJ, Tetlow GA, Wang J, Armstrong JJ, Leigh MS, et al. Measurement, reconstruction, and flow-field computation of the human pharynx with application to sleep apnea. *IEEE Trans Biomed Eng* 2010;57:2535-48.
69. Huang Y, White DP, Malhotra A. Use of computational modeling to predict responses to upper airway surgery in obstructive sleep apnea. *Laryngoscope* 2007;117:648-53.
70. Goodfellow I, Bengio Y, Courville A, editors. Deep learning. Cambridge (UK): London MIT Press; 2016.
71. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, editors. Advances in Neural Information Processing Systems 25 (NIPS 2012); 2012 Dec 3-6; Lake Tahoe, NV. Red Hook (NY): Curran Associates, Inc.; 2012. p. 1097-105.
72. Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJ, Išgum I. Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Trans Med Imaging* 2016;35:1252-61.
73. Chen L, Bentley P, Rueckert D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *Neuroimage Clin* 2017;15:633-43.

74. Qi D, Hao C, Lequan Y, Lei Z, Jing Q, Defeng W, et al. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. *IEEE Trans Med Imaging* 2016;35:1182-95.
75. Somandepalli K, Toutios A, Narayanan SS. Semantic edge detection for tracking vocal tract air-tissue boundaries in real-time magnetic resonance images. *Proc Interspeech* 2017;2017:631-5.
76. Valliappan CA, Mannem R, Ghosh PK. Air-tissue boundary segmentation in real-time magnetic resonance imaging video using semantic segmentation with fully convolutional networks. *Proc Interspeech* 2018;2018:3132-6.
77. Wu Z, Kim YC, Khoo MC, Nayak KS. Evaluation of an independent linear model for acoustic noise on a conventional MRI scanner and implications for acoustic noise reduction. *Magn Reson Med* 2014;71:1613-20.
78. Price DL, De Wilde JP, Papadaki AM, Curran JS, Kitney RI. Investigation of acoustic noise on 15 MRI scanners from 0.2 T to 3 T. *J Magn Reson Imaging* 2001;13:288-93.
79. Kavcic P, Koren A, Koritnik B, Fajdiga I, Groselj LD. Sleep magnetic resonance imaging with electroencephalogram in obstructive sleep apnea syndrome. *Laryngoscope* 2015;125:1485-90.
80. Chen W, Gillett E, Khoo MCK, Davidson Ward SL, Nayak KS. Real-time multislice MRI during continuous positive airway pressure reveals upper airway response to pressure change. *J Magn Reson Imaging* 2017;46:1400-8.
81. Bresch E, Nielsen J, Nayak K, Narayanan S. Synchronized and noise-robust audio recordings during real-time magnetic resonance imaging scans. *J Acoust Soc Am* 2006;120:1791-4.
82. Vaz C, Ramanarayanan V, Narayanan S. Acoustic denoising using dictionary learning with spectral and temporal regularization. *IEEE/ACM Trans Audio Speech Lang Process* 2018;26:967-80.
83. 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;135:1019-22.
84. Traser L, Burdumy M, Richter B, Vicari M, Echternach M. The effect of supine and upright position on vocal tract configurations during singing: a comparative study in professional tenors. *J Voice* 2013;27:141-8.
85. Santos JM, Wright GA, Pauly JM. Flexible real-time magnetic resonance imaging framework. *Conf Proc IEEE Eng Med Biol Soc* 2004;2:1048-51.
86. Holbrook AB, Santos JM, Kaye E, Rieke V, Pauly KB. Real-time MR thermometry for monitoring HIFU ablations of the liver. *Magn Reson Med* 2010;63:365-73.
87. Takemoto H, Kitamura T, Nishimoto H, Honda K. A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions. *Acoust Sci Technol* 2004;25:468-74.
88. Nunthayanon K, Honda E, Shimazaki K, Ohmori H, Inoue-Arai MS, Kurabayashi T, et al. Use of an advanced 3-T MRI movie to investigate articulation. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2015;119:684-94.
89. Eastwood PR, Malhotra A, Palmer LJ, Kezirian EJ, Horner RL, Ip MS, et al. Obstructive sleep apnoea: from pathogenesis to treatment: current controversies and future directions. *Respirology* 2010;15:587-95.
90. Muzumdar H, Arens R. Diagnostic issues in pediatric obstructive sleep apnea. *Proc Am Thorac Soc* 2008;5:263-73.
91. Fogel RB, Malhotra A, White DP. Sleep. 2: pathophysiology of obstructive sleep apnoea/hypopnoea syndrome. *Thorax* 2004;59:159-63.
92. Moon IJ, Han DH, Kim JW, Rhee CS, Sung MW, Park JW, et al. Sleep magnetic resonance imaging as a new diagnostic method in obstructive sleep apnea syndrome. *Laryngoscope* 2010;120:2546-54.
