Image quality in non-gated versus gated reconstruction of tongue motion using Magnetic Resonance Imaging: A comparison using automated image processing

C. Alvey, C. Orphanidou, J. Coleman, A McIntyre, S. Golding and G. Kochanski

Introduction

The development of improved coil and gradient technology has increased the range of applications for which MRI is beneficial. The use of magnetic resonance (MR) in imaging the vocal tract and measuring articulatory motion is one of the applications benefiting from this development.

Initial studies were based on static image acquisition [1,2]. It was later concluded that static MRI is not representative of running speech and is more like hyper-articulated speech [3]; therefore, techniques for acquiring real-time (non-gated) data were developed, providing a more realistic view of articulator movements during speech.

Gated MRI is a well-established technique for examining the structure and function of the heart. Two main types of sequences are used; one to examine the motion of the heart during systole and diastole (functional or 'cine' imaging), and one to examine the structure of the heart walls, valves, and associated anatomy. Modifications to the technique involve 'tagging' or highlighting tracts of muscle to explore movement, which have been used successfully to evaluate motion analysis [4]. Cine MRI, in particular, has been used to determine airway volume in sleep apnoea both in adults [5] and children [6].

Both non-gated [2,6,7,8,9,10] and gated cine MRI [11,12,13,14,19] techniques have been successfully used to capture vocal tract characteristics and tongue movement during speech. Diffusion tensor (DT) static MRI has also recently been successfully used for imaging the human tongue [18] and showed promise for use in imaging tongue motion as it shows the direction of the muscle fibres. Successful vocal tract and tongue imaging has been reported using both gated and non-gated techniques. Non-gated MRI requires a compromise between temporal and spatial resolution; cine MRI, on the other hand, depends heavily that the timing and motions of the articulatory motions be accurately repeatable. While most literature on MRI for the vocal tract addresses the technical challenges common to both techniques, i.e. the synchronization of the audio and image acquisition as well as the high intensity noise caused by the scanner, the present study aims to systematically compare non-gated and gated cine MRI acquisition techniques and to assess the advantages and disadvantages of each one for imaging tongue motions during speech. The same subjects, text and experimental conditions were used with both approaches. Another feature of this work is the use of longer utterances. While most previous research used mono-syllabic [1, 8, 9, 10, 11, 15] or disyllabic utterances [12,13], we study 4- and 6-syllable utterances. These longer utterances are presumably less easily reproducible and so put the gated MRI sequences to a more stringent test.

In this paper, we investigate how gated cine MRI sequences compare to non-gated MRI sequences in terms of image quality for imaging tongue motions associated with speech.

Methods

Four subjects were recruited, three male and one female, aged 23-35. Since the use of the gated sequence requires subjects to speak in time to a metronome, the subjects had a short practise session outside the scanner a few days before the MRI session. In this five-minute practise, the subjects read the experimental phrases to the beat of a metronome and chose a comfortable metronome rate. The phrases were 4-6 syllables long, 6 in each of 4 metrical patterns: *susu*, *usus*, *uusuus* and *suusuu*, where *"s"* is a stressed syllable and *"u"* is unstressed. The subjects were instructed to synchronise a stressed syllable with the metronome beats. (A related acoustic-only experiment is reported in [14].)

We used a 1.5 Tesla MRI unit (Signa HDx, GE Medical Systems, Milwaukee, Wis.) to obtain sagittal images through the mouth and neck region in 4 subjects. Ethical approval for the experiment was granted by the local regional ethics committee prior to commencement of the experiment. The phrases were presented to the subjects on a 25 inch monitor through the control room window which they could see in a mirror fixed to the head coil.

A non-magnetic gradient microphone (Phone-Or Corp, Or-Yehuda, Israel) was placed at approximately 5 cm from the subject's mouth. We recorded the speech, though in this experiment it was only used to confirm that no gross errors were made. The female subject was short-sighted so the phrases were communicated to her through the intercom prior to each trial.

The subjects spoke out the 24 phrases in random order for one imaging technique, and then (after a break) again for the other imaging sequence. Non-gated acquisition happened first for three subjects and second for one subject. For the gated sequence acquisition, the subjects read a sentence repeatedly in time with the beat of an electronic metronome until the MRI image acquisition stopped. In the non-gated sequence acquisition, the subjects repeated the phrases until instructed to stop. This resulted in 10–12 repetitions for the non-gated sequences and 18–20 repetitions for the gated sequences.

