TITLE: Shortening total acquisition times for high speed dynamic speech magnetic resonance imaging using temporally sparse navigation

Authors: Riwei Jin, Xi Peng, Brad Sutton

Department of Bioengineering, University of Illinois Urbana-Champaign

Beckman Institute for Advanced Science and Technology

Introduction

Dynamic magnetic resonance imaging (MRI) has been largely used in the speech imaging field, especially in capturing articulatory changes in both structural and functional scales which provide opportunities to study swallowing activities, linguistics and language variabilities. Recently developed methods utilizing the Partial Separability (PS) model[1] enable structural-quality dynamic MRI with amazing frame rates, up to 166 frames per second for full 3D vocal tract coverage.[2,3] But long acquisition times up to 10 minutes are required for such high spatiotemporal resolution which might not be acceptable in some cases such as dynamic speech imaging in children.

Here, we explore the trade-offs of the temporal frame rate and the overall acquisition time in the PS model-based 2D dynamic MRI speech imaging framework. We achieve this by sparsely sampling the temporal navigator and associating each navigator acquisition with multiple k-space lines.

Methods

We implemented a low rank constraint and spatial regularized partial separability (PS) model for the rank images.[2,3] The PS model assumes that strong spatiotemporal correlation exists. In a dynamic speech imaging experiment, we can express our desired spatiotemporal image $f(\mathbf{r},t)$ as the product of low rank constrained navigator datasets and imaging datasets through PS model:

$$f(\mathbf{r},t) = \sum_{l=1}^{L} \Psi_l(\mathbf{r})\phi_l(t)$$

where L is the model order, $\{\Psi_l(\mathbf{r})\}_{l=1}^{L}$ denotes a set of spatial basis functions, and $\{\phi_l(t)\}_{l=1}^{L}$ denotes a set of temporal basis functions. In this work, we did 2D imaging and used a L=10 model order.

We sparsely sample two sets of (k, t)-space data in an interlaced manner: an imaging dataset (from which we will estimate $\{\Psi_l(\mathbf{r})\}_{l=1}^{L}$) and a temporal navigator dataset (which provides estimates of $\{\phi_l(t)\}_{l=1}^{L}$).[1] For the navigator dataset, we used a spiral-trajectory to sample k-space. For the imaging dataset, a Cartesian trajectory with random phase encoding to acquire imaging data with high spatial resolution and structural image quality. Figure 1 shows a simplified pulse sequence diagram for the MRI sequence that samples random k_y lines and uses a spiral navigator. In this work, we sample a temporal navigator only every 4-imaging phase encodes, and associate each imaging phase encode with the navigator that is closest in time.

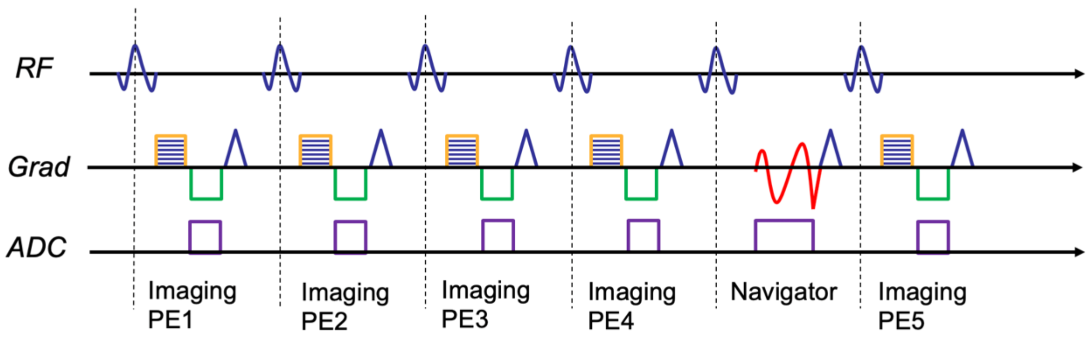

Figure 1. A simplified pulse sequence diagram illustrating (k, t)-space sampling patterns. The navigator dataset is acquired using a spiral trajectory and is only acquired once every 4 imaging data acquisitions. The imaging dataset is acquired using a Cartesian trajectory with random phase encoding to provide structural quality images.

