Received: 8 May 2023 | Revised: 11 July 2023 | Accepted: 10 August 2023

DOI: 10.1002/mrm.29843

TECHNICAL NOTE

Magnetic Resonance in Medicine

Speech production real-time MRI at 0.55 T

Yongwan Lim | **Prakash Kumar** | **Krishna S. Nayak**

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

Correspondence
Prakash Kumar, Ming Hsieh Department of Electrical and Computer Engineering, Viterbi School of Engineering, University of Southern California, 3740 McClintock Ave, EEB 400, University of Southern California, Los Angeles, CA 90089, USA.
Email: prakashk@usc.edu

Funding information
National Science Foundation, Grant/Award Number: 1828736

Abstract

Purpose: To demonstrate speech-production real-time MRI (RT-MRI) using a contemporary 0.55T system, and to identify opportunities for improved performance compared with conventional field strengths.

Methods: Experiments were performed on healthy adult volunteers using a 0.55T MRI system with high-performance gradients and a custom 8-channel upper airway coil. Imaging was performed using spiral-based balanced **SSFP** and gradient-recalled echo (GRE) pulse sequences using a temporal finite-difference constrained reconstruction. Speech-production RT-MRI was performed with three spiral readout durations (8.90, 5.58, and 3.48 ms) to determine trade-offs with respect to articulator contrast, blurring, banding artifacts, and overall image quality.

Results: Both spiral GRE and bSSFP captured tongue boundary dynamics during rapid consonant-vowel syllables. Although bSSFP provided substantially higher SNR in all vocal tract articulators than GRE, it suffered from banding artifacts at TR > 10.9 ms. Spiral bSSFP with the shortest readout duration (3.48 ms, TR = 5.30 ms) had the best image quality, with a 1.54-times boost in SNR compared with an equivalent GRE sequence. Longer readout durations led to increased SNR efficiency and blurring in both bSSFP and GRE.

Conclusion: High-performance 0.55T MRI systems can be used for speech-production RT-MRI. Spiral bSSFP can be used without suffering from banding artifacts in vocal tract articulators, provide better SNR efficiency, and have better image quality than what is typically achieved at 1.5 T or 3 T.

KEYWORDS

0.55 T, real-time MRI, speech production, spiral, susceptibility artifact

1 | INTRODUCTION

Real-time MRI (RT-MRI) has deepened our understanding of the complex spatio-temporal coordination of vocal articulators during speech production.[1,2] Applications in speech science have included phonetic and phonological phenomena,[3–5] spoken language acquisition and speech disorders,[6–8] dynamics of vocal tract shaping during speech and vocal performance,[9] articulatory modeling,

A preliminary version of this work was presented at ISMRM 2022, Abstract #0755.

motor control, biometrics,[10–13] and speech synthesis and recognition technology.

Prior work has used RT-MRI at 1.5 T and 3 T, typically using spiral gradient-recalled echo (GRE) at 1.5 T[14–17] and 3 T,[18] or radial GRE at 3 T.[19,20] The main challenge is susceptibility at air–tissue interfaces, which increases linearly with field strength and can cause blurring and signal loss at speech articulator boundaries.[21,22] The off-resonance by air–tissue interfaces can theoretically be up to 9.41 parts per million, corresponding to about 600 Hz and about 1.2 kHz at 1.5 T and 3 T, respectively.[23] For this reason, current speech RT-MRI studies are most often conducted using very short readouts (2.5 ms) on 1.5T commercial MRI scanners.

Spiral acquisitions are used widely at 1.5 T due to high scan efficiency and resilience to motion artifacts; they have produced high spatial (1–3 mm) and temporal (12–40 ms) resolution when combined with advanced image reconstruction.[2] Lower-field imaging with high-performance gradients[24] offers the advantage of reduced off-resonance, enabling longer spiral readouts[24] with less blurring artifacts. It also makes balanced (bSSFP) feasible, which at higher field strengths is obscured by banding artifacts in the regions of interest. The latter is especially important because bSSFP provides superior SNR efficiency compared with traditional GRE.[25] Although earlier work[24,26] demonstrates the feasibility of long readout spiral GRE sequences for speech imaging at 0.55 T, we explore the feasibility of bSSFP and compare its performance against that of GRE.

In this work, we develop and demonstrate 2D midsagittal RT-MRI of speech production at 0.55 T[24] using spiral-based pulse sequences in conjunction with a custom-designed upper airway coil and constrained image reconstruction. The bSSFP and GRE imaging pulse sequences with varying readout durations are optimized based on the relaxation parameters of tongue muscle measured at 0.55 T on 2 healthy volunteers and based on analysis from Bloch simulations. We evaluate both bSSFP and GRE sequences in vivo and determine the trade-offs quantitatively and qualitatively with respect to articulator contrast, blurring and banding artifacts, SNR, and tongue boundary sharpness.

