The British Journal of Radiology, 85 (2012), e1083–e1092

Towards clinical assessment of velopharyngeal closure using MRI: evaluation of real-time MRI sequences at 1.5 and 3 T

A D SCOTT, PhD, R BOUBERTAKH, PhD, M J BIRCH, PhD and M E MIQUEL, PhD

Clinical Physics, Barts Health NHS Trust, London, UK

Objective: The objective of this study was to demonstrate soft palate MRI at 1.5 and 3 T with high temporal resolution on clinical scanners.
Methods: Six volunteers were imaged while speaking, using both four real-time steady-state free-precession (SSFP) sequences at 3 T and four balanced SSFP (bSSFP) at 1.5 T. Temporal resolution was 9–20 frames s^{-1} (fps), spatial resolution 1.6×1.6×10.0–2.7×2.7×10.0 mm^3. Simultaneous audio was recorded. Signal-to-noise ratio (SNR), palate thickness and image quality score (1–4, non-diagnostic–excellent) were evaluated.
Results: SNR was higher at 3 T than 1.5 T in the relaxed palate (nasal breathing position) and reduced in the elevated palate at 3 T, but not 1.5 T. Image quality was not significantly different between field strengths or sequences (p=NS). At 3 T, 40% acquisitions scored 2 and 56% scored 3. Most 1.5 T acquisitions scored 1 (19%) or 4 (46%). Image quality was more dependent on subject or field than sequence. SNR in static images was highest with 1.9×1.9×10.0 mm^3 resolution (10 fps) and measured palate thickness was similar (p=NS) to that at the highest resolution (1.6×1.6×10.0 mm^3). SNR in intensity–time plots through the soft palate was highest with 2.7×2.7×10.0 mm^3 resolution (20 fps).
Conclusions: At 3 T, SSFP images are of a reliable quality, but 1.5 T bSSFP images are often better. For geometric measurements, temporal should be traded for spatial resolution (1.9×1.9×10.0 mm^3, 10 fps). For assessment of motion, temporal should be prioritised over spatial resolution (2.7×2.7×10.0 mm^3, 20 fps).
Advances in knowledge: Diagnostic quality real-time soft palate MRI is possible using clinical scanners and optimised protocols have been developed. 3 T SSFP imaging is reliable, but 1.5 T bSSFP often produces better images.

Received 12 January 2012
Revised 16 April 2012
Accepted 1 May 2012

DOI: 10.1259/bjr/32938996

Approximately 450 babies born in the UK every year have an orofacial cleft [1], the majority of which include the palate [2]. While a cleft palate is commonly repaired surgically at around 6 months [3], residual velopharyngeal insufficiencies require follow-up surgery in 15–50% of cases [4]. This residual defect results in an incomplete closure of the velopharyngeal port, which in turns leads to hypernasal speech. Assessment of velopharyngeal closure in speech therapy is commonly performed using X-ray videofluoroscopy or nasendoscopy [5, 6]. While nasendoscopy is only minimally invasive, it may be uncomfortable and provides only an *en face* view of the velopharyngeal port. In contrast, X-ray videofluoroscopy is non-invasive and produces an image which is a projection of the target anatomy. Additional information may be obtained from projections at multiple angles [5, 7], but anatomical structures may overlie each other. Furthermore, soft tissue contrast, such as that from the soft palate, is poor, although it may be improved using a barium contrast agent coating [8] at the expense of making the procedure more invasive and unpleasant. Arguably the greatest drawback of X-ray videofluoroscopy is the associated ionising radiation dose, which carries increased risk in paediatric patients [9].

An increasing number of research studies have used MRI to image the soft palate [10–13] and upper vocal tract [14–17]. In contrast to X-ray videofluoroscopy and nasendoscopy, MRI provides tomographic images in any plane with flexible tissue contrast. As a result, MRI has been used to obtain images of the musculature of the palate at rest and during sustained phonation [10, 18, 19]. It has also been used to image the whole vocal tract at rest or during sustained phonation [20–27] and with a single mid-sagittal image dynamically during speech [13, 15–17, 28–35].

For assessment of velopharyngeal closure, dynamic imaging with sufficient temporal resolution and simultaneous audio recording is required. Audio recording during imaging is complicated by the loud noise of the MRI scanner, and both the safety risk and image degradation caused by using an electronic microphone within the magnet. As a result, optical fibre-based equipment with noise cancellation algorithms must be used [36].

In order to fully resolve soft palate motion, Narayanan et al [30] suggested that a minimum temporal resolution

Address correspondence to: Andrew D Scott, Clinical Physics, 4th Floor Dominion House, 60 St Bartholomew's Close, St Bartholomew's Hospital, London EC1A 7BE, UK. E-mail: a.scott@qmul.ac.uk
ADS is funded by the Barts and the London Charity. MEM is partly funded (20%) by the Barts and the London National Institute for Health Research Cardiovascular Biomedical Research Unit.

of 20 frames s^{-1} (fps) is required. A similar conclusion was reached by Bae et al [13], based on measurements of soft palate motion extracted from X-ray videofluoroscopy. Using segmented MRI, Inoue et al [35] demonstrated that changes in the velar position that were evident at acquired frame rates of 33 fps were not observed at 8 fps. However, MRI is traditionally seen as a slow imaging modality and achieving sufficient temporal resolution at an acceptable spatial resolution is challenging. Furthermore, as the soft palate is bordered on both sides by air, the associated changes in magnetic susceptibility at the interfaces make images prone to related artefacts.

Dynamic MRI of the vocal tract has been performed using both segmented [17, 33, 37] and real-time acquisitions [13, 15, 16, 28, 31, 38]. Segmented acquisitions [39] acquire only a fraction of the *k*-space data required for each image during one repetition of the test phrase and, hence, require multiple identical repetitions. While these segmented techniques permit high temporal and spatial resolutions [35], they require reproducible production of the same phrase up to 256 times [34], leading to subject fatigue. Differences between repeats of up to 95 ms in the onset of speech following a trigger have also been demonstrated [36].

In contrast to segmented techniques, real-time dynamic methods permit imaging of natural speech, but require extremely rapid acquisition and often advanced reconstruction methods. The turbo spin echo (TSE) zoom technique [40] has been used to perform real-time MRI of the vocal tract [29, 31] and is available as a clinical tool. The zoom technique excites a reduced field of view in the phase encode direction, hence allowing a smaller acquisition matrix and shorter scan for a constant spatial resolution. While such spin echo-based techniques are less susceptible to magnetic field inhomogeneity related signal dropout artefacts than other sequences, the frame rates achieved with these sequences are limited to 6 fps [31]. Gradient echo-based techniques have also been used to achieve similar temporal resolution [12, 41, 42] in the upper vocal tract, but are often used at much higher frame rates in other MRI applications such as cardiac imaging [43, 44]. A number of gradient echo sequence variants exist. Fast low-angle shot (FLASH) type sequences [45] spoil any remaining transverse magnetisation at the end of every sequence repetition (TR). In contrast, steady-state free-precession (SSFP) sequences are not spoiled [46] and the remaining transverse magnetisation is used in the next TR to improve the signal-to-noise ratio (SNR), but renders the images sensitive to signal loss in the presence of motion. Balanced SSFP (bSSFP) sequences include additional gradients to bring the transverse magnetisation completely back into phase at the end of every TR [47, 48]. The result is that bSSFP sequences have high SNR and are less sensitive to motion than SSFP sequences, but are more sensitive to field inhomogeneities, which cause bands of signal dropout.

