# Improved Imaging of Lingual Articulation Using Real-Time Multislice MRI

**Yoon-Chul Kim, PhD**[*], **Michael I. Proctor, PhD**, **Shrikanth S. Narayanan, PhD**, and **Krishna S. Nayak, PhD**

Ming Hsieh Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089-2564

## Abstract

**Purpose**—To develop a real-time imaging technique that allows for simultaneous visualization of vocal tract shaping in multiple scan planes, and provides dynamic visualization of complex articulatory features.

**Materials and Methods**—Simultaneous imaging of multiple slices was implemented using a custom real-time imaging platform. Midsagittal, coronal, and axial scan planes of the human upper airway were prescribed and imaged in real-time using a fast spiral gradient-echo pulse sequence. Two speakers of native English produced voiceless and voiced fricatives /f/-/v/, /θ/-/ð/, /s/-/z/, /ʃ/-/ʒ/ in symmetrical maximally contrastive vocalic contexts /a_a/, /i_i/, and /u_u/. Vocal tract videos were synchronized with noise-cancelled audio recordings, facilitating the selection of frames associated with production of English fricatives.

**Results**—Tongue grooving was observed from the coronal slice intersecting the post-alveolar region to be most pronounced during fricative production in back vowel contexts, and more pronounced for sibilants /s/-/z/ than for /ʃ/-/ʒ/. The axial slice best revealed differences in dorsal and pharyngeal articulation; voiced fricatives were observed to be produced with a larger cross-sectional area in the pharyngeal airway. Partial saturation of spins provided accurate location of imaging planes with respect to each other.

**Conclusion**—Real-time MRI of multiple intersecting slices can provide valuable spatial and temporal information about vocal tract shaping, including details not observable from a single slice.

## INTRODUCTION

Real-time magnetic resonance imaging (RT-MRI) has proven to be a valuable 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 covers the entire vocal tract from the lips to the glottis, but it does not provide any information about articulation in parasagittal regions, such as grooving/doming of the tongue, asymmetries in tongue shape, and lateral shaping of

[*]Correspondence to: Yoon-Chul Kim, Ph.D., 3740 McClintock Ave, EEB 400, University of Southern California, Los Angeles, CA 90089-2564, Phone: (213) 740-0530, Fax: (213) 740-4651, yoonckim@usc.edu.

the pharyngeal airway – information which is critical to our understanding of the goals of speech production.

For example, the articulation of English fricatives (consonants produced with a turbulent airstream), typically involves the formation of a groove in the tongue to channel the airstream into a small constriction. Different fricative consonants are produced with fine differences in the shape and length of the tongue groove, as well as the location and size of the constriction (5,6). Previous studies used three-dimensional (3D) MRI to investigate the geometry of the entire vocal tract (or 3D tongue shape), but the imaging typically involved long scan times, making these techniques unsuitable for capturing the dynamics of vocal tract shaping during fluent speech. Such techniques are limited to the acquisition of static vocal tract shapes during artificially sustained production of continuant speech sounds such as vowels (7), fricatives (5,6) and liquid consonants (8–10). To investigate vocal tract shaping during natural speech production, a recent RT-MRI study used a single coronal slice, acquired before or after a single midsagittal slice (11). The study investigated lateral tongue shaping in a coronal slice during the production of English fricatives /s/ and /ʃ/ in the vocalic contexts /i_i/, /a_a/, /i_a/, /a_i/. The single-slice RT-MRI technique, however, was unable to visualize features observable in the other imaging orientation at the same time. For example, it was not possible to monitor the location of the tongue tip constriction from the coronal imaging slice alone during production of /s/.

The current study is motivated by the observation that simultaneous acquisition of multiple slices (e.g., midsagittal, coronal, axial) in real-time would allow for more complete visualization of vocal tract dynamics. Multislice RT-MRI technique is not new and has been demonstrated in a variety of applications. Multislice RT-MRI allowed for simultaneous visualization of the myocardial segments from apical, mid, basal short axis views in cardiac imaging (12). Interleaved RT-MRI with three orthogonal scan planes has been shown to provide a rapid and accurate localization of scan plane orientation on a moving subject in fetal imaging (13). RT-MRI with arbitrary multiple scan planes has been demonstrated on a parallel axial multislice imaging of the brain and a single midsagittal slice imaging of the upper airway (14). The technique enabled monitoring of uncued swallowing onsets during functional MRI acquisition. Recently, RT-MRI of one midsagittal and two axial slices in the upper airway has been demonstrated in patients with obstructive sleep apnea (15).

In this work, we apply multislice RT-MRI to the capture of vocal tract shaping during fluent speech. To the best of our knowledge, orthogonal multislice real-time MRI during speech has not been described. The proposed slice-interleaving technique utilizes rapid switching of the scan planes and thus is well suited to virtually simultaneous multislice real-time imaging of the vocal tract. We demonstrate the effectiveness of the technique in 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 (5,6,16).

## MATERIALS AND METHODS

MRI experiments were performed 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 (Photos illustrating experimental setup are available at http://sail.usc.edu/span/technology). Each subject was screened and provided informed consent in accordance with institutional policy.

