Interspeech 2024
1-5 September 2024, Kos, Greece

State-of-the-art speech production MRI protocol for new 0.55 Tesla scanners

Prakash Kumar[1], *Ye Tian*[1], *Yongwan Lim*[1], *Sophia X. Cui*[2], *Christina Hagedorn*[3], *Dani Byrd*[1], *Uttam K. Sinha*[4], *Shrikanth Narayanan*[1], *Krishna S. Nayak*[1]

[1]University of Southern California, United States [2]Siemens Medical Solutions, United States Inc. [3] City University of New York College of Staten Island, New York, United States [4] Keck School of Medicine, University of Southern California, United States

prakashk@usc.edu, knayak@usc.edu

Abstract

Real-time magnetic resonance imaging (RT-MRI) is a safe and powerful tool for studying vocal tract dynamics during speech production. Emerging low- and mid-field MRI platforms bring new capability including higher frame rates, improved tissue contrast, and reduced blurring compared to conventional 1.5T and 3T MRI. Here, we present a state-of-the-art speech production MRI protocol tailored to the 0.55T platform. This includes 2D mid-sagittal vocal tract RT-MRI, simultaneous multislice RT-MRI with 3 parallel slices, tagged RT-MRI for functional evaluation of internal tongue deformation, biofeedback that allows participants to see their scan in real-time, and 3D static imaging. Imaging performance is comparable and, in several cases, superior to 1.5T and 3T MRI, making 0.55T an exciting new platform for speech research.

Index Terms: magnetic resonance imaging, biofeedback, real-time imaging

1. Introduction

Speech production research has benefitted tremendously from non-invasive dynamic imaging. This has allowed scientists to observe and investigate the kinematics of vocal tract articulators, including the tongue, velum, and lips during human speech production. Many imaging modalities have been used for speech including x-ray, ultrasound, and magnetic resonance imaging (MRI) [1]. Among these, MRI offers specific advantages due to the lack of ionizing radiation and the ability to image in arbitrary scan planes that other modalities cannot achieve [2]. Its primary limitation has been cost and access. While conventional MRI scans are too slow to capture the movement of articulators, real-time (RT-MRI) techniques have been able to resolve speech at high temporal and spatial resolutions (e.g., 83 frames/second, 2.4 mm^2 spatial resolution), which are adequate to capture speech events including tongue and velum movements, consonant and vowel constrictions, pharyngeal shaping, and laryngeal height [3]. RT-MRI capability has been enabled by imaging technology including non-Cartesian sampling, off-line constrained reconstruction, and deep learning [4].

Speech production RT-MRI has been extensively studied at 1.5T and 3T. Established methods include 2D RT-MRI [5], 3D imaging of sustained sounds (continuants), tagged RT-MRI for imaging of internal tongue deformation [6], and 3D RT-MRI [7]. Data from speech production MRI have been published in open datasets [8, 9], and MRI images have been used to examine the complexity of vocal tract shaping in healthy speakers [10], cross-sectional studies, longitudinal studies, and in clinical populations such as glossectomy [11] and apraxia [12].

A speech production imaging protocol has recently been developed and refined that leverages the strengths of 0.55T and mitigates the challenges [13], achieving superior image sharpness compared to established protocols at 1.5T [14]. At higher field strengths, magnetic susceptibility at air-tissue interfaces causes blurring at articulator-air boundaries, which requires additional corrections to image reconstruction or short-readout sequences and limits the capabilities of image acquisition [15]. At lower field strengths, this susceptibility is reduced and enables longer spiral readouts without inducing blurring. Furthermore, the balanced steady-state free precession (bSSFP) contrast becomes feasible, which has superior signal-to-noise ratio efficiency compared to the gradient recalled echo sequence used at 1.5T [16].

Other advantages to mid-field systems include reduced specific absorption rate (SAR) constraints and reduced acoustic noise levels. Low SAR requirements allow for radiofrequency pulse designs that have higher energy, including higher flip angles and simultaneous multislice (SMS) encoding [17, 18, 19]. It has been shown that 0.55T systems also have reduced acoustic noise by 20dB [20] and are more comfortable and amenable to real-time interaction at the scanner, where participants can have conversations with members in the operator room or hear their own speech.