Both TSE and the gradient echo techniques discussed here sample in a rectilinear or Cartesian fashion, where one line of *k*-space is sampled in each echo. However, for real-time speech imaging, the highest acquired frame rates have been achieved by sampling *k*-space along a spiral trajectory [15, 16, 30, 49]. While spiral imaging is an efficient way to sample *k*-space and is motion-resilient, it is prone to artefacts, particularly blurring caused by magnetic field inhomogeneities and off-resonance protons (*i.e.* fat) [50]. Recently, one group successfully used spiral imaging with multiple saturation bands and an alternating echo time (TE) to achieve an acquired real-time frame rate of 22 fps [13, 16]. The saturation bands were used to allow a small field of view to be imaged without aliasing artefacts. The alternating TE was used to generate dynamic field maps which were incorporated into the reconstruction to compensate for magnetic field inhomogeneities. However, such advanced acquisition and reconstruction techniques are only available in a small number of research centres.

The aim of this work is to optimise and demonstrate high-temporal-resolution real-time sequences available on routine clinical MRI scanners for assessment of soft palate motion and velopharyngeal closure. Consequently, radial and spiral acquisitions were excluded and the work focuses on Cartesian gradient echo sequences with parallel imaging techniques. As more clinical MRI departments now have 3 T scanners, imaging was performed at both 1.5 and 3 T to enable comparisons. At each field strength, we optimised sequences and implemented four combinations of spatial and temporal resolution in six subjects with simultaneous audio recordings.

Methods and materials

Subjects

Six normal adult subjects for the main study (five male, one female; median age 35 years, range 29–56 years) and four additional subjects for preliminary imaging were recruited from the staff of our institution with informed consent according to ethics committee approval. None of the subjects had linguistic training or known speech disorders. Of the main study participants, three were native speakers of British English, one French, one Arabic and one bilingual Urdu/British English. Information regarding previous dental work was obtained from all subjects as the presence of metallic objects may have a detrimental effect on local magnetic field homogeneity.

Imaging

Imaging was performed using a 1.5 T Philips Achieva, software release 1 (Philips Healthcare, Best, Netherlands) and a 3 T Philips Achieva, software release 2, using a 16-channel neurovascular coil in both cases. Mid-sagittal two-dimensional (2D) images were planned from rapid scout images with the shim volume carefully positioned to cover the oral and nasal cavities down to the level of the epiglottis, extending laterally approximately three times the slice thickness. A short dynamic acquisition (~1 s) was performed at the planned geometry to confirm correct positioning of the slice and shim volume.

The preliminary imaging in four subjects was performed to determine suitable sequences and parameters at 1.5 and 3 T (these data were not included in the subsequent analysis). This included a comparison of real-

Table 1. Sequence parameters common to all sequences

Parameter	Value
Sequence type	bSSFP, 1.5 T/SSFP, 3 T
Flip angle	30°, 1.5 T/15°, 3 T
Field of view	300×240 mm^2
Slice thickness	10 mm
Reconstruction matrix	256×256
Parallel imaging	SENSE
Partial Fourier factor	0.625

bSSFP, balanced SSFP; SENSE, sensitivity encoding; SSFP, steady-state free-precession.

time SSFP, bSSFP and FLASH type sequences with otherwise similar imaging parameters. As a result of these tests, bSSFP sequences were used at 1.5 T and SSFP sequences were used at 3 T to image the six subjects in the main study. All of these subjects were imaged at both field strengths. In order to demonstrate the trade-off between spatial and temporal resolution, four sequences were implemented on each scanner and acquired in a random order in each subject while they performed the speech task. In-plane spatial resolution varied from 1.6×1.6 to 2.7×2.7 mm^2 for temporal resolutions of 9–20 fps. Parameters common to all sequences are given in Table 1 and those modified between the four sequences are given in Table 2. At 1.5 T, a flip angle of 30° was used, whereas at 3 T a flip angle of 15° was used to account for changes in specific absorption rate between field strengths and saturation effects between the different sequences. TE and TR were minimised for every sequence to make full use of the hardware at both field strengths. Three sequences were initially designed at 1.5 T with sensitivity-encoding (SENSE) acceleration factors of 3.0 (Sequences 1, 3 and 4). However, the initial tests suggested that SNR was low with 1.6×1.6 mm^2 spatial resolution (Sequence 1), and therefore a modified version of Sequence 1 with lower spatial resolution and lower SENSE acceleration factor was also included (Sequence 2). Sequences at 3 T were designed to match those at 1.5 T in temporal and spatial resolution, despite the different hardware and sequence variant used. Therefore, the SENSE acceleration factors vary between the sequences at 3 T. Alternatively, SENSE acceleration factors could have been maintained between 3 T sequences, to ensure that SNR values were comparable. However, the key imaging parameters in assessing soft palate function are spatial and temporal resolution; consequently, it was decided to match these variables between equivalent 1.5 and 3 T sequences.

Speech task

The noise reduction algorithms used in the MRI microphone system require an initial period of approximately 8 s of sound to optimise. As a result, the subject was instructed to begin the speech task via the scanner intercom system after this initial period. For each real-time acquisition, the subjects completed the following speech task:

''Scan x'' where x is the index of the acquisition in progress.

Counting from 1 to 10.

The nonsense words: ''zu-nu-zu'', ''zi-ni-zi'' and ''za-na-za'', similar to those used elsewhere [13], which include a nasal sound between vowel sounds.

The sound /ɑ/ (as in ''arm'') and /i/ (as in ''cheese'') sustained briefly.

Simultaneous audio recording

Audio recordings of speech during imaging were made using a FOMRI II dual-channel MRI-compatible fibre optic microphone system (Optoacoustics, Or Yehuda, Israel) with a real-time digital processing system, similar to that described elsewhere [13, 36]. The processed, noise-cancelled speech recording and an audible trigger signal, corresponding to the beginning of each imaging frame, were recorded digitally. The audible trigger signals were extracted using a cross-correlation based algorithm written in Matlab (The Mathworks, Natick, MA) and these triggers were used to create movie files with synchronised audio and video.

Analysis

To provide a comparison between sequences and field strengths, the images were evaluated using measurements of SNR, palate thickness and a visual scoring system. Images were viewed and static SNR measurements were made using Osirix (v. 3.9.4, 32-bit; www.osirix-viewer.com [51]). More complex analysis was performed using in-house software written in Matlab.

