Improved 3D real-time MRI with Stack-of-Spiral (SOSP) trajectory and variable density randomized encoding of speech production

Ziwei Zhao, Yongwan Lim, Dani Byrd, Shrikanth Narayanan and Krishna S. Nayak
University of Southern California, Los Angeles, CA, United States

Target Audience: Linguists; MR engineers who focus on non-Cartesian and dynamic imaging.

Purpose: To develop and evaluate an improved sampling strategy for three-dimensional (3D) real-time (RT) MRI of the entire vocal tract, with high spatio-temporal resolution during natural speech. RT-MRI has emerged as an efficient tool to understand the complex spatio-temporal coordination of vocal tract articulators during speech production[1]. Speech RT-MRI faces fundamental trade-offs between spatial resolution, temporal resolution, spatial coverage, signal-to-noise ratio (SNR), and artifacts, which are common to all fast MRI. Since 2004 2D mid-sagittal RT-MRI has been extensively applied to speech production research[2] and is considered to be a standard and mature method. Recently, 3D RT-MRI has provided an opportunity to study the entire vocal tract volume in montion[3]. Lim et al.[4] demonstrated the feasibility of 3D RT-MRI of natural speech, with full vocal tract coverage, 2.4×2.4×5.8 mm^3 spatial resolution and 61ms temporal resolution. Inspired by this study, we explore and evaluate improved (k,t) data sampling strategies and constrained reconstruction options for 3D RT-MRI with improved spatio-temporal resolution during natural speech tasks.

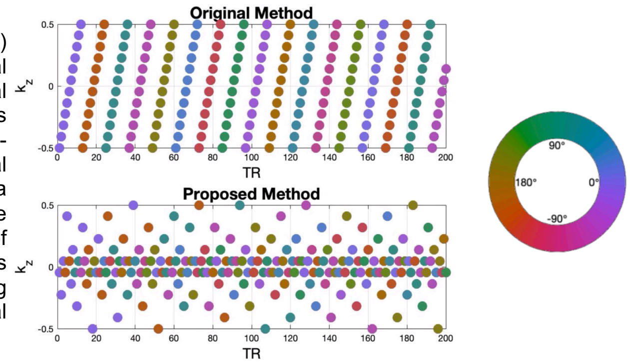

Figure 1: Sampling Patterns. k_z vs. time plots illustrate sampling strategies of the original method from Lim et al[4] (top) and the proposed method (bottom). Each dot represents one spiral arm in the k_x - k_y plane; color represents its initial angle as shown by the color map shown at the right.

Methods: <u>Data sampling and reconstruction</u> **Figure 1** illustrates a proposed sampling strategy compared to the prior work of Lim et al.[4]. During each TR, we acquire one spiral arm in the k_x - k_y plane to achieve 13-fold acceleration. A pseudo-golden angle increment is used in the k_x - k_y plane, and Cartesian sampling is employed along the k_z direction. In Lim's method[4], k_z steps are sampled in a linear order. Spiral patterns are tilted with a golden-angle increment after 12 phase encodings (corresponding to full k_z sampling). In the proposed method, we applied a rotated a golden-angle in the k_x - k_y plane for each phase encoding step. k_z was sampled randomly according to a variable density function. Image reconstruction was performed by solving the following constrained optimization:

$$arg \min_{f(\boldsymbol{r},t)} \|A(f) - \boldsymbol{b}\|_2^2 \ + \lambda_1 \|TV(f)\|_1 + \lambda_2 \|D_t(f)\|_1$$

where $f(\boldsymbol{r},t)$ is the dynamic image time series to be reconstructed, and the vector $\boldsymbol{r}\, \epsilon (x,y,z)$ represents image domain spatial coordinates. $\boldsymbol{b}$ is multi-coil k-t space measurement data. A refers to coil sensitivity encoding as well as Fourier operator along each time frame in 3D volume. Isotropic total variation (TV) and first-order finite difference (D_t) constraints were applied along spatial and temporal dimensions, respectively. Reconstruction was implemented in MATLAB and using the Berkeley Advanced Reconstruction Toolbox[5]. The regularization parameters λ_1 and λ_2 were chosen visually based on the image quality in sagittal views and time-intensity plots. We chose λ_1 = 0.008, λ_2 = 0.03.