2 | METHODS

2.1 | Experimental methods

Experiments were performed using a whole-body 0.55T system (prototype MAGNETOM Aera; Siemens Healthineers, Erlangen, Germany) equipped with stock "Aera XQ" high-performance shielded gradients (45 mT/m amplitude, 200 T/m/s slew rate). We used a custom 8-channel upper airway coil[27] that has four elements on each side of the jaw. This provides SNR 1.71-fold to 2.35-fold superior to a 16-channel head/neck coil over the vocal tract articulators of interest including the lips, tongue, and velum.[27] The imaging protocol was approved by the institutional review board of the University of Southern California. Four healthy adult volunteers (1 female and 3 males; age 25–34 years) were scanned, after providing written informed consent.

All volunteers conducted two stimuli. The English stimulus "/loo/−/lee/−/la/−/za/−/na/−/za/" was read repeatedly at a fast speaking rate to produce alternating consonant and vowel sounds.[28] These consonant-vowel syllables involve constrictions of the tongue tip (e.g., [l], [z], [n]) and relatively extreme postures at the tongue body and back (e.g., "ee" [i], "ah" [a], "oo" [u]). The second stimulus was to count "1" through "5" in English at natural and fast speaking rates.

2.2 | Imaging methods

We measured T_1, T_2, and T_2^* in vivo in the tongue muscle at 0.55 T (1 female and 1 male; age 25–26 years) using protocols listed in Supporting Information Table S1, similar to those described by Campbell-Washburn et al.[24] With the relaxation parameters, we conducted a Bloch simulation for bSSFP and GRE pulse sequences. This allowed us to estimate the signal of the tongue muscle at 0.55 T and optimize the imaging parameters. Specifically, the steady-state signal of bSSFP and GRE were simulated as a function of TR, flip angle (FA), and off-resonance (**Δf**) and used to get in the best range of the imaging parameters. Experiments used a GRE FA of 15º and bSSFP FA of 35º, which are slightly higher than the theoretical optimum values, to provide robustness to RF transmit imperfections. Note that delivering a lower-than-optimal FA incurs a stronger SNR penalty compared with delivering a higher-than-optimal FA for both sequences.

Two-dimensional spiral-out bSSFP and GRE pulse sequences were implemented within the RTHawk real-time imaging platform (HeartVista, Inc., Los Altos, CA, USA).[29] Both pulse sequences used the same readout gradients: a uniform density spiral trajectory with the zeroth-order gradient moments nulled. We tested sequences with N-shot spiral trajectories ($N = 5, 8,$ and 13), with corresponding imaging parameters listed in Table 1. The rotation angle between consecutive spiral interleaves was determined by bit-reversed orders to improve temporal incoherence. For bSSFP, through-slice equilibration compensation[30,31] was used to mitigate eddy current–induced signal oscillations by inducing

TABLE 1 Imaging parameters for spiral gradient-recalled echo (GRE) and spiral balanced SSFP (bSSFP) speech production real-time MRI (RT-MRI) at 0.55 T. Common parameters for all acquisitions are FOV = 28 × 28 cm^2, in-plane resolution = 2.3 × 2.3 mm^2, matrix = 121 × 121, slice thickness = 6 mm, and sampling interval = 2 μs.

Sequence	Spiral GRE			Spiral bSSFP		
TR (ms)	11.59	8.29	6.16	10.95	7.40	5.30
Readout duration (ms)	8.90	5.58	3.48	8.90	5.58	3.48
Flip angle (°)	15°			35°		
No. of interleaves for Nyquist sampling	5	8	13	5	8	13
No. of interleaves per frame for reconstruction	1	2	2	1	2	2
Temporal resolution (ms/frame)	11.59	16.58	12.32	10.95	14.80	10.60

a dephasing of ±18° along the slice direction with unbalanced slice-selection rephasing gradients.

A standard sparse SENSE reconstruction with a temporal finite difference constraint[17] was implemented in *MATLAB* (MathWorks, Natick, MA, USA) and performed offline. The gradient impulse response function (GIRF)[32,33] was estimated from a phantom-based measurement[30] and used to correct B_0 and linear eddy current artifacts for spiral trajectories before reconstruction.

