Received: 16 July 2019 | Revised: 13 November 2019 | Accepted: 2 December 2019

DOI: 10.1002/mrm.28144

NOTE

Magnetic Resonance in Medicine

Improved real-time tagged MRI using REALTAG

Weiyi Chen[1] | **Nam Gyun Lee**[2] | **Dani Byrd**[3] | **Shrikanth Narayanan**[1,3] | **Krishna S. Nayak**[1,2]

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

[2]Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, California

[3]Department of Linguistics, Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, California

Correspondence
Weiyi Chen, 3740 McClintock Avenue, EEB 400, University of Southern California, Los Angeles, CA 90089-2564.
Email: weiyic@usc.edu

Funding information
This work was supported by NIH Grant R01DC007124 and NSF Grant 1514544

Objectives: To evaluate a novel method for real-time tagged MRI with increased tag persistence using phase sensitive tagging (REALTAG), demonstrated for speech imaging.

Methods: Tagging is applied as a brief interruption to a continuous real-time spiral acquisition. REALTAG is implemented using a total tagging flip angle of 180° and a novel frame-by-frame phase sensitive reconstruction to remove smooth background phase while preserving the sign of the tag lines. Tag contrast-to-noise ratio of REALTAG and conventional tagging (total flip angle of 90°) is simulated and evaluated in vivo. The ability to extend tag persistence is tested during the production of vowel-to-vowel transitions by American English speakers.

Results: REALTAG resulted in a doubling of contrast-to-noise ratio at each time point and increased tag persistence by more than 1.9-fold. The tag persistence was 1150 ms with contrast-to-noise ratio >6 at 1.5T, providing 2 mm in-plane resolution, 179 frames/s, with 72.6 ms temporal window width, and phase sensitive reconstruction. The new imaging window is able to capture internal tongue deformation over word-to-word transitions in natural speech production.

Conclusion: Tag persistence is substantially increased in intermittently tagged real-time MRI by using the improved REALTAG method. This makes it possible to capture longer motion patterns in the tongue, such as cross-word vowel-to-vowel transitions, and provides a powerful new window to study tongue biomechanics.

KEYWORDS
real-time MRI, speech production, tag persistence, tagged MRI, tongue deformation

1 | INTRODUCTION

MR tagging is an established technique for measuring regional muscle function, such as in myocardium[1] and skeletal muscle.[2-4] It is performed using a preparation pulse that spatially tags tissue by saturating or inverting parallel strips or orthogonal grid lines within the imaging plane. The deformation of these tag lines can be measured using dynamic imaging and then used to evaluate the properties of the imaged object, such as myocardial motion,[1] limb maneuver,[2] eye-ball rotation,[3] brain deformation,[5] and tongue movement.[6] The measured deformations can be further used to quantify muscle mechanics, such as strain[7] and torsion[8] in the heart, and to develop atlases of motion within the tongue.[9]

MR tagging typically has been combined with CINE imaging, which requires multiple repetitions. For example, 144 voluntary head rotations for studying traumatic brain injury in Sabet et al,[5] and more than 135 repeated left-to-right eye movements for extraocular muscle motion in Piccirelli et al.[3] Notably, CINE tagged MRI has been successfully applied to image tongue muscle deformation during speech production.[4,10] However, multiple repetitions were required; note, e.g., 16 speech utterances for each slice in Parthasarathy et al.[10] The CINE process lengthens scan time and is best suited to highly repeatable motions such as cardiac contraction. Robust deployment of tagged CINE MRI to noncardiac applications has been hindered by variability in motion, such as normal and natural token and type variability during speech production[11] or voluntary effort discrepancy during body movement.[2]

An alternative to tagged CINE strategies is to use real-time (RT) MRI approaches that do not require repetition or synchronization. Tagged RT MRI has been proposed to reduce scan time for cardiac application[12-15] and, recently, to eliminate the need for repetitions in imaging speech production.[16] These techniques are able to provide adequate spatiotemporal resolution, and tag persistence has been the primary limitation. For example, tag persistence was demonstrated to be around 650 ms at 1.5T for the tongue.[16] This duration is sufficient for imaging the production of single syllables in American English. However, this is not sufficient for longer utterances in which the tongue motions of interest may be on the order of 1 s, such as vowel-to-vowel transition across words or geminate (i.e., phonologically long) articulations. Tongue deformation patterns occurring in the formation and release of lingual constrictions over these longer speech intervals are not well understood due to limitations of other articulometry methods, which have focused on point-tracking or tongue-/lip-surface imaging, yet the biomechanical underpinnings of internal tongue deformations remains an important aspect of characterizing speech production and speech motor control, particularly for the complex lingual hydrostat.[17]

Several methods can potentially extend the duration of tag persistence. For spoiled gradient sequence (SPGR), fewer excitation pulses per image can improve contrast,[18] but this will increase the readout time and will be limited by the substantial off-resonance in air-tissue boundaries.[19] A balanced steady-state free precession (bSSFP) sequence yields improvement for contrast-to-noise ratio (CNR) and tag persistence by improving tissue signal-to-noise ratio.[20] However, large off-resonance can introduce banding artifacts. Complementary SPAtial Modulation of Magnetization (CSPAMM) uses 2 consecutive scans with opposite tagging radiofrequency polarity and ramped imaging flip angles to cancel out the tag fading.[21] This technique requires 2 repetitions and, therefore, requires gating. REALTAG, proposed by Derbyshire et al,[22] uses a total flip angle (TFA) of 180° for the tagging pulse and phase-sensitive reconstruction to correct the rectified inverted tags. This method prolongs tag persistence by roughly a factor of 2 without increasing scan time, and it is compatible with tagged RT-MRI approaches. However, REALTAG uses a fixed linear fitting to the phase within an automatically selected static region of interest (ROI),[22] which is not suitable for speech imaging.