Subjects lay in supine position inside the MRI magnet bore, and their upper airways were imaged using RTHawk (HeartVista, Inc., Los Altos, CA), a custom real-time imaging platform (17). The head of each subject was immobilized by padding it with memory foam to minimize head motion. Figure 1(a) illustrates the temporal ordering of the slice/spiral-interleaf in the case of three slices. Note that slice switching occurs at every repetition time (TR). In this work, TR was defined to be the time interval between two subsequent excitation pulses, and was approximately 6 ms in our setup. This 1-TR slice alternation degrades temporal resolution by a factor of the number of slices (see Figure 2). However, it has advantages of i) achieving virtually simultaneous imaging of multiple scan planes and ii) enabling the increase in frame rates by the use of the sliding window technique (i.e., adjacent image frames overlapped in time) (18). Multiple slices of interest were prescribed and their slice geometries were saved during real-time localization in RTHawk. Pulse sequences were programmed to allow for loading of the specified slices of interest, and to facilitate temporal interleaving during slice acquisition. The slices of interest were imaged in time-interleaved fashion during lingual articulation.

The spiral fast gradient echo pulse sequence consisted of slice-selective excitation, spiral readout, rewinder, and spoiler gradients (1,4,11). Imaging parameters were: slice thickness = 6 mm, flip angle = 15°, receiver bandwidth = ±125 kHz, field-of-view (FOV) = $20 \times 20$ cm$^2$. We used two imaging protocols: **(Protocol 1)** 2-slice (midsagittal/axial) interleaved imaging, 13-interleaf spiral, temporal resolution = 156 ms, in-plane resolution = $2.4 \times 2.4$ mm$^2$, image pixel dimension = $84 \times 84$, TR = 6.004 ms. **(Protocol 2)** 3-slice (midsagittal/oblique-coronal/axial) interleaved imaging, 9-interleaf spiral, temporal resolution = 163 ms, in-plane resolution = $3.0 \times 3.0$ mm$^2$, image pixel dimension = $68 \times 68$, TR = 6.028 ms. Note that temporal resolution is similar in both protocols but the number of slices and the spatial resolution are different.

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/ (see Table 1) were elicited in symmetrical maximally-contrastive vocalic contexts /a_a/, /i_i/, and /u_u/ (19) using pseudo-word stimuli. Subject 1 is an adult male native speaker of Australian English. Subject 2 is an adult 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. Audio recordings of subjects' speech were acquired synchronously at 20 kHz sampling rate during RT-MRI acquisition. MRI-gradient acoustic noise signals were canceled during post-acquisition processing (20).

Gridding reconstructions (21) were performed in Matlab (Mathworks, South Natick, MA) to reconstruct image frames for each slice. Root sum-of-squares reconstructions from the two anterior coil elements were performed to obtain final images. Parallel imaging was not used. The reconstructed MRI videos were synchronized with the noise-cancelled audio recordings (20). The synchronization facilitated the qualitative analysis of the vocal tract shapes.

## RESULTS

Figure 3 shows example image frames captured from Subject 1 when the two-slice interleaved RT-MRI protocol was applied. In Fig 3a, the axial slice reveals vocal tract shape features that are not seen in the midsagittal slice. The axial vocal tract cross-sectional shape is characterized by lateral expansion of the pharyngeal airway, lingual advancement, and formation of a groove in the tongue dorsum during the articulation of sibilant fricative /ʃ/.

Time-interleaved imaging of multiple intersecting slices resulted in the partial saturation of spins in the soft tissue (see the dark line in each slice in Fig 3). This allows for accurate

location of the imaging planes with respect to the vocal tract anatomy. 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 reveals that the axial slice is located in the middle of the pharyngeal airway and is 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. It is important to note that the saturation effect degrades the quality of visualization in the intersected region. For example, the saturation band seen in the axial slice in Fig 3a lies in the center of the tongue groove. This may result in overestimation of the degree of tongue grooving. It is anticipated that the saturation effect will be mitigated when T1 of the tissue is shorter or flip angle is lower.

Figure 4 compares two-slice (midsagittal/axial) image frames captured from Subject 1, during the production of English sibilant fricative pairs /ʃ/ and /ʒ/ in intervocalic contexts /a_a/, /i_i/, and /u_u/. Midsagittal images show that constriction locations, where the tongue blade is close to the post-alveolar region in the hard palate, are similar in voiceless /ʃ/ and voiced /ʒ/ and in all intervocalic contexts. Axial images reveal differences in vocal tract shaping between voiceless /ʃ/ and voiced /ʒ/ primarily characterized by the slightly increased area in the pharyngeal airway (see the arrows) during production of voiced fricative /ʒ/ in comparison to its voiceless equivalent /ʃ/, consistent with previous findings (5,6).