2.3 | Analysis methods

For SNR comparisons, a fully sampled (via viewsharing) linear gridding reconstruction of the spiral data with a Kaiser-Bessel kernel function (kernel size = 3, oversampling ratio = 2) was used, and SNR of the tongue was compared among all acquisitions as well as a comparable 1.5T GRE acquisition downloaded from a publicly available data set.[34] SNR was measured using two separate regions of interest and accounting for scaling factors.[35,36]

Image sharpness along the tongue boundaries was computed by a multistep process including (1) finding the air–tissue boundary using an established segmentation method[37]; (2) fitting a sigmoid error function to intensity data along the grid lines perpendicular to the air–tissue boundary locations; and 3) calculating the slope of the fitted sigmoid function. Sharpness scores were then computed by solving for the ratio between the mean slope of the sigmoid function and the noise SD of the background region of the reconstructions. Measurements are reported using mean and SD across scans from all 4 subjects.

3 | RESULTS

Relaxation times of vocal tract articulators at 0.55 T were estimated as ranges from 2 subjects with $T_1 = 650 \sim 662$ ms, $T_2 = 42 \sim 45$ ms, and $T_2^* = 31$ ms. Simulations of steady-state tongue muscle signal at 0.55 T are shown in Figure 1A, where bSSFP provides a relatively consistent signal profile and optimal FA of about 29° for simulated TR values of 6 to 12 ms. At optimal flip angles, bSSFP has a signal gain of 47% over GRE at TR = 6 ms and 25% at TR = 12 ms for the tongue muscle. In Figure 1B, bSSFP has a passband with a relatively uniform signal for a broad range of resonance offsets (Δf). At 0.55 T, a TR ≤ 4.54 ms is required for artifact-free imaging near air–tissue boundaries (Δf = 220 Hz), which is substantially longer than the requirement at 1.5 T (TR ≤ 1.67 ms for Δf = 600 Hz).

Figure 2 and Supporting Information Video S1 show reconstructed images from 1 representative subject acquired using bSSFP and GRE with varying readout durations. Although longer GRE readouts provide higher tongue SNR, they suffer from increased blurring (green arrows). Shorter readouts have lower but adequate tongue SNR and sharper visualization of dynamic articulator boundaries.

For bSSFP, one can readily observe the same blurring increase with readout length as in GRE; however, the more dominant artifact is banding due to off-resonance. Longer readouts result in a longer TR and a narrower bSSFP passband, causing obvious banding artifacts (yellow arrows). In extreme cases of long TR, this can affect visualization of some articulator boundaries (red arrows).

Figure 3 and Supporting Information Video S2 show images and intensity versus time plots from 1 representative subject acquired using GRE and bSSFP. These were reconstructed using 2 interleaves out of 13 ($R = 6.5$) per frame at temporal resolutions of 11.4 ms and 10.1 ms, respectively. Notice the short-duration (50–60 ms) constrictions between the tongue tip and the alveolar ridge illustrated by red arrows that span only for 5–6 frames. Both sequences were able to capture the rapid tongue motion, which indicates that the effective temporal resolution is on the order of 10 ms. bSSFP exhibited a slightly clearer delineation of the tongue boundary than

