31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Dynamic Imaging of Speech and Swallowing with MRI

Bradley P. Sutton, Charles Conway, Youkyung Bae, Cornelius Brinegar, Zhi-Pei Liang, David P. Kuehn

Abstract— **Dynamic imaging with MRI holds great promise for visualizing soft tissue structures in the oropharyngeal region during speech and swallowing studies. However, MRI suffers from historically slow acquisition speed and sensitivity to significant magnetic susceptibility differences in this region. In this work, we describe our efforts in creating high temporal resolution, serial acquisitions of the muscles of the oropharyngeal region. We describe our imaging approach that leads to acquisition rates of up to 21 frames per second. Additionally, we compare the serial acquisition scheme to gated acquisitions that suffer from temporal blur due to limited repeatability of the dynamic action.**

I. Introduction

DYNAMIC imaging with MRI holds great promise for visualizing soft tissues in the oropharyngeal region during swallowing and speech. Proper function of the muscles in this region can have significant impact on self-sustaining life functions for swallowing and quality of life issues in speech. Although the two application areas of speech and swallowing differ in the desired diagnostic assessments, dynamic visualizations of soft tissue contrast are necessary to study the normal and pathological biomechanics of many of the same muscles and structures, including the tongue, velum, levator muscle, and pharyngeal wall.

However, there are many significant challenges to be overcome in order for MRI to impact these application areas. First, the gold standard for swallowing studies is x-ray video fluoroscopy, which can easily provide images at a rate of hundreds of frames per second. In contrast, MRI studies have provided serial acquisitions at a rates up to 10 fps [1]. Although providing fast, high signal images, a video fluoroscopic study does not allow for unobstructed visualization of the soft tissue structures as it provides a sagittal projection through the oropharyngeal region, with high contrast bone structures impeding the viewing of underlying soft tissue structures. Additionally, video fluoroscopy involves ionizing radiation, which can limit research studies. While MRI allows for visualization of arbitrary planes of soft tissue structures, it continues to suffer from frame rates that are too slow to visualize muscles in motion. As an example, retraction of the velum by the levator veli palatini muscle is important in both speech and swallowing events to separate the nasal and oral cavities. Retraction can occur in as little as 100 ms, meaning that frame rates of 10 fps would not provide visualizations of the motion, only of the endpoints of the motion.

High imaging frame rates have been achieved in MRI by gating the acquisition. Gated acquisitions acquire parts of the data over several repetitions of the motion and reconstruct a time series of images showing the average motion. This is in contrast to serial acquisitions that acquire all the data for each image sequentially, providing a complete time series of the motion over only one repetition of the motion, albeit at lower frame rates. In [2], mid-sagittal movies were made using gating for /pa/, /ta/, and /ka/ sounds showing the movements of the velum at 30 frames per second in cleft palate subjects. The subjects repeated the sounds 128 times in order to obtain the entire data set to construct one average movie of the articulation. The speech samples were kept simple in order to maximize reproducibility across the repetitions. In fact, NessAiver et al. reported significant variations in the repeatability of simple complete words, such as “golly” [3]. They showed variations of 50 – 95 ms in onsets of auditory features in both syllables. This inconsistency across repetitions can significantly reduce the ability to resolve temporal events and spatial structures when using a gating acquisition, and hence results in a reduction in the effective temporal and spatial resolutions.

Another acquisition challenge for MRI of the oropharyngeal region is magnetic susceptibility. Magnetic susceptibility of air and tissue differ significantly resulting in large magnetic field inhomogeneity around air/tissue interfaces. This is especially problematic for speech imaging of the tongue and small structures such as the velum. The tongue, velum, and palate may be in contact at their resting position, with no air/tissue interfaces between them. When active, these three structures separate, resulting in air/tissue interfaces surrounding the surfaces of interest. The new air/tissue interfaces result in large deviations in the magnetic field in the region.

Differences in magnetic susceptibility result in image distortions that are dependent on the pulse sequence used to acquire the images. For the spiral acquisitions used in the

Manuscript received April 7, 2009.

