Real-time speech MRI: a comparison of Cartesian and non-Cartesian sequences

Andreia C. Freitas[1,2], Marzena Wylezinska[1,2], Malcolm J Birch[2], Steffen E. Petersen[1], and Marc E. Miquel[1,2]

[1]*William Harvey Research Institute, Queen Mary University of London, London, United Kingdom,* [2]*Clinical Physics, Barts Health NHS Trust, London, United Kingdom*

INTRODUCTION: Orofacial clefts are common congenital deformities that encompass disorders of the lips, soft and hard palate, affecting 1 to 7 per 1000 newborns worldwide [1]. Velopharyngeal insufficiency, incomplete closure of the soft palate (velum) with the pharyngeal wall, is common in cleft patients and leads to hypernasal speech. Clinical assessment of velopharyngeal closure and surgery repair planning is mostly achieve with x-ray videofluruoscopy and/or nasendoscopy [2]. There has been an increased interest in using real-time MRI to dynamically image the vocal tract [3-7] and eliminate the use of ionizing radiation. However, obtaining the required temporal resolution to reliably assess velar motion (suggested 20 fps [3]) while maintaining image quality is a major challenge. Recently, non-Cartesian sequences have been proposed to improve spatial-temporal resolution [5-7]. This study compares the performance of non-Cartesian (radial and spiral) real-time sequences to a previously suggested Cartesian protocol [8] at 1.5T, regarding the image quality/frame rate required to assess velopharyngeal closure.

METHODS: Five healthy subjects (2M, 3F, mean: 40 years) were imaged using a 1.5T Philips Achieva and a 16-channel neurovascular coil. Three non-Cartesian sequences were optimized with different spatial-temporal resolution sets (Table 1 sequences 1-3) in order to match previously published Cartesian protocols [8]. Two additional non-Cartesian sequences (4-5) were developed to investigate additional image quality/frame rate improvement. Cartesian acquisitions were performed using a balanced steady state free precession (bSSFP) sequence (30° flip angle and 10 mm slice thickness). Non-Cartesian acquisitions were performed using a fast low angle shot (FLASH) sequence, (10° flip angle and 10 mm slice thickness). Subjects were imaged while performing a speech sample consisting of counting (1 to 10), non-sense verbalization ('za-na-za', 'zu-nu-zu', 'ze-ne-ze') and sustained phonations (/a/, /i/). Audio was simultaneously recorded using a fiber-optic MR-compatible microphone (FOMRI II, Opto-accoustics).

Sequence	Resolution (mm^2/fps)	Acquisition	TE/TR (ms)	FOV (mm^2)	Sliding window	SENSE factor
1	1.9×1.9 10 fps	Cartesian	1.5/2.9	270	-	×2.4
		Radial	2.3/5.1	180	×5.0	-
		Spiral	1.0/5.1	190	×2.0	-
2	2.2×2.2 15 fps	Cartesian	1.4/2.8	270	-	×3.0
		Radial	2.1/4.7	180	×6.0	-
		Spiral	1.0/5.0	190	×3.0	-
3	2.7×2.7 20 fps	Cartesian	1.2/2.5	270	-	×3.0
		Radial	1.9/4.1	180	×6.0	-
		Spiral	1.0/4.8	190	×4.0	-
4	1.9×1.9 22 fps	Radial	2.3/5.0	170	×9.0	-
		Spiral	1.0/5.1	190	×4.0	-
5	1.5×1.5 25 fps	Radial	2.7/5.9	170	×16.0	-
		Spiral	1.0/6.3	190	×6.0	-

Table 1 - Acquisition parameters

ANALYSIS: Velum signal homogeneity (signal intensity/signal standard deviation) and thickness were measured for both the relaxed (regular breathing) and elevated (sustained /a/) positions. In addition, signal to noise ratio (SNR) was measured on the intensity-time profiles, where a profile (Fig. 1 h) selected over the main direction of velar motion is displayed over time. Qualitative assessment of image quality was performed visually using a classification scale ('1=Non-diagnostic' to '5=Excellent').

RESULTS: Example images acquired at the relaxed and elevated positions can be seen in Fig. 1. Velar signal homogeneity of spiral acquisitions was higher than for Cartesian at sequences 1-3 ($p<0.0005$), at both velar positions. No significant difference in signal homogeneity was found between radial and spiral in sequences 1-5. As expected, measured palate thickness was greater in the elevated position for all cases and no significant difference was found between sequences/trajectories pairs. Example intensity-time plots can be seen in Fig. 2 and SNR measurements are summarized in Table 2. An increase in SNR was found for spiral ($p<0.0005$) acquisition in sequence 3 compared to sequence 1. As expected with pixel size decrease, a decrease in SNR was observed with sequences 4-5. However, no significant difference was found between sequence 4 and 1 for both radial and spiral, allowing doubling the fps while maintaining spatial resolution and SNR. For sequences 1-3, a significantly higher SNR was measured with non-Cartesian than with Cartesian acquisition (Table 2). For sequences 4-5, spiral acquisition provided higher SNR than radial. Mean image quality scoring (good intra-observer agreement κ=0.60) obtained for spiral (4.30±0.46) acquisitions was significantly higher than for Cartesian (2.47±0.61, $p<0.005$) and radial (2.60 ±0.79, $p<0.0005$). In total, 8 images (12%) were scored as '5=Excellent', all acquired with spiral.

