

High-frame-rate real-time imaging of speech production

Sajan Goud Lingala, Yinghua Zhu, Yoon-Chul Kim, Asterios Toutios, Shrikanth Narayanan, and Krishna Nayak

Sparse sampling and constrained reconstruction enable 80-frames/second real-time magnetic resonance imaging, providing new insights into the dynamics of vocal-tract shaping.

Real-time magnetic resonance imaging (RT-MRI) involves the rapid and continuous acquisition of MRI images of a dynamically evolving physiological process. It is emerging as a powerful tool to noninvasively visualize complex spatiotemporal dynamics for in vivo applications such as cardiac cine MRI (movement

of cardiac muscles, chambers, and valves to assess heart function), functional MRI (assessment of brain activity during an ongoing task), and flow MRI (tracking the blood flow).^{1,2} Our work seeks to develop and apply RT-MRI methods to understand human speech production,³⁻⁵ which involves complex and intricate coordination between the lungs, diaphragm, chest wall, larynx, pharynx, vocal cords, tongue, lips, soft palate (velum), teeth, jaw, and the nasal cavity.⁶ MRI offers unique advantages over competing modalities such as x-ray fluoroscopy, computed tomography, and ultrasound; it provides noninvasive safe imaging of arbitrary image planes, and can visualize deep soft tissue

Continued on next page

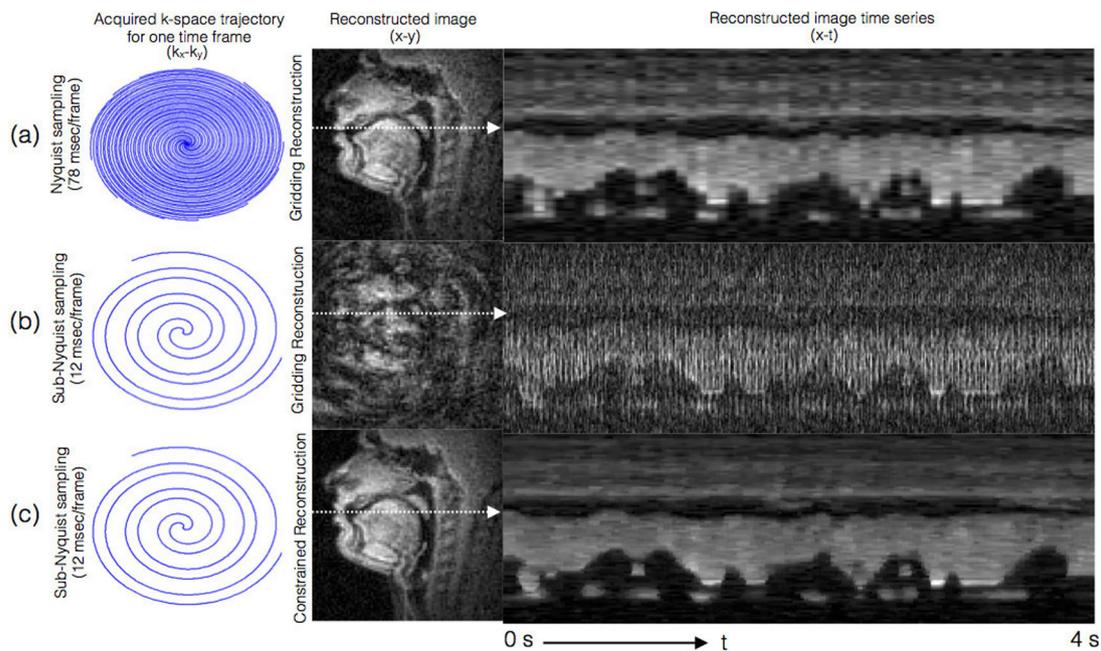


Figure 1. Sub-Nyquist sampling combined with constrained reconstruction improves the true time resolution in real-time magnetic resonance imaging (RT-MRI) of speech. (a) Images reconstructed using Nyquist sampling result in time resolution of 78ms/frame. (b) Sub-Nyquist sampling allows for significantly improved time resolution, at the expense of aliasing artifacts. (c) Constrained reconstruction addresses this tradeoff by resolving the aliasing at the native time resolution of 12ms/frame. Note the apparent advantage of the increase in time resolution of (c) compared with (a) in terms of crispness along the time axis (t). *k*-space: Spatial frequency domain.