Figure 5 compares three-slice (midsagittal/oblique-coronal/axial) image frames captured during the production of English fricatives /f/, /θ/, /s/, and /ʃ/ in vocalic contexts /a_a/, /i_i/, and /u_u/ by Subject 2. For the low vowel context /a_a/, the coronal images reveal that tongue grooving in the post-alveolar region (hollow arrows) is most pronounced in the dental fricative /θ/, typically produced with the tongue tip contacting the back of the upper teeth, and more pronounced for sibilant /s/ than for /ʃ/ (consonants produced by forming constrictions in the alveolar and post-alveolar regions, respectively). 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, consistent with the findings of Shadle et al. (22). Axial slices reveal that vowel context effects on dorsal articulation and pharyngeal aperture are more pronounced for the labiodental fricative /f/ than for the post-alveolar fricative /ʃ/ (see solid arrows). Most importantly, because these data were acquired using the RT-MRI protocol, temporal changes in vocal tract formation can be monitored in all imaging slices (see Figure 6 and Movie 2 in the supplementary material). For example, tongue tip approximation towards, and subsequent release from the alveolar target during the production of /s/ is observed in the midsagittal image sequence, and the time-course of tongue groove (see the arrow in Fig 6) formation can be observed in the coronal image sequence.

## 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 detailed analysis of consonant production using 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 (5,6).

The proposed RT-MRI technique compromises temporal resolution for greater spatial coverage and higher signal-to-noise ratio (SNR) from longer $T_1$ recovery time. Compared to

conventional single midsagittal slice imaging, the technique can lead to increased motion artifacts in the image frames that correspond to the events for which there are rapid movements in the articulators.

In the imaging of intersecting slices, the spatial extent of the excited spins highly varies depending on the scan plane orientation. For example, the midsagittal slice of the upper airway typically has larger spatial extent of the excited spins than the axial slice. Our real-time acquisition setup is based on the use of a single spiral acquisition protocol when dealing with imaging multiple scan planes. When acquisition protocols are independently designed to be optimized for each scan plane, it will lead to more time-efficient sampling at each slice and thus potentially improve overall spatio-temporal resolution.

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. Since the setup is flexible in terms of slice selection, it should be straightforward to apply multiple parallel sagittal slice imaging, as originally described by Shadle et al. (23) (note that they did not demonstrate real-time imaging of parallel sagittal slices, and that their technique required many repetitions of a target utterance). Multiple parallel sagittal images could provide data for the extraction of dynamics in the vocal tract area functions. Note that parallel sagittal slice imaging does not exhibit the traces (i.e., partial saturation of the spins) that are seen in imaging of multiple intersecting slices.

In conclusion, time-interleaved real-time multislice MRI was applied to dynamic imaging of the vocal tract during 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 cross-linguistic studies of multiply-articulated consonants such as fricatives and liquids, where detailed knowledge about the shaping of the tongue in multiple orientations is critical to our understanding of speech production.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## REFERENCES

1. 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. 2004; 115(4):1771–1776. [PubMed: 15101655]

2. Sutton BP, Conway CA, Bae Y, Seethamraju R, Kuehn DP. Faster dynamic imaging of speech with field inhomogeneity corrected spiral fast low angle shot (FLASH) at 3 T. J Magn Reson Imaging. 2010; 32:1228–1237. [PubMed: 21031529]

3. Demolin D, Hassid S, Metens T, Soquet A. Real-time MRI and articulatory coordination in speech. Comptes Rendus Biologies. 2002; 325(4):547–556. [PubMed: 12161933]

4. Bresch E, Kim YC, Nayak KS, Byrd D, Narayanan SS. Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging. IEEE Signal Processing Magazine. 2008:123–132.

5. Narayanan SS, Alwan AA, Haker K. An articulatory study of fricative consonants using magnetic resonance imaging. J Acoust Soc Am. 1995; 98(3):1325–1347.

6. Proctor MI, Shadle CH, Iskarous K. Pharyngeal articulation in the production of voiced and voiceless fricatives. J Acoust Soc Am. 2010; 127(3):1507–1518. [PubMed: 20329851]

7. Baer T, Gore JC, Gracco LC, Nye PW. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. J Acoust Soc Am. 1991; 90(2):799–828. [PubMed: 1939886]

8. Alwan A, Narayanan S, Haker K. Toward articulatory-acoustic models for liquid consonants based on MRI and EPG data. Part II: The rhotics. J Acoust Soc Am. 1997; 101(1078–1089)

9. Narayanan S, Alwan A, Haker K. Toward articulatory-acoustic models for liquid consonants based on MRI and EPG data. Part I: The laterals. J Acoust Soc Am. 1997; 101:1064–1077. [PubMed: 9035398]

10. Zhou X, Espy-Wilson CY, Boyce S, Tiede M, Holland C, Choe A. A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. J Acoust Soc Am. 2008; 123(6):4466–4481. [PubMed: 18537397]

11. Bresch E, Riggs D, Goldstein L, Byrd D, Lee S, Narayanan SS. An analysis of vocal tract shaping in English sibilant fricatives using real-time magnetic resonance imaging. Proceedings of Interspeech. 2008:2823–2826.