MR Technique

Subjects were screened for safety prior to entering the MR environment, and in all cases a 4 channel NV Array coil was used. The volunteers were positioned supine, with the neck in the neutral position. The head was not rotated; it was supported by a firm foam pillow but not restrained, though the volunteers were asked to keep it as fixed as possible. The volunteers were given an alarm button.

In all cases a test of the microphone and gradient coil detectors was conducted prior to data acquisition and the volunteers were given clear instructions both prior to and during the experiment. The sessions typically lasted about 40 minutes, including one short break.

Localiser scans were obtained in three anatomical planes. From these, the gated and non-gated sequences were planned.

In both sequence types, the volunteers were provided with a set of pneumatically driven headphones and ear plugs, through which a metronome could be heard. This was set to a comfortable rate and comfortable level for each volunteer, and served to maintain speech rhythm. For the gated sequences the metronome acted as the trigger for scanner activation. The MR acquisition parameters are shown in tables 1 and 2.

Image Analysis

To compare the two types of MRI sequences, we developed an approach based on automatic image processing. The process was implemented using Python routines and the SAOimage ds9 software [17]. We fit the proton density with a parametric model. Later, we test how well the

resulting images represent known anatomical facts: that the proton density in the airway is essentially zero, that the tongue surface is fairly flat where we imaged it, and that the tongue at the resolution we imaged it (1.4 mm per pixel) can be approximated as a uniform mass of muscle.

To begin with, we defined nine "measurement lines" referenced to stable and easily identifiable features that would allow us to sample equivalent anatomical regions despite individual variations. The nine lines were drawn automatically, based on manually marked anatomical references; this way the measurement lines could be consistently placed across different individuals. As the basic reference, we used a straight line along the back of the throat from the velic opening down to the level of the epiglottis. As reference points to set the distance scale along that line, we chose the highest point of the palatal arch and the second cervical disk. To reference distances perpendicular to that line, we used the upper incisor as an anchor point. Figures 1 and 2 show sample images, measurement lines and the reference points.

Data within two pixels of each measurement line were used in the analysis. This data was fitted to a function that represents a central region of low density with a higher density region on either side. On the tongue side, this density function is an s-shaped increase to a uniform plateau (Fig 3, right). On the palate/back of throat side, the function yields a smooth density increase from the airway to the tissue followed by a decrease as the measurement point proceeds into the cervical vertebrae or palate and nasopharynx (Fig 3, left).

The function is an eight-parameter non-linear equation along each line, where the parameters set proton densities in three regions, the locations of two edges, their widths, and one final parameter sets the rate at which density decreases into the vertebrae/nasopharynx. It can be written as $\rho(x) = \rho_c + \rho_0 \cdot s((x_0 - x)/\sigma_0, \eta) + \rho_1 \cdot s((x - x_1)/\sigma_1, 0)$, where ρ_c is the proton density in the airway, ρ_0 and ρ_1 are the proton densities on the two edges of the airway, x_0 and x_1 are the positions of the edges, and σ_0 and σ_1 control the widths of the air-tissue interfaces. Subscript 1 refers to the tongue edge and subscript 0 to the palate or back of the throat, as appropriate.

The function s generates an air-tissue interface and takes the form $s(a, \eta) = \exp((2+\eta) \cdot a)/(1 + \exp(2 \cdot a))$; for $\eta = 0$, s approaches zero for large negative a, and approaches 1 for large positive a. This approximates a density step at the air-tongue interface, blurred somewhat by motion and/or curvature of the tongue's surface. For $-2 < \eta < 0$, s reaches a peak near $a = 1$, then begins to decline back toward zero as a continues to increase. This approximates the change in proton density as the measurement line goes from the airway through tissue, then into bone and/or sinus.

We fitted the equation to the data using a simulated annealing approach, maximizing the posterior probability of the model (MAP) on the assumption of exponentially distributed errors. (To a reasonable approximation, this boils down to minimising the sum of the absolute values of the errors.) The procedure involves putting constraints on the parameters ($-2 < \eta \leq 0$, $\rho_c > 0$, $\rho_0 > 0$, $\rho_1 > 0$, $\sigma_0 > 0$, $\sigma_1 > 0$, $L > x_1 > x_0 > 0$), but we checked that these constraints are rarely invoked when the air-tissue boundaries are clear to the eye.