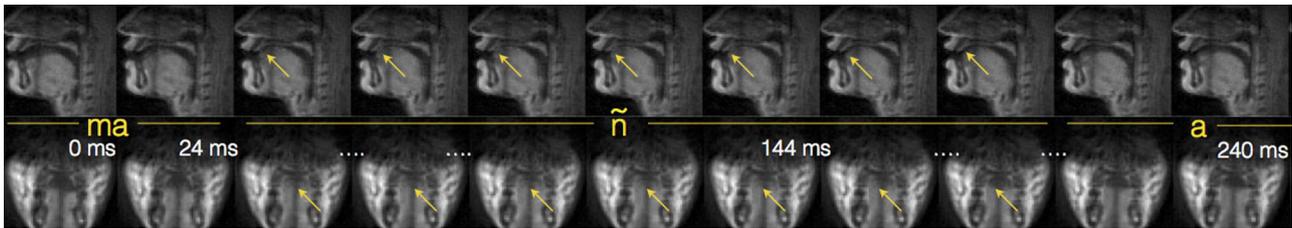


Figure 2. Simultaneous RT-MRI of mid-sagittal and coronal planes (top and bottom rows, respectively) at a time resolution of 24ms/frame. This exceptionally high frame rate enables visualization of the articulatory timing events associated with the sound ‘ \tilde{n} ,’ which are contact of the tongue with the roof of the hard palate, and narrowing of the airway, in the mid-sagittal and coronal planes, respectively.

structures. However, MRI is notoriously slow due to fundamental physical limitations. This results in a challenging tradeoff between spatial and temporal resolution, and signal-to-noise ratio.

Our first approaches to imaging the upper airway during speech production relied on using short spiral interleaves to rapidly scan the k-space (spatial frequency domain).³⁻⁵ In comparison to the widely used 2D-fourier-transform Cartesian approach, spirals are highly time-efficient and are resilient to motion artifacts (errors). We implemented the sequences within a customized real-time imaging environment that allowed for interactive imaging.⁷ On a modern 1.5T MRI scanner equipped with high-speed gradients, we were able to image up to a time resolution of 78ms with a spatial resolution of 2.4mm², and reconstruct at 24 frames/second using a sliding window technique.

We have recently made further advances in the capabilities of RT-MRI.⁸ In addition to fast spirals, we used constrained reconstruction and advances in upper-airway radiofrequency coil design to improve the speed and quality of RT-MRI. We used a customized eight-channel upper-airway receiver coil that has four elements on either side of the jaw. The coil’s design enables high sensitivity over all the important articulators (lips, tongue, epiglottis, velum), thereby greatly increasing the signal-to-noise ratio in these regions in comparison with coils developed for other body parts (such as the neurovascular or head-and-neck coil).

We used constrained reconstruction to improve the native time resolution by reconstructing images at sub-Nyquist sampling levels. As depicted in Figure 1(a), an artifact-free image reconstruction requires Nyquist sampling in k-space, which corresponds to acquisition with a temporal footprint of 78ms (or 13 spiral interleaves with a repetition time of 6ms). Sub-Nyquist sampling improved the temporal footprint to 12ms (two interleaves), but led to increased image artifacts, as shown in Figure 1(b). We resolved this issue by exploiting prior knowledge of the dynamic image time series to have sparse temporal pixel time profiles under the finite difference operation along time.