In this work, we apply a novel REALTAG approach to intermittently tagged RT-MRI and demonstrate its successful application to the study of human speech production. We use a spatial low-pass filter to extract and compensate for smooth image phase and isolate tag lines. We evaluate this method using both simulations and in vivo studies of American English vowel-to-vowel transitions. In experiments focused on the human tongue, the proposed method extended tag persistence by a factor of 1.9×.

2 | METHODS

2.1 | Imaging methods

A combinatorial 1-3-3-1 SPAMM sequence was used with a TFA of 180° (proposed) and a TFA of 90° (original) implementation.[16] The tagging pulses were intermittently triggered to place a 2D grid of tag lines with 1-cm spacing during the continuous SPGR imaging. The imaging parameters were: 13 spiral interleaves, field of view 20 cm, slice thickness 7 mm, readout duration 2.49 ms, echo time/repetition time 0.71 ms/5.58 ms, and in-plane resolution isotropic 2 mm.

Gradient delay was accounted for during prescan. First-order shimming was interactively performed in a carefully selected ROI, including both the oral tongue-air interface and the pharyngeal tongue root, while the subject was instructed to sustain a relaxed open mouth posture. In practice, we found this shimming procedure to be crucial to reduce off-resonance at the tongue surface boundary and to effectively enhance the performance of phase sensitive reconstruction. Concomitant field correction[23] and image unwarping that takes into account gradient nonuniformity[24] were applied with gridding reconstruction. Gridding reconstruction with a sliding 13-interleaf window of 1 repetition time was performed coil-by-coil, providing 179 frames/s, with 72.6 ms temporal window width. ESPIRiT[25] was used to estimate coil sensitivity during prescan for coil combining.

Figure 1 illustrates the phase sensitive reconstruction pipeline. After gridding and coil combining, we selected a k-space center and implemented a low pass filter by zero padding. We denote the tag spacing as $\Delta_{tag} = \alpha\delta$, where δ is

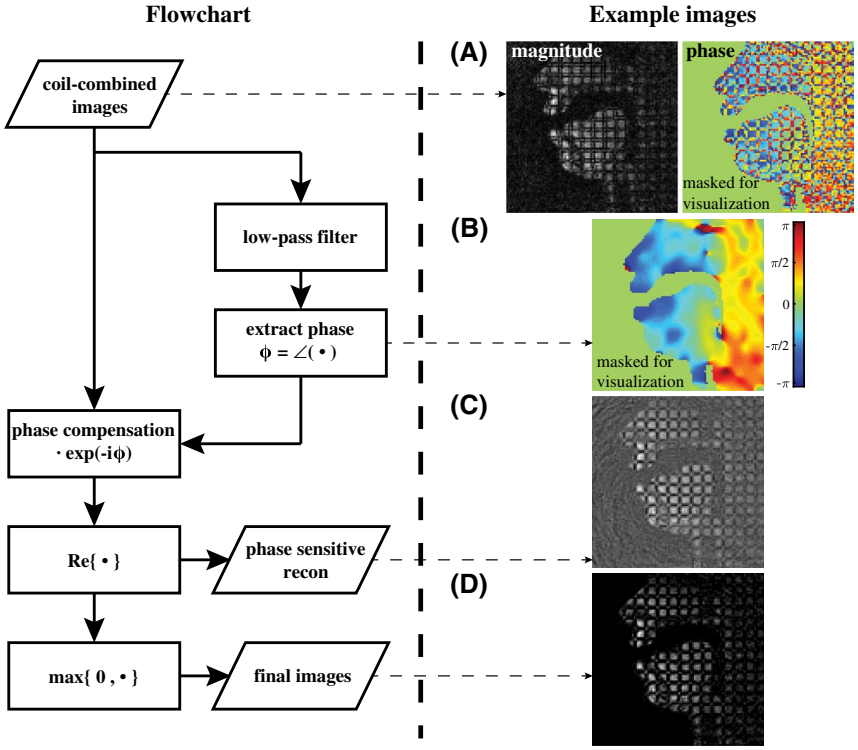

FIGURE 1 Phase sensitive reconstruction flowchart (left) and example images from intermediate steps (right). Low-resolution and tag-line-free phase (B) was estimated for each time frame using a synthesized k-space center after gridding and coil combining (A). The lowest tag harmonic frequencies are located at $\pm 2k_{max}/\alpha$, where α is the scaling factor between the tag spacing and the image resolution. The width of the low pass filter $W = 2k_{max}/\alpha$ was set as one-half of this frequency. Example image (B) used $W = 20$ for a 100×100 k-space with 2 mm image resolution and 1 cm tag spacing. Final images (D) were generated by taking the nonnegative values from phase sensitive reconstruction (C) for better visualization. Note the bright spots existing in the intersection of tag lines in (D) due to double inversion

