Visualization of Vocal Tract Shape Using Interleaved Real-time MRI of Multiple Scan Planes

Yoon-Chul Kim, Michael I. Proctor, Shrikanth S. Narayanan, Krishna S. Nayak

Department of Electrical Engineering, University of Southern California, CA, USA

yoonckim@usc.edu, mproctor@usc.edu, shri@sipi.usc.edu, knayak@usc.edu

Abstract

Conventional real-time magnetic resonance imaging (RT-MRI) of the upper airway typically acquires information about the vocal tract from a single midsagittal scan plane. This provides insights into the dynamics of all articulators, but does not allow for visualization of several important features in vocal tract shaping, such as grooving/doming of the tongue, asymmetries in tongue shape, and lateral shaping of the pharyngeal airway. In this paper, we present an approach to RT-MRI of multiple scan planes of interest using time-interleaved acquisition, in which temporal resolution is compromised for greater spatial coverage. We demonstrate simultaneous visualization of vocal tract dynamics from midsagittal, coronal, and axial scan planes in the articulation of English fricatives.

Index Terms: speech production, vocal tract shaping, real-time MRI, spiral imaging, English fricatives

1. Introduction

Real-time magnetic resonance imaging (RT-MRI) has proven to be a viable non-invasive method with which to capture the shaping of the vocal tract during speech production [1-4]. Current methods for RT-MRI of the upper airway typically involve imaging a single midsagittal slice. Midsagittal imaging encompasses the entire vocal tract from the lips to the glottis. However, it does not visualize several important features in vocal tract shaping such as grooving/doming of the tongue, asymmetries in tongue shape, and lateral shaping of the pharyngeal airway. Recent studies of vocal tract dynamics using RT-MRI involved the acquisition of a single coronal slice, after the acquisition of a single midsagittal slice [5]. The studies investigated characteristics of the lateral tongue shape such as grooving/doming of the tongue, which is not available from midsagittal but from coronal slice, in articulatory studies of English fricatives under different vowel-consonant-vowel contexts.

The current study is motivated by the observation that it would be desirable to acquire vocal tract dynamics from multiple slices (e.g., midsagittal, coronal, axial) in real time, to allow for visualization of tissue dynamics in all image planes. Multi-slice RT-MRI methodology has been previously applied to parallel multi-slice imaging of the beating heart [6]. Recently, Sutton and colleagues demonstrated interleaved RT-MRI of a parallel axial multi-slice imaging of the brain and a single midsagittal slice imaging of the upper airway [7]. The technique made it possible to monitor swallowing during functional MRI (fMRI) acquisition so as to allow for retrospective discarding of "corrupted" fMRI data resulting from the subject's physiological motion. The application of RT-MRI to vocal tract imaging through simultaneous imaging of a midsagittal and coronal slice, for example, would reveal vocal tract dynamics from two different views although it sacrifices temporal resolution.

In this paper, we introduce a slice-interleaving technique, which utilizes rapid scan plane switching rate and thus is well suited to simultaneous multi-slice real-time imaging. We also demonstrate the advantage of imaging intersecting slices (e.g., midsagittal and axial slices). We apply the technique to the articulatory study of English fricatives – a class of consonants for which detailed knowledge of the precise configuration of the vocal tract in all dimensions is critical to a proper understanding of speech production [8,9].

2. Methods

2.1. MR Imaging Setup

MRI experiments were performed at the Los Angeles County Hospital on a Signa Excite HD 1.5 T scanner (GE Healthcare, Waukesha, WI) with gradients capable of 40 mT/m amplitudes and 150 mT/m/ms slew rates. A body coil was used for radiofrequency (RF) transmission. A custom 4-channel upper airway receiver coil array was used for RF signal reception. In the 4-channel receiver coil array, two coil elements are anterior and the other two coil elements are posterior to the head and neck. Each subject was screened and provided informed consent in accordance with institutional policy.