In this work, we adapt prior work on speech protocols at 0.55T [13] and introduce new imaging techniques that take advantage of mid-field hardware. We demonstrate single slice 2D RT-MRI, simultaneous multislice RT-MRI, tagged RT-MRI, biofeedback that allows participants to see their scan in real-time, and 3D static imaging of the vocal tract at 0.55T.

2. Methods

2.1. Experimental Methods

Experiments were conducted on a prototype 0.55T scanner (MAGNETOM Aera; Siemens Healthineers, Erlangen, Germany) equipped with "Aera XQ" high-performance shielded gradients (45mT/m amplitude, 200T/m/s slew rate). Pulse sequences were designed using the Pulseq framework [21] and the RTHawk framework (Vista.ai, Inc., Los Altos, CA, USA) [22]. The Pulseq framework is an open-source, vendor-agnostic framework that allows pulse sequences to be run on many commercial scanners, and pulse sequences can be shared to increase accessibility. Forty-five participants were imaged under protocols approved by our Institutional Review Board, after providing written informed consent. Ten participants had undergone a glossectomy procedure with lingual tissue reconstruction using a flap of a different composition than tongue muscle.

Figure 1 shows an example experimental setup. In the scanner room, an MRI-compatible TV display is connected to an HDMI-Fiber optic converter box, which is fed through the

10.21437/Interspeech.2024-2263

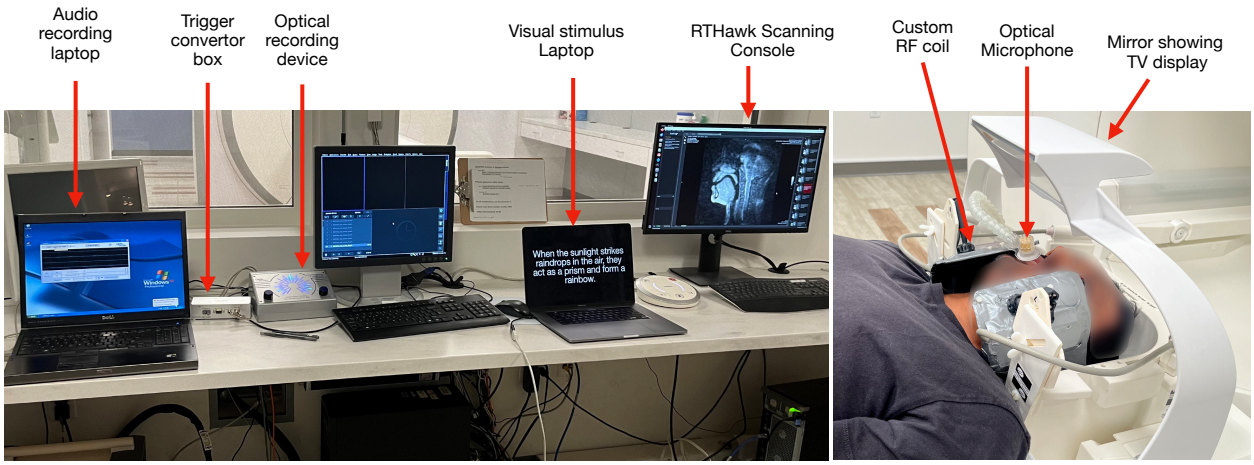

Figure 1: *Experimental setup for 0.55T MRI of speech production. MRI operator room (left) and volunteer positioned on the scanner table prior to table shift (right). A custom upper airway radiofrequency (RF) coil is used to provide optimal signal-to-noise ratio from vocal tract articulators. An MRI-compatible optical microphone is attached to the coil array and placed in front of the mouth of the participant. A display input is connected to an HDMI to fiber optic convertor box and then fed through the waveguide to a TV inside the scanner room. The participant inside the scanner room can see the TV inside the room using a lookout mirror that is placed in front of their eyes and pointed at the TV.*

waveguide to the operator room and connected to a laptop which can display stimulus or biofeedback to the user. A custom 9-channel upper-airway radiofrequency coil is used to measure signals from the vocal articulators. Custom speech coils are able to achieve 3.5x better performance at the imaging of speech articulators compared to a traditional head/neck coil [23]. An optical microphone (Optoacoustics Ltd., Moshav Mazor, Israel) is also attached to the coil and fed through the waveguide to a laptop running software to record audio from the participant. Finally, a trigger signal is sent from the scanner to the recording device such that when scans are activated, the audio recording is automatically turned on and synchronized to the MR signal.