Static signal-to-noise ratio and thickness measurements

SNR was measured as:

$$\frac{S_p - S_a}{\sigma_a} \quad (1)$$

where S_p and S_a are the mean signal in regions of interest (ROI) drawn in the palate and the air immediately

Table 2. Variable sequence parameters

Parameter	Sequence 1	Sequence 2	Sequence 3	Sequence 4
Acquired in-plane spatial resolution	1.6×1.6 mm^2	1.9×1.9 mm^2	2.1×2.1 mm^2	2.7×2.7 mm^2
Temporal resolution	111 ms/9 fps	104 ms/10 fps	71 ms/14 fps	49 ms/20 fps
SENSE factor 1.5 T	3.0	2.4	3.0	3.0
SENSE factor 3 T	2.1	1.8	2.3	2.4
TE/TR 1.5 T	1.6/3.2 ms	1.5/2.9 ms	1.4/2.8 ms	1.2/2.5 ms
TE/TR 3 T	1.1/2.3 ms	1.0/2.2 ms	1.2/2.1 ms	0.9/2.0 ms
Acquisition matrix	192×154	160×128	144×116	112×88

fps, frames per second; SENSE, sensitivity encoding; TE, echo time; TR, repetition time.

adjacent to the palate, respectively, and σ_a is the standard deviation of the signal in the ROI drawn in air. SNR measurements were made in two frames where there was minimal or no motion: the first while the subject performed nasal breathing, before starting the speech task (relaxed position); and the second while the subject sustained the /a/ sound (elevated position). At these two static positions, the thickness of the soft palate was also measured along a straight line following the primary action of the levator veli palatini muscle (demonstrated in Figure 1), subsequently referred to as the reference line.

Intensity–time signal-to-noise ratio

In order to fully capture the motion of the soft palate, the acquisition time for each imaging frame should be short with respect to the motion. Comparison of the performance of the sequences during palate motion was achieved by generating 2D intensity–time plots from image profiles taken along the reference line in each imaging frame. The intensity profiles are stacked next to each other in a time sequential order, creating a 2D representation of palate motion throughout the image acquisition [29]. SNR was also measured (defined as before) in a short (~2.5–3.5 s) section of each of these intensity–time plots (I-t SNR). ROI were defined in both the palate and oral cavity on the intensity–time plot from Sequence 4 (the highest temporal resolution) for each subject at each field strength and then copied, with translations where necessary, to the other three plots, corresponding to Sequences 1–3.

Visual scoring

Images were rated blindly in a randomised order on a four-point scale by two independent scorers (MRI physicists) with 6 and 15 years of experience in MRI. The scorers were instructed to rate the images based on a combination of how well the soft palate could be delineated from the surroundings and how well velopharyngeal closure could be assessed. One of the scorers also rated all the images a second time (77 days later) in order to provide data for intraobserver variability measures. The scale was defined as described below and the scorers were shown example data sets for each score:

(1) Very poor/non-diagnostic—the palate is masked by noise or artefact in a majority of images, and image quality is insufficient to assess velopharyngeal closure.
(2) Adequate—the palate is visible in the images but is poorly delineated in most frames, and an assessment of velopharyngeal closure could be performed, albeit with some uncertainty.
(3) Good—the palate could be segmented in the majority of frames, interpolating across some noise/artefact. Velopharyngeal closure could be determined with confidence.
(4) Excellent—the palate is clearly delineated in a large majority of frames and velopharyngeal closure could be determined with a high level of certainty.

Statistics

All statistical analysis was performed using SPSS v. 19 (IBM, New York, NY) and $p<0.05$ was considered a statistically significant difference. Continuous variables were tested for normality and paired data was compared using a two-tailed paired *t*-test. Where there were more than two measures of the same variable (*i.e.* comparisons between all four sequences) repeated-measures analysis of variance (ANOVA) was used to detect a difference between the measures (sequences) and in the case of significant results, multiple Bonferroni corrected paired *t*-tests were used to search for statistically significant pairs. Visual image quality score was compared between sequences using a Friedman test for related samples and between field strengths using a Wilcoxon signed-rank test.

Results

Preliminary imaging

The results of the preliminary tests at 1.5 T suggested that bSSFP sequences were optimal, due to their high SNR. In the relaxed palate, SNR was measured at 7.3 *vs* 5.5 using bSSFP and SSFP, respectively, in one subject, and in another SNR was measured at 9.1 *vs* 3.9 using bSSFP and FLASH, respectively. At 3 T, artefacts caused by field inhomogeneities were too severe to use bSSFP sequences. In one subject imaged at 3 T, SNR in the relaxed palate was measured as 10.8 *vs* 10.6 *vs* 7.8 using bSSFP, SSFP and FLASH, respectively. As a result of the lower SNR using FLASH, SSFP sequences were used at 3 T.

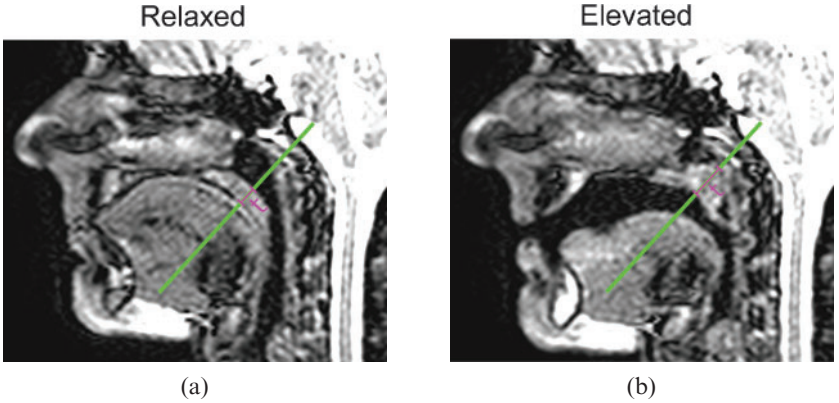

Figure 1. Example images acquired in one subject at 1.5 T using Sequence 2, both (a) in the relaxed palate position and (b) in the elevated palate position. The reference line along the primary direction of palate motion is shown in green on both images, as is the measurement of palate thickness (t).