the image resolution, and α is a scaling factor. The tag pattern can be written as a finite cosine series having a fundamental frequency of $1/\Delta_{tag}$.[26] In k-space, modulating by this frequency results in replicating the image k-space at certain harmonic frequencies.[27] These harmonic frequencies are located at

$$k_{tag} = N/\Delta_{tag} = 2Nk_{max}/\alpha \quad (1)$$

with $N = \pm 1, \pm 2, \ldots$, and $2k_{max}$ is the k-space extent (see Supporting Information Figure S1, which is available online). The width of the low pass filter was chosen in a way such that it only passes the k-space within one-half of the lowest harmonic frequency, that is

$$W = 2k_{max}/\alpha. \quad (2)$$

Low-resolution (LR) and tag-line-free phase was estimated for each frame using the filtered k-space center. The LR phase map was used to compensate for the background image excluding tag lines (see Supporting Information Figures S2 and S3). We enforced nonnegative value in the final images to perceptually increase contrast between dark tag lines and tongue tissue for visualization purposes.

2.2 | Experiments

Experiments were performed on a Signa Excite HD 1.5T scanner with a custom 8-channel upper-airway coil.[28] The pulse sequence was implemented using RTHawk Research v2.3.4 (HeartVista, Inc., Los Altos, CA) custom RT imaging platform.[29] We used a previously described tagged RT-MRI technique with an imaging flip angle of 5°.[16]

The experiment protocol was approved by our Institutional Review Board, and informed consent was obtained from all volunteers.

2.2.1 | Tag persistence

Tag persistence with TFA = 90° (original) and 180° (proposed) was simulated and compared with in vivo experiments. Five healthy volunteers (2 males, 3 females; age, 28-36 years) were scanned to verify tag persistence in the tongue. In this substudy, subjects were instructed to keep their mouth in a closed neutral position and remain still during the scan to minimize off-resonance and motion artifacts.

1D tag with 1-cm spacing was used. A separate scan was used to measure the steady state signal to properly scale

TABLE 1 Vowel-to-vowel transition stimuli[a]

Stimuli	Instructed pronunciation	Target starting posture	Target ending posture	Measured transition duration*
"A pie again"	/paɪ/	Low back	High front	175 ± 25 ms
"A poppy again"	/papɪ/	Low back	High front	232 ± 15 ms
"A pop pip again"	/pap pɪp/	Low back	High front	328 ± 22 ms

[a]Underlines mark target words (column 1) or target vowel sounds (column 2).

*The transition duration is measured as mean ± SD in 10 trials for 2 speakers during speech production for the time from the initiation of the first vowel's constriction to the end of release of the second vowel's constriction.

between simulation and measurements. The coil noise covariance matrix was measured by a separate scan with excitation radiofrequencies turned off to prewhiten the multicoil data and to normalize the result.[30]

CNR_{tag} is defined as the ratio between tag line peak-to-valley difference in the image and standard deviation of the image noise. The peak and valley values were calculated by averaging over the manually selected ROIs. The peak ROI was drawn in two 4 × 6-pixel squares in the bright regions in the tongue; the valley ROI was drawn in three 1 × 16-pixel strips at the center of the dark tag lines.

2.2.2 | Speech production experiments

Two volunteers (1 male, 1 female; age, 28 years), both native American English speakers, were scanned in the second sub-study. The text for the speech stimuli were projected onto a screen visible to the subject in the scanner through a mirror. Audio recording was synchronized with the data acquisition.[31] Upon visual appearance of the task stimuli on the projector screen, the MRI operator and scan subject were instructed, respectively, to push the triggering button and to read the stimulus.[16]

Table 1 details the stimuli used in this experiment. Three stimuli targeting the same vowel-to-vowel transition were placed in the carrier phrase *"a _____ again."* All 3 stimuli involve a transition from (approximately) the same starting vowel sound /a/ to (approximately) the same ending vowel sound /ɪ/. The 3 stimuli are the monosyllable "pie" with diphthong /aɪ/, "poppy" with the 2 vowels /a/ and /i/ in successive syllables in a word, and "pop pip" with the 2 vowels /a/ and /ɪ/ in adjacent words. These English vowels in sequence produce sweeping lingual motions that move the tongue from a low-back (pharyngeal constriction) posture to a high-front (oral palatal constriction) posture. The approximate sweeping or transition time becomes increasingly longer in duration (all else equal) from the monosyllable "pie" to the bisyllable CVCV [CV = consonant-vowel] "poppy" to the cross-word CVCCVC sequence "pop pip." Importantly the words' medial consonant [p] is made with the lips rather than the tongue, so was used near the target vowel-to-vowel transition to minimize coarticulation with (i.e., interference of) other nearby lingual sounds. The tagging module was triggered in close temporal proximity with the onset of the starting mid-central vowel /ə/ (the initial word "a" of the carrier phrase), so that the lingual deformations in the later frames are relative to this fairly neutral vocalic schwa posture of the tongue.