To reduce the risk that the optimization might fail to converge we also add regularization conditions, implemented as Bayesian priors, to weakly constrain the solutions to be reasonable. We verified that, on average, these regularization conditions are weak relative to the data: averaged over the data set, they give a contribution to the log probability of the model which is only 2% as large as the data's contribution, so we expect that most parameters derived from most images would only be weakly affected by the regularization terms.

We ran the optimization eight times for each line with different starting parameters shifting the line by up to two pixels sideways in both directions. We confirmed that the parameters are tightly clustered for well-defined air-tissue interfaces. The eight solutions for the positions of the air-tissue interfaces are shown by crosses and diamonds in Figures 1 and 2. Figure 4 shows the model and the observed proton density data for one of the eight fits for one of the 22,732 combinations of measurement line and image.

Finally, to strengthen the test, we dropped solutions where the estimated width of the airway was in the smallest or largest quartile; this choice dropped cases where the airway was closed (which would thus have ill-defined edges) along with some cases where the optimization failed to converge (thus giving unusually large or small values). It has the added advantage of focussing the data set on the points in the phrases where the tongue motions will be faster than average, thus giving a stronger test of motion blurring.

In the results section, we represent the edges in terms of the width of the density rise from 25% to 75% percent of the tongue's proton density. This representation is more directly interpretable as an anatomical measurement.

Results

Three of the results of the data fitting process are useful measures of the image quality. The two edge widths provide a direct measure of the spatial resolution and smearing of the image, and the proton density in the airway provides a measure of image smearing and noise. Finally, the error about the fit provides another noise measure. In Figures 4, 5, and 6, there are 96 results in each plot box, corresponding to 4 subjects times 24 phrases.

Figure 4 shows the apparent width of the air-tissue interface at the tongue surface. (Each measurement within the box plot is the width, averaged over all 20 frames in the image sequence for that phrase.) Non-gated and gated images provide similar values, differing by much less than one pixel. Similar results are obtained for the palate/throat interface.

Figure 4 shows that the resolutions of the two types of image sequence are similar, and nearly independent of position along the tongue. The grand average for the 25%—75% width of the tongue-air interface is 1.5±0.5 pixels for the non-gated sequences and 1.7±0.5 pixels for gated sequences. (The "±" symbol is used to denote the standard deviation of the measurement.) The values obtained are somewhat larger than the estimated blurring of the air-tongue interface caused by the curvature of the tongue across the 7mm width of the image slice, so these widths are probably dominated by motion blurring and/or the resolution of the MRI sequences employed.

The difference between the mean non-gated and gated resolution is significant at $P < 0.001$ in a paired-sample t-test, even if we assume maximal correlations between parameters derived from the same measurement line in the same image, but the difference is not large – it amounts to only a 10% difference in resolution and rather less than one pixel.

Figure 5 shows the apparent proton density in the airway as a fraction of the average of the two edge densities; the true density is near zero. (Like Figure 4, points in the box-plots are averages over a phrase's image sequence.) The gated images show substantially smaller values than the non-gated images.

If we consider the smearing of the image into empty regions, Figure 5 shows substantially better performance for the gated sequences: they indicate a density within the airway equal to 13%±6% of the tissue density, while the non-gated images show a spuriously large density equal to 26%±5% of the tissue density. (The difference in the means is significant at $P \ll 10^{-6}$.) Pairing like with like, the gated sequences give 52% as much smearing as the same measurement line in the non-gated images.

Figure 6 compares noise between the two types of images by comparing the weighted mean absolute difference between the data and the above model of the proton density. (This includes small regularization terms discussed above.) It shows substantially better performance for gated sequences with a mean error of 1.11±.15 vs. 1.53±0.26 (in arbitrary units). The difference in the means is significant at $P < 10^{-4}$ in a paired t-test.

Discussion

Our study confirms the findings of others [10, 20], that gated MR images are of superior image quality than non-gated images for regions, when repetitive motion is practical. The difference between tissue and airway is larger due to a lower background in the airway and the noise is lower, leading to 62% better signal-to-noise ratio for the gated image, relative to the non-gated image. However, these differences in the performance should be seen in the context of three large difference between the techniques. First, non-gated imaging is limited to 4 frames per second with our scanner; even state-of-the-art techniques [12] have only produced 8 independent frames per second, while gated sequences are limited only by the inverse of the free induction decay time. As a result, gated sequences can give better cinÈ sequences for motions that repeat one or more times per second. Second, to make gated sequence practical, the subject must be able to repeat the motions fairly precisely. Third, data collection is substantially slower for gated sequences because subjects need to repeat the motions.