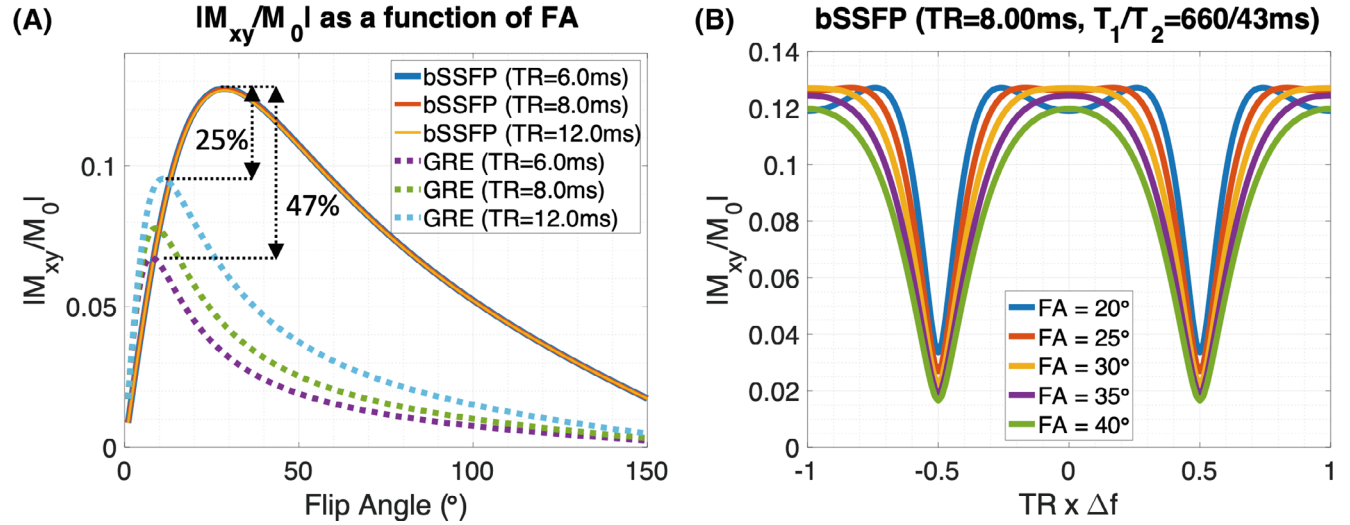

FIGURE 1 Simulation results of steady-state balanced SSFP (bSSFP) and gradient-recalled echo (GRE) signal from human tongue muscle at 0.55 T ($T_1/T_2 = 660/43$ ms) as a function of flip angle (A) and dephasing (TR × Δf) (B). (A) bSSFP offers a consistent signal intensity profile (and optimal flip angle [FA] of ~29°) independent of TR simulated (6–12 ms) and offers the highest possible SNR efficiency. GRE provides a signal intensity profile that increases with TR but has lower peak SNR efficiency when compared with bSSFP. (B) bSSFP signal exhibits a dependence on resonance offset and TR (TR × Δf), resulting in well-known banding artifacts.

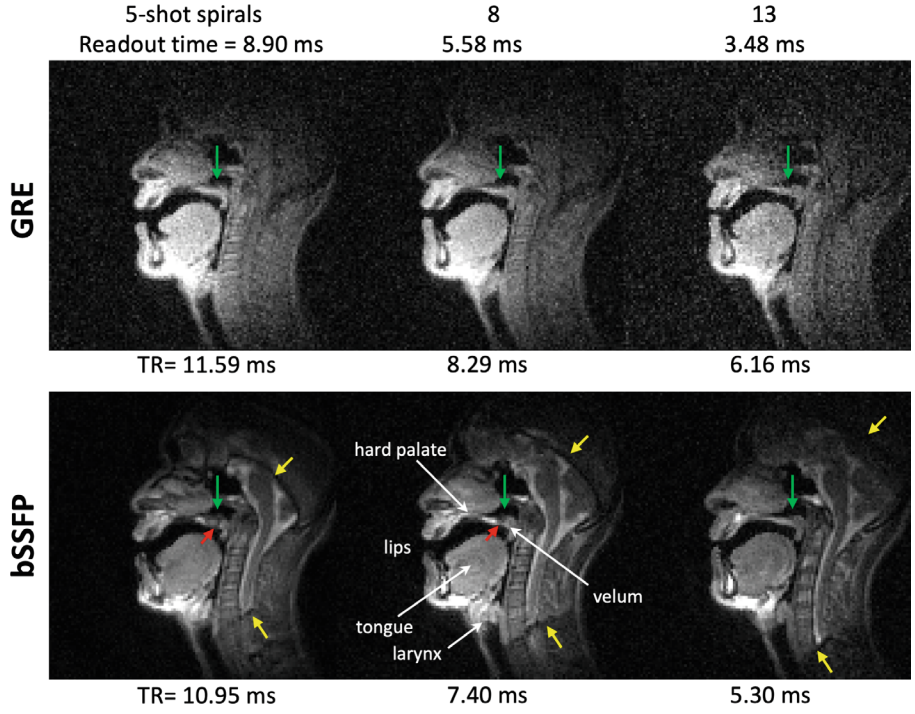

FIGURE 2 Comparison of bSSFP and GRE for 5-shot, 8-shot, and 13-shot spirals. Both GRE and bSSFP images exhibit sharper tissue boundaries when using shorter readouts due to reduced off-resonance blurring (green arrows). GRE with longer readouts shows higher image quality likely due to SNR, which is expected based on simulations in Figure 1A. bSSFP images offer substantially better image quality and tissue contrast compared with GRE. However, bSSFP with a longer readout and TR introduces characteristic banding artifact (yellow arrows) that can be especially severe around moving air–tissue boundaries such as the one across the velum (red arrows). The video for this example is provided as Supporting Information Video S1.

GRE. This is likely due to the better SNR achieved by bSSFP.