To enable biofeedback, the display signal from the RTHawk scanning console is used as the stimulus signal. The open-source Open Broadcasting Software (OBS)[1] was used to display the relevant portion of the scan and any additional text stimuli or target markers on the display. These markers and text can be updated in real-time by the scanner operator by typing or clicking and dragging the mouse using the OBS software. Before being presented with a specific stimulus, scan participants were given a 1-minute "free-exploration" period, during which they could speak freely, become acquainted with visualization of their vocal tract in real-time, and receive education about basic vocal tract anatomy locations (e.g., lips, tongue tip).

2.2. Acquisition Methods

Table 1 defines the MRI protocols used in this study, including real-time spiral images and 3D static sequences.

[1]https://github.com/obsproject/ obs-studio

2.2.1. 2D RT-MRI (single slice, simultaneous multislice, and tagged)

Pulse sequences were adapted from an established method that optimized bSSFP real-time speech imaging at 0.55T [13] and had a 2.4 mm in-plane resolution and a 6 mm slice thickness. For simultaneous multislice imaging, a blipped-controlled aliasing in parallel imaging was used to achieve 3 parallel sagittal slices with 6 mm slice thickness and a 10 mm slice gap [24]. For tagged imaging, a 180-degree tag pulse was used prior to imaging for the operator to place grid-lines on the tissue image [6]. Tag pulses were periodically played to the participant in the scanner, and the participant responded to the sound of the MRI scanner by reacting to the repetitive sound, repeating a single word or phrase when the tag pulse was played.

2.2.2. 3D Static Imaging

Static images were adapted from product sequences at 1.5 Tesla. A T2-weighted turbo spin echo and T1 Dixon vibe sequence that measures two echoes to generate separate fat and water images were used. With separate fat and water images, reconstructed lingual tissue from glossectomy surgery can be identified due to the difference in tissue composition. Static sequences had higher resolution than dynamic sequences, with $<$ 1 mm in-plane resolution, but had larger scan times (3 mins per scan).

2.3. Reconstruction Methods

2.3.1. Online Reconstruction

Real-time low-latency reconstruction was implemented in RTHawk by using a gridding algorithm that interpolates non-Cartesian samples into a Cartesian grid before the inverse Fast-

