



# Improved Real-time MRI of Oral-Velar Coordination Using a Golden-ratio Spiral View Order

Yoon-Chul Kim, Shrikanth S. Narayanan, Krishna S. Nayak

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

yoonckim@usc.edu, shri@spi.usc.edu, knayak@usc.edu

## Abstract

In speech research using real-time magnetic resonance imaging (RT-MRI), frame reconstruction is typically performed with a constant temporal resolution. However, a flexible selection of temporal resolution is desirable because of natural variations in speaking rate and variations in the speed of different articulators. In this work, a novel spiral golden-ratio temporal view order was applied to nasal RT-MRI studies in imaging a mid-sagittal slice of the upper airway. Compared to conventional spiral bit-reversed temporal view order scheme, the proposed golden-ratio scheme provides less temporal blurring in the depiction of rapid tongue tip motion with a selection of high temporal resolution. It also provides higher signal-to-noise ratio (SNR) in the depiction of relatively slow velar motion with a selection of low temporal resolution.

**Index Terms:** speech production, vocal tract shaping, real-time MRI, golden ratio, spiral trajectory

## 1. Introduction

Real-time magnetic resonance imaging (RT-MRI) is a promising technology that provides insight into the dynamics of vocal tract shaping during natural speech production [1-4]. In speech RT-MRI experiment, MRI data acquisition is performed simultaneously with speech signal recording in the magnet. Dynamic frames of a 2D mid-sagittal slice of human upper airway are obtained after image reconstructions. Air-tissue boundary detection in the vocal tract regions of interest (ROIs) is performed at each image frame [5] to extract the shape of vocal tract which covers from the lips to the glottis. Adaptive noise cancellation removes loud MRI gradient noise from the audio recordings and produces clean speech signals [6]. Speech signals synchronized with video frames are sources for further articulatory and acoustic analysis. The processes of frame reconstruction, segmentation, and analysis are performed retrospectively after acquiring MRI data in real time.

In speech production, speaking rate is dependent on the subjects and their speech tasks. The motion of articulators (e.g., lips, jaw, velum, and tongue) in the production of sustained sounds or during pauses between the words is slow. The motion of the tongue in the production of stop consonants or diphthongs is rapid. Tract variables such as tongue tip constriction, lip aperture, and velum aperture are timely controlled and coordinated to produce target words. The speeds of articulators may differ depending on the sequences of the vowels and consonants being produced.

Real-time speech MRI often adopts an interleaved spiral fast gradient echo sequence, which provides the most time-efficient way of capturing the dynamics of vocal tract motion during speech. However, current real-time spiral speech MRI protocol lacks in temporal resolution compared to other speech imaging technologies such as electromagnetic articulometry (EMA) and ultrasound. In spiral RT-MRI, higher temporal resolution can be achieved by lowering spatial resolution or designing longer spiral readout with a fewer number of spiral interleaves. Lower spatial resolution imaging may lose details of fine structures such as the epiglottis or lead to more difficulties in resolving the narrowing of the airway, e.g., the constriction between the alveolar ridge and tongue tip in certain sound productions such as English fricatives /s/ and /ʃ/ or vowel /i/. Lengthening spiral readout would cause images to be more susceptible to blurring in the air-tissue boundaries due to a large amount of resonance offset from air-tissue magnetic susceptibility. Higher temporal resolution can also be achieved by reconstructing images from the data within shorter temporal window (i.e., sampling below the Nyquist rate). This introduces spatial aliasing artifacts. Advanced image reconstruction methodologies based on parallel imaging, temporal filtering, or compressed sensing may eliminate these artifacts.

In addition, current speech RT-MRI protocols rely on imaging with a constant temporal resolution and may lack in providing flexibility in selecting temporal resolution. Higher temporal resolution is necessary for the frames that reflect rapid articulator motion. Lower temporal resolution is sufficient for the frames that correspond to static postures. Winkelmann et al. [7] demonstrated the effectiveness of radial golden-ratio sampling scheme in its flexibility in choosing temporal resolution from continuously acquired MR data, and applied radial golden-ratio sampling scheme in real-time cardiac MRI. The golden-ratio sampling may be suited for real-time speech MRI, in which the motion of articulators significantly varies in time and it is difficult to determine an appropriate temporal resolution a priori in the MR acquisition design process.

In this paper, we apply a golden-ratio acquisition scheme to real-time spiral speech MRI (see the schematic diagram in Fig. 1). We describe a spiral golden-ratio acquisition scheme and compare it with a conventional bit-reversed 13-interleaf uniform density spiral which is routinely used in speech RT-MRI data collection. From nasal sound studies performed with the proposed golden-ratio scheme, we seek to retrospectively select temporal resolution and demonstrate an improved depiction of oral-velar coordination, which potentially provides better insight into understanding differences in the timing of articulators under different syllable conditions [8].