One of the main results of this paper is that the necessary reproducibility is relatively easy to achieve with a modest amount of practice, even for relatively complex four- and six-syllable utterances. While even our short practise session might be too much for some clinical populations, the actual task is just speaking to a beat, and many people may already have some experience. Possibly, the imaging might work with even less practise, especially for simpler utterances. Our results support [19] in suggesting that gated MRI imaging may be be useful in assessing the articulatory abilities of patients of reconstructive surgery such as repair of a cleft palate. However, it should be noted that our related acoustic work, [14], indicates that some subjects have difficulties speaking to a metronome, so further work may be needed.

Research [20] also suggests that tasks like simultaneous tapping of two fingers give timing errors small enough (circa 20 ms) so that good imaging of finger motions with gated imaging techniques may be possible. We suspect that tapping a finger to a metronome could also be learned rapidly enough (within minutes) that gated-cine imaging of wrist and hand motion may be clinically practical.

Conclusion

We have compared optimized non-gated and gated cine-MRI sequences using objective measures of image quality that can be automatically computed on a large set of images. The gated sequences we tested gave lower noise images and a better discrimination between tissue and air, but showed slightly lower spatial resolution. The performance of gated sequences can be limited by the reproducibility of the motion, but we have shown that good reproducibility can be obtained for tongue motions in speech, even in poly-syllabic utterances. We suggest that the technique may also be useful for hand and wrist motions.

The image processing techniques themselves are nearly completely automated and highly robust, reliably finding air-tissue interfaces wherever they are obvious to a human inspector. Human inspection and marking of just one image was necessary, so long as the subject remained reasonably stationary within the MRI scanner.

For imaging tongue movements, gated sequences permit a higher frame rate and a higher signal-to-noise ratio than non-gated sequences, with negligible (less than 1 pixel) difference in spatial resolution. Where repetitive speech is not feasible, non-gated sequences may be adequate for measuring slow movements. However, if repetition can be tolerated, gated sequences can give slightly better spatial resolution, better signal, and much better temporal resolution.

References

1. Baer T., Gore J.C., Boyce S. and Nye P.W. (1987) "Application of MRI to the analysis of speech production", Magnetic Resonance Imaging, Vol. 5, pp. 1-7.

2. Bresch E., Nielsen J., Nayak K. and Narayanan S., (2006) "Synchronized and noise-robust audio recordings during realtime magnetic resonance scans," *J. Acoust. Soc. Am.*, Vol. 120(4), pp. 1791-1794.

3. Soquet A., Lecuit V., Metens T. and Demolin D.(2002) "Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI," *Speech Communication* 36, pp. 169-180.

4. Ozturk C., McVeigh (2000) "Four-dimensional B-spline based motion analysis of tagged MR images: introduction and in vivo validation" *Phys. Med. Biol* 45, pp. 1683-1702.

5. Abbott M.B., Donnelly L.F., Dardzinski B.J., Poe S.A., Chiney B.A.,and Amin R. S. (2004) "Obstructive Sleep Apnea: MR Imaging Volume Segmentation Analysis," *Radiology* Vol. 232(3) pp. 889-89.

6. Donnelly L.F. (2005) "Obstructive Sleep Apnea in Pediatric Patients: Evaluation with Cine MR Sleep Studies," Radiology Vol. 236(3) pp. 768-778.

7. Crary M.A., Kotzur I.M., Gauger J., Gorham M. and Burton S. (1996) "Dynamic Magnetic Resonance Imaging in the Study of Vocal Tract Configuration," *Journal of Voice*, Vol. 10(4), pp 378-388.

8. Demolin D., Lecuit V., Metens T., Nazarian B. and Soquet A. (1998). "Magnetic Resonance Measurements of the Velum Port Opening," Proceedings of the Fifth Inational Conference on Spoken Language Processing (ICSLP). Sydney, pp 425-429. http://www.isca-speech.org/archive/icslp_1998/i98_0532.html (checked 5/2008).