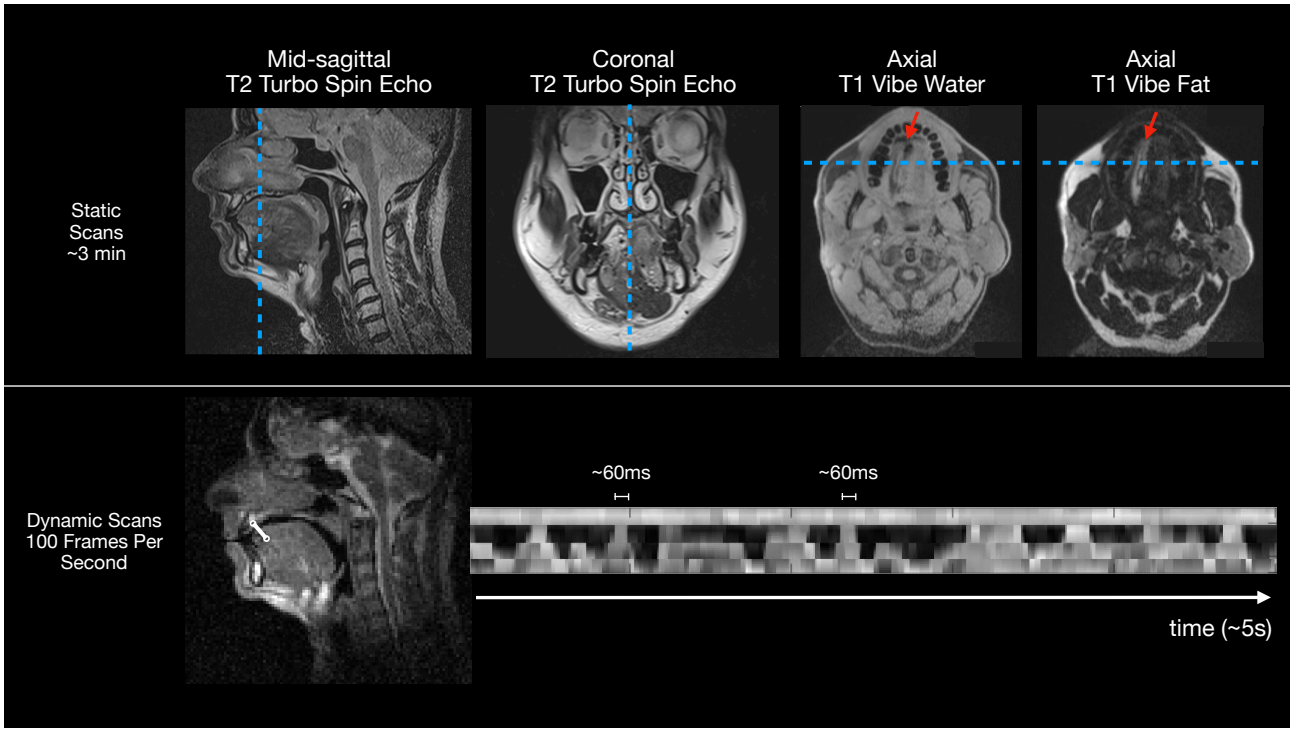

Figure 2: *Representative examples of static and dynamic imaging from a patient with tongue cancer after glossectomy and lingual reconstruction using a radial forearm flap. The static scan provides clear visualization of all relevant anatomic features in multiple orientations, and includes water/fat separated images, which allow clear delineation of the reconstructed tissue from tongue tissue (red arrows). Dynamic imaging at 100 frames per second resolves fast movement of the tongue tip as illustrated by the ability to resolve short alveolar constriction events in the intensity vs. time profiles. See also supplemental file 2D_real-time.mp4.*

Fourier Transform and view-sharing to acquire an aliasing-free image. This reconstruction enables fast scan plane switching and biofeedback, where the scanning console is mirrored onto a TV screen that the participant in the scanner can see and interact with.

2.3.2. *Offline Reconstruction*

Raw data were reconstructed by solving a constrained sparse-SENSE cost function with a temporal and spatial finite-differences constraint, and a non-linear conjugate gradient descent solver [25]. Spatial and temporal regularization parameters were chosen by doing a grid search of possible combinations and qualitatively assessing the image for temporal and spatial blurring. 2D RT-MRI images were reconstructed with 2 spiral arms per frame (100 frames per second), and SMS RT-MRI images were reconstructed with 4 spiral arms per frame (46 frames per second).

Table 1: *MRI scan protocol parameters*

Scan	Resolution	Time
2D Real-time Spiral	2.4 x 2.4 x 6.0 mm^3	100 FPS
2D Real-time SMS	2.4 x 2.4 x 6.0 mm^3	46 FPS
2D Real-time Tagging	2.4 x 2.4 x 6.0 mm^3	100 FPS
T2 Static TSE sagittal	0.5 x 0.5 x 3.5 mm^3	3 min scan time
T2 Static TSE coronal	0.5 x 0.5 x 3.5 mm^3	3 min scan time
T1 Static Vibe Dixon	0.9 x 0.9 x 1.5 mm^3	3 min scan time

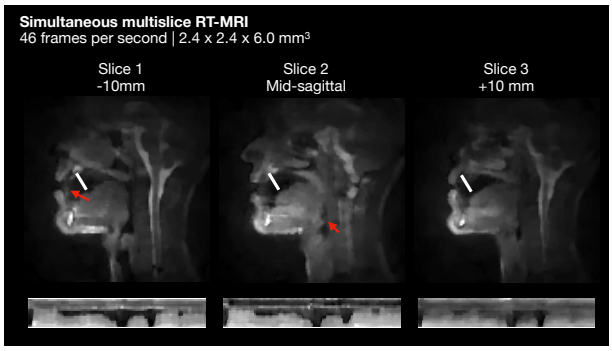

Figure 3: *Simultaneous multislice RT-MRI during speech production of a "count 1-10" stimulus. Intensity line profiles show that temporal dynamics of the tongue tip are resolved. Compared to single-slice RT-MRI, the images have additional spatial blurring around the airway, which may be due to imperfect slice separation during reconstruction. See also supplemental file SMS_real-time.mp4.*

3. Results

Figure 2 shows data of an individual who had undergone treatment for tongue cancer with lingual reconstruction using a radial forearm flap. Static imaging is able to resolve vocal tract features with high spatial resolution, and dynamic images can resolve tongue tip motion, as shown by intensity vs. time profiles.