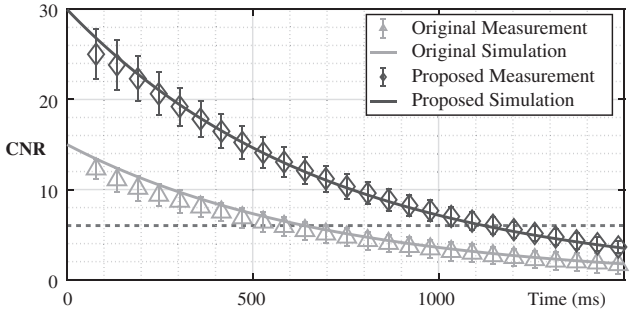

FIGURE 2 Simulated and measured tag CNR decay for TFA = 90° in the original implementation and with TFA = 180° with REALTAG in the proposed method. (Top) Solid lines show simulations; symbols and error bars show mean and standard deviation of the measurement, respectively. The in vivo CNR measurement with 5 subjects is consistent with the simulation. Dashed horizontal line indicates CNR = 6. CNR drops below 6 around 600 ms for TFA = 90°; it reaches the same level around 1150 ms for TFA = 180° with REALTAG

3 | RESULTS

Figure 2 compares simulated and measured tag CNR decay for TFA = 90° in the original implementation and TFA = 180° with REALTAG in proposed method. The plot shows that the experimental measurements are consistent with simulations. The starting CNR is doubled in the proposed method, as expected with homodyne detection compared with magnitude detection. The tag persistence is, therefore, prolonged with the same decay rate. CNR drops below 6 around 600 ms for TFA = 90°, while it reaches the same CNR level around 1150 ms for TFA = 180° with REALTAG. Note that a minimum CNR threshold of 6 has been used for both cardiac[32] and speech applications.[16]

Figure 3 compares tag persistence between the original method and the proposed method in 5 subjects (from our first

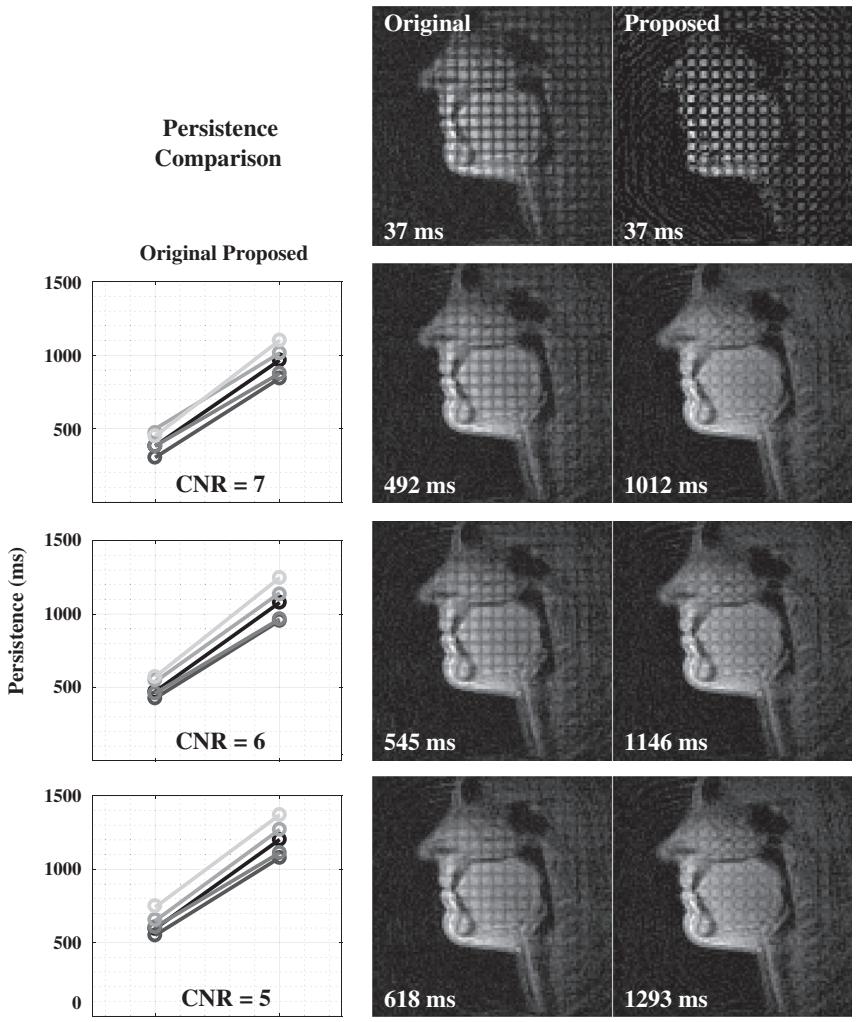

FIGURE 3 Time threshold improvement by using REALTAG in 5 in vivo scans with 1 cm tag spacing. The 2nd, 3rd, and 4th rows compare the 2 methods with CNR thresholds of 7, 6, and 5, respectively. For all 5 subjects, the proposed method significantly improved the persistence by a factor of >1.9×

sub-study). The 2nd, 3rd, and 4th rows compare the 2 methods with CNR thresholds of 7, 6, and 5, respectively. In all subjects and all cases, the proposed method significantly improved the persistence by a factor of >1.9×. For instance, the 3rd row shows the tag persistence was prolonged from 545 ms to 1146 ms with a CNR threshold at 6.