12. Nayak KS, Pauly JM, Nishimura DG, Hu BS. Rapid ventricular assessment using real-time interactive multislice MRI. Magn Reson Med. 2001; 45:371–375. [PubMed: 11241692]

13. Neustadter DM, Chiel HJ. Imaging freely moving subjects using continuous interleaved orthogonal magnetic resonance imaging. Magn Reson Imaging. 2004; 22:329–343. [PubMed: 15062928]

14. Paine TL, Conway CA, Malandraki GA, Sutton BP. Simultaneous dynamic and functional MRI scanning (SimulScan) of natural swallows. Magn Reson Med. 2011; 65:1247–1252. [PubMed: 21360741]

15. Shin, LK.; Holbrook, AB.; Chang, CE.; Santos, JM.; Fischbein, NJ.; Capasso, R.; Kushida, CA. Real-time MRI with synchronous polysomnography of the upper airway in patients with obstructive sleep apnea. Proceedings of the 19th Annual Meeting of ISMRM, Montreal; 2011. (abstract 4634)

16. Shadle CH. The effect of geometry on source mechanisms of fricative consonants. Journal of Phonetics. 1991; 19:409–424.

17. Santos, JM.; Wright, GA.; Pauly, JM. Flexible real-time magnetic resonance imaging framework. Proceedings of the 26th Annual Meeting of IEEE EMBS; 2004. p. 1048-1051.

18. Riederer SJ, Tasciyan T, Farzaneh F, Lee JN, Wright RC, Herfkens RJ. MR fluoroscopy: Technical feasibility. Magn Reson Med. 1988; 8(1):1–15. [PubMed: 3173063]

19. Öhman SEG. Coarticulation in VCV utterances: spectrographic measurements. J Acoust Soc Am. 1966; 39:151–168. [PubMed: 5904529]

20. Bresch E, Nielsen J, Nayak KS, Narayanan S. Synchronized and noise-robust audio recordings during real-time MRI scans. J Acoust Soc Am. 2006; 120(4):1791–1794. [PubMed: 17069275]

21. Jackson J, Meyer CH, Nishimura D, Macovski A. Selection of convolution function for Fourier inversion using gridding. IEEE Trans Med Imaging. 1991; 10(3):473–478. [PubMed: 18222850]

22. Shadle CH, Tiede M, Masaki S, Shimada Y, Fujimoto I. An MRI study of the effects of vowel context on fricatives. Proc Inst of Acoust. 1996; 18(9):187–194.

