

This article was downloaded by: [University of Maryland, Baltimore], [Maureen Stone]

On: 16 October 2012, At: 12:11

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Computer Methods in Biomechanics and Biomedical Engineering

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/gcmb20>

Single syllable tongue motion analysis using tagged cine MRI

Devrim Unay^a, Cengizhan Ozturk^b & Maureen Stone^c

^a Department of Biomedical Engineering, Bahcesehir University, Istanbul, Turkey

^b Institute of Biomedical Engineering, Bogazici University, Istanbul, Turkey

^c University of Maryland School of Dentistry, Baltimore, MD, USA

Version of record first published: 12 Oct 2012.

To cite this article: Devrim Unay, Cengizhan Ozturk & Maureen Stone (2012): Single syllable tongue motion analysis using tagged cine MRI, *Computer Methods in Biomechanics and Biomedical Engineering*, DOI:10.1080/10255842.2012.723697

To link to this article: <http://dx.doi.org/10.1080/10255842.2012.723697>



PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Single syllable tongue motion analysis using tagged cine MRI

Devrim Unay^{a*}, Cengizhan Ozturk^b and Maureen Stone^c

^aDepartment of Biomedical Engineering, Bahcesehir University, Istanbul, Turkey; ^bInstitute of Biomedical Engineering, Bogazici University, Istanbul, Turkey; ^cUniversity of Maryland School of Dentistry, Baltimore, MD, USA

(Received 24 July 2011; final version received 20 August 2012)

The complicated muscle activity of the human tongue and the resultant surface shapes can give us important clues about speech motor control and pathological tongue motion. This study uses tagged magnetic resonance imaging to provide a 2D surface deformation analysis of the tongue, as well as a 4D compression–expansion analysis, during utterances of four different syllables (/ba/, /ta/, /sha/ and /ga/). All speech tasks were performed several times to confirm the repeatability of the motion analysis. The results showed that the tongue has unique motion patterns for utterances of different syllables, and these differences, which may not be observed by a simple surface analysis, can be examined thoroughly by a 4D motion model-based analysis of the tongue muscles.

Keywords: tongue; magnetic resonance tagging; tissue motion analysis; 4D motion tracking; speech physiology

1. Introduction

In order to understand speech motor control and the nature of various speech disorders, we need to know the relationship between muscle activity and the resultant surface shapes of the tongue. The tongue is a highly flexible organ with a complicated muscle architecture and comprises entirely of muscles, fat and connective tissue (Stone 1991). The detailed motion of these muscles (e.g. during speech) cannot be easily explained by the deformation patterns seen on the surface of the tongue.

Tongue muscle fibres, which run longitudinally and transversely, determine the direction of the tongue deformation. The contraction of longitudinal muscles will shorten and draw back the tongue while the contraction of transverse muscles will flatten and extend it. Moreover, the tongue can be bent, twisted and tensed by different contraction combinations of these muscles (Stone 1990).

Speech disorders may result from disruption of phonation, articulation or resonance. A speech disorder is a symptom or sign – not a diagnosis. It cannot be properly treated if the aetiology is not known. Hence abnormal tongue movements can manifest serious underlying disorders (Stone et al. 2001). Tongue movement abnormalities can be neurological as well as anatomical. Tongue-tie (ankyloglossia), for example, refers to a reduced mobility of the tongue due to restricted lingual fraenum (a small fold of tissue under the tongue checking or limiting its movements). However, damage to the hypoglossal nerve (cranial nerve XII) will also cause abnormal tongue movements.

The tongue shows different displacement patterns for different syllables. By identifying these different patterns, we could begin understanding the muscular activity of the tongue during continuum of speech. In this paper, we present 2D surface deformation and 4D motion analyses of the tongue using tagged cine magnetic resonance imaging (MRI) during four consonant-to-vowel syllables (/ba/, /ta/, /sha/ and /ga/).

2. Background

2.1 Anatomy of the human tongue

The tongue is part of the supra-laryngeal motor system, and its control is critical for respiration, chewing and swallowing of food, and speaking in humans. In humans, the tongue forms both the lower surface of the oral cavity and the anterior margin of the pharynx. Furthermore, the posterior section of the tongue has a rounded surface contour in the midsagittal plane, whereas in all the other mammals it is long, thin and positioned entirely within the oral cavity. The tongue has many inter-digitised muscles and no bones or joints. It has been described as a ‘muscular hydrostat’ (Smith and Kier 1985, 1989), meaning that its internal musculature provides both its shape and motion. It is widely believed that it is capable of providing this support due to its incompressibility and constant volume. Thus, contraction in one direction is compensated by expansion in another (Smith and Kier 1985, 1989). The tongue is mostly composed of interlacing masses of intrinsic and extrinsic skeletal muscle fibres (Dick 2000). Many of the intrinsic