Figure 4 shows representative images by the 2 methods during the production of the target speech stimuli (second sub-study). Figures 4A-C show the subject speaking "a pie again," "a poppy again," and "a pop pip again," respectively, progressing in time from the starting to the ending vowel constriction postures. For both methods in each stimulus, the arrows in the audio waveform indicate the relative timing of the representative images. In all stimuli, both the original and proposed method were able to visualize the deformation for the starting vowel /a/ (indicated by horizontal compressed bi-concave rectangles), as it was within the persistence time (396-486 ms). However, the results differed across stimuli for the ending vowel /ɪ/ (evidenced by vertically compressed bi-concave rectangles).

In the diphthong case (Figure 4A), the vocalic tongue movement spanned a relatively short duration as the 2 vowel postures are in the same syllable; so, both methods can capture the ending deformation (orange). In (Figure 4B), the 2 vowel sounds are in 2 successive syllables with 1 intervening segment, with the result that the ending vowel /ɪ/ occurred around the cutoff time for the conventional method. The new proposed method, denoted with blue dashed lines in (Figure 4B), provides clear internal tongue deformation through the vowel transition endpoint, while the tag line in the original method becomes obscure by this timepoint. Finally, in (Figure 4C), the 2 vowels were located in 2 words with 2 intervening segments between the vowels, and consequently were further in time from each other. Crucially, while the tag lines in the original method nearly disappear by the time of the second vowel's constriction, the proposed method is able to maintain the tag lines with CNR > 8 (gold) through the vocalic articulation of the second syllable.

4 | DISCUSSION

We have demonstrated intermittent tagging with REALTAG to extend the tag persistence in RT-MRI. This approach can

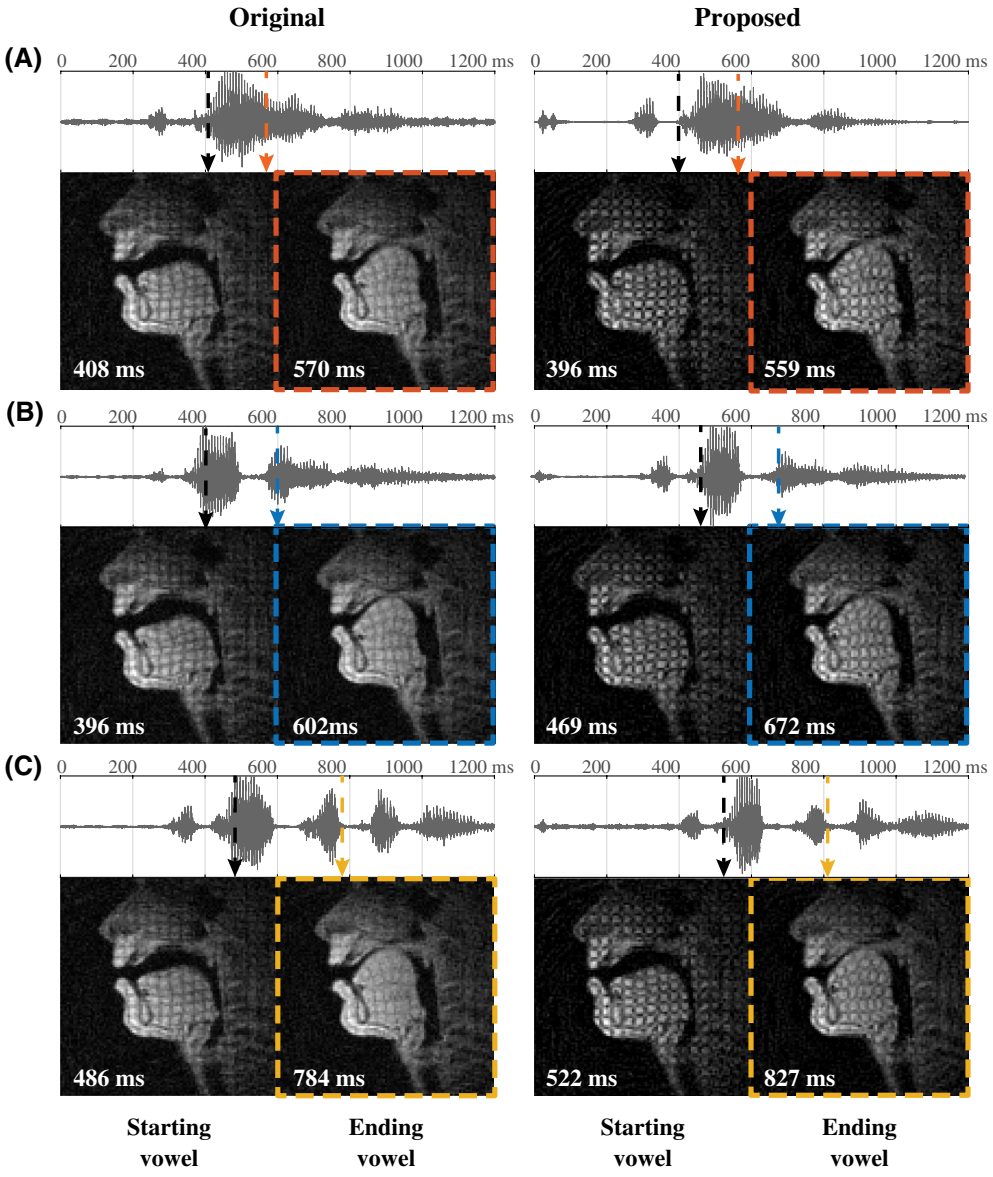

FIGURE 4 Representative images during the speech stimuli production. A-C, Show the subject speaking "a pie again," "a poppy again," and "a pop pip again," respectively, progressing from the starting to the ending target vowel vocal tract constrictions. Arrows in the audio waveform indicate the timepoints of the selected images. The proposed method captures the tongue deformation of the starting and the ending vowels for all 3 stimuli (right: red, blue, and gold), while CNR by the original method drops below threshold in the latter 2 cases and becomes obscure (left: blue and gold)