The subjects lay in supine position inside the MRI magnet bore, and their upper airways were imaged using custom real-time imaging software [10]. The head of each subject was immobilized by padding it with memory foam to minimize head motion. Spiral fast gradient echo pulse sequence was used for real-time imaging [3,4,5]. Imaging parameters were: slice thickness = 6 mm, flip angle = 15°, receiver bandwidth = ±125 kHz. We used two imaging protocols (see Table 1 for other imaging parameters). Note that temporal resolution is similar in both protocols but the number of slices and the spatial resolution are different, illustrating trade-offs between the number of scan planes and spatial resolution.

2.2. Interleaved RT-MRI

Figure 1(a) illustrates the temporal ordering of the slice/spiral-interleaf in the case of interleaved RT-MRI of two slices. Note that the slice-switching occurs at every repetition time (TR). TR was approximately 6 ms in our setup. This 1-TR slice alternation degrades temporal resolution by a factor of the number of slices. However, it has advantages of i) performing virtually simultaneous imaging of multiple scan planes and ii) being capable of increasing video frame rates via sliding
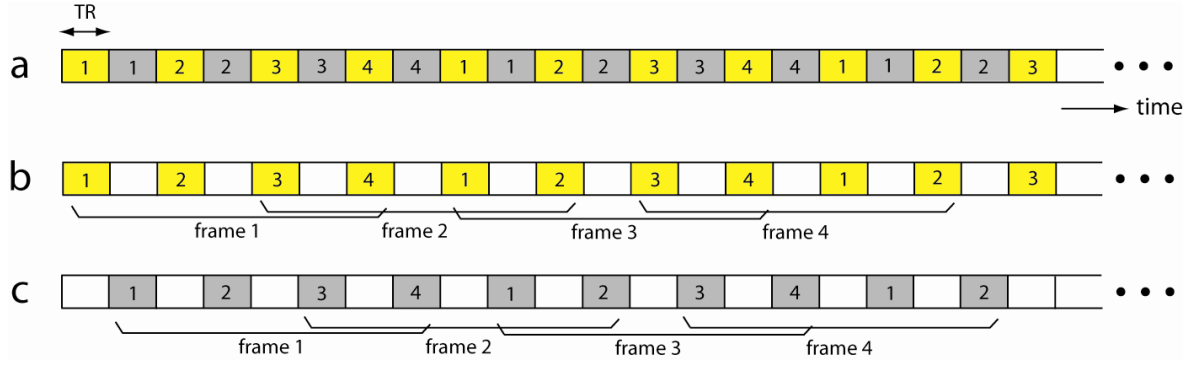
Figure 1: *Interleaved real-time MR imaging technique. (a) Slice and spiral-interleaf order is shown. Slice 1 (e.g., midsagittal) and Slice 2 (e.g., axial) are denoted by yellow and gray colors within block, respectively. Spiral interleaf index is indicated by number within each block. Generalization to more than two slices is straightforward. Slice switching occurs at every TR. The case for 4-interleaf spiral imaging is illustrated. 13- (Protocol 1 in Table 1) and 9-interleaf (Protocol 2 in Table 1) spirals were used in the in-vivo experiments. Frame reconstruction by sliding window technique (see Ref. [3]) is shown for (b) Slice 1 and (c) Slice 2. Frames 1, 2, 3, and 4 in (b) and (c) are practically simultaneous.*

window technique (i.e., adjacent image frames overlapped in time). Two (or three) slices of interest were prescribed and saved during real-time localization. The slice orientation information obtained from the localization was used for the actual interleaved RT-MRI acquisitions.

2.3. Stimuli/Audio Setup

Imaging data were acquired from two healthy subjects while they produced intervocalic fricative consonants. English fricative pairs /f/-/v/, /θ/-/ð/, /s/-/z/, /ʃ/-/ʒ/ were elicited in symmetrical maximally-contrastive vocalic contexts /a_a/, /i_i/, and /u_u/ [11] using pseudo-word stimuli. Subject M1 is a male native speaker of Australian English. Subject V2 is a male native speaker of Indian English (also a native speaker of Tamil).