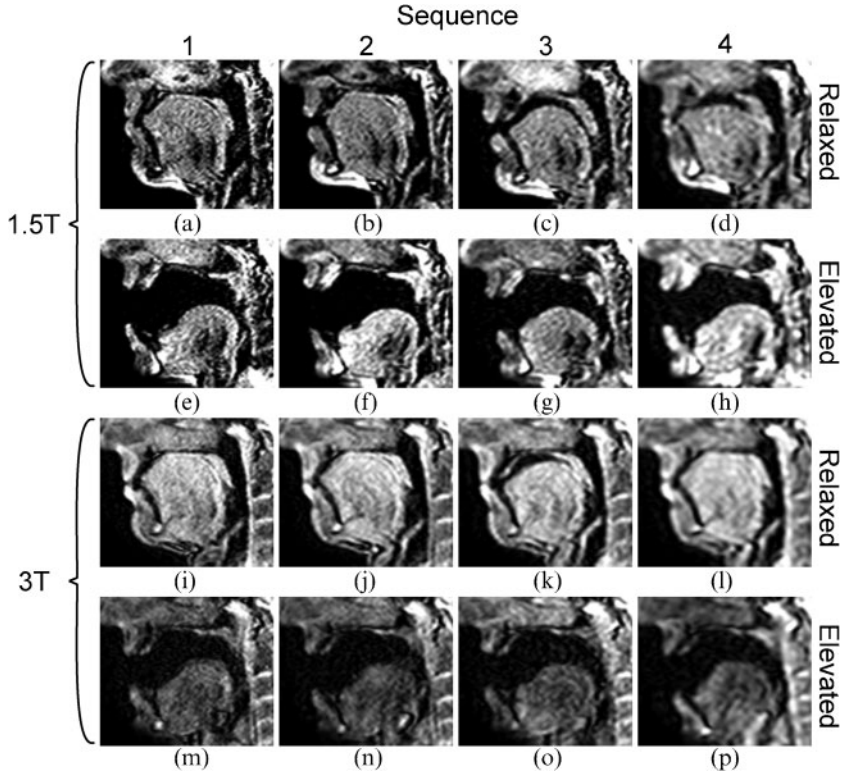

Figure 2. Example static images at (a–h) 1.5 T and (i–p) 3 T using all sequences at the (a–d, i–l) relaxed and (e–h, m–p) elevated palate positions. Note the small differences in the configuration of the oral tract in the relaxed position, most notably at 1.5 T.

Main study

Images were obtained with simultaneous audio recording for all six subjects using all four sequences at both field strengths, and synchronisation of the audio with the video was performed successfully. Four of the six subjects were imaged at 3 T before 1.5 T and two at 1.5 T before 3 T. The mean time between scans was 12 ± 6 days (range 0–16 days).

Example images acquired using all four sequences at both field strengths in one subject, both while in the relaxed position and in the elevated position, are shown in Figure 2. The equivalent movies with synchronised audio recordings (M1–M4 for Sequences 1–4 at 1.5 T and M5–M8 for the equivalents at 3 T) are available at http://mriphysics.net/speech/BJR.html. The static images clearly show the reduction in in-plane spatial resolution from Sequence 1 to Sequence 4 at both field strengths. Comparing the images acquired in the relaxed position, the 3 T images (Figure 2i–l) demonstrate a more homogeneous signal between tissues and a higher signal than at 1.5 T (Figure 2a–d). However, the equivalent images acquired with the palate in the elevated position (windowed identically to the relaxed images in Figure 2) show considerable signal loss at 3 T (Figure 2m–p) but not at 1.5 T (Figure 2e–h). Measurements of SNR and the thickness of the soft palate were made in one example image acquired in the relaxed position from each acquisition and one corresponding image acquired while the subject sustained the /ɑ/ sound in five of the six subjects. In the remaining subject, the /ɑ/ sound was pronounced as an /a/ sound (as in "bad") and the palate was not fully elevated. In this subject an example frame from the /i/ sound was used for the elevated position instead.

Static signal-to-noise ratio

Table 3 gives the mean SNR values obtained using each sequence at each field strength in the relaxed and elevated palate positions. Using a paired *t*-test, the SNR at 3 T is

Table 3. Signal-to-noise ratio in the relaxed and elevated palate positions

	1.5 T				3 T			
Sequence	Relaxed	Elevated	Ratio	*p*-value	Relaxed	Elevated	Ratio	*p*-value
1	5.7 (1.9)	5.1 (1.7)	1.2 (0.6)	NS	9.8 (2.8)	5.0 (1.0)	2.0 (0.6)	<0.01
2	9.1 (2.1)[a]	8.4 (1.8)[a]	1.1 (0.3)	NS	13.7 (2.8)	6.5 (1.2)	2.2 (0.7)	<0.005
3	6.9 (1.7)	7.9 (1.2)[a]	0.9 (0.2)	NS	11.8 (3.9)	6.8 (1.9)	1.9 (0.8)	<0.01
4	7.1 (1.6)[a]	8.1 (3.6)	1.1 (0.7)	NS	9.0 (3.6)	7.1 (1.8)	1.4 (0.9)	NS
p-value	<0.0005	<0.05	–	–	<0.005	NS	–	–

NS, not significant.

Data are mean values, with standard deviation in parentheses. Relaxed and elevated refer to the signal-to-noise ratio (SNR) in a representative static frame and while the subject sustained the /ɑ/ sound. Ratio is the quotient of the relaxed and elevated SNR values. *p*-values refer to one-way repeated-measures analysis of variance tests between SNR values in each column (bottom row) and a paired *t*-test between the SNR in the relaxed and elevated position (columns 5 and 9).

[a] $p<0.05$ pairwise comparison with Sequence 1.

Table 4. Soft palate thickness in the relaxed and elevated positions

	1.5 T		3 T	
Sequence	Relaxed (mm)	Elevated (mm)	Relaxed (mm)	Elevated (mm)
1	9.3 (1.1)	13.0 (1.7)	10.1 (1.4)	14.2 (2.1)[a]
2	9.5 (2.0)	13.2 (2.0)	9.6 (1.0)	13.9 (1.8)
3	11.6 (1.5)	13.9 (1.5)	10.5 (1.2)	14.7 (2.9)
4	11.8 (1.7)	14.5 (1.5)	11.1 (2.4)	15.5 (2.1)[a]
p-value	<0.05	NS	NS	<0.05

NS, not significant.
Data are mean values, with standard deviation in parentheses. No significant pairs were found in the relaxed data at 1.5 T.
[a]$p<0.05$ elevated thickness in Sequence 1 *vs* Sequence 4 at 3 T.

significantly greater than that at 1.5 T ($p<0.0005$) for the relaxed data, and it is significantly greater at 1.5 T than at 3 T for the elevated data ($p<0.05$). Repeated-measures one-way ANOVA demonstrated significant differences in SNR between sequences at 1.5 T in the relaxed and elevated palate positions ($p<0.0005$ and $p<0.05$ respectively), and at 3 T in the relaxed position ($p<0.005$). Paired comparisons (Bonferroni corrected) highlighted that in the relaxed position at 1.5 T, the SNR in Sequence 1 was significantly lower than in both Sequence 2 and Sequence 4 ($p<0.05$ in both cases). At the elevated position at 1.5 T, the SNR in Sequence 1 was significantly lower than the SNR in Sequences 2 and 3 ($p<0.05$ in both cases). In the relaxed position at 3 T no significant differences in SNR were found in any of the pairs of sequences using multiple comparisons (Bonferroni corrected; p=NS). At 1.5 T, the mean relaxed/elevated SNR ratio was 1.10 ± 0.49 with no significant differences (p=NS) between the relaxed and elevated positions in any of the sequences. At 3 T, the mean ratio was 1.90 ± 0.74, with significant differences for Sequence 1 ($p<0.01$), Sequence 2 ($p<0.005$) and Sequence 3 ($p<0.01$).