*Corresponding author. Email: devrim.unay@bahcesehir.edu.tr

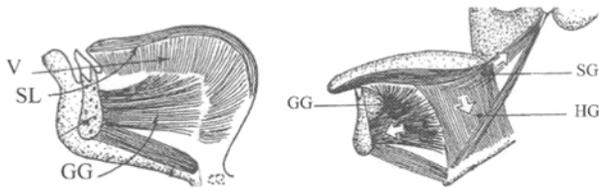


Figure 1. Musculature of the human tongue with midsagittal (left) and lateral sagittal (right) views. The intrinsic muscles include the superior longitudinalis (SL) and verticalis (V). The extrinsic muscles include genioglossus (GG), styloglossus (SG) and hyoglossus (HG) (Dick 2000).

muscles, such as verticalis (V) and superior longitudinalis (SL), and the extrinsic muscles, such as genioglossus (GG), styloglossus (SG) and hyoglossus (HG), can be observed in Figure 1.

The intrinsic muscles originate and insert within the tongue and are not attached to any bones. Their muscle fibres run in all three planes (longitudinal, transverse and vertical). The extrinsic muscles of the tongue originate from the bones of the skull or the soft palate. The classic view of the tongue is that the intrinsic muscles affect tongue shape, allowing it to thicken, thin, lengthen and shorten during speech and swallowing (Dick 2000), whereas the extrinsic muscles alter the tongue's position by protruding, retracting or moving it from side to side (Marieb and Hoehn 2006). More recent data suggest that all muscles shape and move the tongue (Stone and Lundberg 1996).

2.2 Imaging the human tongue

Imaging the movement of the tongue is a difficult task because of its complex shape and composition as well as its location in the vocal tract. The selected technique must be able to image the tongue without interfering in its normal patterns of motion and must be capable of recording its rapid motion.

Different research groups have used different techniques to determine the vocal tract changes during speech, including X-ray (Mermelstein 1973), X-ray micro-beam (Kiritani et al. 1975), electropalatography (Palmer 1973), electromagnetic articulography (Stone et al. 1992), electromyography (MacNeilage and Sholes 1964; Honda and Kusakawa 1997), ultrasound (Akgul et al. 1999) and MRI (Baer et al. 1987; Demolin et al. 1998; Badin et al. 2008).

X-ray, the most well-known technique for vocal tract studies, has the risk of over-exposure and produces data loss by mapping 3D space into 2D images. Although the X-ray micro-beam technique reduces the exposure, data loss still exists in the images. Electropalatography records the tongue's contact with the hard palate and allows real-time recording, but has the main disadvantage of capturing no information when the tongue is not in contact with the hard

palate. Electromyography, used to study the electrical activity of tongue muscles, is not commonly used due to its invasiveness, unpleasantness and difficult interpretation. Ultrasound, using the reflective properties of sound waves, images the surface of the tongue efficiently. It is very fast and poses no known health risks to the patient, but provides low-quality images of only the surface of the tongue and no information from any vocal tract reference structures.

2.3 Magnetic resonance imaging

MRI is based on exciting magnetic dipoles, mainly protons of water molecules within the body, and observing the subsequent effects (Stark and Bradley 1999). A unique motion analysis technique in MRI is tagging, where temporary magnetic fiducial markers are created and tracked within tissues. Magnetic tags are spatially encoded magnetic saturation planes, which temporarily mark material points within human tissues. In other words, they are voids in the regular MRI created by spatially selected pre-saturation pulses. So, when a tagged tissue is imaged after a certain evolution time, the shape changes of tags reflect the underlying tissue motion. Several tagging approaches have been proposed in the literature with the main motivation of measuring the cardiac motion (Zerhouni et al. 1988; Axel and Dougherty 1989). The parallel plane stripe pattern and the combination of two orthogonal plane tags forming a grid are the most common types. Basic tag sequences of this type are now integrated within the pulse sequence libraries of all clinical MRI machines (McVeigh 1996; Kerwin and Prince 2000).

The tagging operation may be considered to be a spatially selective excitation involving the combined use of radio frequency (RF) pulses and gradients. Grid tags are created by the successive creation of two sets of parallel stripe tags with tagging gradient pulses oriented orthogonal to each other. Once created, the encoded tagging pattern decays over time as the magnetisation recovers by longitudinal relaxation with time constant T_1 . Muscles have a T_1 of approximately 850 ms, so potentially, tags can persist throughout short syllables.