Stimuli were presented in-scanner, using a mirror-projector setup. Each subject uttered the pseudo-word stimuli at a normal speech rate. Audios were recorded synchronously at 20 kHz sampling rate during RT-MRI acquisition. MRI-gradient acoustic noise signals were canceled during post-acquisition processing [12].

	Protocol 1	Protocol 2
Number of spiral interleaves	13	9
Temporal resolution of each slice	156 ms	163 ms
In-plane spatial resolution	2.4×2.4 mm^2	3.0×3.0 mm^2
Imaging field of view	20×20 cm^2	20×20 cm^2
Image pixel dimension	84×84	68×68
TR (repetition time)	6.004 ms	6.028 ms
Number of slices	2 (midsagittal/ axial)	3 (midsagittal/ coronal/axial)

Table 1: *Imaging parameters tested in our in-vivo experiments in the articulatory study of English fricatives, demonstrating trade-offs among temporal resolution, spatial resolution, and number of slices.*
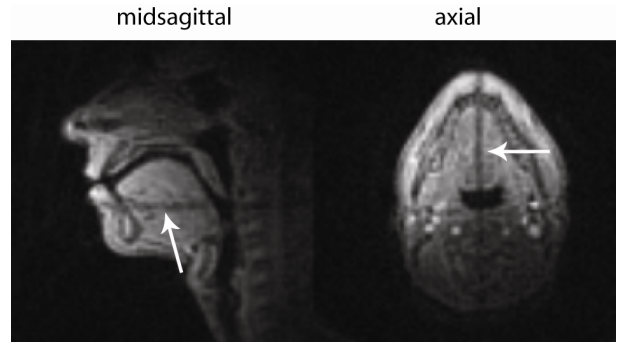

Figure 2: *Example midsagittal/axial frame captured using Protocol 1 when Subject M1 was stationary. Partially saturated spins (see the arrows) indicate where one slice is located from the other slice.*

2.4. Video Reconstruction

Gridding reconstructions were performed to reconstruct image frames [13] individually for each slice. Root sum-of-squares reconstructions from the two anterior coil elements were performed to obtain final images. The reconstructed MRI videos were synchronized with the noise-cancelled audios [4]. The synchronization facilitated the qualitative analysis of the vocal tract shapes.

3. Results

Figure 2 illustrates a single time frame captured when two-slice interleaved RT-MRI acquisition was applied. Note the partial saturation of spins (i.e., the dark line in each slice) in the soft tissue, which provides useful indication of the relative position of the slices. The saturation effect is attributed to the fact that the spins in the intersection between the two scan planes undergo more frequent (i.e., every 1-TR) tip-downs by the RF excitation pulses than the spins which are not in the intersected region. The midsagittal slice image indicates that the axial slice is prescribed to be located in the middle of the pharyngeal airway and be perpendicular to the pharyngeal wall. The axial slice image indicates that the midsagittal scan plane lies in the midline of the vocal tract in that axial section.
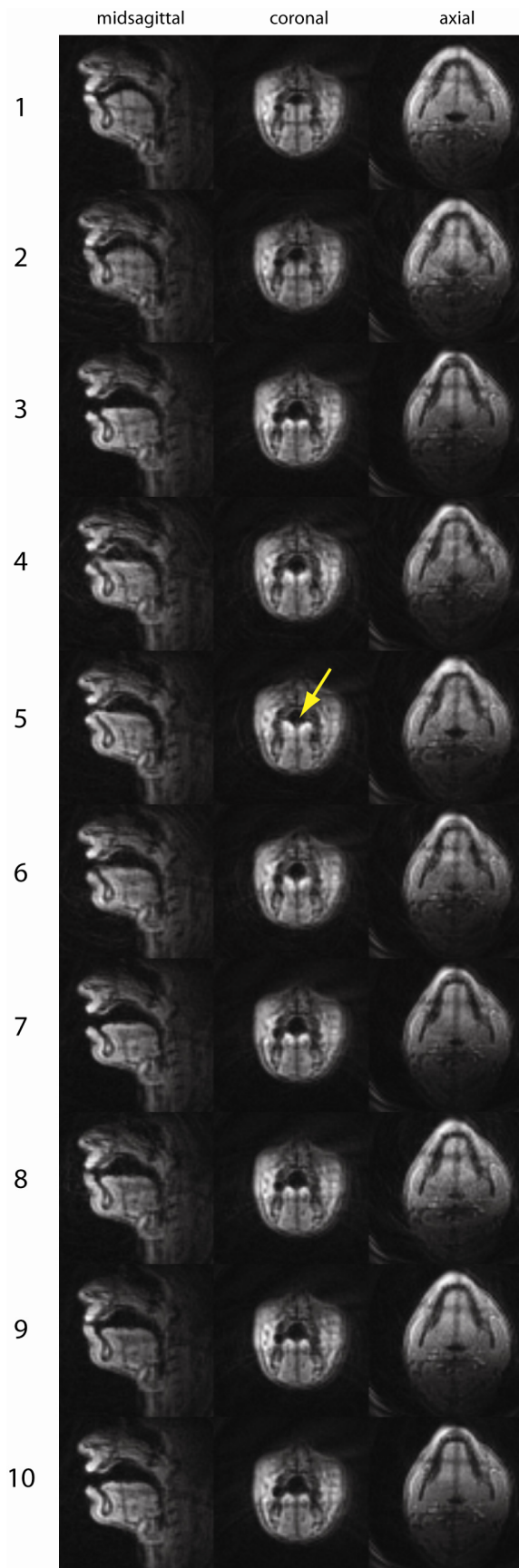
	midsagittal	coronal	axial