23. Shadle CH, Mohammad M, Carter JN, Jackson PJB. Multi-planar dynamic magnetic resonance imaging: new tools for speech research. International Congress of Phonetics Sciences. 1999:623–626.
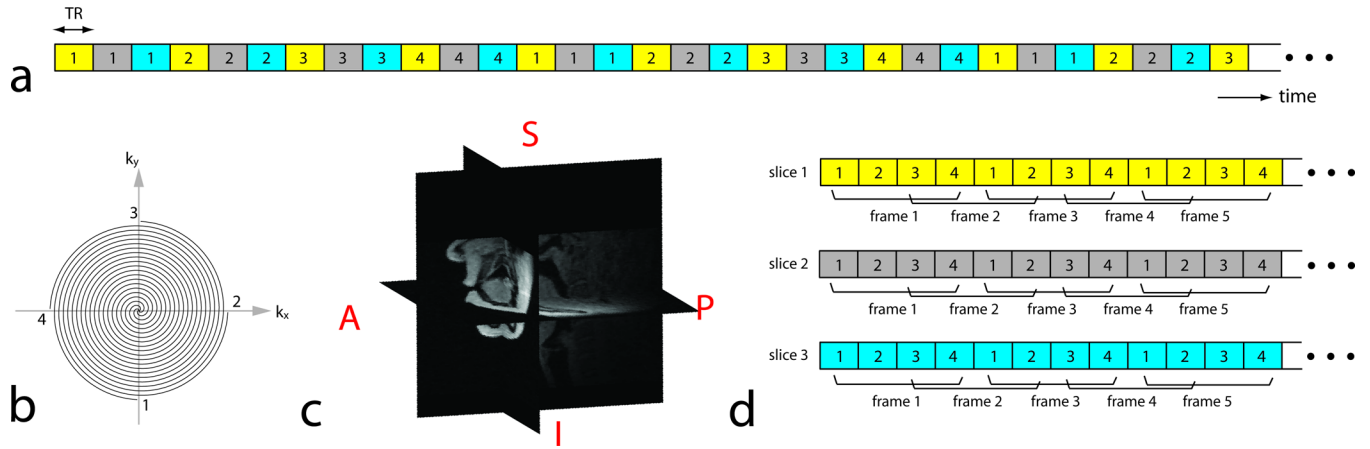
**Figure 1.**
Illustration of the proposed acquisition and reconstruction scheme in real-time MRI of the upper airway during speech. (a) Slice and spiral-interleaf order is shown. Slice 1 (yellow), slice 2 (gray), and slice 3 (blue) are acquired at every TR. Spiral interleaf index is indicated by number within each block. (b) 4-interleaf spiral trajectory with the index at the end of the spiral interleaf. 