provide 2× initial CNR and >1.9× longer persistence without repetitions and is suitable for investigations of natural speech production. We illustrated a method for phase sensitive reconstruction that provides better image quality compared with the existing intermittent tagging method.[16] We demonstrate a usable imaging window of 1150 ms at 1.5T, with imaging CNR $\geq$ 6. The proposed method is able to capture internal tongue deformation during American English vowel-to-vowel transition in separate words. This method provides a powerful new tool for imaging muscle movement in natural speech production and other similar RT applications, particularly where CINE imaging is not applicable or suboptimal due to its repetition requirements and the natural variability of human action.

Recently, RT speech MRI has been performed at a broad range of field strengths from 0.55T[33] to 3T[34-38] with adequate image quality. The proposed method could be particularly useful at low field strengths where both muscle T_1 is shorter and the baseline signal-to-noise ratio is lower. The T_1 is approximately 30% shorter at 0.55T compared with 1.5T.[39] This will cause approximately 30% shorter persistence but can be compensated by the 2x improvement achieved using REALTAG.

Derbyshire et al used bSSFP sequence in their original CMR REALTAG to further extend the tag persistence.[22] Furthermore, bSSFP provides superior contrast between myocardium and blood, which offers additional improvement

on image quality. However, in our implementation we did not use bSSFP. In our experience, banding artifacts inevitably appeared in tongue tip and tongue body boundary due to the 9.4 ppm (600 Hz at 1.5T) off-resonance at the air-tissue boundary. Shorter repetition time and less-efficient acquisition have to be used to mitigate signal nulling. This topic remains for future work considering trading among spatiotemporal resolution and banding artifact removal.

Another difference between our method and the original CMR REALTAG is phase estimation. CMR REALTAG uses a fixed linear fitting to the phase within an automatically selected static ROI.[22] We did not directly adopt this pipeline for 2 reasons. First, linear phase is a reasonable model for cardiac imaging as the myocardial ROI is distant from the surface coils relative to the diameter of the coils.[22] Our custom 8-channel coil assembly wraps around the subject's jaw anteriorly to both lateral sides and was designed to be in close proximity to the upper airway (see Lingala et al[28]). Therefore, the linear phase assumption does not hold. Second, speech production involves rapid and irregular tongue movement relative to the phased array coils, introducing constantly varying image phase. Hence, the assumption of a fixed image phase is not valid. A dynamic approach has been used to effectively depict and track articulators' phase for off-resonance correction during speech production.[40] For all of these reasons, we use frame-by-frame estimates of smooth phase using the images themselves. See Supporting Information File1 for detailed discussion on this topic. See Supporting Information Video S1 for a comparison between the original and the proposed REALTAG.

The design of the low pass filter determines the accuracy of the estimated phase. In our experience, a low-pass filter that is 10-20% wider than the designed width can be used to estimate the phase without any perceivable artifacts in the tag lines. This may outperform the current choice due to a more accurate homodyne detection[41] but inevitably introduces invisible phase errors from the tag information in the hormonic peaks. This trade-off between phase compensation and errors can be potentially resolved by using other advanced image phase estimation methods, such as ESPIRiT with virtual conjugate coils (VCC-ESPIRiT).[42] VCC-ESPIRiT enforces smooth image phase (therefore, avoiding phase errors in tag lines), while implicitly imposing data consistency from the entire k-space rather than only the synthesized center.

We enforced nonnegative value in the final images to perceptually increase contrast between dark tag lines and tongue tissue. This process also eliminated the bright background noise from the phase sensitive reconstruction, and, therefore, provided better visualization. It is worth noting that speech-scientist observers reported only 1 qualitative drawback with the REALTAG approach, which was the bright dots at the intersection of tag lines during earlier phases (<400 ms), due to double-inversion. This would not be present in 1D tagging, and for 2D tagging could be easily read through with practice.

5 | CONCLUSIONS

We demonstrated improved RT tagged MRI with substantially increased tag persistence using a novel REALTAG method suitable for RT imaging where background phase needs to be estimated frame-by-frame. Tag persistence was roughly 1150 ms compared with 550 ms with the prior conventional approach. This enables capturing longer motion patterns in speech production, such as lingual vowel-to-vowel transition, and provides a powerful new window to study tongue muscle function.

ACKNOWLEDGMENTS

We thank HeartVista, Inc. for supporting on RTHawk Research system. We acknowledge Asterios Toutios, Sarah Harper, and Tanner Sorensen of the Speech Production and Articulation kNowledge (SPAN) group at the University of Southern California, for audio data collection and insightful discussions.