Figure 3: *Representative midsagittal (left column), coronal (center), and axial (right) frames captured using Protocol 2 during the utterance /baθaba/ by Subject V2. Time evolution is from top to bottom.*

Figure 3 illustrates the change in vocal tract formation in the utterance /baθaba/ by Subject V2. The data clearly reveal the location of the tongue tip constriction for the dental fricative /θ/ (see the midsagittal slice image in frame 5) and the mid-lingual tongue grooving in the coronal plane during the production of /θ/ in the low vowel context (see the yellow arrow). The intensity from the tongue surface is bright (see the coronal slice images in frames 3-9) due to the through-plane motion of the tongue tip in coronal slice imaging. Tongue tip closure towards, and release from the dental target during the production of /θ/ can be seen in the midsagittal image sequence.

Figure 4 compares three-slice (midsagittal / coronal / axial) image frames captured from Subject V2, during the production of English fricatives /s/, /z/, /θ/, and /ð/ in vocalic contexts /a_a/, /i_i/, /u_u/. From the coronal slice images, the highest degree of tongue grooving is observed in /a_a/. The axial imaging orientation reveals that the largest lateral expansion in the pharyngeal airway is observed in the front vowel context of /i_i/. The midsagittal slice images show the tongue tip constriction locations for each fricative pair to be very similar in each vowel context, but reveal the highly context-dependent nature of the tongue body shaping during fricative production. The data also reveal the slightly increased area of the pharyngeal airways during production of voiced fricatives in comparison to their voiceless equivalents, consistent with previous findings [8,9].

4. Discussion

Interleaved RT-MRI of multiple scan planes has been shown to provide valuable information about lingual articulation and vocal tract shaping, compared to conventional single-slice midsagittal RT-MRI. The proposed technique allows for more detailed analysis of consonant production using more natural stimuli, without requiring subjects to sustain articulatory postures for artificially long durations, as is typically required in conventional static three-dimensional MR imaging studies [8,9].

The proposed RT-MRI technique compromises temporal resolution for greater spatial coverage. Compared to conventional single midsagittal slice imaging, the technique produced increased artifacts as evidenced by temporal motion blurring from spiral acquisition. Noting the fact that the spatial range of the excited spins is smaller in axial and coronal slices than midsagittal, reduced field-of-view imaging can be applied to axial and coronal slices to improve temporal resolution in these slices. Golden-ratio spiral temporal interleaf ordering scheme can also be applied to the proposed technique [14]. This can facilitate a more flexible retrospective selection of temporal resolution and imaging field-of-view, and potentially improve overall image quality.