9. Demolin D., Metens T. and Soquet A. (2000), "Real Time MRI and Articulatory Coordinations in Vowels", *Proceedings of the Fifth Seminar on Speech Production: Models and Data*, Kloster, Munich: Universit‰t M¸nchen, pp. 93-96.

10. Engwall, O. (2004). "From non-gated MRI to 3D tongue movements," *Proceedings of the Eighth International Conference on Spoken Language Processing* (ICSLP 2004), vol. II, pp. 1109-1112. Jeju Island, Korea, October 4-8. http://www.speech.kth.se/ctt/publications/papers04/icslp2004_realtime.pdf (checked 5/2008).

11. Stone M., Davis E.P., Douglas A.S., NessAiver M., Gullapalli R., Levine W.S. and Lundberg A. (2001) "Modeling the motion of the internal tongue from tagged cine-MRI images," *J. Acoust. Soc. Am.*, Vol. 109(6), pp 2974-2982.

12. Parthasaranthy V., Prince J. L., Stone M., Murano E.Z., NessAiver M. (2007) "Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing," *J. Acoust. Soc. Am.*, Vol. 121(1), pp. 491-504.

13. NessAiver M.S., Stone M., Parthasaranthy V., Kahana Y. and Paritsky A. (2006) "Recording High Quality Speech During Tagged Cine-MRI Studies Using a Fiber Optic Microphone," *Journal of Magnetic Resonance Imaging*, Vol. 23, pp. 92-97.

14. Kochanski G. and Orphanidou C. (2008) "What marks the beat of speech?" *J. Acoust. Soc. Am.*, Vol. 123(5) pp. 2780-2791.

15. Mathiak K., Klose U., Ackermann H., Hertrich I., Kincses W-E., Grodd W. (2000) "Stroboscopic Articulography using Fast Magnetic Resonance Imaging," *International Journal of Language and Communication Disorders,* Vol. 35(3), pp. 419-425. doi:10.1080/136828200410663 .

16. Stemerman D.H., Krinsky G.A., Lee V.S., Johnson G., Yang B.M., and Rofsky N.M. (1999) "Thoracic Aorta: Rapid Black-Blood MR Imaging with Half-Fourier Rapid Acquisition with Relaxation Enhancement with or without Electrocardiographic Triggering," *Radiology* Vol 213(1) pp. 185-191.

17. Joye W.A. and Mandel E. (2003) "New Features of SAOImage DS9", *Astronomical Data Analysis Software and Systems XII ASP Conference Series*, Vol. 295. In H.E. Payne, R.I. Jedrzejewski and R. N. Hook (eds), p. 489.

18. Gaige T.A., Benner T., Wang R., Wedeen VJ, Gilbert R.J. (2007) "Three dimensional myoarchitecture of the human tongue determined in vivo by diffusion tensor imaging with tractography." *J. Magn. Reson. Imaging*, 26(3), pp 654-661.

19. Kane A.A., Butman J.A., Mullick R., Skopec M., Choyke P. (2002) "A new method for the study of velopharyngeal function using gated magnetic resonance imaging." *Plast. Reconstr. Surg.*, 109(2), pp 472-481.

20. Semjen, A. and Ivry, R. B. (2001) "The coupled oscillator model of between-hand coordinations in alternate-hand tapping: a reappraisal," *J. of Experimental Psychology: Human Perception and Performance* 27(2), 251-265.

Appendix – Detailed description of data analysis

1) For each group of images, one reference image was selected and it was marked with a set of manually selected "skull regions". These were 6-12 circular regions of approximately 1 centimetre radius that (a) were in anatomical regions that would be expected to move with the skull, (b) showed relatively high contrast structure, and (c) didn't suffer from image wrapping.

A group of images was defined as all data collected while the volunteer remained lying down in the MRI machine under constant imaging conditions. If the subject came out, we started another group. Typically, each MRI session corresponded to three groups of images.

2) For each skull region, a cross-correlation analysis was performed between a reference image and other images in the sequence. (The reference image was selected arbitrarily.) For each skull region and each image in the sequence, this yields a position offset from the reference image. (We also compute a statistic related to how many cross-correlations are bigger than 90% of the maximum value. This is later used to compute a weight factor for each skull region's offset.)