Palate thickness

Table 4 summarises the measurements of the soft palate thickness. In the relaxed position the mean palate thickness was 10.5 ± 1.9 mm at 1.5 T and 10.3 ± 1.6 mm at 3 T (p=NS), whereas in the elevated position, the mean thickness was 13.6 ± 1.7 mm at 1.5 T and 14.6 ± 2.2 mm at 3 T (p=NS). As expected, the measured palate thickness is greater in all cases at the elevated position. Using repeated-measures one-way ANOVA there were significant differences between sequences at 1.5 T in the relaxed thicknesses ($p<0.05$), but no significant differences between the individual pairs of sequences, and significant differences at 3 T in the elevated data ($p<0.05$), with a significant difference between Sequence 1 and Sequence 4 (14.2 ± 2.1 *vs* 15.5 ± 2.1 mm, $p<0.05$).

Table 5. Intensity–time signal-to-noise ratio measurements

	1.5 T	3 T
Sequence	I-t SNR	I-t SNR
1	5.4 (3.4)	6.0 (2.4)
2	5.8 (3.2)	6.3 (2.3)
3	8.2 (4.2)[a]	7.4 (1.8)
4	9.7 (4.1)[a]	8.4 (2.1)

I-t SNR, signal-to-noise ratio measured in a short (~2.5–3.5 s) section of the intensity–time plot.
Data are mean values, with standard deviation in parentheses. No pairwise significant differences were found in the 3 T data. *p*-value was not significant between signal-to-noise ratio at 1.5 and 3 T.
[a]$p<0.05$ when compared with Sequence 1.

Assessment of dynamic image quality

Intensity–time signal-to-noise ratio

Figure 3 shows intensity–time plots for each sequence at both field strengths in one example subject. The increased temporal fidelity when increasing from 9 fps or 10 fps in Sequences 1 and 2, to 14 and 20 fps in Sequences 3 and 4 respectively is evident; in Figure 3a–d there are sharp step changes in the palate position between time frames which are not present in Figure 3g, h. Table 5 summarises the SNR measurements from the intensity–time plots (I-t SNR). There was no significant difference between the I-t SNR at 1.5 and 3 T. At both field strengths, there was a significant difference in I-t SNR ($p<0.05$ and $p<0.005$ at 1.5 and 3 T, respectively) between sequences. I-t SNR increases in each case from Sequence 1 to Sequence 4, although not significantly. However, significant differences were found between Sequence 1 and both Sequences 3 and 4 at 1.5 T ($p<0.05$).

Visual scoring

Intraobserver agreement in the scoring system was good (Cohen's $\kappa=0.68$), with differences in 11 cases (of 48 analysed). The maximum absolute intraobserver difference in score was 1 with no significant difference between the initial and repeat scores (p=NS). Interobserver agreement was excellent (Cohen's $\kappa=0.80$) with differences in 7 cases (of 48 analysed) and a maximum absolute interobserver difference of 1. Of these 7 cases, one scorer rated the acquisition more highly in 6 cases, and the paired comparison was borderline significant ($p=0.06$). There was no clear trend in the subject or sequence that the differences occurred in.

The results of the image quality scoring are presented as histograms in Figure 4. The differences in the median image score between sequences (range of median scores 2.5–3.5) and field strengths (3.0 at both 1.5 and 3 T) are small and not statistically significant (p=NS in both cases). However, the median score ranges from 3.0 to 3.5 at 1.5 T and 2.5 to 3.0 at 3 T. In addition, at 3 T only one acquisition was rated as 4 and no acquisitions were rated

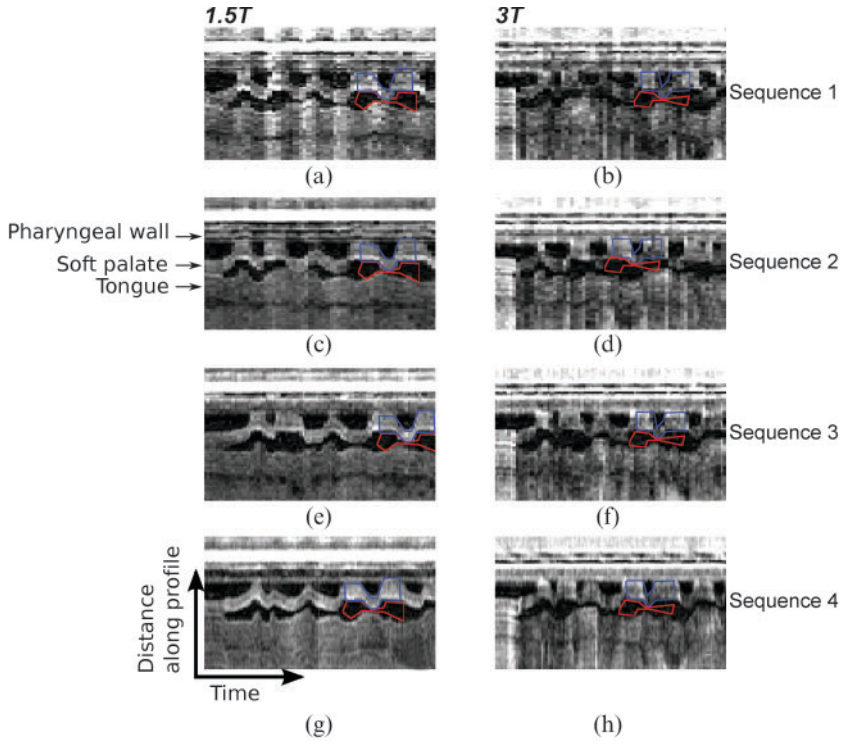

Figure 3. Example intensity–time plots for all sequences in one subject at (a, c, e, g) 1.5 T and (b, d, f, h) 3 T covering the first 5.7 s of speech. These plots are generated by producing an intensity profile along the reference line (see Figure 1). Intensity profiles from adjacent time frames are stacked side by side to represent the time axis in the *x* direction. Increased temporal fidelity is evident moving in the vertical direction, going from (a–d) 9 or 10 fps, to (e, f) 14 fps and (g, h) 20 fps. The regions of interest (ROIs) used for the signal-to-noise ratio measured in a short section of the intensity–time plot are superimposed in blue (signal ROI) and red (noise ROI).

as 1, whereas at 1.5 T 11 acquisitions were rated as 4 and 4.5 acquisitions were scored 1 (one acquisition was rated as 1 by the first scorer and 2 by the second).

In the subjects studied, image quality did not depend on the extent of prior dental work. Only two subjects had images scored 1, one of whom had extensive dental work (two bridges, one crown and six fillings), while the other had only minor fillings. In two subjects, all acquisitions were scored as 4 at 1.5 T, and in a third subject, 3 of the 4 acquisitions were scored 4 (Sequence 2 was scored 3). One of these high-scoring subjects had substantial dental work (3 crowns and 3 fillings) and the other two had only minor fillings.