93. Donnelly LF, Surdulescu V, Chini BA, Casper KA, Poe SA, Amin RS. Upper airway motion depicted at cine MR imaging performed during sleep: comparison between young patients with and those without obstructive sleep apnea. *Radiology* 2003;227:239-45.
94. Darquenne C, Elliott AR, Sibille B, Smales ET, DeYoung PN, Theilmann RJ, et al. Upper airway dynamic imaging during tidal breathing in awake and asleep subjects with obstructive sleep apnea and healthy controls. *Physiol Rep* 2018;6:e13711.
95. Cheng S, Butler JE, Gandevia SC, Bilston LE. Movement of the tongue during normal breathing in awake healthy humans. *J Physiol* 2008;586:4283-94.
96. Bilston LE, Gandevia SC. Biomechanical properties of the human upper airway and their effect on its behavior during breathing and in obstructive sleep apnea. *J Appl Physiol* (1985) 2014;116:314-24.
97. Schwab RJ, Pasirstein M, Pierson R, Mackley A, Hachadorian R, Arens 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;168:522-30.

98. Kim AM, Keenan BT, Jackson N, Chan EL, Staley B, Poptani H, et al. Tongue fat and its relationship to obstructive sleep apnea. *Sleep* 2014;37:1639-48.
99. Byrd D, Tobin S, Bresch E, Narayanan S. Timing effects of syllable structure and stress on nasals: a real-time MRI examination. *J Phon* 2009;37:97-110.
100. Hsieh FY, Goldstein L, Byrd D, Narayanan S. Truncation of pharyngeal gesture in English diphthong [a]. In: Bimbot F, Cerisara C, Fougeron C, Gravier G, Lamel L, Pellegrino F, editors. 14th Annual Conference of the International Speech Communication Association (INTERSPEECH 2013); 2013 Aug 25-29; Lyon, FR. Red Hook (NY): Curran Associates, Inc.; 2013. p. 968-72.
101. Smith C, Proctor MI, Iskarous K, Goldstein L, Narayanan S. Stable articulatory tasks and their variable formation: tamil retroflex consonants. In: Bimbot F, Cerisara C, Fougeron C, Gravier G, Lamel L, Pellegrino F, editors. 14th Annual Conference of the International Speech Communication Association (INTERSPEECH 2013); 2013 Aug 25-29; Lyon, FR. Red Hook (NY): Curran Associates, Inc.; 2013. p. 2006-9.
102. Hagedorn C, Proctor M, Goldstein L, Wilson SM, Miller B, Gorno-Tempini ML, et al. Characterizing articulation in apraxic speech using real-time magnetic resonance imaging. *J Speech Lang Hear Res* 2017;60:877-91.
103. Bresch E, Narayanan S. Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing. *J Acoust Soc Am* 2010;128:EL335-41.
104. Echternach M, Popeil L, Traser L, Wienhausen S, Richter B. Vocal tract shapes in different singing functions used in musical theater singing: a pilot study. *J Voice* 2014; 28:653.e1-e7.
105. Beer AJ, Hellerhoff P, Zimmermann A, Mady K, Sader R, Rummeny EJ, et al. Dynamic near-real-time magnetic resonance imaging for analyzing the velopharyngeal closure in comparison with videofluoroscopy. *J Magn Reson Imaging* 2004;20:791-7.
106. Perry JL, Sutton BP, Kuehn DP, Gamage JK. Using MRI for assessing velopharyngeal structures and function. *Cleft Palate Craniofac J* 2014;51:476-85.
107. Anagnostara A, Stoeckli S, Weber OM, Kollias SS. Evaluation of the anatomical and functional properties of deglutition with various kinetic high-speed MRI sequences. *J Magn Reson Imaging* 2001;14:194-9.
108. Zu Y, Narayanan SS, Kim YC, Nayak K, Bronson-Lowe C, Villegas B, 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;139:1312-9.
109. Amin MR, Lazarus CL, Pai VM, Mulholland TP, Shepard T, Branski RC, et al. 3 Tesla turbo-FLASH magnetic resonance imaging of deglutition. *Laryngoscope* 2012;122: 860-4.
110. Breyer T, Echternach M, Arndt S, Richter B, Speck O, Schumacher M, et al. Dynamic magnetic resonance imaging of swallowing and laryngeal motion using parallel imaging at 3 T. *Magn Reson Imaging* 2009;27:48-54.
111. Zhang S, Olthoff A, Frahm J. Real-time magnetic resonance imaging of normal swallowing. *J Magn Reson Imaging* 2012;35:1372-9.
112. Freitas AC, Ruthven M, Boubertakh R, Miquel ME. Real-time speech MRI: commercial Cartesian and non-Cartesian sequences at 3T and feasibility of offline TGV reconstruction to visualise velopharyngeal motion. *Phys Med* 2018;46:96-103.
113. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; 521:436-44.
114. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88.