We found that the RMS skull motion within a group was 2.2 pixels (0.3 cm) with a RMS rotation of 0.4 degree; most of the subjects apparently moved their heads very little. The mean weighted fitting error was 1 pixel between skull regions of the same image.

3) Using the offsets of the skull regions, a linear transform is computed that matches the skull on each image to the reference image. This was a weighted regression. We minimised the city-block distance between measured and predicted offsets (rather than a Euclidean distance) so that the optimization would be robust against failures of the cross-correlation analysis. (Here, we collect the error measure for each optimization for later use in trimming away bad data.)

4) On the reference image, a "jaw region" is manually selected, centred approximately on where the mandible intersects the image plane (i.e. the mental protuberance). The selected region was approximately 1.5 cm in radius.

5) A similar cross-correlation analysis was performed to compute the jaw position in each image relative to the reference image.

6) Anatomical reference points (tooth, palate, back of the throat and vertebra) are marked on the reference image.

7) The endpoints of the measurement lines are computed based on the anatomical reference points, suitably transformed into each image's coordinate system.

8) The tongue ends of the "back" measurement lines were moved up and down so that they would better track the structure of the tongue. This was done by making a simple kinematic model of jaw motion (as a lever with a fulcrum at the hinge) and then assuming that the tongue structures move with the jaw on average. This allowed us to transform the relevant measurement line endpoints to follow the average jaw-motion-induced motion of the tongue tissue. Typically, the relevant endpoints were moved about half as far as the measured jaw motion.

We did not transform the endpoints of the "top" measurement lines for two reasons: (a) the tissue motion caused by jaw motion was almost along the direction of the measurement lines, so it would have little or no effect, and (b) our primary aim was to measure the airway width.

9) (Steps 9, 10, and 11 are replicated eight times. In each replication, the measurement line is shifted sideways, starting two pixels to one side on the first replication and ending two pixels to the other side on the last iteration.)

By moving the reference line, we use slightly different data for the nonlinear optimization below. This, combined with different, randomly chosen starting parameters for each iteration allows us to estimate how trustworthy the results of the nonlinear optimization are: if all eight replications yield nearly identical results, they are presumably trustworthy. On the other hand, if the results vary dramatically from replication to replication, the results are presumably not precise and trustworthy.

10) We collect the image data within two pixels of the replication's reference line and fit it to the equation in the "Image Analysis" section. Each proton density measurement was given a weight decreasing linearly from one (if it were on the replicantion's measurement line) to zero (if it were two or more pixels away).

11) The nonlinear optimization was conducted with the "mcmc" and "mcmc_helper" python modules that can be found at http://sourceforge.net/projects/speechresearch in the "gmisclib" subdirectory. These conduct a Markov-Chain Monte-Carlo optimization, and they were run to implement a simulated annealing approach.

One of the outputs of the optimization procedure is the weighted RMS error between the model of proton density (equations in the "Image Analysis" section) and the observed data. This tells you how well the density along a measurement line is represented by a function that starts high (in the tongue), drops in the airway, rises up again (at the back of the throat or palate), and then optionally falls. We store this for later use.

12) We then collected all the results for each combination of measurement line and subject together, then dropped all replications where the width of the airway was in the smallest or largest quartile.

Since our goal was to measure the sharpness of the air-tissue interfaces, it was appropriate to drop data from those images where the airway was closed. (Our procedure cannot estimate the sharpness of the air-tissue interface if there isn't one.) Similarly, for a few cases, the tongue is occasionally drawn down and back far enough that the front-most measurement line "falls off" the tongue.

And, finally, we wished to test the imaging technique in the most difficult case, when the tongue is moving as fast as possible. Generally, the most rapid tongue motion occurs near the motion's

midpoint.

13) Each of the surviving replications of each measurement line on each image yields a set of eight parameters that controls the best-fit model of proton density.

The central proton density is one of these parameters: we then analyze its values with standard statistical techniques. However, note that while the individual images are (nearly) statistically independent, the replications within an image are correlated. To compensate for this, we treat each group of replications as being tightly correlated: this is conservative (since two replications do not share 100% of their data), so the resulting error bars and confidence levels will likewise be somewhat conservative.

To make the data more easily interpretable, we compute the distance over which the model rises from the minimum proton density seen in the airway to the tongue density, and then take the distance between the 25% and 75% points on that rise. (This distance is proportional to σ_1 so long as the airway is not narrow and so long as σ_0 is small. However, our computed distance is a better match to normal measurement procedures in cases where the edges are fuzzy or the airway is narrow.) Again, we treat each block of replications as being tightly correlated and compute normal statistics.