Discussion

The dynamic motion of the soft palate can be imaged during normal speech using MRI sequences and hardware widely available in clinical radiology departments. Furthermore, images obtained at a variety of combinations of spatial and temporal resolution can be synchronised with simultaneously acquired audio recordings with relative ease. In this study, the images obtained were compared using measures of both SNR and thickness of the soft palate, and an image quality score based on visual assessment. Despite the relatively low number of subjects and large variations in image quality between them, we were able to demonstrate differences between bSSFP imaging at 1.5 T and SSFP imaging at 3 T. For a given sequence, SNR in the soft palate was greater at 3 T than at 1.5 T when imaging the subject at rest, breathing nasally with the palate in the relaxed position. However, for all the sequences tested, at 3 T the SNR is much reduced when the palate is in the elevated position (mean relaxed/elevated ratio 1.87 ± 0.74), whereas SNR is little changed between the two positions at 1.5 T (ratio 1.09 ± 0.49). This can be attributed to both the SSFP sequence used at 3 T, which has a high sensitivity to motion, and to the more difficult task of shimming at the higher field strength. In fact, the improvements in SNR at 3 T, where the SENSE acceleration factors were also lower, are all but eliminated by the reduction in signal intensity that occurs in the moving anatomy with the SSFP sequence; the I-t SNR is similar between 1.5 and 3 T (across all sequences 7.3 ± 3.9 *vs* 7.0 ± 2.2, respectively; p=NS).

The visual image scoring provided further insight into the differences between bSSFP at 1.5 T and SSFP at 3 T. While the median image score was the same between 1.5 and 3 T (both 3.0), in all but one acquisition the images acquired at 3 T were scored as 2 or 3, whereas those acquired at 1.5 T were scored as 1 or 4 in the majority of cases (15.5 cases of 24). All of the images scored as 1 (4.5 cases) were acquired in two subjects and were caused by a poor shim in both cases, which is particularly detrimental to the quality of bSSFP images. Extensive metallic dental work is likely to reduce the quality of the shim, and consequently result in poor images, but image quality did not appear to be related to dental work in this study. However, any related image degradation will depend on the imaging plane and further optimisation is likely to be necessary when targeting other vocal tract anatomy. The wide variation in image quality at 1.5 T is also reflected in the high standard deviation of the I-t SNR when compared with the equivalent measurements at 3 T (3.9 *vs* 2.2, respectively). Further work will address these issues at 1.5 T by optimising the shimming procedure. This may be partly achieved by considering dynamic field maps, which could be used to retrospectively correct for off-resonance effects.

Differences in image quality between the sequences were more subtle than those between 1.5 and 3 T or between subjects, and the number of subjects was insufficient to demonstrate a statistical difference between

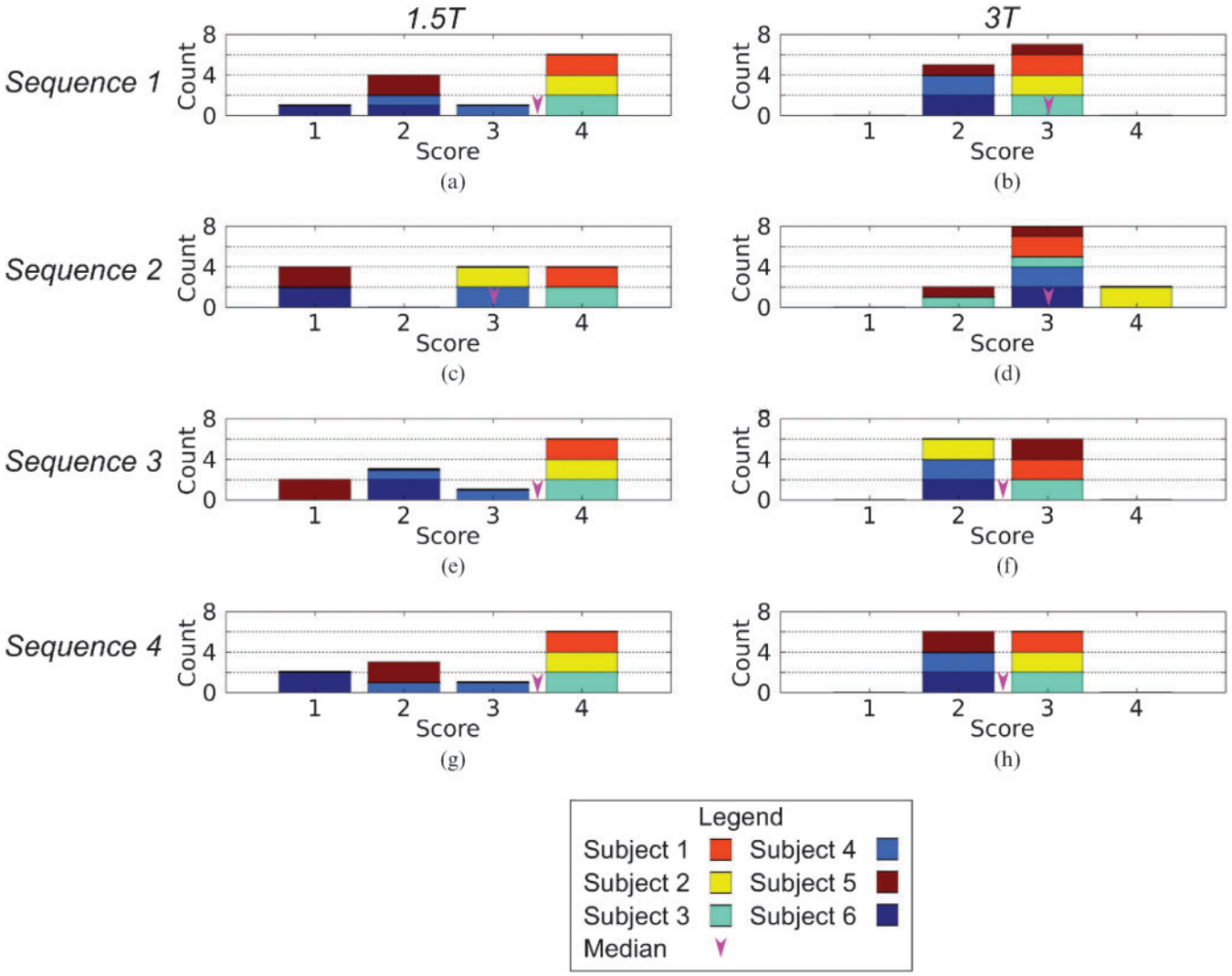

Figure 4. Histograms of image quality score by field strength, imaging sequence and imaged subject. Values from each scorer were included independently, hence the sum of the score bins in each histogram is 12 (6 subjects × 2 scorers for each sequence at both field strengths).

sequences. However, the absence of significant image quality differences between sequences demonstrates that images can be successfully obtained at a range of combinations of temporal and spatial resolution. For accurate geometric measurements of the palate, high spatial resolution is desirable but, in these initial comparisons, at 1.6×1.6 mm^2 (Sequence 1) SNR was lower (significantly at 1.5 T, $p<0.05$) than at 1.9×1.9 mm^2 (Sequence 2). Also, despite the lower spatial resolution, the measured palate thickness was similar between Sequences 1 and 2 (p=NS, 1.5 and 3 T).