In this paper, we employ a general 4D B-spline-based motion model for motion analysis. B-spline-based methods are used because they provide several advantages such as parametric continuity, compact representation of the information, local support and differentiability. Several B-spline-based techniques have been employed in the past to describe deformed tags: Moulton et al. (1996) utilised them to describe the deformed tag planes as surfaces, but they used a global polynomial basis function expansion for the final forward tracking. Amini et al. (1998) segmented the tags using a 2D-coupled B-snake grid, whereas Radeva et al. (1997) extended this approach to 3D. Later, a detailed, step-by-step implementation of the B-spline-based 4D

(3D + time) motion field fitting method was presented by Ozturk and McVeigh (1999, 2000).

The application of MRI tagging for tongue motion analysis has been a less-studied topic. Stone et al. (2001) utilised a similar tagging protocol as ours, but employed a much simpler motion reconstruction scheme that was limited to 2D strain analysis. Takano and Honda (2006) performed in-plane displacement and velocity analyses of tagged cine-MRI data to speculate on the 3D deformation of the tongue during articulation of a single vowel. Parthasarathy et al. (2007) used harmonic phase (HARP) processing to measure tongue motion from tagged cine-MRI. The advantages of the HARP processing are that it does not require prior contouring of the tagging pattern, and the resulting motion measurements are not restricted to the tag intersections. Nevertheless, they presented analyses of a single utterance and employed a pseudo-3D motion analysis approach. Fujita et al. (2007) constructed a 3D physiological model of the tongue based on tagged MRI data from a normal subject, and verified its prediction accuracy of pathological tongue movements in a glossectomy case. Buchaillard et al. (2008) proposed to use a 3D finite-element model of the oral cavity to study the relationship between the strain levels of the tongue and the activated tongue muscles. They found good correlation in the case of single muscle activation but limited correlation when multiple muscles were activated, which is probably due to the intertwined nature of the tongue fibres deep in the tongue body. As a result, the authors underlined the necessity of good knowledge of tongue muscle activations to correctly interpret strain maps computed from tagged MRI. Recently, Stone et al. (2010) have employed principal components analysis and clustering for tongue motion categorisation based on velocity fields computed from tagged cine-MRI data of normal and pathological cases.

In view of the above literature review, this study relies on semi-interactive extraction of the tagging patterns in tagged cine-MRI sequences, modelling of the spatio-temporal displacement field using a 4D cubic B-spline model and computation and visualisation of 3D strains for multiple vocal sounds.

3. Methods

3.1 Subjects, speech material and data acquisition

A single volunteer, a 26-year-old male, native speaker of English with no dental fillings, was used for this study. The subject was given proper informed consent. Four consonant-to-vowel syllables (/ba/, /ta/, /sha/ and /ga/) were chosen as the speech material for their large tongue motions and short durations. Synchronisation between the subject and triggering of the tags is achieved using an in-house device, which simultaneously sends an electrical signal to the scanner to trigger the tagging-imaging sequence and creates an auditory cue to be heard by the

subject. The delay between the trigger and the auditory cue was adjustable, in order to assure the correct tag timing at the tongue rest (prephonatory) position. Analysis prior to imaging showed that the response of the subject to the auditory cue was consistently delayed, and hence triggering of the tags was realised 300 ms after the auditory cue.

The subject repeated the syllables three times (only twice for /ga/) to provide repeatability data. At each repetition, 16 phases of a single slice were acquired in 'real-time' mode all within a second (where the imaging takes around 989 ms, while the remaining time is utilised for tagging and machine reset for the next repetition), a complete slice set could be acquired for each syllable at each imaging session, and the patient was instructed to breath freely in between the acquisitions. In the course of each imaging session, the consonants were all released at time phase 2, the vowel maximum was reached at time phase 5 and later the tongue moved forward to rest for the next repetition or for an inhalation.

Imaging the tongue during speech presents the additional challenge of dealing with motion during the image acquisition process. Segmented k-space imaging is an approach, which was developed for heart imaging permitting the acquisition of a set of images at multiple cardiac phases over the course of several heartbeats in a single electrocardiography (ECG)-gated, breath-held scan (McVeigh and Atalar 1992). This is achieved by partitioning the k-space data matrix into several so-called segments. The region corresponding to each k-space segment is acquired repeatedly for the duration of a heartbeat, or as for tongue imaging for the duration of syllable, providing multiple snapshot images; and the successive segments are acquired in successive heartbeats, or in successive repeats of the syllables. The temporal resolution can be improved by reducing the segment size (i.e. reducing the number of k-space lines in a segment), but note that this is at the price of an increased total imaging time (increased number of heart beats or repeats of syllables). The acquisition times for segmented k-space imaging can also be reduced through the use of echo-planar, spiral and parallel imaging techniques. For further reading, several review articles exist for a general introduction of these different imaging techniques with a focus on their cardiac implementations (Pettigrew et al. 1999; Reeder and Faranesh 2000).