B. Sutton and C. Conway are with the Bioengineering Department at the University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (Corresponding author B. Sutton: 217-244-5154; fax: 217-265-0246; e-mail: bsutton@illinois.edu).

Y. Bae and D. Kuehn are with the Speech and Hearing Science Department at the University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA.

C. Brinegar and Z.-P. Liang are with the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA.

current work, the distortion is a radial blurring of the point-spread-function. For a small structure such as the velum, this can result in difficulties in visualizing the entire structure since this blurring will be significant along all of its surfaces when viewed mid-sagittally.

Several additional acquisition challenges exist for the application areas. For swallowing, MRI provides challenges for the subject in the constrained positioning required for performance of the task. The subject must often be placed in a supine position. However, studies have shown that gravity can have effects on the behavior of muscles in both swallowing [4] and speech [5]. The significance of these differences for the functional performance is still an ongoing area of research. Finally, for speech imaging with co-recorded acoustic signals, there is a significant challenge in accurately aligning the audio track with the resulting MRI time series. We have developed a custom time-code generation software to align the MRI images very accurately, within 10 μs.

In the current work, we will demonstrate a high serial frame rate acquisition using a spiral FLASH sequence with a reduced field of view acquisition. Additionally, we will compare the effective temporal resolution between the serial and gated acquisitions using the same data set. This will demonstrate the temporal blur introduced by variability in the timing of the repetitions of the motions. We will also show preliminary results of incorporating spatial temporal models into the acquisition and reconstruction. Examples will be shown from both speech and swallowing studies.

II. Materials and Methods

We use a multi-shot spiral FLASH acquisition to optimize the imaging rate and take advantage of the full gradient performance of our Siemens 3 T Allegra MRI scanner. The spiral acquisition was designed using the analytical design method of [6] for a regular density multi-shot spiral and maximum gradient amplitude of 40 mT/m and a maximum slew rate of 400 mT/m/ms.

As mentioned above, magnetic susceptibility causes image-blurring distortions with spiral acquisitions. The magnitude of blurring is dependent on the length of the data acquisition readout, ie. the duration of a shot of the spiral acquisition. Magnetic susceptibility distortions can be reduced through two mechanisms: multiple shots and reduced field of view techniques. Multishot acquisitions require several separate acquisitions, or shots, to cover the entire data space to make an image. Although using a high number of shots will reduce blurring from magnetic susceptibility, it will also reduce the frame rate of our acquisition by requiring repetition of all the pulse sequence elements (RF pulse, slice select gradient and refocuser and spoiler gradients) in each shot. This results in a tradeoff between image acquisition rate and image quality through susceptibility distortions.

Another mechanism to reduce the data readout and magnetic susceptibility distortions is through the use of a reduced field of view or outer volume suppression (OVS) technique, similar to techniques used in cardiac MRI [7]. With OVS, static portions of the image are saturated. Those regions do not contribute to the dynamic signal and do not need to be encoded. By reducing the field of view for dynamic imaging, the data sampling requirements are lessened and shorter readouts can be achieved with the same number of shots.

For this work, we used a 6-shot acquisition and a magnetic susceptibility corrected reconstruction [8, 9]. The inclusion of magnetic field inhomogeneity correction in the reconstruction allowed us to obtain the imaging speed of a 6-shot acquisition but with the image quality of a 12-shot acquisition, as shown in Figure 1. The magnetic susceptibility reconstruction required a field map, which was obtained by alternating the echo time of sequential images by 0.5 ms. This was necessary to get an accurate field map with the dynamically changing air/tissue interfaces. As an example of the dynamic nature of the field map, Figure 2 shows the field map in two states of the dynamic acquisition during a speech study.

The spiral acquisition sequence that was used in the current study was a 6-shot acquisition with TE alternating between 0.9 and 1.4 ms (for dynamic field map estimation), TR was 6.7 ms and a single 6 mm slice was acquired in the midsaggital plane with an in-plane resolution of 1.875 mm and a field of view of 12 cm. The natural frame rate for this acquisition was approximately 21 fps.