We acquired a 256-matrix size with a 24 cm field of view and 6-mm mid-sagittal slice, for a spatial resolution of 0.94×0.94×6 mm. We reduce the temporal resolution of our model by 4 which provides 43 images per second instead of 108 since we run only one temporal navigator every 4-imaging phase encode lines. The temporal navigator determines the temporal resolution of the PS model. This frame rate is still high enough to visualize the articulatory motions during speech. Furthermore, the total duration of the scan is reduced to 62.5% when acquiring the same number of full frames of data for fitting the PS model. In this case, we acquired 45 full frames of data in 66 sec (instead of 105 sec), resulting in 2880 reconstructed images from the PS model at 43 frames per second.

Results

Figure 2 and 3 shows the results from the 2D single-slice experiment where the subject counted to four. Figure 2 shows the temporal profile taken along a vertical strip across the tongue which displays tongue movement across time. From Figure 3 we can observe the different position of tongue and lip in three different time instances.

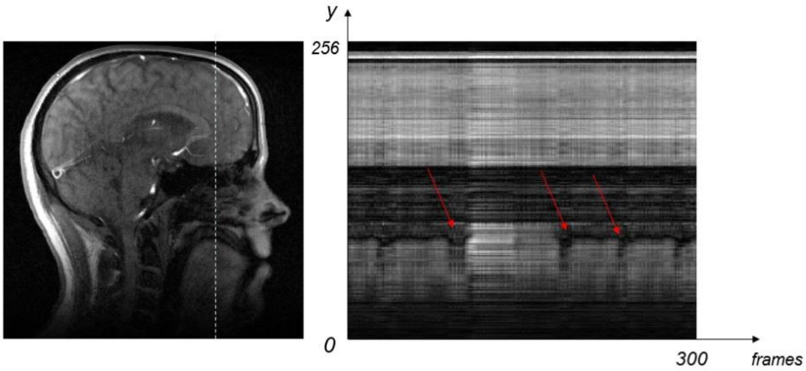

Figure 2 Temporal profile taken in 300 frames (about 7 seconds) along a vertical strip which shows tongue movement.

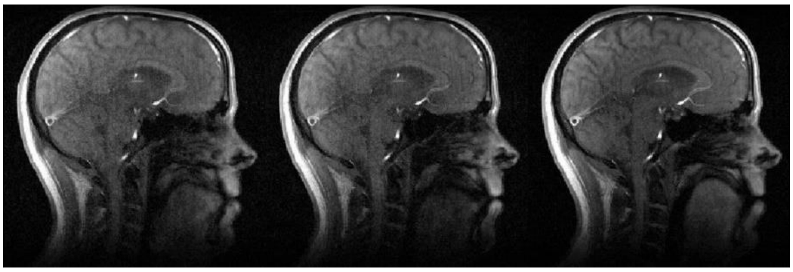

Figure 3 Articulatory and tongue motions at three different time instances

Conclusion

The PS-model framework enables high quality and high spatiotemporal resolution for dynamic imaging of speech with MRI. We have traded total acquisition time for temporal resolution in the method by acquiring multiple random phase encode lines per temporal navigator, enabling reasonable total acquisition times.

Research reported in this publication was supported by the National Institute Of Dental & Craniofacial Research of the National Institutes of Health under Award Number R01DE027989. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Reference

1. Liang Z-P. Spatiotemporal imaging with partially separable functions. In Proceedings of IEEE International Symposium on Biomedical Imaging, Washington D.C., USA, 2007. pp. 988–991.
2. Fu, M. , Zhao, B. , Carignan, C. , Shosted, R. K., Perry, J. L., Kuehn, D. P., Liang, Z. and Sutton, B. P. (2015), High‐resolution dynamic speech imaging with joint low‐rank and sparsity constraints. Magn. Reson. Med., 73: 1820-1832. doi:10.1002/mrm.25302
3. Fu, M. , Barlaz, M. S., Holtrop, J. L., Perry, J. L., Kuehn, D. P., Shosted, R. K., Liang, Z. and Sutton, B. P. (2017), High‐frame‐rate full‐vocal‐tract 3D dynamic speech imaging. Magn. Reson. Med., 77: 1619-1629. doi:10.1002/mrm.26248