Image acquisition was carried out on a 1.5 Tesla General Electric Signa Cardiac Scanner with a standard MRI head coil. A modified gated multi-echo spoiled gradient echo (SPGR) sequence was used in 'real-time' mode (all echos forming an image acquired immediately one after another) with the following parameters: a repetition time of 10.3 ms, an echo time of 1.5 ms, a bandwidth of 125 kHz, a flip angle of 10°, an echo train length of 8, an imaging matrix of 128 × 96, a field-of-view of 30 × 15 cm (corresponding to a pixel resolution in the midsagittal plane as 2.34 × 1.56), a tag spacing of 5 mm and a slice thickness of 7 mm.

Three sagittal slices (left, mid and right) and six axial slices (s1–s6, where s1 refers to the axial slice closest to the upper tongue) were collected during 16 time frames (62 ms each) for each syllable repetition. Hence, a total of 144 images (6 + 3 = 9 slices \times 16 time frames in each) were collected for each syllable repetition, while the whole database consisted of 1584 images (three repetitions for /ba/, /ta/ and /sha/ and two repetitions for /ga/). The tagging operation was applied in both the x and y spatial directions resulting in a grid of horizontal and vertical tag lines in each image.

3.2 Image processing and analysis

The measured tag deformation at a single tag point contains only a unidirectional component of its past motion, from tagging to imaging time. In order to achieve a full tracking of any point through time, the information coming from different tagging sets has to be combined and interpolated in space and time. The classical analysis of tagged MRI contains these steps: (1) segmentation of the region of interest which is routinely done interactively or semi-automatically; (2) detection of the tag points for each slice, tag orientation and time frame and (3) fitting a motion field (or mechanical tissue model) using three (or two for 2D analysis) orthogonal 1D displacement data-sets coming from all the tag detected points.

We have axial and sagittal groups of images for each syllable both with a grid of tags. We have chosen to use only one tag direction (horizontal) from the sagittal images and two tag directions (vertical and horizontal) from the axial images to get the necessary 3D set to compute the information of tongue deformation.

The surface of the tongue was interactively contoured (segmented) for each slice (three sagittal and six axial slices for each image group) and every time frame (16 time frames for each slice). This interactive contouring was done using a custom software program (XBS) (Shechter et al. 1999). The 3D view of axial and sagittal surface contours for the first time phase (at 0 ms) of syllable /sha/ is displayed in Figure 2.

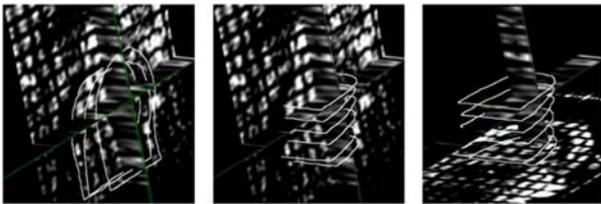


Figure 2. Interactive contouring using XBS, which allows to put and view axial and sagittal surface contours overlaid on any original or re-sliced 3D image data-set. Three sagittal contours for the first time frame (0 ms) of syllable /sha/ with sagittal image are on the left, six axial contours with sagittal image are in the middle and the same axial contours with an axial image are on the right.

The tongue surface contours in axial and sagittal views form a closed loop. In the sagittal views, the loop was closed at the level of the top-most axial image plane. Each image was then masked using XBS, to include only the region of interest. On the resultant masked images of the tongue, tags were detected automatically for all image planes and time frames using the ‘Find tags’ software program developed for tracking motion of the heart by Guttman et al. (1997). Detected points on each tag line were automatically registered in 3D space and time using the image header information (Figure 3). Automatic registration was mostly successful, but there was mismatch or superfluous deformation on some of the resulting tags. These errors were observed when there were large and rapid tongue motions between time frames, and they occurred mainly because the motion of the tongue is much greater than that of the heart for which the software was developed.

In order to track the motion of the tongue, an adapted version of the motion-tracking algorithm named tagged-tissue tracking developed by Ozturk and McVeigh (1999, 2000) was used. This algorithm combined all tagging information coming as 1D displacements from each tagging set over time into a comprehensive 4D B-spline-based motion field. In the end, the trajectory of any point is known from the tagging time onward. This 4D motion field permits calculation of 3D displacements, velocities and strains for any material point inside the tongue at any time.