ORCID

Weiyi Chen https://orcid.org/0000-0001-5116-8645
Nam Gyun Lee https://orcid.org/0000-0001-5462-1492
Dani Byrd http://orcid.org/0000-0003-3319-5871
Shrikanth Narayanan http://orcid.org/0000-0002-1052-6204
Krishna S. Nayak https://orcid.org/0000-0001-5735-3550

REFERENCES

1. Ibrahim el-SH. Myocardial tagging by cardiovascular magnetic resonance: evolution of techniques–pulse sequences, analysis algorithms, and applications. *J Cardiovasc Magn Reson*. 2011;13:36.
2. Moerman KM, Sprengers AM, Simms CK, Lamerichs RM, Stoker J, Nederveen AJ. Validation of continuously tagged MRI for the measurement of dynamic 3D skeletal muscle tissue deformation. *Med Phys*. 2012;39:1793–1810.
3. Piccirelli M, Luechinger R, Sturm V, Boesiger P, Landau K, Bergamin O. Local deformation of extraocular muscles during eye movement. *Invest Opthalmol Vis Sci*. 2009;50:5189.
4. Woo J, Stone M, Suo Y, Murano EZ, Prince JL. Tissue-point motion tracking in the tongue from cine MRI and tagged MRI. *J Speech Lang Hear Res*. 2014;57:S626–S636.
5. Sabet AA, Christoforou E, Zatlin B, Genin GM, Bayly PV. Deformation of the human brain induced by mild angular head acceleration. *J Biomech*. 2008;41:307–315.
6. Brown EC, Cheng S, McKenzie DK, Butler JE, Gandevia SC, Bilston LE. Tongue and lateral upper airway movement with mandibular advancement. *Sleep*. 2013;36:397–404.
7. Bazille A, Guttman MA, McVeigh ER, Zerhouni EA. Impact of semiautomated versus manual image segmentation errors on myocardial strain calculation by magnetic: resonance tagging. *Invest Radiol*. 1994;29:427–433.

8. Azhari H, Buchalter M, Sideman S, Shapiro E, Beyar R. A conical model to describe the nonuniformity of the left ventricular twisting motion. *Ann Biomed Eng*. 1992;20:149–165.
9. Woo J, Xing F, Stone M, et al. Speech map: a statistical multimodal atlas of 4D tongue motion during speech from tagged and cine MR images. *Comput Methods Biomech Biomed Eng Imaging Vis*. 2019;7:361–373.
10. Parthasarathy V, Prince JL, Stone M, Murano EZ, Nessaiver M. Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing. *J Acoust Soc Am*. 2007;121:491–504.
11. 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.
12. McVeigh ER, Epstein F. Myocardial tagging during real-time MRI. In: *Annual Reports of the Research Reactor Institute*, Kyoto University. 2001;2284–2285.
13. Sampath S, Derbyshire JA, Atalar E, Osman NF, Prince JL. Real-time imaging of two-dimensional cardiac strain using a harmonic phase magnetic resonance imaging (HARP-MRI) pulse sequence. *Magn Reson Med*. 2003;50:154–163.
14. Pan L, Stuber M, Kraitchman DL, Fritzges DL, Gilson WD, Osman NF. Real-time imaging of regional myocardial function using fast-SENC. *Magn Reson Med*. 2006;55:386–395.
15. Ibrahim E-S, Stuber M, Fahmy AS, et al. Real-time MR imaging of myocardial regional function using strain-encoding (SENC) with tissue through-plane motion tracking. *J Magn Reson Imaging*. 2007;26:1461–1470.
16. Chen W, Byrd D, Narayanan S, Nayak KS. Intermittently tagged real-time MRI reveals internal tongue motion during speech production. *Magn Reson Med*. 2019;82:600–613.
17. Hagedorn C, Proctor M, Goldstein L. Automatic analysis of singleton and geminate consonant articulation using real-time magnetic resonance imaging. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2011.
18. Tang C, McVeigh ER, Zerhouni EA. Multi-shot EPI for improvement of myocardial tag contrast: comparison with segmented SPGR. *Magn Reson Med*. 1995;33:443–447.
19. Lingala SG, Sutton BP, Miquel ME, Nayak KS. Recommendations for real-time speech MRI. *J Magn Reson Imaging*. 2016;43:28–44.
20. Herzka DA, Guttman MA, McVeigh ER. Myocardial tagging with SSFP. *Magn Reson Med*. 2003;49:329–340.
21. Fischer SE, McKinnon GC, Maier SE, Boesiger P. Improved myocardial tagging contrast. *Magn Reson Med*. 1993;30:191–200.
22. Derbyshire JA, Sampath S, McVeigh ER. Phase-sensitive cardiac tagging - REALTAG. *Magn Reson Med*. 2007;58:206–210.
23. King KF, Ganin A, Zhou XJ, Bernstein MA. Concomitant gradient field effects in spiral scans. *Magn Reson Med*. 1999;41:103–112.
24. Markl M, Bammer R, Alley MT, et al. Generalized reconstruction of phase contrast MRI: analysis and correction of the effect of gradient field distortions. *Magn Reson Med*. 2003;50:791–801.
25. Uecker M1, Lai P, Murphy MJ, Virtue P, et al. ESPIRiT - an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magn Reson Med*. 2014;71:990–1001.
26. Shinnar M, Leigh JS. Inversion of the Bloch equation. *J Chem Phys*. 1993;98:6121.
27. Osman NF, Kerwin WS, McVeigh ER, Prince JL. Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. *Magn Reson Med*. 1999;42:1048–1060.
28. 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;77:112–125.
29. Santos JM, Wright GA, Pauly JM. Flexible real-time magnetic resonance imaging framework. In: The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2005:1048–1051.
30. Kellman P, McVeigh ER. Image reconstruction in SNR units: a general method for SNR measurement. *Magn Reson Med*. 2005;54:1439–1447.
31. Bresch E, Nielsen J, Nayak K, Narayanan S. Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans. *J Acoust Soc Am*. 2006;120:1791–1794.
32. Markl M1, Scherer S, Frydrychowicz A, Burger D, Geibel A, Hennig J. Balanced left ventricular myocardial SSFP-tagging at 1.5T and 3T. *Magn Reson Med*. 2008;60:631–639.
33. Bhattacharya I, Ramasawmy R, Restivo M, Campbell-Washburn A. Dynamic speech imaging at 0.55T using single shot spirals for 11ms temporal resolution. In: Proceedings of the 27th Annual Meeting of ISMRM, Montreal, 2019. Abstract 440.
34. 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.
35. Fu M, Barlaz MS, Holtrop JL, et al. High-frame-rate full-vocal-tract 3D dynamic speech imaging. *Magn Reson Med*. 2017;77:1619–1629.
36. Burdumy M, Traser L, Burk F, et al. One-second MRI of a three-dimensional vocal tract to measure dynamic articulator modifications. *J Magn Reson Imaging*. 2017;46:94–101.
37. Ruthven M, Freitas AC, Boubertakh R, Miquel ME. Application of radial GRAPPA techniques to single- and multislice dynamic speech MRI using a 16-channel neurovascular coil. *Magn Reson Med*. 2019;82:948–958.
38. Lingala SG, Lim Y, Kruger S, Nayak K. Improved spiral dynamic MRI of vocal tract shaping at 3 Tesla using dynamic off resonance artifact correction. In: Proceedings of the 27th Annual Meeting of ISMRM, Montreal, 2019. Abstract 441.
39. Campbell-Washburn A, Herzka D, Kellman P, Koretsky A, Balaban R. Image contrast at 0.55T. In: Proceedings of the 27th Annual Meeting of ISMRM, Montreal, 2019. Abstract 1214.
40. 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.
41. Noll DC, Nishimura DG, Macovski A. Homodyne detection in magnetic resonance imaging. *IEEE Trans Med Imaging*. 1991;10:154–163.
42. Uecker M, Lustig M. Estimating absolute-phase maps using ESPIRiT and virtual conjugate coils. *Magn Reson Med*. 2017;77:1201–1207.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