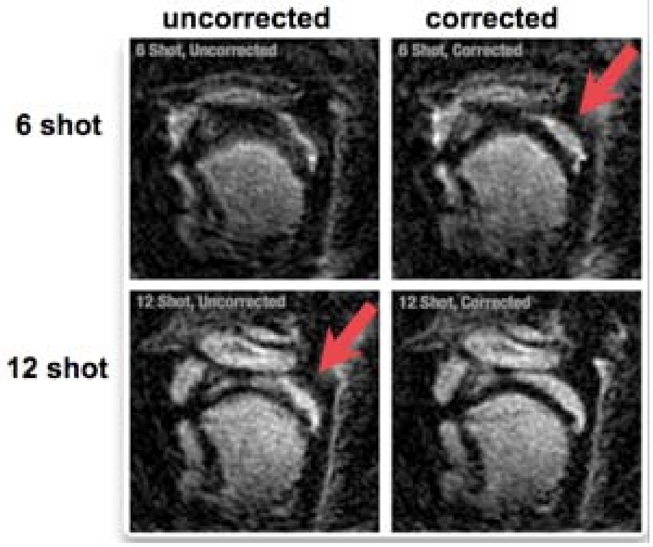

Figure 1: Comparison of 6 and 12-shot acquisitions with and without susceptibility-corrected reconstructions.

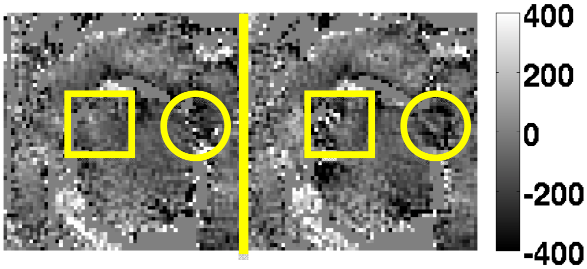

Figure 2: Field map in Hz with the velum and tongue tip shown at two different orientations during a dynamic speech imaging experiment. The velum is in the circle and the tongue tip is in the square box

Both speech and swallowing experiments will be shown. For speech experiments, subjects were instructed to repeat the speech fragment /zanaza/ several times, paced by a metronome playing through the headphones with beeps at a rate of 2 Hz. The subjects transitioned syllables with every metronome beat, followed by a skip of 0.5s, repeating the 3 syllable phrase every 2 seconds.

To allow for comparison of the serial acquisition with a gated acquisition, the speech fragment was repeated 11 times. For the serial reconstruction, images were reconstructed across sequential groups of 6 shots to yield the acquired frame rate of 21 fps. These 6 shots in the serial reconstructions come from the same repeat of the speech sample, with no mixing of data from different repeats and no interpolation. For the retrospectively gated reconstruction, the closest 6 shots to the desired frame time (also set to be 21 fps) across all repetitions were found. This results in a random distribution of shots across all eleven repetitions; however, the temporal spread of the shots relative to the desired frame timing is very small.

III. Results

We analyzed the average temporal blur of both the serial and the gated image reconstructions for the speech imaging data. The average temporal blur for the serial acquisition and reconstruction was 37.6 ms with an average temporal separation of images of 46.7 ms. Note that temporal blur is defined by the time window over which the data for an image is collected.

In contrast to this, the gated acquisition resulted in an average absolute timing difference between the center of the desired image and the center of each ADC in the image of only 1.72 ms with a maximum difference of 4.2 ms. The temporal blur is slightly larger as the length of the data readouts were 4.27 ms, making the average temporal blur approximately 6 ms. For the gated acquisition, the temporal separation was set by the desired frame rate of 21 fps to match the serial acquisition. Figure 3 shows the distribution of shots across repetitions for one of the dynamic images.