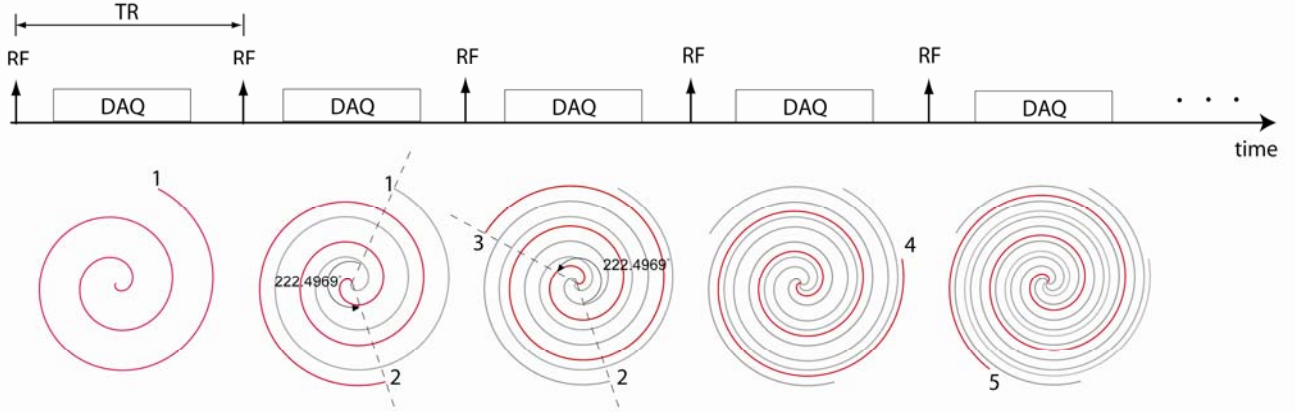


Figure 1. Schematic diagram of real-time continuous MRI data acquisition (DAQ) using a golden-ratio spiral view order. A sequence of only first five TRs is shown. (Top) Pulse sequence diagram. (Bottom) Accumulation of spiral interleaves in the sampling of k-space as time elapses. Every spiral interleaf acquired during current DAQ period is indicated in red color. It is noted that temporal resolution can be controlled by the number of adjacent TRs chosen when reconstructing a frame. In the golden-ratio sampling scheme, next spiral interleaf never overlaps with previously acquired interleaves. Hence, sampling density (i.e., imaging field-of-view) increases as the number of adjacent TRs used for a frame reconstruction increases.

## 2. Materials and Methods

### 2.1. Simulation

A simulation study was performed to illustrate either a penalty or benefit of spiral golden-ratio view order. The spiral trajectory design was based on the imaging protocol routinely used in RT-MRI speech research conducted by Speech Production and Articulation kNowledge (SPAN) group at the University of Southern California. The design parameters were: 13-interleaf uniform density spiral (UDS),  $20 \times 20$  cm<sup>2</sup> field of view (FOV),  $2.4 \times 2.4$  mm<sup>2</sup> in-plane spatial resolution, maximum gradient amplitude = 40 mT/m, maximum slew rate = 150 mT/m/ms, and pseudo bit-reversed temporal view order (see Ref. [9] for details). The pseudo bit-reversed temporal view order was applied in order to shorten the spiral interleaf angle gaps in a few adjacent interleaves at any time point and thus reduce motion artifacts. Spiral golden-ratio view order was performed by sequentially incrementing the spiral interleaf angle by the golden-ratio angle  $360^\circ/1.618 = 222.4969^\circ$  at every TR (see Fig. 1). Because golden-ratio sampling results in non-uniform angle spacing of spiral interleaves in k-space, unaliased FOV was defined to be the reciprocal of the maximum sample spacing in k-space.

### 2.2. Experimental Methods

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

A mid-sagittal scan plane of the upper airway was imaged in supine position using custom real-time imaging software [10]. The spiral trajectory design followed those described in the Simulation section. The imaging protocol was: slice thickness = 5 mm, repetition time (TR) = 6.004 ms, receiver bandwidth =  $\pm 125$  kHz. The golden-ratio view order scheme

was compared with the pseudo bit-reversed 13-interleaf UDS method on the same mid-sagittal scan plane. During each RT-MRI scanning session, there were 6 stimulus slides which are sequentially presented as “type bow know five”, “type bone oh five”, “type toe node five”, “type bone know five”, “type tone oh five”, and “don’t carry an oily rag like that”. The presentation of the stimuli was controlled by Microsoft Powerpoint, in which there was a 1 second pause between the adjacent slides. The subject was instructed to look at the screen through the mirror that is attached to the coil array and speak the words at a normal speech rate.