Motion of the tongue is described as a 4D cubic B-spline forward motion field, Φ , with the original tagging plane being the reference plane:

$$\Phi(\vec{P}, t) = \sum_u \sum_v \sum_w \sum_t B_u(x)B_v(y)B_w(z)B_t(t)\vec{C}_{u,v,w,t}. \quad (1)$$

Here, P refers to a point on the tag plane, B is the B-spline basis function in 4D (x , y , z and t) and C denotes the control points of the B-spline function. By this, we can have a parametric representation of the material

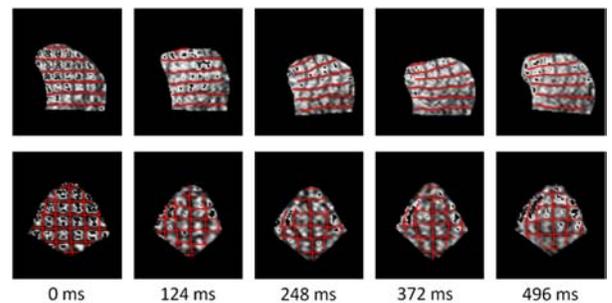


Figure 3. Detected tags displayed on masked MRI for syllable /sha/: horizontal tags on midsagittal slice (top row), and horizontal and vertical tags on top-axial slice (bottom row).

coordinate displacement field. Once the motion field is created, the displacement gradient (g) at any point can be calculated by

$$g_{i,j}(x,y,z) = \frac{\partial U_i}{\partial X_j}, \quad (2)$$

where U_i is the i th component of the material displacement field and X_i is the material coordinate position vector. This gradient can be calculated by directly differentiating the 4D B-spline motion field as

$$g_{x,x} = \sum_u \sum_v \sum_w \sum_t B'_u(x)B'_v(y)B'_w(z)B'_t(t)\vec{C}'_{u,v,w,t}, \quad (3)$$

where B' and C' are the derivatives of the B-spline basis function and the 3D control points of the field, respectively (see Ozturk and McVeigh 1999, 2000 for details of computations). Then from the displacement gradient, we can compute the local Lagrange strain tensor (ε) at any location by

$$\varepsilon = \frac{1}{2}(g^T + g + g^t g). \quad (4)$$

The computed strain can be transformed to a local surface-based coordinate system to display circumferential, radial and longitudinal components. Accordingly, the compression–extension at any location (point) can be calculated by the difference of the strain values of the selected point for two consecutive time frames.

4. Results

4.1 Error analysis

The motion-tracking algorithm, generating a final 4D motion field for the tongue, was originally developed for heart motion analysis, and therefore a detailed error analysis is needed to evaluate the accuracy of motion fit.

In motion-field fitting, the tag points are first mapped backward (to their material locations at tagging time) and then forward (the material points found are then brought back to their corresponding tag point locations). This allows us to assess the goodness of the fit using two types of residual errors: inverse error is simply the difference of the observed and the computed displacements calculated at tag points via backward mapping only. Forward error is calculated by backward mapping the observed tag point to the reference plane, then from that position forward tracking to the original time using the motion field, and taking the difference between the observed and computed tag points. In a sense, inverse error assesses the suitability of cubic B-spline basis functions for the underlying motion, and forward error gives an overall assessment of the fit and includes the backward error (Ozturk and McVeigh 2000). In Figure 4, the inverse error plot of the first repetition of syllable /sha/ is given for both tagging directions. Average errors for horizontal tags increase rapidly in the first four time frames, and then slowly decrease. However, average errors of vertical tags show a smoother behaviour. Errors of vertical tags are higher in the beginning and at the end, indicating more displacement in vertical tags than horizontal tags for those time frames, which may be due