Tongue SNR from linear reconstructions (reported using mean ± SD across subjects) was 7.67 ± 1.17 for the shortest TR-GRE acquisitions, compared with 12.22 ± 1.21 for the 1.5T speech open data set.[34] As we increase the GRE TR, the SNR improves to 12.06 ± 1.57, albeit with undesirable blurring due to off-resonance. In bSSFP images, SNR was 11.84 ± 2.1 using the shortest TR and 13.49 ± 1.64 using the longest TR, with the latter matching the SNR of the tongue achieved at 1.5 T.[34] Tongue boundary sharpness scores were 12.03 ± 1.03 for the shortest TR bSSFP scan, and 6.82 ± 0.61 for the shortest TR-GRE scan. For comparison, the tongue boundary sharpness from a representative high-quality 1.5T sample of the speech open data set[34] is 8.5.

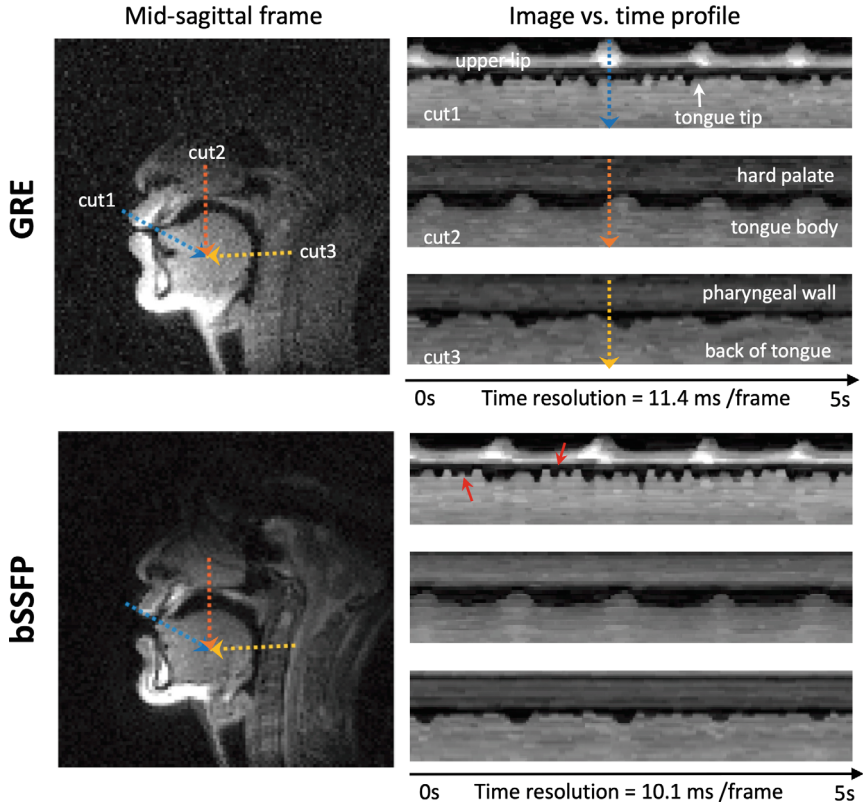

FIGURE 3 Comparison of balanced SSFP (bSSFP) and gradient-recalled echo (GRE) images and intensity versus time plots for a stimulus "loo-lee-laa-za-na-za" spoken at a fast rate. Images were acquired using 13-shot spiral GRE and bSSFP sequences, and each frame was reconstructed using the constrained reconstruction from two spirals per frame ($R = 6.5$). Three intensity profiles are placed at the tongue boundaries: the tongue tip, body, and back of the tongue (blue, orange, and yellow dashed lines, respectively). bSSFP images exhibit sharp articulator boundary definition and the ability to resolve rapid constrictions (red arrows). The video for this example is provided as Supporting Information Video S2.

4 | DISCUSSION AND CONCLUSION

We have demonstrated 2D midsagittal RT-MRI of speech production at 0.55 T using spiral-based bSSFP and GRE pulse sequences along with a custom upper airway coil. We characterize image quality and artifact tradeoff across varying readout durations for both bSSFP and GRE sequences. Initial studies on 4 healthy volunteers show that short-TR bSSFP with 10.1-ms temporal resolution can offer superior SNR and tissue contrast compared with GRE, with quality deemed to be superior to that commonly reported at 1.5T and 3T field strengths. We attribute this improvement to the significantly reduced off-resonance from air–tissue boundaries, supported by the 1.6-times improvement in tongue sharpness.