Similarly, we treat the RMS proton density error between the model and observations. Our model is sufficiently flexible to capture normal variations in proton density along the measurement line, so this statistic is largely a measure of noise in the image. (This statistic also includes a contribution from muscular structure in the tongue, but we estimate this to be relatively small compared to imaging noise.)
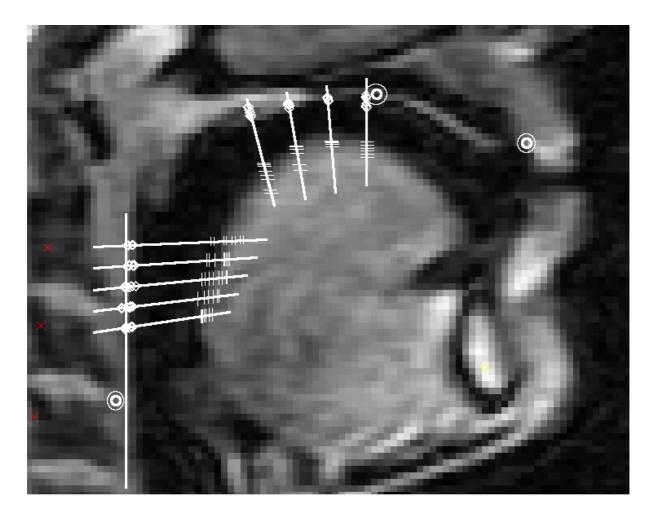
Figure 1: Gated image of the vocal tract.   The phase-encoding direction is horizontal. The anatomical anchor points for the measurement lines are the double white circles on the upper incisor, palate, and at the base of the 2nd cervical vertebra: the last anchor is the vertical line tracing the back of the throat. We measure the density along the nine lines pointing radially into the tongue. Along each of those lines are marked the eight computed solutions for the air-tissue interface positions; a perpendicular line marks each estimate of the tongue edge, and diamonds (often overlapping) mark estimates of the other edge of the airway.
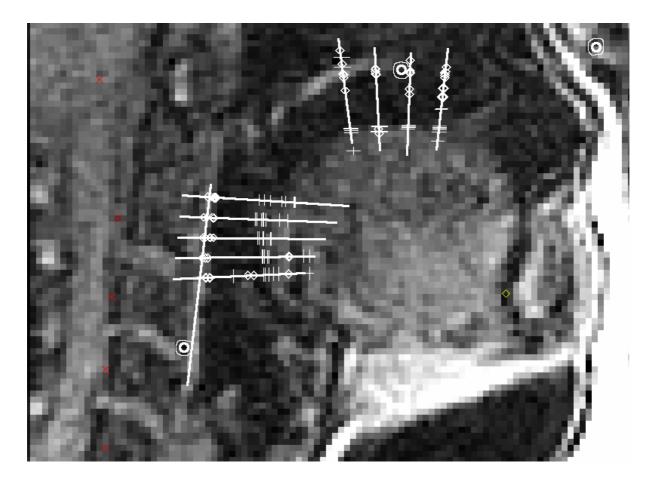
Figure 2: Non-gated image of the vocal tract, displayed as per Figure 1. (In this image, the back of the head is wrapped over the lips [right], but not the top and back of the tongue, where we analyze the data. Plotted as per Figure 1.
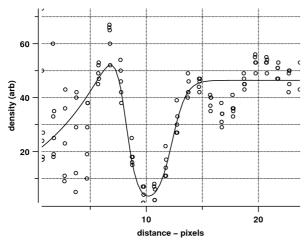
Figure 3: Density vs. position data along a measurement line (circles) also showing a fit of the model to the data (line). The tongue is to the right, the centre of the airway is at approximately 10 pixels, and the spine is to the left. This plot is of measurement line 3 on a gated image.
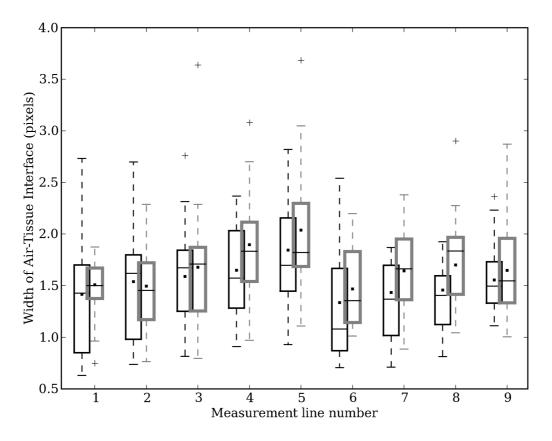
Figure 4: Fitted widths of the air-tongue interface for the nine measurement lines. Line 1 is just above the epiglottis; line 5 is just below the velar opening; lines 6-9 are progressively farther forward on the palate. In each pair, the value from the non-gated image is to the left. These are box plots where each point corresponds to a single image. The central line marks the median, and the central dot marks the mean.
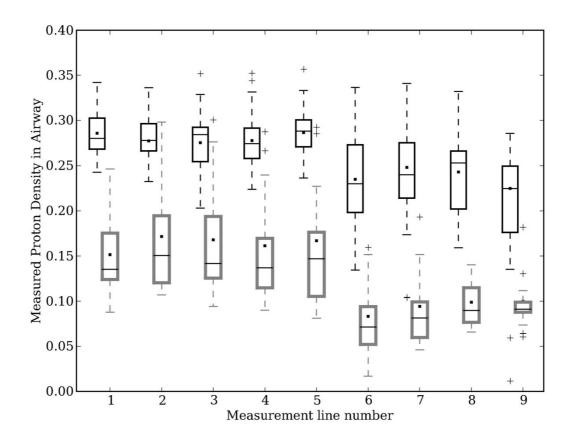
Figure 5: Fitted proton density in the airway as a fraction of the tissue density. Values from non-gated images are to the left and above in each pair.
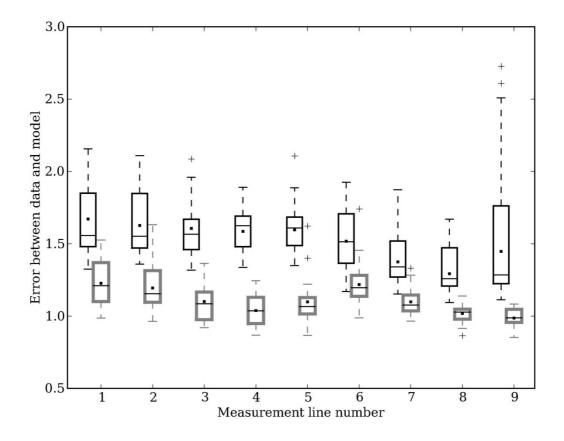
Figure 6: Weighted RMS error between the data and the model, plotted as per Figure 4.

