High-resolution Full-vocal-tract 3D Dynamic Speech Imaging

Maojing Fu[1,2], Joseph Holtrop[2,3], Jamie Perry[4], David Kuehn[5], Zhi-Pei Liang[1,2], and Bradley Sutton[2,3]
[1]*Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States,* [2]*Beckman Institute for Advanced Science and Technology, Urbana, IL, United States,* [3]*Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States,* [4]*Communication Sciences and Disorders, East Carolina University, NC, United States,* [5]*Speech and Hearing Science, University of Illinois at Urbana-Champaign, IL, United States*

INTRODUCTION

Dynamic MRI is important for speech-related studies because of its potential to investigate velopharyngeal motion and acoustic properties jointly in real time.[1] However, many of its applications are limited by trade-offs in imaging speed, spatial coverage and spatial resolution. In particular, high-frame-rate, three-dimensional (3D) dynamic speech MRI with full-vocal-tract coverage remains a challenging goal for speech researchers. Previously a partial separability (PS) model-based method[2] with a cone-navigator-based acquisition[3,4] was demonstrated for accelerated speech MRI, but higher imaging speed is desired.[5] This work further accelerates PS model-based acquisition by merging navigator and imaging data collection, achieving an imaging speed of 150 frames per second (fps). In addition, it integrates an optimized navigation strategy with a randomized 3D encoding scheme to capture speech dynamics from an imaging volume covering the entire vocal tract with a spatial resolution of 2.0 mm × 2.0 mm × 5.0 mm.

METHODS

The PS model assumes l^{th}-order partial separability of the desired image, $I(\mathbf{r},t)=\sum_{\ell=1}^{L}\psi_\ell(\mathbf{r})\phi_\ell(t)$, where $\psi_\ell(\mathbf{r})$ and $\phi_\ell(t)$ represent the l^{th} spatial and the l^{th} temporal basis functions, respectively.[2] $I(\mathbf{r},t)$ can be reconstructed from highly-undersampled data, $d(\mathbf{k},t)$, by jointly enforcing the partial separability and the spatiotemporal total-variation (TV) constraints.[5] The image reconstruction problem can be formulated as:

$$\hat{\mathbf{I}} = \arg\min \|\mathbf{d} - \mathrm{E}(\mathbf{I})\|_2^2 + \lambda \mathrm{TV}_{st}(\mathbf{I}),$$

where $\mathrm{E}(\bullet)$ is an imaging operator encompassing both sparse sampling and parallel imaging, λ is a regularization parameter and $\mathrm{TV}_{st}(\bullet)$ is a spatial-temporal TV operator. An algorithm based on half-quadratic regularization has been applied to solve this optimization problem effectively.[4]

The PS model-based data acquisition scheme sparsely samples the (**k**, *t*)-space to obtain two data sets: a navigator data set with high temporal resolution and an imaging data set with high spatial resolution. Previous work acquired these two data sets within two separate TRs, which prevented greater acceleration due to concerns over gradient performance and slew rate limits. Alternatively, this work chooses to accelerate data acquisition using a "self-navigating" approach that collects both data sets within one single TR.[6] As illustrated in Fig. 1, a navigator acquisition precedes the associated imaging acquisition after each RF excitation. The navigator data are acquired using an optimized 3D reverse-cone trajectory that enjoys twofold benefits - simultaneously navigating across all imaging slices and minimizing the mismatch in TE between both data sets. The imaging data are acquired in 3D random encoding order with a Cartesian trajectory to reduce susceptibility effects.

To demonstrate the practical utility of the proposed method, experiments were performed on a Siemens Trio 3T scanner with a 12-channel head receiver coil. A FLASH sequence integrating a reverse-cone navigation (TE = 3.85 ms) and a Cartesian imaging acquisition (TE = 4.32 ms) was developed to acquire data over a 256 mm × 256 mm × 40 mm FOV covering the entire vocal tract. The reverse-cone trajectory was designed by applying prephasing gradient to a reversed 3D cone trajectory.[3] During data acquisition, a volunteer subject produced repetitive /loo/-/lee/-/la/-/za/-/na/-/za/ sounds at a natural speaking rate in accordance with the local IRB. Reconstructions had a matrix size of 128 × 128 × 8, a spatial resolution of 2.0 mm × 2.0 mm × 5.0 mm and a frame rate of 150 fps (TR = 6.67 ms) covering the entire vocal tract.