At low field using a GRE contrast, the expected SNR loss in the tongue is about 0.5 times when combining factors due to the polarization scaling linearly with B_0 (0.55 T vs. 1.5 T) and the shorter T_1 at 0.55 T (700 ms vs. 950 ms). Our measured loss of 0.62 times (7.67 vs. 12.22) matched our expectations. For bSSFP imaging, simulations predicted a 1.88-times improvement in SNR compared with GRE for the shortest TRs, but we observed a 1.57-times improvement (12.06 vs. 7.67). This is likely due to signal loss from on-resonance magnetization transfer, which can cause a 30%–50% loss in short-TR bSSFP imaging.[38] With bSSFP imaging, we achieved comparable SNR in the tongue to that of 1.5 T (0.96 times for short TR, 1.12 times for long TR). The shortest TR is optimal, as it is banding-free and still has adequate signal, unlike longer TR bSSFP, which produces undesirable banding artifacts.

The use of a dedicated multiple channel receiver coil array is beneficial over standard head coils because it provides an SNR boost due to higher loading, as the coils can be placed closer to the articulators as well as lower noise due to highly localized elements. This has been demonstrated at 3 T,[39] 1.5T,[17] and recently at 0.55 T.[27] The SNR boost is extremely important at the low field because SNR is lower than at higher field strengths. Also, the highly localized elements can offer opportunities for reduced encoding via a smaller FOV. It may be possible to replicate some of the reduced encoding effects using a modern beam-forming coil combination approach[40] as well.

Shimming is crucial for speech RT-MRI, especially for bSSFP methods, at any field strength. The bSSFP acquisition requires spins to fall within $-1/2TR < \Delta f < 1/2TR$ even during dynamic movements, as the field map is dynamic due to the constantly changing air cavity. At the articulator boundary, Δf can theoretically go up to 220 Hz at 0.55 T, requiring TR $\leq$ 4.54 ms. Although meeting this banding-free condition is challenging, we have found it feasible with proper shimming to image with TR > 4.54 ms and still have minimal banding artifacts over the vocal articulators. We recommend conducting linear shimming with a local shimming box placed over the mouth while in

open postures (as opposed to a neutral closed position) to ensure a good banding profile across the FOV. Afterward, we adjust the shimming profile, including B_0 manually, to minimize signal loss due to off-resonance in the velum while keeping the characteristic banding curve (Figure 2 yellow arrows) outside of the upper airway region.

bSSFP is sensitive to B_0 drift, which is known to happen in some settings of gradient heating.[41] Sequences like spiral bSSFP make aggressive use of gradients, which lead to some gradient distortions that are mitigated in this work using GIRF-corrected trajectories. They also lead to gradient heating, which leads to dynamic changes in the GIRF, without any significant negative effect on spiral image quality.[42] This often leads to B_0 drift, which can require periodic center frequency adjustment during long acquisitions. On our system, the maximum rate of B_0 drift observed has been 5 Hz per minute.

Eddy currents are an important consideration in bSSFP and can lead to unbalanced phase that perturbs the steady state. In this work, partial dephasing was used, which is a simple solution and slightly reduces the width of the bSSFP spectral passband. GIRF-based precompensation[30] is another promising option, which has added complexity but maintains the spectral bandwidth.

ACKNOWLEDGMENTS

We acknowledge grant support from the National Science Foundation (#1828736) and research support from Siemens Healthineers. We also thank Helmut Stark for developing the 0.55T upper-airway RF coil, and Felix Munoz for evaluating the coil. We thank Dani Byrd for help with linguistic qualitative analysis of reconstructed images. We also thank the Speech Production and Articulation kNowledge (SPAN) group at the University of Southern California for collaboration and useful discussions.

ORCID

Yongwan Lim https://orcid.org/0000-0003-0070-0034
Prakash Kumar https://orcid.org/0000-0002-0685-5160
Krishna S. Nayak https://orcid.org/0000-0001-5735-3550