<u>*In-vivo* Experiments</u> Experiments were performed on a 1.5 T Signa Excite HD scanner (GE Healthcare, Waukesha, WI), using the real-time imaging platform (RT-Hawk, Heart Vista Inc, Los Altos, CA). Experiments used the body coil for RF transmission and a custom eight-channel upper airway coil for signal reception. Two healthy adult volunteers were scanned. Speaker 1 (male native Chinese speaker, English as a second language) was scanned while reading the English stimuli: "/loo/-/lee/-/la/-/za/-/na/-/za/" repeated twice at a natural rate. Speaker 2 (male American English speaker) was scanned with a 3D sequence using the original and the proposed sampling patterns, as well as 2D three-slice sequences as reference[6]. The stimuli for Speaker 2 include 4 sentences and were each spoken twice, once at a natural rate and once at a speaking rate of approximately 1.5× the initial rate. All stimuli were read in the scanner using a mirror projector setup used for display[7].

Results: Figure 2 shows the stimuli results for the original 3D and proposed 3D methods, along with the interleaved 2D multi-slice method (reference). Fast lip opening and closing in a speeded rate (blue dotted box) can be better captured using proposed method. Images are significantly blurred for tongue body movements in the original method (pink, yellow and green arrows).

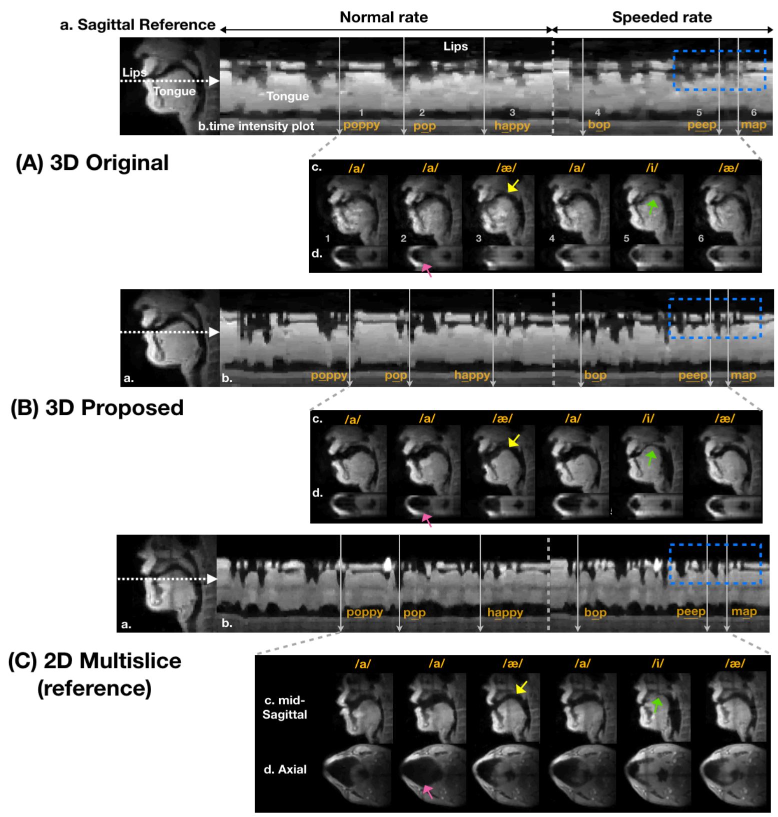

Figure 2. Stimuli results. A. 3D original, B. 3D proposed and C. 2D multi-slice method. Top row for A-C: a. sagittal reference and b. time intensity plots during two repeated trials at normal and speeded rates. Gray arrows 1-6 indicate specific frames below. Second row for A-C: Six representative frames in c. sagittal and d. axial views that capture /a/, /æ/, /i/ vowels.

Discussion: In this study, four other sampling schemes were also designed, which allowed pairwise comparisons for a single change of sampling strategy. Due to the space constraints, they are not shown here. Our proposed method, the randomized variable density stack-of-spiral sampling, enables retrospective selection of temporal resolution. We found that better temporal resolution enables capturing fast movements of certain articulators, while a relatively high SNR can be preserved with adequate temporal resolution.

Conclusion: We have demonstrated 3D RT-MRI of the vocal tract with improved spatio-temporal resolution achieved by using randomized variable density stack-of-spiral sampling, combined with a spatially and temporally constrained reconstruction. Improved capture and visualization of several speech tasks is achieved notably during fast lip and tongue movements.

Acknowledgements: NIH Grant R01DC007124 and NSF Grant 1514544.

References: [1] Lingala, et al., JMRI (2016) [2] Niebergall, et al., MRM (2013) [3] Burdumy, et al., JMRI (2017) [4] Lim, et al., MRM (2019) [5] Uecker, et al., Proc. Intl. Soc. Mag. Reson. Med. (2015) [6] Kim, et al., JMRI (2012) [7] Lingala, et al., INTERSPEECH (2016)