Gridding reconstructions were performed using temporal windows of 8TRs, 13TRs, 21TRs, and 34TRs with a frame update rate of 1 TR. Gridding reconstructions consisted of gridding the convolution of k-space data on the spiral trajectories with a  $6 \times 6$  Kaiser-Bessel kernel onto  $2 \times$  grids followed by taking 2D inverse fast Fourier transform (FFT) and deapodization [11]. Since the images reconstructed from the posterior coil elements had severe aliasing artifacts in the anterior portions of the mid-sagittal upper airway, sum-of-squares reconstructions from the two anterior coil elements were performed to obtain final images. For comparison of image quality in different temporal window reconstructions, image frames were reconstructed from the data in which the centers of each temporal window were aligned.

Temporal Window	Temporal Resolution	Unaliased FOV	
		Conventional Pseudo Bit-reversed 13-interleaf UDS	Golden-Ratio
8 TRs	48 ms	6.7 cm, 10.0 cm	10.6 cm
13 TRs	78 ms	20.0 cm	17.1 cm
21 TRs	126 ms	20.0 cm	27.6 cm
34 TRs	204 ms	20.0 cm	44.7 cm

Table 1: A relationship between temporal window (i.e., temporal resolution) and unaliased FOV when a temporal window is selected retrospectively. Listed are the unaliased FOVs on four different temporal windows that were considered in the experimental studies.

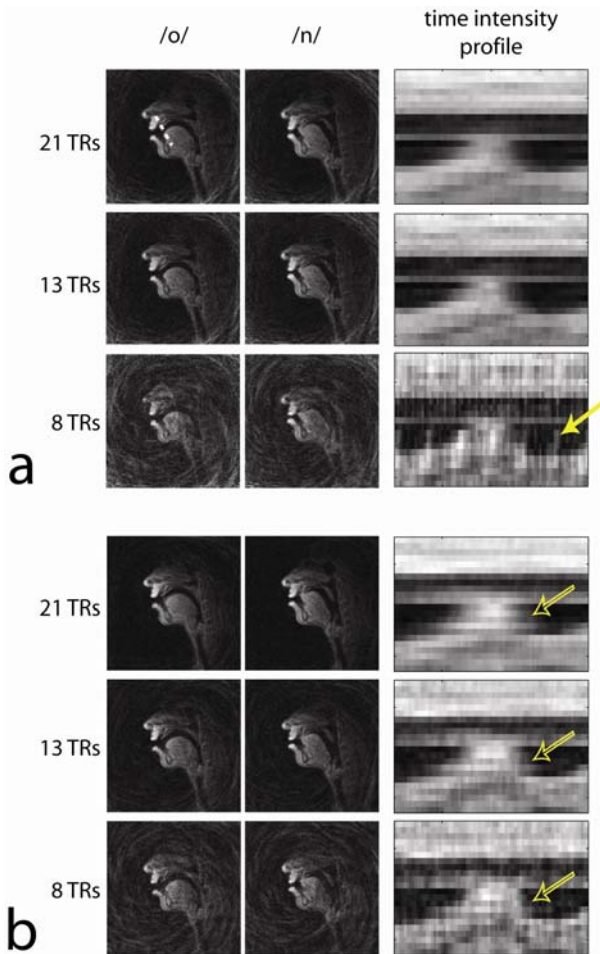


Figure 2. *Qualitative comparison of mid-sagittal images captured at /o/ and /n/ sound production and time intensity profiles from the prescribed line (see the dashed line in the upper left image in (a)) in the articulation of /ono/ for the (a) conventional spiral and (b) golden-ratio spiral acquisition schemes. In (a) and (b), reconstruction results from 8TRs, 13TRs, and 21TRs are compared to demonstrate the effect of temporal window length on the degree of aliasing artifacts and temporal blurriness in the depiction of the tongue tip motion.*

### 3. Results

Table 1 shows a penalty or benefit of the golden-ratio spiral view order sampling. The golden-ratio sampling has larger and consistent unaliased FOV compared to the conventional spiral sampling when undersampling with an 8TRs of temporal window. When sampling with a 13TRs of temporal window, the conventional 13-interleaf UDS sampling provides larger unaliased FOV than the golden-ratio sampling. When sampling with a 21TRs of temporal window, the golden-ratio sampling exhibits larger unaliased FOV than the conventional spiral sampling. These varying FOV characteristics are well manifested in the mid-sagittal images in Fig. 2. It is noted that aliasing artifacts outside the vocal tract ROIs are substantially removed from the 21 TRs mid-sagittal images in Fig. 2(b) but they are not removed in Fig. 2(a).