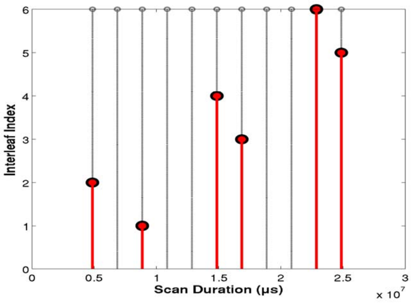

Figure 3: Distribution of the 6 shots across the repetitions of the speech sample. The red lines indicate the temporal location across repeats for each shot with the height indicating the shot number. The gray lines show timings of the desired frame time across repeats.

The temporal blurs above describe only the sequence timing contributions to the temporal blur in the data. Additional temporal blur resulted in the gated acquisitions due to deviations in timing of the execution of aspects of the motions in the repeats. Since the gated acquisition results in an average motion over all of the repeats, any deviations in timing of the motion will show up as temporal blur. This is visualized in Figure 4 and Figure 5, which show a dynamic time series of images from both the serial and gated acquisitions, respectively. The images shown span the first /za/ of the speech fragment. Notice that the delineations of the tongue and velum are much sharper in the serial acquisition, representing a better capture of the temporal dynamics of the motion.

In Figure 6 we show examples of dynamic images acquired during a swallow of a bolus of water that was delivered via a straw.

IV. Discussion

As can be seen in Figure 5, although gated acquisitions provide the potential of increased temporal resolution over serial acquisitions, the repeatability of the task must be analyzed. The effective temporal resolution will be a combination of both the acquisition temporal resolution and the temporal alignment of repeats of the motion. Serial acquisitions allow for more natural motions. They also allow a subject to perform a task at a self-determined pace and without much prompting or practice. This will result in visualizing motions that are more natural.

Future work for this project will include spatiotemporal models in the reconstruction to allow for increased frame rates. In the spatiotemporal models, navigator data can be used to provide information about the dynamics of the data and this information can be used to train a model for reconstruction [10, 11]. The potential result is an increase in frame rate equivalent to the number of shots. However, an additional navigator acquisition must be added to the sequence for each shot. Modification of the sequence presented in this work dropped the frame rate only slightly,

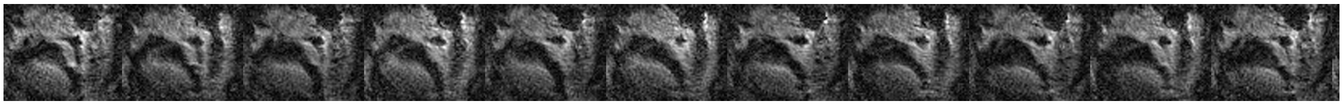

Figure 4: Dynamic imaging of speech sample /za/ with serial acquisition, showing images from a single repetition of the speech fragment.

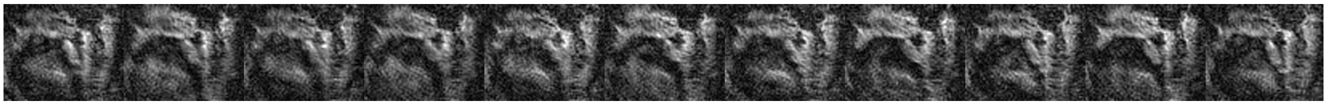

Figure 5: Dynamic imaging of speech sample /za/ with gated acquisition using 11 repetitions of the speech fragment.

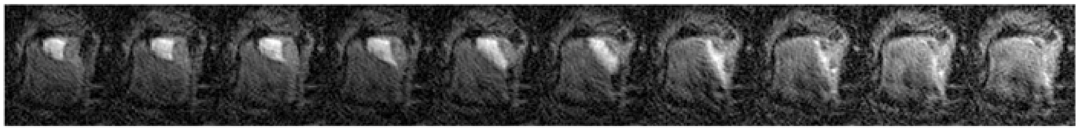

Figure 6: Dynamic imaging of swallowing a water bolus.