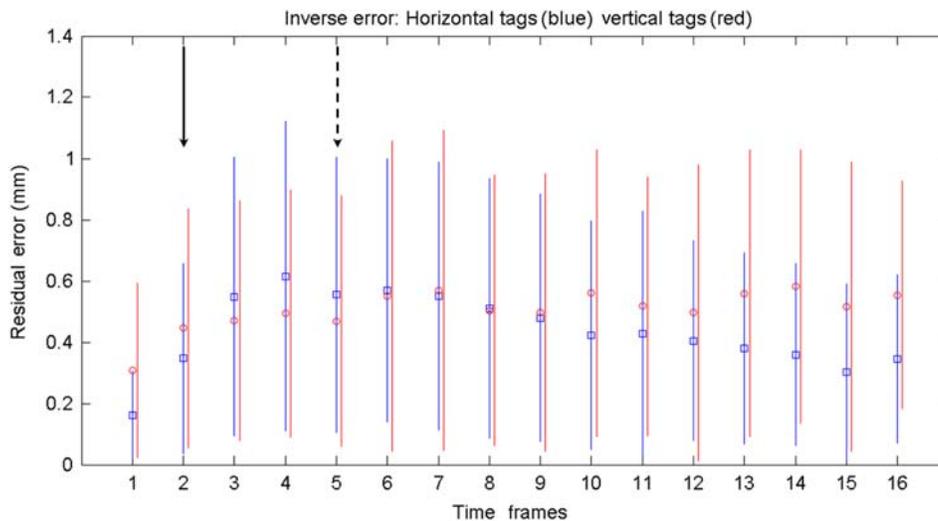


Figure 4. The inverse error plot of the first repetition of syllable /sha/. The horizontal axis represents the time frames with 62-ms temporal resolution and the vertical axis represents the residual errors in millimetres. The errors are displayed as tick marks with vertical lines corresponding to mean \pm standard deviation, respectively. The lines with squares indicate the horizontal tags, whereas those with circles belong to the vertical tags. Consonant release is indicated by the solid arrow, and the maximum vowel displacement by the dashed arrow.

to compression or expansion in the upper tongue. Similarly in Figure 5, the forward error plot of the first repetition of the syllable /sha/ is displayed. A smooth behaviour is observed in the forward errors in general, except for the second and third time frames of time smoothed result. For all the tags in all time frames, we observe that both inverse (1.20 mm) and forward (2.25 mm) errors are less than one pixel, which are comparable to the results that are obtained in a typical cardiac tagging experiment. These errors can be attributed to inaccuracies in tag tracking (method originally developed for heart motion analysis, possible inconsistencies in manual corrections and decaying nature of tags), variations in multiple repetitions of utterances and noise in the images.

4.2 2D displacement analysis

We analysed the kinematics of tongue motion by measuring the displacements in the front, tip and back of the tongue during the same four syllables. The three reference points were selected from the front, top and back portions of the tongue (Figure 6). The reference points were selected from the maximally moving regions of the tongue by intersecting the tag lines with the surface contours at each time frame. Note that the asymmetrical pincushion distortion observed in the original images, which can be attributed to different variations in local magnetic fields, is consistent through out all the time frames, and therefore does not affect the outcome of our motion analysis.

In this part of the study, 2D analysis was carried out without any motion field fitting. Assuming there is no through-plane motion, we can identify matching points by intersecting tag lines with surface contours. Since tags

move with the tissue, we followed the same material points over time and calculated the total displacement by integrating the distance between matching points with respect to initial time frame (i.e. the distance between a material point at any time frame and its original location at the initial time frame).

In Figure 7, total displacement values of three selected points from different regions of the tongue for four syllables and different repetitions are given. The vowel /a/ has a low back tongue position, and the four consonants represent different vocal tract configurations. The /b/ is a labial sound; that is the primary articulators are the lips, not the tongue. As such the tongue does not produce a constriction during /b/ and thus its motion during /b/ is more heavily influenced by the vowel than during the other three consonants. For the other three consonants and the vowel /a/, the primary articulator is the tongue. The /t/ uses the tongue tip positioned against the alveolar ridge. The /sh/ is made using the tongue blade (immediately posterior to the tongue tip) and the palatal slope. The /g/ is made by contacting the tongue body with the velum. In these syllables the consonants were released at the 2nd time frame, maximum vowel displacement was reached at the 5th time frame and the breath following the syllable was initiated at time frames 9 or 10. The front region is the maximum moving part of the tongue for syllables /ba/, /ta/ and /sha/. This reflects the lowering of the jaw for these sounds and the release of the anterior tongue for the latter two, seen particularly in the ‘front’ region. The ‘top’ region is lowered for /ga/, consistent with the release of velar contact. For the three syllables with lingual consonants – /ta/, /sha/ and /ga/ – considerable tongue motion is visible (20 mm). The /ba/ is different, however. After the lips open