Time-varying intensity profiles from the 13-interleaf UDS and golden-ratio methods are compared in Figure 2. As seen in the undersampled 8-TRs case, the conventional 13-interleaf UDS sampling produces aliasing artifacts which significantly

vary with respect to time while the golden-ratio sampling produces rather incoherent aliasing with respect to time. This is attributed to the inconsistent FOVs (i.e., 6.7 cm, 10.0 cm) from the conventional 13-interleaf UDS as shown in Table 1. In the 13TRs case, aliasing is removed in the 13-interleaf UDS, but it is seen in the golden-ratio result. The time intensity profile from the 8TRs exhibits the sharpest transition of the tongue tip motion (compare the hollow arrows in 8TRs, 13TRs, and 21TRs results in Fig. 2(b)).

Nasal speech imaging studies were performed using the golden-ratio acquisition and the results are shown in Figure 3. Time intensity profiles are shown from the tongue tip and velum in the production of nasal consonants in three different syllable conditions at onset, coda, and juncture geminate. It is noted that time intensity profiles available from multiple temporal resolution videos facilitate a proper selection of temporal resolution on each articulator. Among the four temporal resolutions considered, tongue tip dynamics is depicted clearly with least temporal blurring from a selection of 48 ms temporal resolution. Velum lowering is depicted clearly with sufficient SNR from a selection of 126 ms temporal resolution.

### 4. Discussion

A new acquisition scheme that adopts a spiral golden-ratio view order has been demonstrated in real-time speech MRI. The golden-ratio scheme has been compared with the pseudo bit-reversed 13-interleaf UDS acquisition, which is routinely used in RT-MRI speech research. When retrospectively sampling the real-time data, the spiral golden-ratio view order provides the following advantages. First, the golden-ratio method provides larger and consistent unaliased FOV when undersampling real-time data for higher temporal resolution. Second, FOV can be increased virtually to infinity when taking infinitely long temporal window. This is due to the fact that a next spiral interleaf never overlaps with previously acquired spiral interleaves. Flexibility in the selection of FOV allows for retrospective adjustment of FOV when imaging the subjects with different head size. Third, spiral interleaves are evenly distributed for any choices of the number of spiral interleaves at any time point, and hence a parallel imaging reduction factor can be flexibly chosen and applied to dynamic golden-ratio data. Auto-calibration with full FOV coil sensitivity maps is easily achievable at any time point by utilizing fully-sampled temporal window data centered on that time point.

Although not addressed in this paper, the MRI gradient noise produced by the golden-ratio spiral pulse sequence was incoherent and non-periodic in nature. Its sound characteristic was different from the MRI gradient noise produced by the conventional bit-reversed 13-interleaf UDS pulse sequence, which has a periodicity in the gradient noise sound [6]. This non-periodic gradient noise sound resulting from the golden-ratio sequence poses a new challenge in MRI gradient noise cancellation. The development of a new noise cancellation technique for the audio recorded from the MR scan will be explored in future work.

We have focused on an example of nasal sound production in which knowledge of the timing of the tongue tip and velum coordination is important in modeling of speech production mechanism [8, 12]. The articulation involves a rapid tongue tip motion and relatively slow velum motion when producing nasal consonant /n/. The golden-ratio acquisition and image reconstruction with variable temporal resolution was effective at improving overall image quality of dynamics of both the