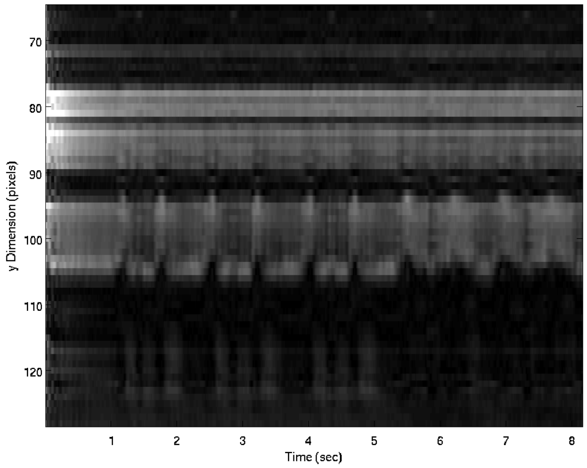

Figure 7: Spatiotemporal reconstruction applied to speech imaging. A strip of pixels through the velum is shown as a function of the acquisition time.

down to 18 fps. With the 6-shot acquisition, the spatiotemporal reconstruction can provide over 100 fps in the image reconstruction. For a preliminary speech experiment, a time series is shown in Figure 7 for a strip through the velum with this spatiotemporal reconstruction applied.

V. Conclusion

There are still several challenges to be addressed in order for MRI to become a useful research and diagnostic tool in biomechanical studies of the oropharyngeal region in speech and swallowing. Natural speech and swallowing functions can suffer from a lack of repeatability, necessitating serial acquisitions instead of gated acquisitions that can suffer additional blur from deviations in the timing of the execution of the action across several repetitions. However, we have shown that adequate frame rates and good image quality are possible using serial acquisitions with MRI.

VI. References

1. Narayanan, S., K. Nayak, S. Lee, A. Sethy, and D. Byrd. An approach to real-time magnetic resonance imaging for speech production. *The Journal of the Acoustical Society of America*, 2004. 115(4): p. 1771-1776.
2. Shinagawa, H., T. Ono, E. Honda, S. Masaki, Y. Shimada, I. Fujimoto, T. Sasaki, A. Iriki, and K. Ohyama. Dynamic analysis of articulatory movement using magnetic resonance imaging movies: methods and implications in cleft lip and palate. *Cleft Palate Craniofac J*, 2005. 42(3): p. 225-30.
3. NessAiver, M.S., M. Stone, V. Parthasarathy, Y. Kahana, and A. Paritsky. Recording high quality speech during tagged cine-MRI studies using a fiber optic microphone. *J Magn Reson Imaging*, 2006. 23(1): p. 92-7.
4. Honda, Y. and N. Hata. Dynamic imaging of swallowing in a seated position using open-configuration MRI. *J Magn Reson Imaging*, 2007. 26(1): p. 172-6.
5. Stone, M., G. Stock, K. Bunin, K. Kumar, M. Epstein, C. Kambhamettu, M. Li, V. Parthasarathy, and J. Prince. Comparison of speech production in upright and supine position. *J Acoust Soc Am*, 2007. 122(1): p. 532-41.
6. Glover, G.H. Simple analytic spiral K-space algorithm. *Magn Reson Med*, 1999. 42(2): p. 412-415.
7. Le Roux, P., R.J. Gilles, G.C. McKinnon, and P.G. Carlier. Optimized outer volume suppression for single-shot fast spin-echo cardiac imaging. *J Magn Reson Imaging*, 1998. 8(5): p. 1022-32.
8. Noll, D.C., C.H. Meyer, J.M. Pauly, D.G. Nishimura, and A. Macovski. A homogeneity correction method for magnetic resonance imaging with time-varying gradients. *IEEE Trans Med Imaging*, 1991. 10(4): p. 629-637.
9. Schomberg, H. Off-resonance correction of MR images. *IEEE Trans Med Imaging*, 1999. 18(6): p. 481-495.
10. Liang, Z.-P. Spatiotemporal imaging with partially separable functions, in *International Symposium on Biomedical Imaging*. 2007. p. 988-991.
11. Brinegar, C., Y.-J.L. Wu, L. Foley, T. Hitchens, Q. Ye, C. Ho, and Z.-P. Liang. Real-time cardiac MRI without triggering, gating, or breath holding, in *IEEE Engineering in Medicine and Biology Conference*. 2008. p. 3381-3384.