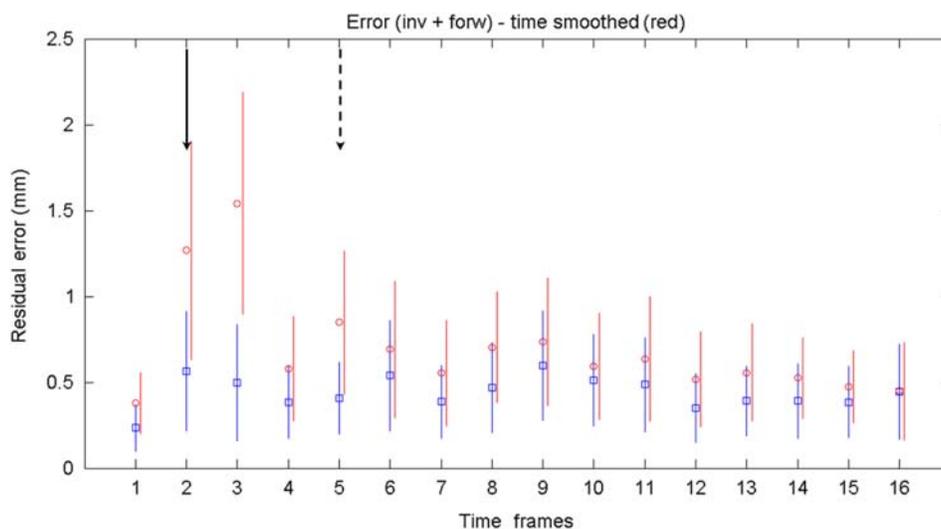


Figure 5. Forward error plot of the first repetition of the syllable /sha/. The errors are displayed as tick marks with vertical lines corresponding to mean \pm standard deviation, respectively. The lines with circles are error plots calculated from time-smoothed (4D) control points. Consonant release is indicated by the solid arrow, and the maximum vowel displacement by the dashed arrow.

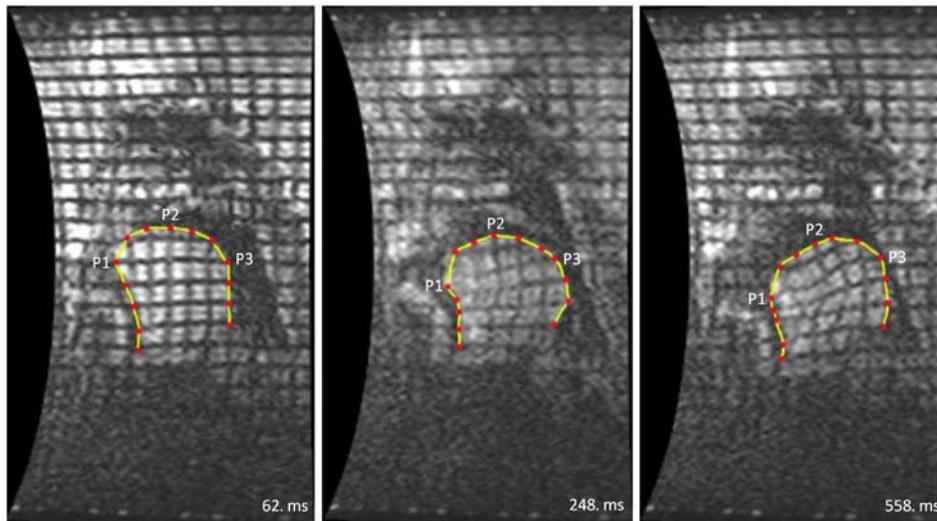


Figure 6. The contour of the tongue overlaid on the corresponding sagittal images of the human head (lips facing to the left) acquired at 62 (left), 248 (middle) and 558 ms. The contour is seen in yellow with contour points as red circles. The three selected points from the front (P1), top (P2) and back (P3) portions of the tongue are indicated.

the tongue moves very little into the /a/. This is because the tongue shape for /a/ is mostly in place during the /b/ and mostly jaw opening distinguishes the tongue position

between /b/ and /a/. After the 7th time frame (372 ms), the tongue begins moving forward to open the pharynx for the breath that starts at the 10th time frame.

