Reconstructing 3D dynamics based on complementary 2D acquisitions: a preliminary case study on speech imaging

Xiaoguang Lu[1], Peter Speier[2], Hasan Cetingul[1], Marie-Pierre Jolly[1], Michaela Schmidt[2], Mariappan Nadar[1], Frank Sauer[1], and Edgar Mueller[2]
[1]Corporate Technology, Siemens Corporation, Princeton, New Jersey, United States, [2]Siemens AG, Erlangen, Germany

Purpose: MRI has been widely used to study anatomical structures and their dynamics. While segmented and navigated 3D imaging is available for some applications, most existing MRI studies still rely on 2D slices for practical reasons, among them ease of use, robustness, flexible balance between spatial and temporal resolutions, and total exam time. Fluoroscopic 2D image streams can be generated with frame rates up to 30Hz, which is fast enough for freeze framing most macroscopic anatomical motion. Due to the sequential acquisition of the 2D slices, all dynamic processes must be taken into account when constructing consistent, time-resolved 3D image sequences. We propose an automatic method to selectively align and synchronize densely acquired 2D slices into a sequence of 3D volumes through iterative non-rigid registration based on correlation analysis with anchor images.

Methods: We acquire two orthogonal stacks of 2D slices. Then we choose slices from one stack as anchors to reconstruct consistent volumes by selecting and aligning slices from the other stack. A series of consecutive anchors cover one repetition of the dynamics. Each anchor corresponds to a different dynamic phase and is used to reconstruct one 3D volume in a sequence representing the anatomical dynamics. The methodology is generally applicable to all repeatable freeze frameable dynamic processes. Here we demonstrate it for 3D speech dynamics imaging. One volunteer was asked to repeatedly pronounce 'blah' while two stacks of slices with quasi-continuous coverage were acquired, one in sagittal (SAG) and one in axial (TRA) orientation. We used a fluoroscopic 2D radial bSSFP sequence prototype with Compressed Sensing reconstruction (flip angle = 45deg, frame rate = 27Hz, resolution = $1.82 \times 1.82 \times 7mm^3$, slice shift=0.05mm/frame). From the SAG acquisition, a set of consecutive slices were selected as the anchors, which were located around the middle line of the face and covering one repetition of the speech paradigm. Normalized correlation was calculated for each TRA slice at the intersection line segment with the SAG anchor. The algorithm selected the TRA slices that were best correlated with the SAG anchor within a small temporal neighborhood corresponding to the average duration of the speech paradigm. Residual misalignment among selected TRA slices was iteratively corrected through non-rigid registration[1], optimizing the correlation with the SAG anchor. The deformations applied during optimization were constrained by the neighboring selected TRA slices. A 3D volume was obtained by applying scattered interpolation onto an isotropic grid based on the corrected TRA slices. Overall, we reconstructed a time series of volumes according to the set of consecutive anchors.

Results: Figure 1 shows MPRs of the reconstructed volume sequence at two different speech phases. The reconstructed temporal resolution is 27 frames/s, and the inter-slice distance between selected TRA slices (used to reconstruct a volume) is (2.20 ± 0.52) mm. The isotropic 3D grid of each volume is $1.82 \times 1.82 \times 1.82mm^3$.
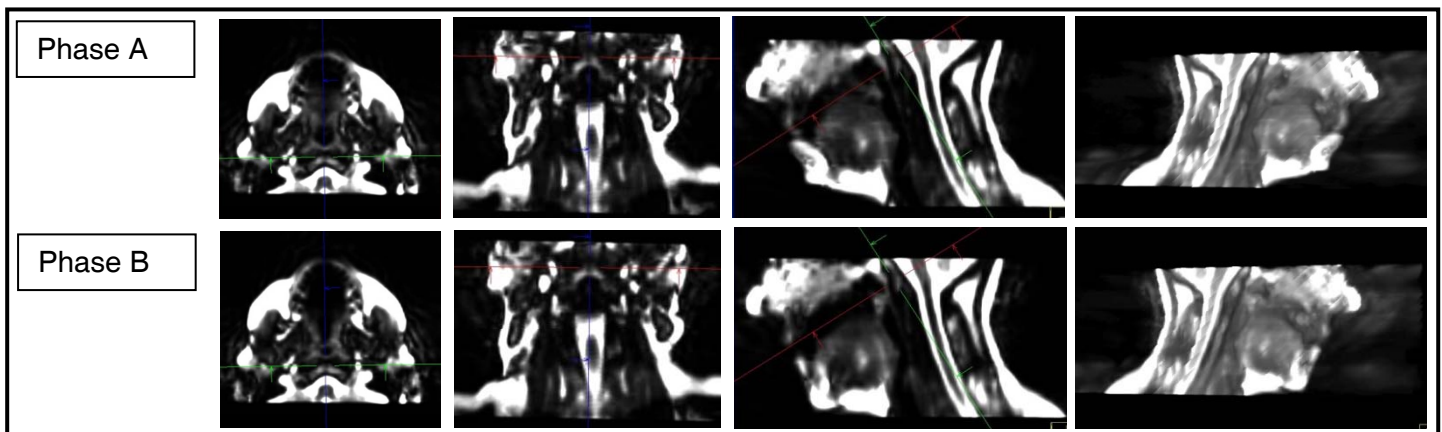

Figure 1. Reconstruction results at two different phases. For each phase, three MPRs and one 3D rendering (right-most) of the volume are presented. Window levels were adjusted to highlight tongue and structures surrounding the vocal tract.

Discussion: The experimental results demonstrate that the proposed method is capable of reconstructing consistent 3D freeze frames of the speech dynamics, capturing the dynamics (such as tongue, jaw, etc.) in 3D and in correspondence to the speech pattern ("blah" in this study) with a good balance between spatial and temporal resolution. This method was extended from a technique designed for 3D cardiac cine reconstruction[2]. However, the acquisition protocol was based on bSSFP with typical cardio-optimized parameters, and turned out to provide sub-optimal contrast for visualization of tongue and throat tissues. Therefore the reconstructed image quality is sub-optimal, but significant improvement can be expected by tailoring the acquisition and reconstruction parameters to the specific application, as done already for 3D cardiac cine. Super-resolution algorithms can be applied for further improvement[3].

Conclusion: The proposed 3D dynamics imaging method is designed for general applicability. The preliminary experimental results demonstrate its capability and potentials. To increase the spatial resolution in the slice normal direction the presented approach can also be applied to reconstruct a volume sequence from SAG slices using TRA anchors in this case study. Both volume sequences reconstructed from SAG and TRA slices can then be combined to generate a single volume sequence with an improved spatial resolution. For each specific application, the acquisition protocols should be adjusted to reach an optimal solution. We plan further evaluation based on more datasets and additional applications.

References: [1] Guetter et al., IEEE ISBI, 2011, P.590. [2] Accepted by SCMR 2014 (anonymized for review). [3] Zhou et al., JASA 133(6): P439.