For dynamic assessment of the motion of the soft palate, high temporal resolution is desirable. In these tests, Sequence 4 was acquired at 2.7×2.7 mm^2 in-plane spatial resolution and 20 fps temporal resolution. The mean I-t SNR was the highest for this sequence at both field strengths ($p<0.05$ between all sequences at 1.5 T and $p<0.05$ *vs* Sequence 1; $p<0.005$ between all sequences at 3 T) and the increased temporal fidelity of this sequence is evident in the intensity–time plots (see Figure 3). Therefore, along with Sequence 2, this sequence should form the basis for future comparative studies in volunteers and patients.

In this work, sequences used at 3 T were designed to match those at 1.5 T in spatial and temporal resolution. Due to variations in the hardware between scanners and the additional time required for balancing in the bSSFP sequence, the SENSE acceleration factors were lower at 3 T than at 1.5 T. While it would have been possible to match all parameters, including the SENSE acceleration factor and trade increased temporal resolution for decreased spatial resolution, matched spatial and temporal resolutions between 1.5 and 3 T were prioritised here instead. Consequently, 3 T images could be acquired at the spatial resolutions used in this study but with improved temporal resolutions. This would be achieved by increasing the SENSE factor to match that used at 1.5 T (3.0 *vs* 2.1–2.4 in Sequences 1, 2 and 4). However, this would also result in some loss of SNR.

In contrast to previous work using Cartesian acquisitions to perform real-time imaging of the soft palate, we have demonstrated frame rates up to and including 20 fps. Earlier studies [28, 41, 42] using spoiled gradient echo techniques have achieved up to 7 fps at 1.6×1.9 mm^2 in-plane resolution using FLASH techniques with partial

Fourier. Cartesian TSE imaging with the ZOOM technique has been applied in the assessment of velopharyngeal closure [12, 29, 31] and has achieved up to 9 fps at $1.6\times3.1\,mm^2$ in-plane resolution. We achieved 9 fps at $1.6\times1.6\,mm^2$ in-plane resolution with consistently adequate or better image quality at 3 T and the highest quality images in 50% of cases at 1.5 T. The highest reported temporal resolution of 22 fps used in imaging the soft palate was achieved using a research only spiral imaging sequence [13, 16]. Despite using widely available imaging sequences, we achieved a similar temporal resolution (20 fps) at a reduced in-plane spatial resolution (2.7×2.7 *vs* $1.9\times1.9\,mm^2$), but without the added complications of the non-Cartesian trajectory.

In conclusion, we have demonstrated real-time imaging of soft palate motion using bSSFP imaging at 1.5 T and SSFP imaging at 3 T with four combinations of temporal–spatial resolution in a small cohort of healthy volunteers. Audio recordings of speech were also made during imaging and synchronised with the images, as is the norm for X-ray videofluoroscopy—a current standard for velopharyngeal assessment. For reliably adequate image quality, SSFP imaging at 3 T is preferable, but only resulted in images with the highest visual score in one acquisition (of 24). For the highest image quality, bSSFP sequences at 1.5 T are superior, but in this work, they also resulted in a number of studies rated very poor/non-diagnostic. From our initial results presented here, when geometric measurements of the soft palate are required Sequence 2 with 10 fps temporal resolution and $1.9\times1.9\times10.0\,mm^3$ spatial resolution was a good choice. However, for evaluation of soft palate motion, temporal resolution should be prioritised, and Sequence 4 with 20 fps temporal resolution and $2.7\times2.7\times10.0\,mm^3$ spatial resolution was preferred. In future, these techniques will be of use in evaluating velopharyngeal closure as part of a comprehensive MRI exam for cleft palate assessment.

References

1. Office of National Statistics. Congenital Anomaly Statistics 2008. Series MB3, no. 23. Accessed 5 July 2012. Available from: www.ons.gov.uk/ons/rel/vsob1/congenital-anomaly-statistics–england-and-wales--series-mb3-/no--23--2008/index.html.
2. Tolarova MM, Cervenka J. Classification and birth prevalence of orofacial clefts. Am J Med Genet 1998;75:126–37.
3. Sommerlad BC. A technique for cleft palate repair. Plast Reconstr Surg 2003;112:1542–8.
4. Bicknell S, McFadden LR, Curran JB. Frequency of pharyngoplasty after primary repair of cleft palate. J Can Dent Assoc 2002;68:688–92.
5. Havstam C, Lohmander A, Persson C, Dotevall H, Lith A, Lilja J. Evaluation of VPI-assessment with videofluoroscopy and nasoendoscopy. Br J Plast Surg 2005;58:922–31.
6. Shprintzen RJ, Golding-Kushner KJ. Evaluation of velopharyngeal insufficiency. Otolaryngol Clin North Am 1989;22:519–36.
7. Golding-Kushner KJ, Argamaso RV, Cotton RT, Grames LM, Henningsson G, Jones DL, et al. Standardization for the reporting of nasopharyngoscopy and multiview videofluoroscopy: a report from an International Working Group. Cleft Palate J 1990;27:337–47; discussion 47–8.
8. Henningsson G, Isberg A. Comparison between multiview videofluoroscopy and nasendoscopy of velopharyngeal movements. Cleft Palate Craniofac J 1991;28:413–17; discussion 417–18.
9. Wald NJ, Berrington de González A, Bridges BA, Easton D, Little MP, Stiller C, et al. Risk of solid cancers following radiation exposure: estimates for the UK population. London, UK: Health Protection Agency; 2011.
10. Ettema SL, Kuehn DP, Perlman AL, Alperin N. Magnetic resonance imaging of the levator veli palatini muscle during speech. Cleft Palate Craniofac J 2002;39:130–44.
11. Demolin D, Delvaux V, Metens T, Soquet A. Determination of velum opening for French nasal vowels by magnetic resonance imaging. J Voice 2003;17:454–67.
12. Drissi C, Mitrofanoff M, Talandier C, Falip C, Le Couls V, Adamsbaum C. Feasibility of dynamic MRI for evaluating velopharyngeal insufficiency in children. Eur Radiol 2011;21:1462–9.
13. Bae Y, Kuehn DP, Conway CA, Sutton BP. Real-time magnetic resonance imaging of velopharyngeal activities with simultaneous speech recordings. Cleft Palate Craniofac J 2011;48:695–707.
14. Baer T, Gore JC, Boyce S, Nye PW. Application of MRI to the analysis of speech production. Magn Reson Imaging 1987;5:1–7.
15. Kim YC, Narayanan SS, Nayak KS. Flexible retrospective selection of temporal resolution in real-time speech MRI using a golden-ratio spiral view order. Magn Reson Med 2011;65:1365–71.
16. 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;32:1228–37.
17. Ventura SM, Freitas DR, Tavares JM. Toward dynamic magnetic resonance imaging of the vocal tract during speech production. J Voice 2011;25:511–18.
18. Kuehn DP, Ettema SL, Goldwasser MS, Barkmeier JC. Magnetic resonance imaging of the levator veli palatini muscle before and after primary palatoplasty. Cleft Palate Craniofac J 2004;41:584–92.
19. Ha S, Kuehn DP, Cohen M, Alperin N. Magnetic resonance imaging of the levator veli palatini muscle in speakers with repaired cleft palate. Cleft Palate Craniofac J 2007;44:494–505.
20. Rokkaku M, Hashimoto S, Imaizumi S, Niimi S, Kiritani S. Measurements of the three-dimensional shape of the vocal tract based on the magnetic resonance imaging technique. Annual Bulletin: Research Institute of Logopedics and Phoniatrics 1986;20:47–54.
21. Lakshminarayanan AV, Lee S, McCutcheon MJ. MR imaging of the vocal tract during vowel production. J Magn Reson Imaging 1991;1:71–6.
22. Wein BB, Drobnitzky M, Klajman S, Angerstein W. Evaluation of functional positions of tongue and soft palate with MR imaging: initial clinical results. J Magn Reson Imaging 1991;1:381–3.
23. Shellock FG, Schatz CJ, Julien PM, Silverman JM, Steinberg F, Foo TK, et al. Dynamic study of the upper airway with ultrafast spoiled GRASS MR imaging. J Magn Reson Imaging 1992;2:103–7.
24. Greenwood AR, Goodyear CC, Martin PA. Measurements of vocal tract shapes using magnetic resonance imaging. Communications, Speech and Vision, IEE Proceedings I 1992;139:553–60.
25. Narayanan SS, Alwan AA, Haker K. An articulatory study of fricative consonants using magnetic resonance imaging. J Acoust Soc Am 1995;98:1325–47.
26. Story BH, Titze IR, Hoffman EA. Vocal tract area functions for an adult female speaker based on volumetric imaging. J Acoust Soc Am 1998;104:471–87.
27. Clement P, Hans S, Hartl DM, Maeda S, Vaissiere J, Brasnu D. Vocal tract area function for vowels using three-