FIGURE S1 Tagging creates k-space replicas at harmonic peaks. The harmonic peaks created by the SPAMM tagging are located at integer multiples of $k_{tag} = 2k_{max}/\alpha$ along the tagged directions (yellow arrows), where $2k_{max}$ denotes

the k-space extent, α is the ratio between the tag spacing and the image resolution

FIGURE S2 Phase of an un-tagged image after low pass filtering with different widths. Dashed lines indicate the designed filter width $W = k_{tag} = 2k_{max}/\alpha$ (in k-space samples). White arrows identify sharp spatial variations in phase due to off-resonance. This variation is suppressed when using a low pass filter with small width (see black arrow in $0.5k_{tag}$). A filter width of $W = k_{tag}$ allows for a frequency smaller than one-half of the first harmonic frequency, minimizing phase contamination from the replicas due to tagging, while provide an adequate estimation of the background phase. Comparison among the $W = k_{tag}, 1.5k_{tag}, 2k_{tag}$ columns indicates that larger filter width does not significantly improve the phase estimation in the tongue. Furthermore, larger filter width introduces unwanted artifacts in the tag lines as shown in Supporting Information Figure S3

FIGURE S3 Phase compensation by the low pass filter with different widths W. A smaller W can cause over-smoothed phase estimation and shading artifacts, indicated by the yellow arrow. A larger W can introduce spurious phase information from the spectral replicas created by tagging and rectify the negative-valued tag lines. The white arrow indicates a bright spot in the tag lines caused by the unwanted phase compensation in the tag lines

VIDEO S1 Reconstruction using the original REALTAG and the proposed improved REALTAG. The blue arrow identifies insufficient phase compensation due to inaccurate phase estimation by the original REALTAG method. (Left) The original REALTAG uses a fixed linear fitting to the phase within an automatically selected static ROI. Linear model is not suitable for the phase in speech imaging, as typically the coils assembly wraps around the subject's jaw anteriorly to both lateral sides and is designed to be in close proximity to the upper airway. In addition, speech production involves rapid and irregular tongue movement relative to the phased array coils, introducing constantly varying image phase. Therefore, it is not reasonable to assume constant phase. (Right) The proposed REALTAG uses frame-by-frame estimates of smooth phase drawn from the images themselves

How to cite this article: Chen W, Lee NG, Byrd D, Narayanan S, Nayak KS. Improved real-time tagged MRI using REALTAG. *Magn Reson Med*. 2019;00: 1–9. https://doi.org/10.1002/mrm.28144