Under the current approach, configuration of multiple scan plane geometries is performed prior to scanning, which does not accommodate interactive adjustment of scan planes on the fly. Real-time adjustments of the scan planes from knowledge of location of the partially saturated spins can provide more accurate prescription of the scan planes. The implementation remains as future work.

Our RT-MRI setup is flexible in terms of slice selection. It is straightforward to apply multiple parallel sagittal slice imaging, as originally described in [15] (note that they did not demonstrate it in time-interleaved manner). Multiple parallel sagittal images acquired using this technique could provide data for the extraction of highly accurate vocal tract area functions. Note that parallel sagittal slice imaging does not
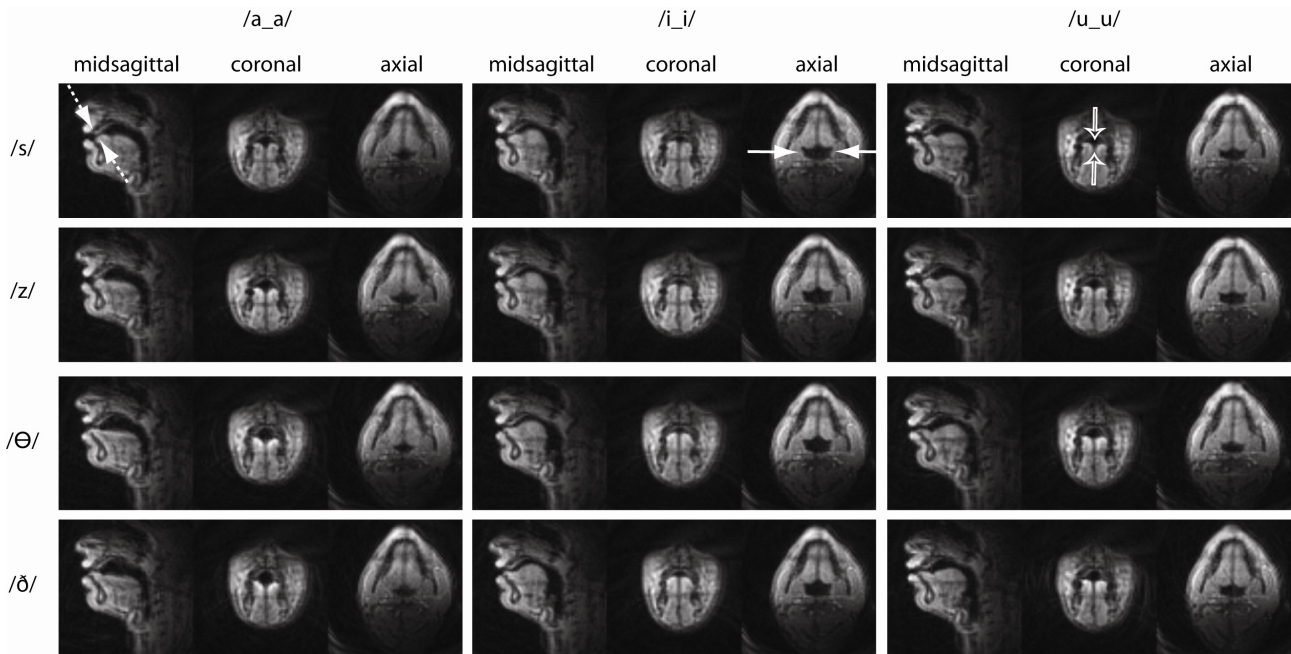
Figure 4: *Three-slice frames captured using Protocol 2 during intervocalic coronal fricative productions. Top row: voiceless alveolar fricative; 2nd row: voiced alveolar fricative; 3rd row: voiceless dental fricative; Bottom row: voiced dental fricative; Each scan plane provides unique vocal tract shaping information that is not available from the other scan planes. For example, the midsagittal slice shows the constriction locations (see the dashed arrows), the coronal slice shows the degree of tongue grooving (see the hollow arrows), and the axial slice shows the degree of lateral expansion of the pharyngeal airway (see the solid arrows).*

exhibit the traces (i.e., partial saturation of the spins) that are seen in imaging of multiple intersecting slices.