Figure 4: *Biofeedback Example. An individual who has undergone glossectomy treatment for tongue cancer (Female, 68 yo) is shown a purple star in the dental region as a visual target for tongue tip placement to pronounce the initial dental fricative in “theme”. An orthographic stimulus is shown at the top of the screen, and the participant’s live scan is shown at the bottom of the screen. The biofeedback scan display is updated in real-time with* < 100 *ms latency. See also supplemental file biofeedback.mp4.*

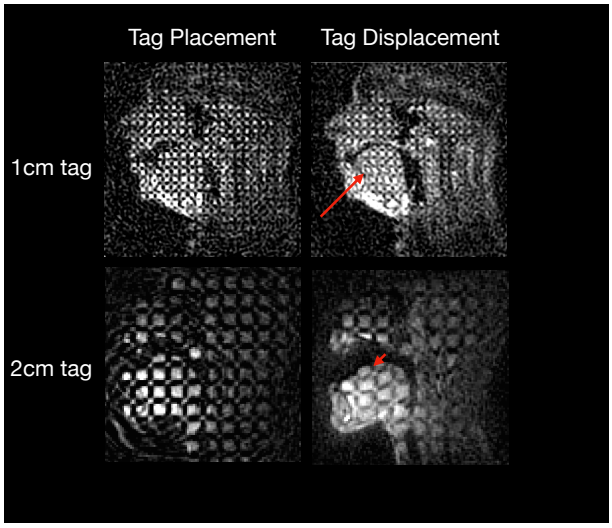

Figure 5: *Tagged RT-MRI reveals tongue deformation. Two participants are shown with 1cm and 2cm tag spacing. Tag lines appear as a perfect 2D grid at the time of placement (left) and are distorted during speech production (right). After 200 ms, we observe that tag lines have faded due to T1 recovery (red arrows). See also supplemental file 2D_tagging.mp4.*

Figure 3 presents real-time simultaneous multislice imaging of three parallel slices along with intensity line profiles. SMS images are able to resolve real-time kinematic features of the tongue tip. It was found that SMS image reconstruction required additional temporal and spatial regularization to remove aliasing in the image; this was not required in single-slice reconstructions, and introduced spatial blurring and an increased background noise.

Figure 4 shows an example of biofeedback in the scanner. In this example, the user’s scan is being reconstructed in real-time using online reconstruction and is displayed to the user via HDMI cable (see Figure 1). The overall latency of the image reconstruction was measured to be $<$ 100 ms, which is fast enough for there to be no noticeable delay. In this study, text is displayed to the user and a purple star is placed on the screen as a “target” to reach with their tongue tip.

Figure 5 shows tagging images using 1cm and 2cm of tag spacing. After placement of the tagging grid, image acquisition is done using the standard 2D spiral real-time acquisition. Because of T1 recovery, the tag lines disappear quickly (< 0.5 s), but tongue-internal motion indicative of strain, for example can still be discerned.

4. Discussion

We demonstrate a speech imaging protocol optimized for 0.55T. It includes 3D static, 2D real-time, and tagged 2D real-time imaging, as well as two new techniques in simultaneous multislice imaging and interactive biofeedback. RT-MRI at 0.55T presents a novel imaging contrast and signal-to-noise performance that matches 1.5T, while also having much sharper images due to reduced inhomogeneities at mid-field [13]. The introduction of SMS RT-MRI provides insights into vocal tract dynamics at planes other than mid-sagittal, while retaining high frame-rates to capture fast speech events. Furthermore, biofeedback with RT-MRI can play a role in speech training and remediation. For glossectomy patients, biofeedback improves substantially upon traditional therapy techniques by providing both the clinician and patient with information about vocal tract movements that are not otherwise observable.

There currently remain limitations with the present protocol. In SMS images, we observed additional blurring, which could be removed by including additional regularization in the constrained reconstruction. The short T1 recovery at 0.55T made the tagged imaging lines fade quickly. Further work may introduce ramping radiofrequency pulses, following previously published cardiac tagging [26], which will improve tag persistence. A limitation of the biofeedback imaging setup is that the stimulus preparation, scanning console, and image reconstruction all used the same computer. This made it difficult to operate simultaneously with one operator and introduced latency for the image reconstruction. This could be avoided by off-loading the image reconstruction to a separate server, using open-source data streaming [27]. In sum, ongoing development of RT-MRI at 0.55T holds enormous potential for improved quality and translational utility of real-time vocal tract imaging for both typically speaking and clinical populations.