13- and 9-interleaf spirals were used in the in-vivo experiments. (c) Example midsagittal, coronal, and axial slice locations for imaging of the vocal tract. (d) Frame reconstructions for slice 1, 2, and 3. Sliding window reconstruction is used to increase frame rates.
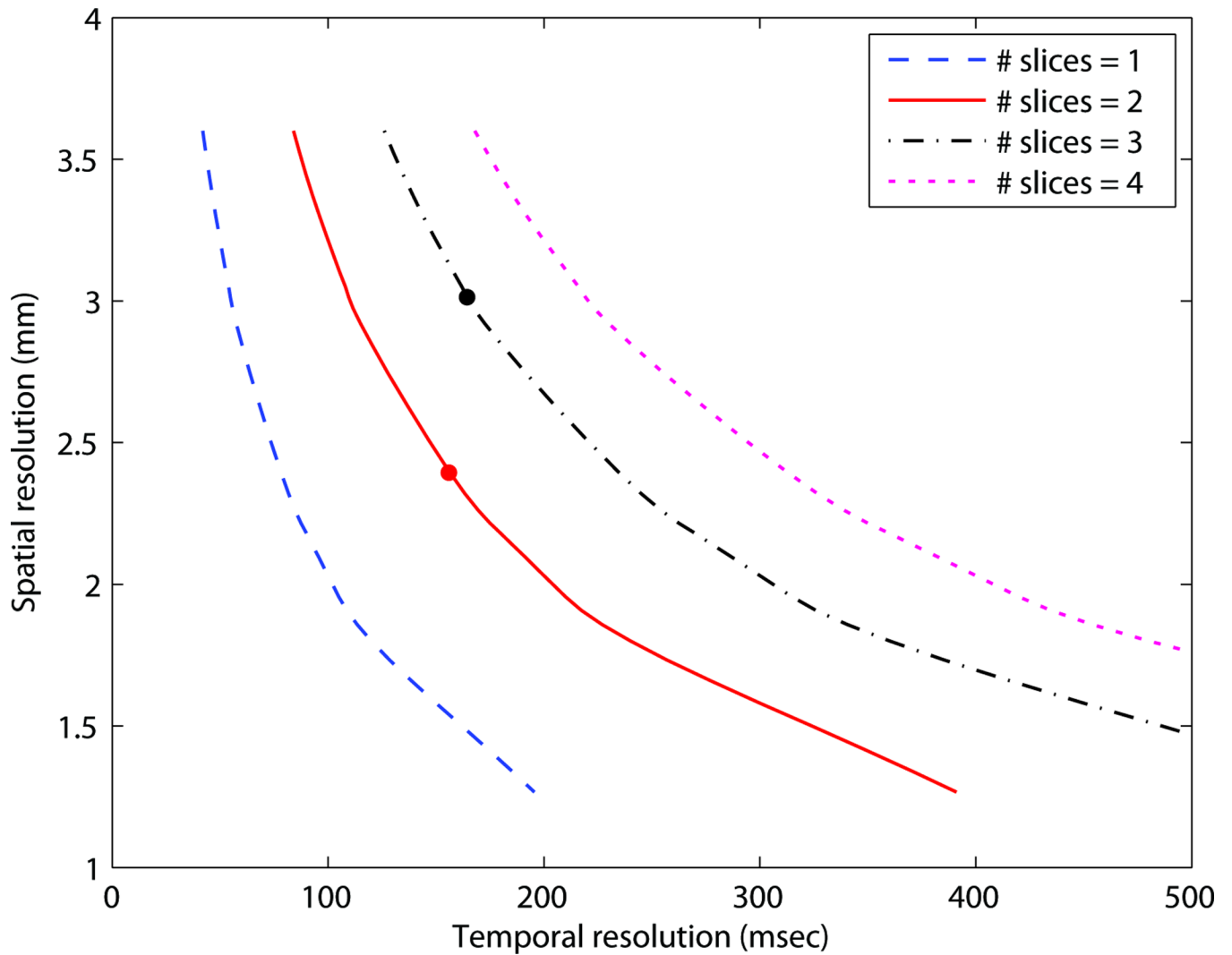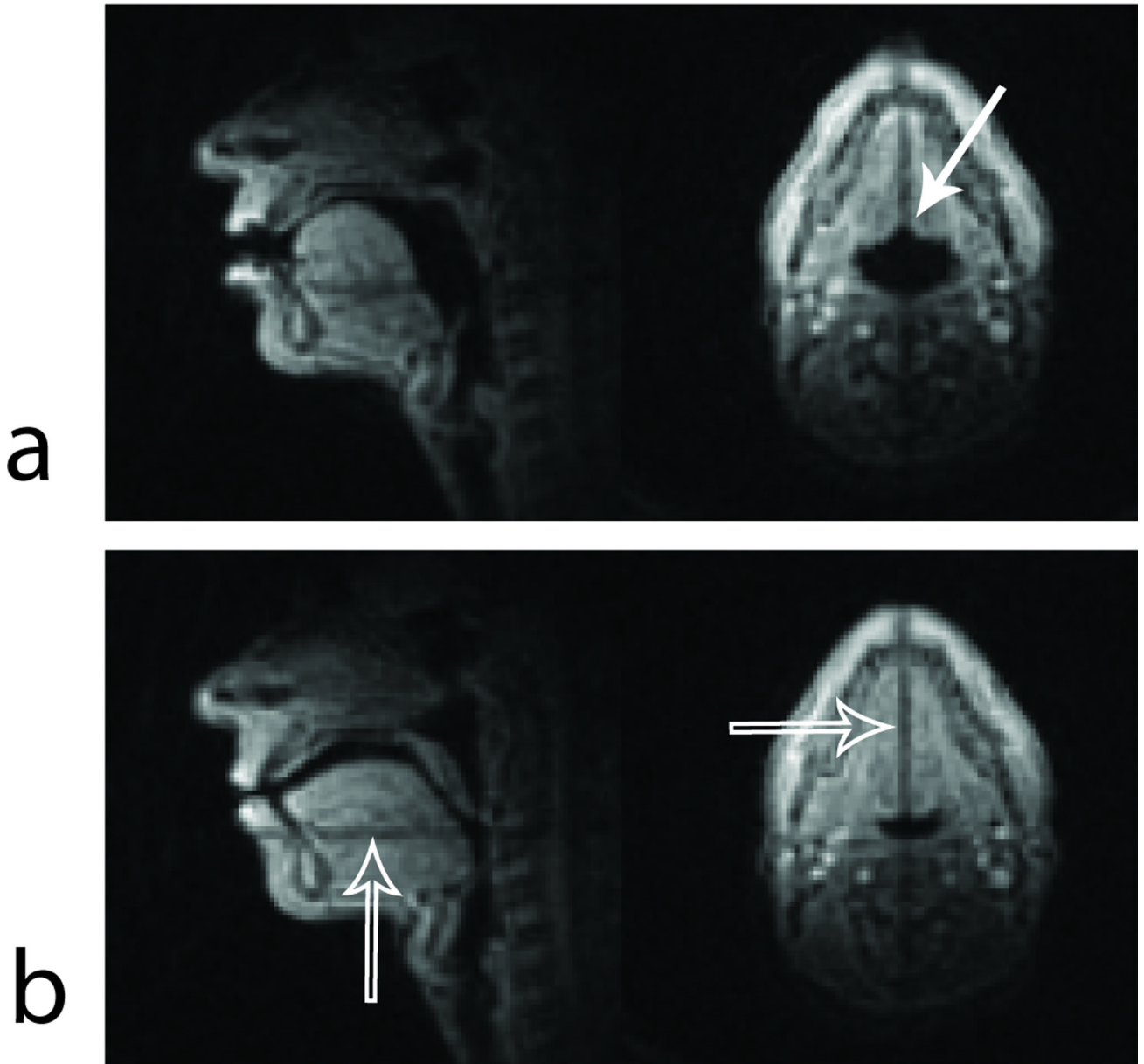
**Figure 2.**
Spatial-temporal resolution tradeoff. Interleaved spiral trajectories were designed with a field-of-view (FOV) of 20 cm and readout duration of approximately 2.5 msec. Spatial and temporal resolutions indicated by the red and black dots were used for time-interleaved imaging of 2 and 3 slices, respectively.