5. Conclusions

We have applied interleaved real-time MRI of multiple slices to dynamic imaging of vocal tract during natural speech production. The proposed approach can provide additional information about vocal tract shape features that can not be seen using only a single slice, and may be useful for articulatory studies of multiply-articulated consonants such as fricatives and liquids, where detailed knowledge about the global shaping of the tongue is critical to our understanding of the goals of speech production.

6. Acknowledgements

7. References

[1]   Demolin, D., Hassid, S., Metens, T., Soquet, A., "Real-time MRI and articulatory coordination in speech", Comptes Rendus Biologies, 325(4):547-556, 2002.

[2]   Sutton, B.P., Conway, C., Bae, Y., Seethamraju, R., Kuehn, D.P., "Faster dynamic imaging of speech with field inhomogeneity corrected spiral fast low angle shot (FLASH) at 3 T", J Magn Reson Imaging, 32:1228-1237, 2010.

[3]   Narayanan, S., Nayak, K., Lee, S., Sethy, A., Byrd, D., "An approach to real-time magnetic resonance imaging for speech production", J Acoust Soc Am, 115(4):1771-1776, 2004.

[4]   Bresch, E., Kim, Y-C., Nayak, K.S., Byrd, D., Narayanan, S., "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging", IEEE Signal Processing Magazine, 123-132, 2008.

[5]   Bresch, E., Riggs, D., Goldstein, L., Byrd, D., Lee S., Narayanan, S., "An analysis of vocal tract shaping in English sibilant fricatives using real-time magnetic resonance imaging", Interspeech, 2008, p2823-2826

[6]   Nayak, K.S., Pauly, J.M., Nishimura, D.G., Hu, B.S., "Rapid ventricular assessment using real-time interactive multislice MRI", Magnetic Resonance in Medicine, 45:371-375, 2001.

[7]   Sutton, B.P., Conway, C.A., Kuehn, D.P., "Simultaneous monitoring of tongue tip movements in functional MRI motor tasks for speech and swallowing studies", Proceedings of ISMRM, Hawaii, USA, April 2009, p20.

[8]   Proctor, M.I., Shadle, C.H., Iskarous, K., "Pharyngeal articulation in the production of voiced and voiceless fricatives", J Acoust Soc Am, 127(3):1507-1518, 2010.

[9]   Narayanan, S., Alwan, A., Haker, K., "An articulatory study of fricative consonants using magnetic resonance imaging", J Acoust Soc Am, 98(3):1325-1347, 1995.

[10] Santos, J.M., Wright, G.A., Pauly, J.M., "Flexible real-time magnetic resonance imaging framework", Proceedings of the 26th Annual Meeting of IEEE EMBS, 47:1048-1051, 2004.

[11] Öhman, S.E.G., "Coarticulation in VCV utterances: Spectrographic measurements", J Acoust Soc Am, 39:151-168, 1966.

[12] Bresch, E., Nielsen, J., Nayak, K.S., Narayanan, S., "Synchronized and noise-robust audio recordings during real-time MRI scans", J Acoust Soc Am, 120(4):1791-1794, 2006.

[13] Jackson, J., Meyer, C.H., Nishimura, D., Macovski, A., "Selection of convolution function for Fourier inversion using gridding", IEEE Trans Med Imaging, 10:473-478, 1991.

[14] Kim, Y-C., Narayanan, S., Nayak, K.S., "Flexible retrospective selection of temporal resolution in real-time speech MRI using a golden-ratio spiral view order", Magnetic Resonance in Medicine, 65(5):1365-1371.

[15] Shadle, C.H., Mohammad, M., Carter, J.N., Jackson, P.J.B., "Multi-planar dynamic magnetic resonance imaging: new tools for speech research", International Congress of Phonetics Sciences, 1999.