We posed the reconstruction as a penalized convex optimization problem, where we penalized rapidly varying pixel time profiles (that usually correspond to alias artifacts and noise) subject to consistency with the acquired data from the eight-channel coil. We used an iterative nonlinear conjugate gradient algorithm to solve the resulting optimization, which resulted in images with excellent spatiotemporal fidelity, as shown in Figure 1(c).

The crispness along the time profile in Figure 1 shows that our approach can dramatically improve the visualization of rapid articulatory movements, for example, the production of consonant clusters. The gains in time resolution can be used in conjunction with other factors such as improved slice coverage and/or spatial resolution. For instance, in Figure 2, we show concurrent mid-sagittal and coronal imaging at a native time resolution of 24ms/frame during the production of the consonant \tilde{n} . This capability allows flexibility in modeling complex spatiotemporal patterns by utilizing information from more than one plane at a high time resolution.

The described imaging approach, along with a synchronized noise-cancelled audio acquisition scheme,⁹ is currently set up as the RT-MRI speech acquisition sequence at our site. We will use it to acquire data for several current and future studies that are targeted to address open questions in the areas of phonetics and phonology, understanding language acquisition and language disorders, and to inform treatment plans in clinical applications such as clefts of lips/palate and oropharyngeal cancer.

This work is supported by the National Institutes of Health under grant NIH/NIDCD R01 DC007124.

Author Information

**Sajan Goud Lingala, Yinghua Zhu, Asterios Toutios,
Shrikanth Narayanan, and Krishna Nayak**
Ming Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, CA

Yoon-Chul Kim
Samsung Medical Center
Seoul, South Korea

References

1. K. S. Nayak and B. S. Hu, *The future of real-time cardiac MRI*, *Curr. Cardiol. Rep.* **7**, pp. 45–51, 2005.
2. S. Zhang, T. K. Block, and J. Frahm, *Magnetic resonance imaging in real time: advances using radial FLASH*, *J. Magn. Reson. Imag.* **31** (1), pp. 101–109, 2010.
3. S. Narayanan, K. S. Nayak, S. Lee, A. Sethy, and D. Byrd, *An approach to real-time magnetic resonance imaging for speech production*, *J. Acoust. Soc. Am.* **115** (5), pp. 1771–1776, 2004.
4. E. Bresch, Y. C. Kim, K. S. Nayak, D. Byrd, and S. Narayanan, *Seeing speech: capturing vocal tract shaping using real-time magnetic resonance imaging*, *IEEE Sig. Process. Mag.* **25** (3), pp. 123–132, 2008.
5. S. Narayanan, A. Toutios, V. Ramanarayanan, A. Lammart, J. Kim, S. Lee, K. S. Nayak, et al., *Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research*, *J. Acoust. Soc. Am.* **136**, pp. 1307–1311, 2014.
6. A. D. Scott, M. Wylezinska, M. J. Birch, and M. E. Miquel, *Speech MRI: morphology and function*, *Phys. Med.* **30** (6), pp. 604–618, 2014.
7. J. M. Santos, G. A. Wright, and J. M. Pauly, *Flexible real-time magnetic resonance imaging framework*, *Proc. 26th Ann. Int'l Conf. IEEE EMBS*, pp. 1048–1051, 2004.
8. S. G. Lingala, Y. Zhu, Y.-C. Kim, A. Toutios, S. Narayanan, and K. S. Nayak, *High spatio-temporal resolution multi-slice real time MRI of speech using golden angle spiral imaging with constrained reconstruction, parallel imaging, and a novel upper airway coil*, *Proc. 23rd Int'l Soc. Magn. Res. Med (ISMRM) Sci. Sess.*, p. 789, 2015.
9. C. Vaz, V. Ramanarayanan, and S. Narayanan, *A two step technique for MRI audio enhancement using dictionary learning and wavelet packet analysis*, *Proc. InterSpeech*, pp. 1312–1315, 2013.