Application of compressed sensing to 3D imaging of the vocal tract for speech MRI

Y-C. Kim[1], J-F. Nielsen[1], S. Narayanan[1], D. Byrd[2], and K. S. Nayak[1]

[1]Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, United States, [2]Department of Linguistics, University of Southern California, Los Angeles, CA, United States

Introduction: Three-dimensional (3D) MR imaging of the vocal tract provides full anatomical information of the vocal tract shaping during speech production [1], but it requires prohibitively long scan time. Previously, 3D imaging of the vocal tract has been performed using a multi-slice image acquisition that consists of acquiring three sets of multiple axial, 45° oblique, and coronal slices, all of which are prescribed from a mid-sagittal scan plane [2]. In this abstract, the feasibility of true 3D imaging of the vocal tract is investigated. To reduce scan time, k-space is undersampled several-fold, and images are reconstructed using the principle of compressed sensing MRI (CS-MRI) [3]. Sustained fricative consonants are tested to demonstrate the effectiveness of the method with an acceleration factor of 3, resulting in a total scan time of 17 sec.

Methods: *(Data acquisition)* Experiments were performed on a GE Signa 3T scanner using a single channel transmit/receive head coil. Imaging parameters were: gradient echo sequence, flip angle = 6°, TR = 5.7 msec, NEX = 1, spatial resolution = 1.0x1.0x3.0 mm^3, FOV = 18x18x15 cm^3. A 15 cm axial excitation slab was used. The readout axis was aligned with the A-P (anterior-posterior) direction. Four different imaging protocols were used: (I) fully sampled (180x180x50 acquisition matrix), (II) 3-fold undersampling, (III) 4-fold undersampling, and (IV) 5-fold undersampling. (ky,kz) space sampling patterns were designed as follows: Two independent and uniformly distributed random numbers corresponding to k-radius and azimuthal angle were each generated to create random (ky,kz) location in polar form. Resulting sampling pattern was similar to 2D Gaussian distribution. From the randomly chosen samples, nearest (ky,kz) grids were selected as sampling points. Protocol (I) (full sampling) was used to image a static posture (mouth held open) lasting 51 sec, whereas each of the undersampled protocols were used to image the vocal shaping during production of the English fricative consonants /s/, /ʃ/, and /θ/.

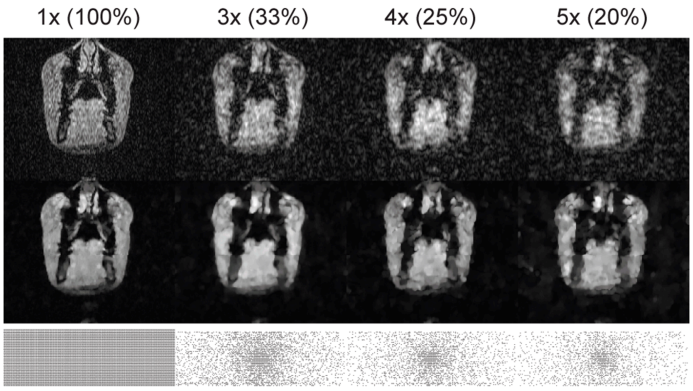

Figure 1: *Sub-sampling result from a representative coronal slice. Simple IFFT reconstruction (top row) and TV iterative reconstruction (middle row). The iterative reconstruction preserves air-tissue boundaries and improves SNR. Background noise due to either aliasing or low SNR has been reduced with the iterative reconstruction. Distribution of associated (ky,kz) variable density random samples used for the sub-sampling test (bottom row).*

(Image reconstruction) Images were reconstructed using total variation (TV) regularized iterative reconstruction, in order to preserve edges such as air-tissue boundaries, reduce aliasing artifacts due to undersampling in k-space, and improve SNR qualitatively [3]. Quantitative evaluation was performed by considering images reconstructed with the iterative reconstruction from fully sampled data as true images and then taking the difference of the true images from the images reconstructed from the sub-sampled data. 3D visualization of tongue shape was realized by manually segmenting tongue ROI in each TV reconstructed image of the coronal slices, stacking the segmented slices, and finally performing 3D volume rendering process.

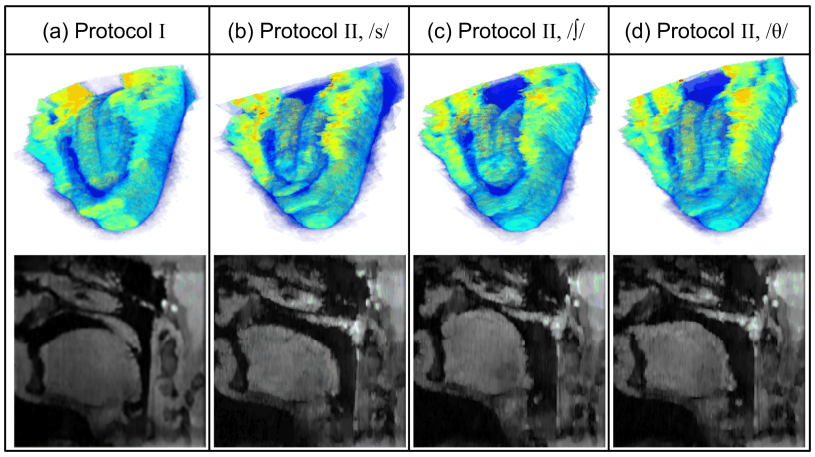

Figure 2: *3D visualization of tongue shape that is constructed after tongue ROI selection when (a) its mouth is opened and (b-d) sounds are produced with English fricative consonants (b)/s/, (c)/ʃ/, and (d)/θ/. (a) is obtained using full k-space data, and (b)-(d) are obtained with three-fold accelerated imaging. Corresponding reconstructed 2D mid-sagittal images are shown.*

Results: Fig. 1 shows images reconstructed from data acquired with Protocol (I). On the upper left is the result of performing a direct IFT on the entire (fully sampled) data set, which serves as the gold standard. Also shown are images reconstructed using TV regularized iterative reconstruction from only a subset of the acquired data. In each case, the subset used for reconstruction is identical to the subset acquired in Protocols (II,III,IV). Fig. 1 shows that the iterative reconstruction produces substantially reduced aliasing artifacts compared to simple IFFT reconstruction in which missing data are filled with zeros. Normalized root mean square errors for the iteratively reconstructed images from 3x, 4x, and 5x sub-sampled data were 1.93, 5.78, and 7.62, respectively. Fig. 2 shows 3D visualization of tongue shape and 2D mid-sagittal slice reconstructed from data acquired with Protocols (I) and (II). The groove of the tongue surface is clearly observed from the 3D images, but it is not seen in a mid-sagittal view (bottom row of Fig. 2). With three-fold accelerated scan, i.e., 17 seconds in scan time, 3D tongue shape is clearly seen in all fricative sounds.

Discussion: The drawback of the regularized iterative reconstruction from the 3D data was the prohibitively long (~6 hours) reconstruction time. Total variation norm has been effective in improving the overall image quality, but higher acceleration factors with reasonable quality of 3D vocal tract imaging may be achieved with more elaborate choices of the sparsifying basis and its associated k-space sampling schemes.

References: [1] Narayanan et al. J.Acoust.Soc.Am., 98:1325-1347, 1995, [2] Engwall, Speech Comm., 41:303-329, 2003, [3] Lustig et al. MRM, 2007 (in press).