# midsagittal     axial



**Figure 3.**
Example midsagittal/axial frames captured using Protocol 1 for Subject 1 during (a) articulation of /ʃ/ in the front vowel context /i_i/ and (b) interspeech pause. The axial slice provides additional vocal tract information such as the cross-sectional area in the pharyngeal airway and grooving of the tongue (see the solid arrow in (a)). Partially saturated spins (see the hollow arrows in (b)) clearly indicate where one slice is located with respect to the other slice. See also Movie1 in the supplementary material.
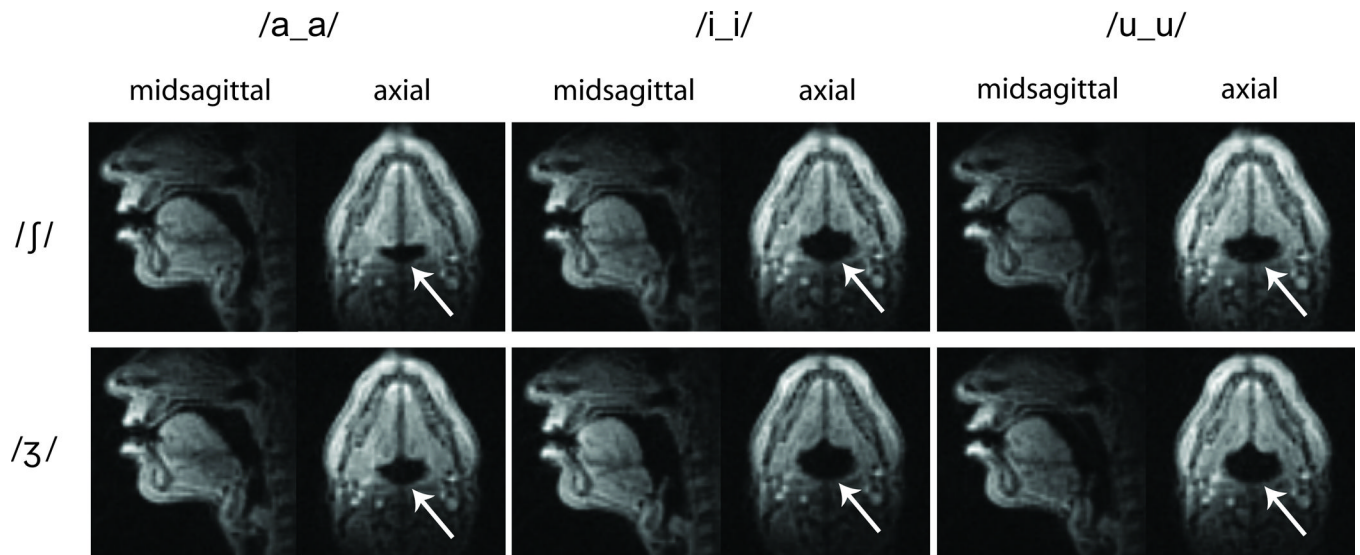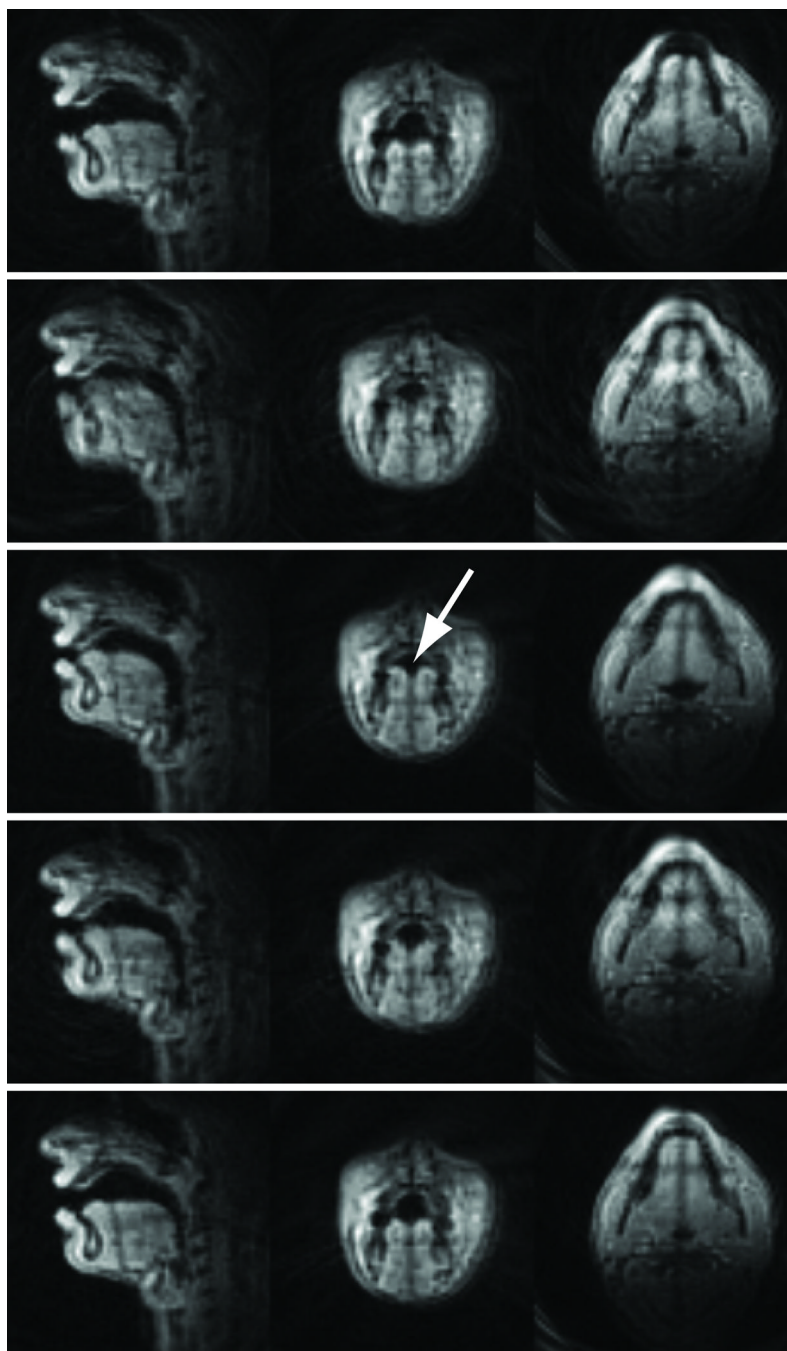
**Figure 4.**
Two-slice frames captured from Subject 1 using Protocol 1 during intervocalic productions of English sibilant fricatives. Top row: voiceless /ʃ/. Bottom row: voiced /ʒ/. The images from the axial slice clearly reveal increased areas in the pharyngeal cross-sectional airway in voiced fricative /ʒ/ (see arrows). For other voiceless/voiced fricative pairs, see Movie 1 in the supplementary material.
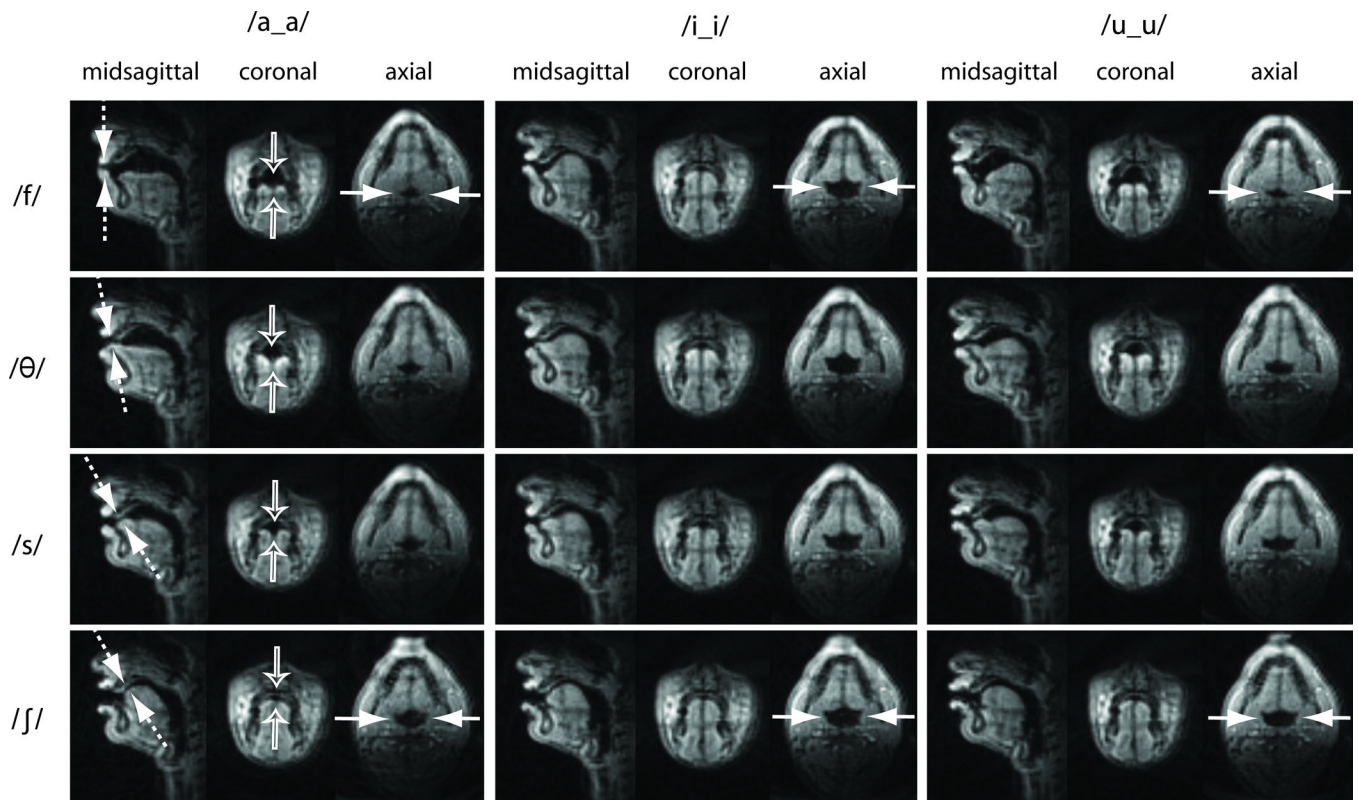
**Figure 5.**
Three-slice frames captured from Subject 2 using Protocol 2 during intervocalic voiceless
fricative productions. Top row: labiodental fricative; 2nd row: dental fricative; 3rd row:
alveolar fricative; Bottom row: post-alveolar fricative. Each scan plane provides unique
vocal tract shaping 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/doming (see the hollow arrows), and the axial
slice shows the degree of aperture and shaping of the pharyngeal airway (see the solid
arrows). See also Movie 2 in the supplementary material.

**Figure 6.**
Representative midsagittal (left column), coronal (center), and axial (right) frames captured using Protocol 2 during the production /asa/ by Subject 2. Time evolution is from top to bottom.

**Table 1**

English Fricatives Imaged in Experimental Studies.

| Place of Articulation | Voicing | IPA Symbol | Example |
|---|---|---|---|
| Labiodental | Voiceless | /f/ | **f**at |
| | Voiced | /v/ | **v**at |
| Dental | Voiceless | /θ/ | **th**igh |
| | Voiced | /ð/ | **th**y |
| Alveolar | Voiceless | /s/ | **s**ink |
| | Voiced | /z/ | **z**inc |
| Post-alveolar | Voiceless | /ʃ/ | mi**ss**ion |
| | Voiced | /ʒ/ | vi**s**ion |

Primary place of articulation, voicing state, and International Phonetic Alphabet (IPA) symbol are listed for each phoneme of interest, along with an example word contrasting the consonant with its voiced/voiceless counterpart.