Gated: The computer controlled trigger device was connected to the scanner ECG leads and the rate set specific for the volunteer. The parameters for the acquisition are shown in table 1

TR	Minimum
TE	Minimum
Pulse sequence	Fast Spoiled Gradient Recalled Echo (FSPGR)
Flip angle	20
Trigger type	Simulated heart beat connected to ECG (determines 'Heart Rate')
Arrhythmia rejection	10%
Trigger delay	Minimum (10 ms)
Gated Phases to reconstruct	20
Phase / Frequency	256 / 128
Matrix	256 sq
Slice thickness	10 mm
FOV	36 cm
NEX	1
Views per segment	8
User selected options	Flow compensation, Gating, Sequential acquisition, Extended Dynamic range
Scanning time	18 – 20 s
Number of images per phase	20
Temporal resolution per image, HR =60 bpm	$1/20^{th}$ second

Table 1 the parameters selected for gated acquisitions

Non-Gated: The parameters selected for the non gated acquisitions are shown in table 2

TR	7.1 ms
TE	Minimum full
Pulse sequence	Fast Spoiled Gradient Recalled Echo (FSPGR)
Flip angle	35
Phases per location	440
Phase / Frequency	256 / 192
Matrix	256sq
Slice thickness	7 mm
FOV	36 cm
NEX	0.5
Phase field of view	0.5
Display Field of View	35 cm
User selected options	Flow compensation, Sequential acquisition, Extended Dynamic range, Fast, Multi
Scanning time	110s
Total number of images acquired	120
Temporal resolution per image	¼ second

Table 2 the parameters selected for non-gated acquisitions