REFERENCES

1. Nayak KS, Lim Y, Campbell-Washburn AE, Steeden J. Real-time magnetic resonance imaging. *J Magn Reson Imaging*. 2022;55:81-99.
2. Lingala SG, Sutton BP, Miquel ME, Nayak KS. Recommendations for real-time speech MRI. *J Magn Reson Imaging*. 2016;43:28-44.
3. Monteserín ML, Narayanan S, Goldstein L. Perceptual lateralization of coda rhotic production in Puerto Rican Spanish. *Proceedings of the Annual Conference of INTERSPEECH*. International Speech Communication Association (ISCA); 2016:2443-2447.
4. Harper S, Goldstein L, Narayanan S. L2 acquisition and production of the English rhotic pharyngeal gesture. *Proceedings of the Annual Conference of INTERSPEECH*. International Speech Communication Association (ISCA); 2016:208-212.
5. Harper S, Goldstein L, Narayanan S. Variability in individual constriction contributions to third formant values in American English /ɹ/. *J Acoust Soc Am*. 2020;147:3905-3916.
6. Zu Y, Narayanan SS, Kim Y-C, et al. Evaluation of swallow function after tongue cancer treatment using real-time magnetic resonance imaging: a pilot study. *JAMA Otolaryngol Head Neck Surg*. 2013;139:1312-1319.
7. Stone M, Langguth JM, Woo J, Chen H, Prince JL. Tongue motion patterns in post-glossectomy and typical speakers: a principal components analysis. *J Speech Lang Hear Res*. 2014;57:707-717.
8. Beer AJ, Hellerhoff P, Zimmermann A, et al. Dynamic near-real-time magnetic resonance imaging for analyzing the velopharyngeal closure in comparison with videofluoroscopy. *J Magn Reson Imaging*. 2004;20:791-797.
9. Lander-Portnoy M, Goldstein L, Narayanan SS. Using real time magnetic resonance imaging to measure changes in articulatory behavior due to partial glossectomy. *J Acoust Soc Am*. 2017;142:2641-2642.
10. Töger J, Sorensen T, Somandepalli K, et al. Test–retest repeatability of human speech biomarkers from static and real-time dynamic magnetic resonance imaging. *J Acoust Soc Am*. 2017;141:3323-3336.
11. Sorensen T, Toutios A, Goldstein L, Narayanan S. Task-dependence of articulator synergies. *J Acoust Soc Am*. 2019;145:1504-1520.
12. Toutios A, Sorensen T, Somandepalli K, Alexander R, Narayanan SS. *Articulatory synthesis based on real-time magnetic resonance imaging data*. In: *Proceedings of the Annual Conference of INTERSPEECH*; 2016:1492-1496.
13. Lammert AC, Shadle CH, Narayanan SS, Quatieri TF. Speed-accuracy tradeoffs in human speech production. *PloS One*. 2018;13:e0202180.
14. 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;80:1467-1474.
15. Lingala SG, Zhu Y, Lim Y, 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;78:2275-2282.
16. Narayanan SS, Nayak KS, Lee S, Sethy A, Byrd D. An approach to real-time magnetic resonance imaging for speech production. *J Acoust Soc Am*. 2004;115:1771-1776.
17. Lingala SG, Zhu Y, Kim YC, Toutios A, Narayanan SS, Nayak KS. A fast and flexible MRI system for the study of dynamic vocal tract shaping. *Magn Reson Med*. 2017;77:112-125.
18. Sutton BP, Conway CA, Bae Y, Seethamraju R, Kuehn DP. Faster dynamic imaging of speech with field inhomogeneity corrected spiral fast low angle shot (FLASH) at 3 T. *J Magn Reson Imaging*. 2010;32:1228-1237.
19. Niebergall A, Zhang S, Kunay E, et al. Real-time MRI of speaking at a resolution of 33 ms: undersampled radial FLASH with nonlinear inverse reconstruction. *Magn Reson Med*. 2013;69:477-485.
20. 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.
21. Lim Y, Lingala SG, Narayanan SS, Nayak KS. Dynamic off-resonance correction for spiral real-time MRI of speech. *Magn Reson Med.* 2019;81:234-246.
22. Lim Y, Bliesener Y, Narayanan SS, Nayak KS. Deblurring for spiral real-time MRI using convolutional neural network. *Magn Reson Med.* 2020;84:3438-3452.
23. Schenck JF. The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds. *Med Phys.* 1996;23:815-850.
24. Campbell-Washburn AE, Ramasawmy R, Restivo MC, et al. Opportunities in interventional and diagnostic imaging by using high-performance low-field-strength MRI. *Radiology.* 2019;293:384-393.
25. Restivo MC, Ramasawmy R, Bandettini WP, Herzka DA, Campbell-Washburn AE. Efficient spiral in-out and EPI balanced steady-state free precession cine imaging using a high-performance 0.55T MRI. *Magn Reson Med.* 2020;84:2364-2375.
26. Bhattacharya I, Ramasawmy R, Restivo MC, Campbell-Washburn AE. Dynamic speech imaging at 0.55T using single shot spirals for 11ms temporal resolution. *Proceedings of the 27th Annual Meeting of ISMRM*. Montreal; 2019:440.
27. Muñoz F, Lim Y, Cui SX, Stark H, Nayak KS. Evaluation of a novel 8-channel RX coil for speech production MRI at 0.55 T. *MAGMA*. 2022;36:419-426. doi:10.1007/s10334-022-01036-0
28. Fu M, Zhao B, Carignan C, et al. High-resolution dynamic speech imaging with joint low-rank and sparsity constraints. *Magn Reson Med.* 2015;73:1820-1832.
29. Santos JM, Wright GA, Pauly JM. Flexible real-time magnetic resonance imaging framework. *Conf Proc IEEE Eng Med Biol Soc*. 2004;2:1048-1051.
30. Bieri O, Markl M, Scheffler K. Analysis and compensation of eddy currents in balanced SSFP. *Magn Reson Med.* 2005;54:129-137.
31. Bruijnen T, Stemkens B, van den Berg CAT, Tijssen RHN. Prospective GIRF-based RF phase cycling to reduce eddy current-induced steady-state disruption in bSSFP imaging. *Magn Reson Med.* 2020;84:115-127.
32. Robison RK, Li Z, Wang D, Ooi MB, Pipe JG. Correction of B_0 eddy current effects in spiral MRI. *Magn Reson Med.* 2019;81:2501-2513.
33. Campbell-Washburn AE, Xue H, Lederman RJ, Faranesh AZ, Hansen MS. Real-time distortion correction of spiral and echo planar images using the gradient system impulse response function. *Magn Reson Med.* 2016;75:2278-2285.
34. Lim Y, Toutios A, Bliesener Y, et al. A multispeaker dataset of raw and reconstructed speech production real-time MRI video and 3D volumetric images. *Sci Data.* 2021;8:187.
35. Constantinides CD, Atalar E, McVeigh ER. Signal-to-noise measurements in magnitude images from NMR phased arrays. *Magn Reson Med.* 1997;38:852-857.
36. Dietrich O, Raya JG, Reeder SB, Reiser MF, Schoenberg SO. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. *Magn Reson Med.* 2007;26:375-385.
37. Kim J, Kumar N, Lee S, Narayanan S. *Enhanced airway-tissue boundary segmentation for real-time magnetic resonance imaging data*. In: *Proceedings of the International Seminar on Speech Production*, 2014:222-225.
38. Bieri O, Scheffler K. On the origin of apparent low tissue signals in balanced SSFP. *Magn Reson Med*. 2006;56:1067-1074.
39. Kim YC, Hayes CE, Narayanan SS, Nayak KS. Novel 16-channel receive coil array for accelerated upper airway MRI at 3 tesla. *Magn Reson Med.* 2011;65:1711-1717.
40. Kim D, Cauley SF, Nayak KS, Leahy RM, Haldar JP. Region-optimized virtual (ROVir) coils: localization and/or suppression of spatial regions using sensor-domain beamforming. *Magn Reson Med.* 2021;86:197-212.
41. Stich M, Pfaff C, Wech T, et al. The temperature dependence of gradient system response characteristics. *Magn Reson Med.* 2020;83:1519-1527.
42. Nussbaum J, Dietrich BE, Wilm BJ, Pruessmann KP. Thermal variation in gradient response: measurement and modeling. *Magn Reson Med.* 2022;87:2224-2238.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Table S1. Acquisition parameters for T_1, T_2, and T_2^* measurement at 0.55 T.

