High-frame-rate Multislice Speech Imaging with Sparse Samping of (k,t)-space

Maojing Fu[1,2], Anthony G. Christodoulou[1,2], Andrew T. Naber[3], David P. Kuehn[4], Zhi-Pei Liang[1,2], and Bradley P. Sutton[2,3]

[1]*Department of 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]*Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States,* [4]*Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Urbana, IL, United States*

INTRODUCTION

Dynamic MRI is a potentially powerful tool for real-time visualization of vocal tract shaping during sound production [1]. However, low imaging speed and limited spatial coverage hamper its ability to investigate spatiotemporal dynamics of soft tissues, such as the tongue tip and the velum. Recently, fast spiral sequences [2] have been developed to improve temporal resolution, but higher speed is needed for visualizing more than one imaging plane. To further speed up image acquisition, we incorporate parallel imaging methods and sparse sampling techniques based on the partial separability (PS) model [3] and compressed sensing [4]. By combining all three of the above strategies, we achieved dynamic speech imaging with 2.2 mm × 2.2 mm × 8.0 mm spatial resolution at a frame rate of 20 fps for a whole stack of 5 slices.

METHODS

The measured data from a dynamic speech imaging experiment can be expressed as: $d_n(\mathbf{k},t) = \int S_n(\mathbf{r}) I(\mathbf{r},t) e^{-i2\pi\mathbf{k}\cdot\mathbf{r}} d\mathbf{r}$, where t is the time of the image frame, $S_n(\mathbf{r})$ is the sensitivity map of the nth coil, $I(\mathbf{r},t)$ is the desired image sequence. We propose to reconstruct $I(\mathbf{r},t)$ from highly under-sampled $(\mathbf{k},t)$-space data by jointly imposing partial separability and spatial-spectral sparsity constraints. The optimization problem can be written as:

$$\hat{\mathbf{I}} = \underset{I(\mathbf{r},t)\in\{\sum_{\ell=1}^{L}\xi_\ell(\mathbf{r})\phi_\ell(t)\}}{\arg\min} \|\mathbf{d} - \mathrm{E}\{\mathbf{I}\}\|_2^2 + \lambda\|\Phi\{\mathbf{I}\}\|_1 ,$$

Where $\mathrm{E}\{\cdot\}$ is an imaging operator that incorporates sparse sampling and parallel imaging, $\Phi\{\cdot\}$ is a temporal Fourier transform operator that maps $I(\mathbf{r},t)$ to $\tilde{I}(\mathbf{r},f)$, and $\|\Phi\{\mathbf{I}\}\|_1$ is a measurement of spatial-spectral sparsity of the image sequence [5]. An algorithm based on half-quadratic regularization with continuation procedure has been proposed to solve this optimization problem [5].

A composite data acquisition scheme has been developed based on the above reconstruction strategy. In this scheme, $(\mathbf{k},t)$-space is highly undersampled in a way that two data sets are acquired: a navigator data set with high temporal resolution and an imaging data set with high spatial resolution. As illustrated in Fig. 1, we obtain the navigator data set with a spiral trajectory in order to obtain distributed k-space coverage while capturing the temporal dynamics of the image sequence. The temporal basis functions $\{\phi_\ell(t)\}_{\ell=1}^{L}$ in the above problem are extracted from this data set using singular value decomposition [3]. The imaging data set is acquired with Cartesian trajectories.

We performed our experiment on a Siemens Trio 3 T scanner with a 12-channel receiver coil. A spiral FLASH sequence (TE = 1.4 ms) and a Cartesian FLASH sequence (TE = 2.3 ms) were developed to acquire data over a 280 mm × 280 mm × 40 mm FOV. Repetitive /za/-/na/-/za/ sounds were acquired from one native English speaker at normal speaking rate. The reconstructions had a matrix size of 128 × 128 × 5, a spatial resolution of 2.2 mm × 2.2 mm × 8.0 mm and a frame rate of ~20 fps for the whole stack of 5 slices (TR = 9.78 ms/slice).

RESULTS

Figure 2 illustrates one mid-sagittal slice at the production of the /n/ in the /na/ sound. The reconstructed sequence clearly demonstrates contact of the tongue tip with the hard palate (alveolar ridge). Figure 3 depicts the formation of an airstream between the tongue tip and the palate with the mid-sagittal slice at the /z/ part of the /za/ sound. This is not observed in other sagittal slices. Figure 4 further reveals "tongue grooving" with images of coronal section posterior to the tongue tip. Obvious concavity and convexity in tongue tip curvature are depicted at /z/ in the /za/ sound and /n/ in the /na/ sound.

CONCLUSION

This work proposes a novel method for multi-slice dynamic speech imaging combining parallel imaging methods, a composite sparse acquisition scheme and a reconstruction strategy exploiting the partial separability and the spatial-spectral sparsity of the speech image sequence. The proposed method enables visualization of spatiotemporal speech dynamics in multiple imaging planes at 2.2 mm × 2.2 mm × 8.0 mm spatial resolution with a frame rate of 20 fps in all 5 slices. This method can lead to precise assessment of oropharyngeal dynamics and vocal tract dynamics in speech.

ACKNOWLEDGMENT

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

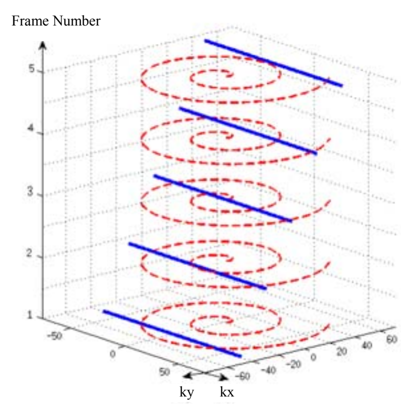

Fig. 1. The $(\mathbf{k},t)$-space sampling strategy that combines a spiral trajectory navigator and a Cartesian trajectory for image encoding. Both are acquired at each frame for each slice.

Fig. 2. Mid-sagittal slice clearly demonstrates soft-tissue structures in the oropharyngeal region during production of the /na/ sound.

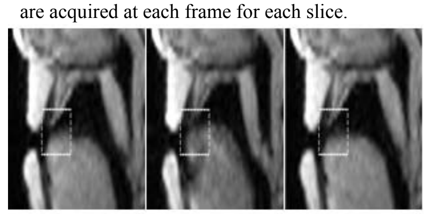

Fig. 3. The tongue tip contacts the palate during /n/ in the /na/ sound while an airstream exists over the center of the tongue during /z/ in the /za/ sound.

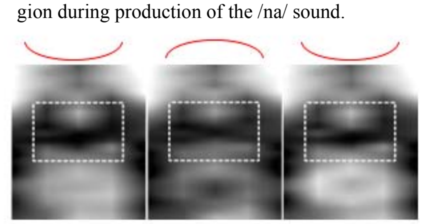

Fig. 4. Images of the coronal section posterior to the tongue tip reveal "tongue grooving". Tongue tip curvature is shown in red for reference.

REFERENCE

[1] Zhu, Y. et al. ISMRM 2011. pp. 4355. [2] Sutton, B. et al. JMRI 2010. pp. 1228-37. [3] Liang, Z.-P. IEEE-ISBI 2007. pp. 988-991. [4] Lustig, M. et al. MRM 2007. pp. 1182-1195. [5] Zhao, B. et al. IEEE-EMBS 2010. pp. 3390-3393.