Pulse sequences developed in the open source Pulseq format, as well as image reconstruction code for these experiments, are available at `https://github.com/usc-mrel/interspeech-rtmri`.

5. Acknowledgements

We acknowledge funding from the National Institutes of Health (U01-HL167613) and National Science Foundation (#1828736).

6. References

[1] E. Bresch, Y.-C. Kim, K. Nayak, D. Byrd, and S. Narayanan, "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging [Exploratory DSP]," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 123–132, 2008.

[2] A. D. Scott, M. Wylezinska, M. J. Birch, and M. E. Miquel, "Speech MRI: Morphology and function," *Physica Medica: European Journal of Medical Physics*, vol. 30, no. 6, pp. 604–618, 2014. [Online]. Available: https://www.physicamedica.com/article/S1120-1797(14)00081-7/abstract

[3] K. S. Nayak, Y. Lim, A. E. Campbell-Washburn, and J. Steeden, "Real-Time Magnetic Resonance Imaging," *Journal of Magnetic Resonance Imaging*, vol. 55, no. 1, pp. 81–99, 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.27411

[4] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep Magnetic Resonance Image Reconstruction: Inverse Problems Meet Neural Networks," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 141–151, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8962949/

[5] M. Fu, B. Zhao, C. Carignan, R. K. Shosted, J. L. Perry, D. P. Kuehn, Z.-P. Liang, and B. P. Sutton, "High-Resolution Dynamic Speech Imaging with Joint Low-Rank and Sparsity Constraints," *Magnetic Resonance in Medicine*, vol. 73, no. 5, pp. 1820–1832, 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4261062/

[6] W. Chen, N. G. Lee, D. Byrd, S. Narayanan, and K. S. Nayak, "Improved real-time tagged MRI using REALTAG," *Magnetic Resonance in Medicine*, vol. 84, no. 2, pp. 838–846, 2020. [Online]. Available: http://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.28144

[7] R. Jin, Y. Li, R. K. Shosted, F. Xing, I. Gilbert, J. L. Perry, J. Woo, Z.-P. Liang, and B. P. Sutton, "Optimization of 3D dynamic speech MRI: Poisson-disc undersampling and locally higher-rank reconstruction through partial separability model with regional optimized temporal basis," *Magnetic Resonance in Medicine*, vol. 91, no. 1, pp. 61–74, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.29812

[8] Y. Lim, A. Toutios, Y. Bliesener, Y. Tian, S. G. Lingala, C. Vaz, T. Sorensen, M. Oh, S. Harper, W. Chen, Y. Lee, J. Töger, M. L. Monteserin, C. Smith, B. Godinez, L. Goldstein, D. Byrd, K. S. Nayak, and S. S. Narayanan, "A multispeaker dataset of raw and reconstructed speech production real-time MRI video and 3D volumetric images," *Scientific Data*, vol. 8, no. 1, p. 187, 2021. [Online]. Available: https://doi.org/10.1038/s41597-021-00976-x

[9] K. Isaieva, Y. Laprie, J. Leclère, I. K. Douros, J. Felblinger, and P.-A. Vuissoz, "Multimodal dataset of real-time 2D and static 3D MRI of healthy French speakers," *Scientific Data*, vol. 8, no. 1, p. 258, 2021. [Online]. Available: https://www.nature.com/articles/s41597-021-01041-3

[10] M. Oh and Y. Lee, "ACT: An Automatic Centroid Tracking tool for analyzing vocal tract actions in real-time magnetic resonance imaging speech production data," *The Journal of the Acoustical Society of America*, vol. 144, no. 4, p. EL290, 2018.

[11] C. Hagedorn, J. Kim, U. Sinha, L. Goldstein, and S. S. Narayanan, "Complexity of vocal tract shaping in glossectomy patients and typical speakers: A principal component analysis," *The Journal of the Acoustical Society of America*, vol. 149, no. 6, pp. 4437–4449, 2021. [Online]. Available: https://asa.scitation.org/doi/10.1121/10.0004789

[12] C. Hagedorn, M. Proctor, L. Goldstein, S. M. Wilson, B. Miller, M. L. Gorno-Tempini, and S. S. Narayanan, "Characterizing Articulation in Apraxic Speech Using Real-Time Magnetic Resonance Imaging," *Journal of Speech, Language, and Hearing Research*, vol. 60, no. 4, pp. 877–891, 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5548083/

[13] Y. Lim, P. Kumar, and K. S. Nayak, "Speech production real-time MRI at 0.55 T," *Magnetic Resonance in Medicine*, vol. 91, no. 1, pp. 337–343, 2024.