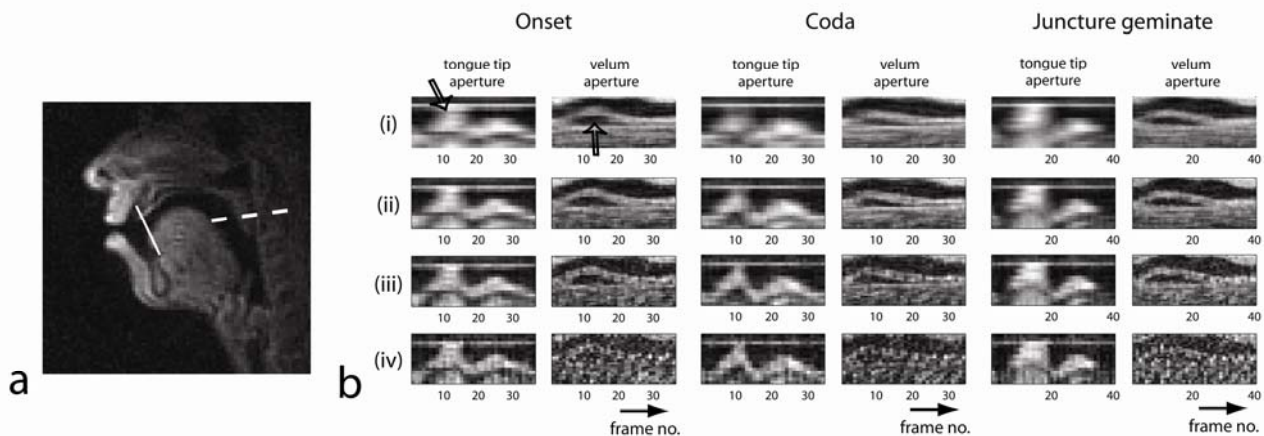


Figure 3. Variable temporal resolution selection from real-time data acquired using the golden-ratio acquisition scheme. (a) A mid-sagittal MR image from the subject. (b) Time intensity profiles for the tongue tip aperture (from the solid prescribed line in (a)) and the velum aperture (from the dashed prescribed line in (a)) for onset, coda, and juncture geminate in the articulations of “bow know”, “bone oh”, and “bone know”. In (b), the hollow arrows in the tongue tip and velum aperture profiles each indicate tongue tip closure onto the alveolar ridge and velar lowering, respectively. Temporal resolutions were (i) 204 ms, (ii) 126 ms, (iii) 78 ms, and (iv) 48 ms.

tongue tip and velum. It is noted that electromagnetic articulometer (EMA) provides high temporal resolution on the order of 5-10 ms and is often used for speech production research, but nasal sound production study using EMA is challenging because the sensors are difficult to attach in the posterior vocal tract regions of interest such as the velum and pharyngeal wall.

In our speech RT-MRI data collection, the MRI scan in a single run typically takes up to 60 seconds and thus the entire video may consist of thousands of frames. Hence, it would be inefficient to perform the vocal tract shape analysis individually on all mid-sagittal vocal tract movies reconstructed from several different temporal resolutions. We are investigating a methodology that automatically assigns temporal resolution region by region in a proper way with knowledge of the motion of articulators and synthesizes multiple videos into a single video. It is expected that this will lead to the efficient analysis of the vocal tract and the improved overall image quality.

## 5. Conclusions

We have demonstrated the application of a spiral golden-ratio temporal view order to speech RT-MRI. The golden-ratio method provided larger and consistent unaliased FOV when retrospectively undersampling real-time data for high temporal resolution than the conventional pseudo bit-reversed 13-interleaf UDS spiral. In nasal speech imaging studies, the proposed method provided improved depiction of rapid tongue tip movement with less temporal blurring and velum lowering with higher signal-to-noise ratio.

## 6. Acknowledgements

This work was supported by NIH Grant R01 DC007124-01. We acknowledge the support and collaboration of the Speech Production and Articulation kNowledge (SPAN) group at the University of Southern California.

## 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., Brinegar, C., Liang, Z-P., Kuehn, D.P., “Dynamic imaging of speech and swallowing with MRI,” *IEEE EMBS*, 6651-6654, 2009.
- [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.S., “Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging”, *IEEE Signal Processing Magazine*, 123-132, 2008.
- [5] Bresch, E., Narayanan, S., “Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images”, *IEEE Trans Medical Imaging*, 28:323-338, 2009.
- [6] 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.
- [7] Winkelmann, S., Schaeffter, T., Koehler, T., Eggers, H., Doessel, O. “An optimal radial profile order based on the golden ratio for time-resolved MRI”, *IEEE Trans Medical Imaging*, 26(1):68-76, 2007.
- [8] Byrd, D., Tobin, S., Bresch, E., Narayanan, S., “Timing effects of syllable structure and stress on nasals: a real-time MRI examination”, *Journal of Phonetics*, 37:97-110, 2009.
- [9] Kim, Y-C., Narayanan, S. and Nayak, K.S., “Flexible retrospective selection of temporal resolution in real-time speech MRI using a golden-ratio spiral view order”, *Proceedings of ISMRM*, Stockholm, Sweden, May 2010, p4967.
- [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] 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.
- [12] Nam, H., Goldstein, L., Saltzman, E., Byrd, D., “TADA: An enhanced, portable Task Dynamics model in MATLAB (A),” *J. Acoust. Soc. Am.* 115(5):2430, 2004.