RESULTS

Figure 2 illustrates the 3D spatio-temporal motion during the production of /loo/-/lee/-/la/-/za/-/na/-/za/ sounds. Articulator gestures from eight sagittal slices at the production of the /la/ sound are demonstrated with high spatial resolution. Transitions of articulator motion are shown with fine temporal details on both sagittal and transverse planes. In addition, the difference in motion pattern between /l/ in the /loo/, /lee/ and /la/ sounds are clearly shown as three line-shaped patterns on a transverse plane that is placed between the tip of the tongue and the alveolar ridge.

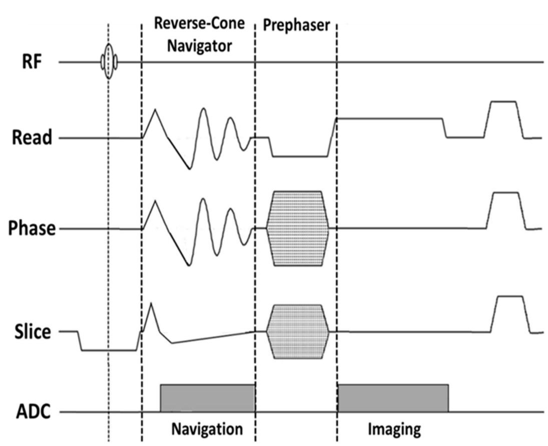

Fig. 1. The accelerated (***k***, t)-space sampling strategy that combines navigator and imaging data acquisition within one TR. The navigator is acquired with a reverse-cone trajectory and imaging data is acquired with random phase encoding order using a Cartesian trajectory.

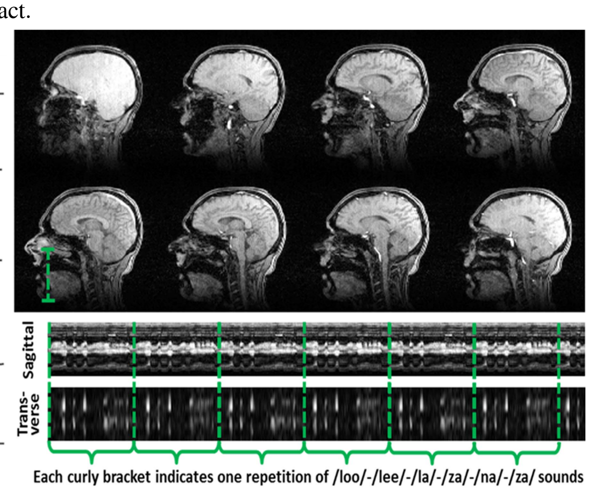

Fig. 2. 3D tongue motion from an imaging volume covering eight consecutive sagittal slices. The bottom rows demonstrate temporal dynamics along a line segment from the mid-sagittal plane, as well as on a transverse plane placed beneath the alveolar ridge.

CONCLUSION

This work enables high-resolution, full-vocal-tract dynamic speech imaging based on a novel acquisition scheme that combines reverse-cone navigation and 3D spatial encoding in one TR. The proposed scheme provides visualization of articulator dynamics over the entire vocal tract at a temporal frame rate of 150 fps and a spatial resolution of 2.0 mm × 2.0 mm × 5.0 mm. This method allows complex 3D soft-tissue structures to be visualized in fine spatial and temporal details, which holds great promise for accurate assessment of anatomical and functional properties of the oropharyngeal tract.

ACKNOWLEDGEMENT This work was supported by a grant from NIH: 1R03DC009676-01A1.

REFERENCE [1] M. Fu, et al. MRM 2014. Preprint. [2] Z.-P. Liang. IEEE-ISBI 2007. pp. 988-991. [3] M. Fu, et al. ISMRM 2012. p. 12. [4] M. Fu, et al. ISMRM 2013. p. 4269. [5] D. Kuehn. Cleft Palate J. 1976. [6] A. Christodoulou, et al. IEEE-TBME 2014. pp. 2451-2457.