[14] S. G. Lingala, B. P. Sutton, M. E. Miquel, and K. S. Nayak, "Recommendations for real-time speech MRI," *Journal of Magnetic Resonance Imaging*, vol. 43, no. 1, pp. 28–44, 2016.

[15] J. Fessler, S. Lee, V. Olafsson, H. Shi, and D. Noll, "Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3393–3402, 2005.

[16] M. C. Restivo, R. Ramasawmy, W. P. Bandettini, D. A. Herzka, and A. E. Campbell-Washburn, "Efficient spiral in-out and EPI balanced steady-state free precession cine imaging using a high-performance 0.55T MRI," *Magnetic Resonance in Medicine*, vol. 84, no. 5, pp. 2364–2375, 2020.

[17] A. E. Campbell-Washburn, J. Varghese, K. S. Nayak, R. Ramasawmy, and O. P. Simonetti, "Cardiac MRI at Low Field Strengths," *Journal of Magnetic Resonance Imaging*, vol. 59, no. 2, pp. 412–430, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.28890

[18] E. Yagiz, P. Garg, K. S. Nayak, and Y. Tian, "Simultaneous multi-slice real-time cardiac MRI at 0.55T," in *ISMRM 32nd Scientific Session*, 2023, p. 1600. [Online]. Available: https://cds.ismrm.org/protected/23MPresentations/abstracts/1600.html

[19] Y. Tian, S. X. Cui, Y. Lim, N. G. Lee, Z. Zhao, and K. S. Nayak, "Contrast-optimal simultaneous multi-slice bSSFP cine cardiac imaging at 0.55 T," *Magnetic Resonance in Medicine*, vol. 89, no. 2, pp. 746–755, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.29472

[20] S. Ponrartana, H. N. Nguyen, S. X. Cui, Y. Tian, P. Kumar, J. C. Wood, and K. S. Nayak, "Low-field 0.55 T MRI evaluation of the fetus," *Pediatric Radiology*, vol. 53, no. 7, pp. 1469–1475, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10276075/

[21] K. J. Layton, S. Kroboth, F. Jia, S. Littin, H. Yu, J. Leupold, J.-F. Nielsen, T. Stöcker, and M. Zaitsev, "Pulseq: A rapid and hardware-independent pulse sequence prototyping framework," *Magnetic Resonance in Medicine*, vol. 77, no. 4, pp. 1544–1552, 2017. [Online]. Available: http://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.26235

[22] J. M. Santos, G. A. Wright, and J. M. Pauly, "Flexible real-time magnetic resonance imaging framework," *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, vol. 26 II, pp. 1048–1051, 2004.

[23] F. Muñoz, Y. Lim, S. X. Cui, H. Stark, and K. S. Nayak, "Evaluation of a novel 8-channel RX coil for speech production MRI at 0.55 T," *MAGMA*, 2022.

[24] A. N. Price, L. Cordero-Grande, S. J. Malik, and J. V. Hajnal, "Simultaneous multislice imaging of the heart using multiband balanced SSFP with blipped-CAIPI," *Magnetic Resonance in Medicine*, vol. 83, no. 6, pp. 2185–2196, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.28086

[25] S. G. Lingala, Y. Zhu, Y. C. Kim, A. Toutios, S. Narayanan, and K. S. Nayak, "A fast and flexible MRI system for the study of dynamic vocal tract shaping," *Magnetic Resonance in Medicine*, vol. 77, no. 1, pp. 112–125, 2017. [Online]. Available: /pmc/articles/PMC4947574/?report=abstract

[26] E.-S. H. Ibrahim, M. Stuber, M. Schär, and N. F. Osman, "Improved myocardial tagging contrast in cine balanced SSFP images," *Journal of Magnetic Resonance Imaging*, vol. 24, no. 5, pp. 1159–1167, 2006. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.20730

[27] M. S. Hansen and T. S. Sørensen, "Gadgetron: An open source framework for medical image reconstruction," *Magnetic Resonance in Medicine*, vol. 69, no. 6, pp. 1768–1776, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.24389