dimensional magnetic resonance imaging. A preliminary study. J Voice 2007;21:522–30.
28. Crary MA, Kotzur IM, Gauger J, Gorham M, Burton S. Dynamic magnetic resonance imaging in the study of vocal tract configuration. J Voice 1996;10:378–88.
29. Demolin D, Hassid S, Metens T, Soquet A. Real-time MRI and articulatory coordination in speech. Comptes Rendus Biologies 2002;325:547–56.
30. 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;115:1771–6.
31. Beer AJ, Hellerhoff P, Zimmermann A, Mady K, Sader R, Rummeny EJ, et al. Dynamic near-real-time magnetic resonance imaging for analysing the velopharyngeal closure in comparison with videofluoroscopy. J Magn Reson Imaging 2004;20:791–7.
32. Stone M, Davis EP, Douglas AS, NessAiver M, Gullapalli R, Levine WS, et al. Modeling the motion of the internal tongue from tagged cine-MRI images. J Acoust Soc Am 2001;109:2974–82.
33. Kane AA, Butman JA, Mullick R, Skopec M, Choyke P. A new method for the study of velopharyngeal function using gated magnetic resonance imaging. Plast Reconstr Surg 2002;109:472–81.
34. Kim H, Honda K, Maeda S. Stroboscopic-cine MRI study of the phasing between the tongue and the larynx in the Korean three-way phonation contrast. J Phonetics 2005;33: 1–26.
35. Inoue MS, Ono T, Honda E, Kurabayashi T, Ohyama K. Application of magnetic resonance imaging movie to assess articulatory movement. Orthod Craniofac Res 2006;9:157–62.
36. NessAiver MS, Stone M, Parthasarathy V, Kahana Y, Paritsky A. Recording high quality speech during tagged cine-MRI studies using a fiber optic microphone. J Magn Reson Imaging 2006;23:92–7.
37. Masaki S, Nota Y, Takano S, Takemoto H, Kitamura T, Honda K. Integrated magnetic resonance imaging methods for speech science and technology. Proceedings of Acoustics '08; 29 June–4 July 2008; Paris, France. Paris, France: Société Française d'Acoustique SFA; 2008. pp. 5083–8.
38. Rasche V, Holz D, Proksa R. MR fluoroscopy using projection reconstruction multi-gradient-echo (prMGE) MRI. Magn Reson Med 1999;42:324–34.
39. Edelman RR, Wallner B, Singer A, Atkinson DJ, Saini S. Segmented turboFLASH: method for breath-hold MR imaging of the liver with flexible contrast. Radiology 1990;177:515–21.
40. Buecker A, Adam G, Neuerburg JM, Glowinski A, van Vaals JJ, Guenther RW. MR-guided biopsy using a T_2-weighted single-shot zoom imaging sequence (Local Look technique). J Magn Reson Imaging 1998;8:955–9.
41. Jager L, Gunther E, Gauger J, Reiser M. Fluoroscopic MR of the pharynx in patients with obstructive sleep apnea. AJNR Am J Neuroradiol 1998;19:1205–14.
42. Anagnostara A, Stoeckli S, Weber OM, Kollias SS. Evaluation of the anatomical and functional properties of deglutition with various kinetic high-speed MRI sequences. J Magn Reson Imaging 2001;14:194–9.
43. Razavi R, Hill DL, Keevil SF, Miquel ME, Muthurangu V, Hegde S, et al. Cardiac catheterisation guided by MRI in children and adults with congenital heart disease. Lancet 2003;362:1877–82.
44. Sievers B, Schrader S, Hunold P, Barkhausen J, Erbel R. Free breathing 2D multi-slice real-time gradient-echo cardiovascular magnetic resonance imaging: impact on left ventricular function measurements compared with standard multi-breath hold 2D steady-state free precession imaging. Acta Cardiol 2011;66:489–97.
45. Haase A, Frahm J, Matthaei D, Hänicke W, Merboldt KD. FLASH imaging: rapid NMR imaging using low flip-angle pulses. 1986. J Magn Reson 2011;213:533–41.
46. Sekihara K. Steady-state magnetizations in rapid NMR imaging using small flip angles and short repetition intervals. IEEE Trans Med Imaging 1987;6:157–64.
47. Scheffler K, Lehnhardt S. Principles and applications of balanced SSFP techniques. Eur Radiol 2003;13:2409–18.
48. Oppelt A, Graumann R, Barfuß H, Fischer H, Hartl W. FISP—a new fast MRI sequence. Electromedica 1986;54: 15–18.
49. Lee S, Bresch E, Adams J, Kazemzadeh A, Narayanan S. A study of emotional speech articulation using a fast magnetic resonance imaging technique. Proceedings of Interspeech-2006; 17–21 September 2006; Pittsburgh, PA. Bonn, Germany: International Speech Communication Association; 2006. p. 1792.
50. Delattre BM, Heidemann RM, Crowe LA, Vallee JP, Hyacinthe JN. Spiral demystified. Magn Reson Imaging 2010;28:862–81.
51. Rosset A, Spadola L, Ratib O. OsiriX: an open-source software for navigating in multidimensional DICOM images. J Digit Imaging 2004;17:205–16.