Video S1. Comparison of balanced SSFP (bSSFP) and gradient-recalled echo (GRE) for 5-, 8-, and 13-shot spirals. Both GRE and bSSFP images exhibit sharper tissue boundaries when using shorter readouts due to reduced off-resonance blurring (green arrows). GRE with longer readouts shows higher image quality likely due to SNR, which is expected based on simulations in Figure 2A. The bSSFP images offer substantially better image quality and tissue contrast compared with GRE. However, bSSFP with a longer readout and TR introduces characteristic banding artifact (yellow arrows) that can be especially severe around moving air–tissue boundaries such as the one across the velum (red arrows).

Video S2. Comparison of balanced SSFP (bSSFP) and gradient-recalled echo (GRE) images and intensity versus time plots for a stimulus "loo-lee-laa-za-na-za" spoken at a speeded rate. Images were acquired using 13-shot-spiral GRE and bSSFP sequences, and each frame was reconstructed using the constrained reconstruction from two spirals per frame ($R = 6.5$). Three intensity profiles are placed at the tongue boundaries—the tongue tip, body, and back of the tongue (blue, orange, and yellow dashed lines, respectively).

How to cite this article: Lim Y, Kumar P, Nayak KS. Speech production real-time MRI at 0.55 T. *Magn Reson Med*. 2024;91:337-343. doi: 10.1002/mrm.29843