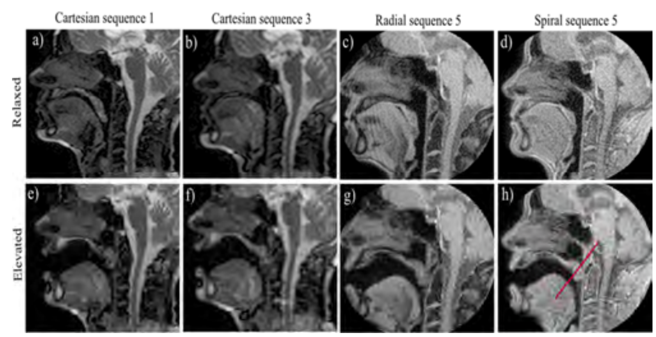

Figure 1- Example images at relaxed (a-d) and elevated (e-h) velar positions

Seq.	Radial	Spiral	Cartesian	*p-value*
1	10.21 (1.74) ‡	12.46 (1.31)	7.10 (1.87)	**<0.005**
2	12.51 (1.92) ‡	13.81(1.23)†	6.67 (2.70)	**<0.0005**
3	13.27 (1.90) ‡	17.68(1.51)*†	6.54 (2.71)	**<0.0005**
4	7.37 (1.02)	11.12 (0.59)	-	**<0.0005**
5	6.98 (1.09)	9.89 (0.94)	-	**<0.005**
p-value	**<0.0005**	**<0.0005**	**NS**	

*Table 2 - Intensity-time mean SNR and standard deviation (*p<0.0005 comparison to sequence 1, † p<0.0005 and ‡ p<0.05 comparison to Cartesian)*

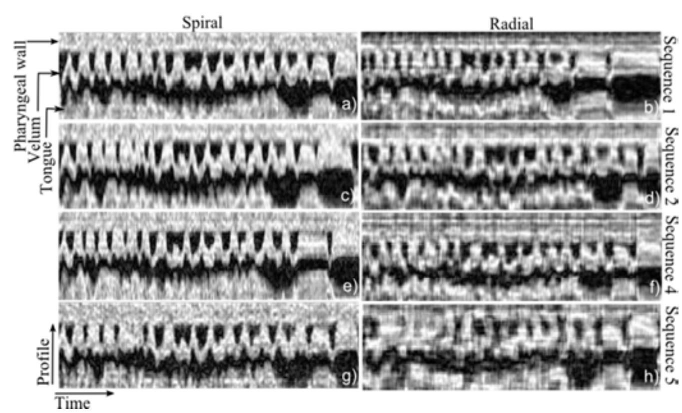

Figure 2- Intensity-time plots for spiral and radial acquisitions

DISCUSSION and CONCLUSION: Results suggest that non-Cartesian real-time sequences are a promising tool to further improve temporal resolution and image quality in dynamic imaging of the soft palate during speech. We found that for all proposed sequences, non-Cartesian (radial and spiral) acquisition provided a higher SNR than Cartesian (Table 2). At higher frame rates of 22 and 25 fps (sequences 4-5), spiral acquisition was optimal and provided images with higher SNR than radial ($p<0.0005$ and $p<0.005$). In addition, spiral acquisitions are intrinsically fast [9] and thus a much lower sliding window acceleration factor was necessary to achieve the desired frame rate than the corresponding radial sequences. This resulted in improved temporal fidelity (Fig. 2 g) while radial acquisition (Fig. 2 h) showed temporal blurring and missed closure events. Spiral sequences presented superior image quality scoring with 32% of cases classified as '5=Excellent', consequently we would recommend their preference for clinical assessment of velopharyngeal closure.

REFERENCES: [1] Mossey. Lancet 2009:374; 1773. [2] Rudnick. Curr Opin Otolaryngol Head Neck Surg 2008:16; 530. [3] Narayanan. J Acoust Soc Am 2004:115; 1771. [4] Beer. JMRI 2004:20; 791. [5] Niebergall. MRM 2013: 69; 477. [6] Sutton. JMRI 2010:32; 1228. [7] Ramanarayanan. J Acoust Soc Am 2013:134; 510. [8] Scott. Br J Radiol 2012: 85; 1083. [9] Delattre. MRI 2010: 28; 862.