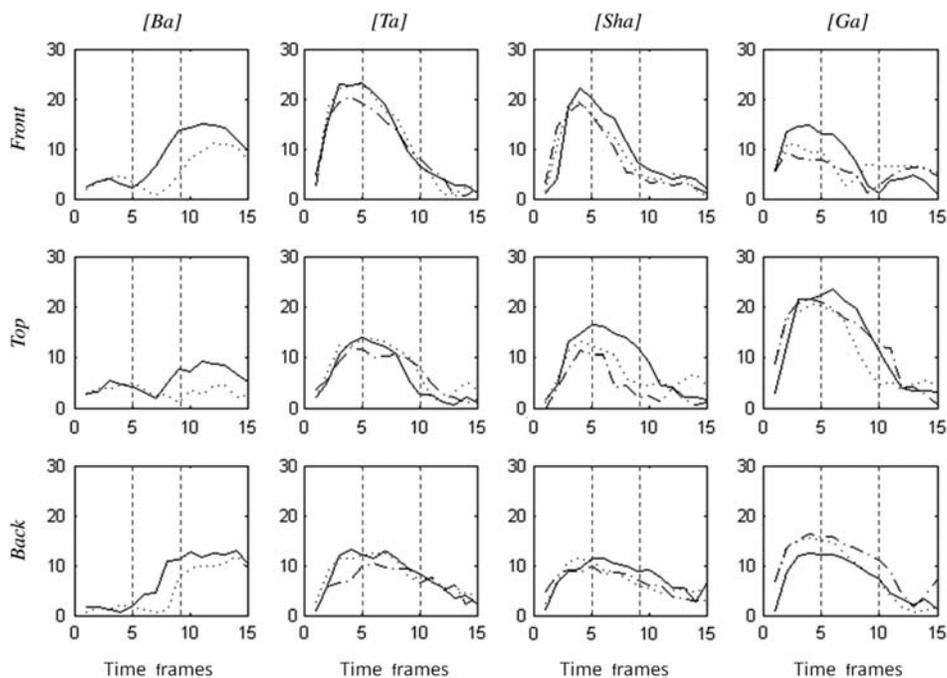


Figure 7. Total displacement values of three selected points from different regions of the tongue for four syllables and different repetitions are seen. Top, middle and bottom rows show the displacements of one reference point each. Each column belongs to one syllable and each graph contains three repetitions (solid line: 1st repetition; dotted: 2nd repetition; dash-dotted: 3rd repetition). Vertical dashed lines in each graph indicate corresponding maximum vowel displacement and the onset of inhalation, respectively. The y-axis is in millimetres and the x-axis is the time-scale where one unit corresponds to 62 ms.

4.3 4D analysis

Tags are equally spaced at the beginning and we know the length between each tag line. So, for each time frame we can calculate the local extensions, as explained before, and display them as a coloured tongue image. These coloured images, with red indicating expansion and blue compression, enable us to observe the movement of the compression regions within time. The final four figures (Figures 8–11) display the colour-coded, sagittal and axial images of the tongue for the first repetitions of the syllables only. Midsagittal and s2-axial slices are considered as sagittal and axial images, respectively, because they were visually the most appropriate slices to observe large tongue surface and internal motion. In each figure, the dashed line on the first sagittal image (0 ms) indicates the position of the corresponding axial slice. Note that in our local extension calculations, this position deforms along with the tongue across frames. Colour-coded images of the other repetitions are not displayed here, but a comparative analysis within repetitions is provided.

Our observations showed that the compression–expansion regions of the tongue change not only from syllable to syllable but also as time passes (Unay 2001). During the syllable /ba/ (Figure 8), we observe sagittal compression in the tongue tip (blue) from time frames 62 to 248 ms. However, the syllable /ba/ shows less tongue deformation than the other syllables because the /b/ is made with the lips, not the tongue. The tongue, therefore, is able to move from its pre-speech rest position into the /a/ with little or no deformation imposed by the /b/. In this case, the tongue moves back by expanding in the top region (red), and compressing anteriorly (aqua/blue), as it moves into the /a/. In the axial images, we observe expansion in the middle region of the tongue (yellow) in the first two time frames,

especially. At the beginning of the utterance, this region expands even to the edges. However, only the right side remains expanded in the end, with the left side of the tongue showing compression. Meanwhile the sagittal images show continued expansion, which reflects inward compression from the sides that causes the backward expansion seen in both the axial and sagittal images. A second repetition of /ba/ (not shown) showed consistent patterns of expansion and compression to the repetition depicted here.

The syllable /ta/ (Figure 9) compresses similarly to /sha/ (Figure 10). One salient difference between /ta/ and /sha/ is that /ta/ has more expansion in the lower back and more compression in the lower front throughout the motion. The /t/ uses a more anterior tongue root than the /sh/, and thus has greater expansion values into the /a/. In both cases the tongue lowers anteriorly, the tip compresses inward and the tongue body expands backward into the pharynx. In the /t/-to-/a/ transition the tip remains extended, and the tongue root remains more anterior throughout the /a/ than in the /sh/-to-/a/ transition. This may be in anticipation of the next repetition of /t/, which requires an extended tip and anterior root. For /t/, the downward motion of the tip is probably due in part to the contraction of genioglossus anterior and assisted by the recoil of the tongue tip as it is released from the palate. Two more repetitions of /ta/ (not shown) presented consistent patterns of expansion and compression to the repetition reported here.

For the syllable /sha/ (Figure 10), the sagittal view depicts variable expansion and compression regions in the sequence between /sh/ and /a/. The axial view shows that expansion occurs in the centre of the tongue body and expands to the posterior region as the tongue moves backward. In sagittal time, frame 2 compression is seen at the extreme upper and posterior edge of the midline

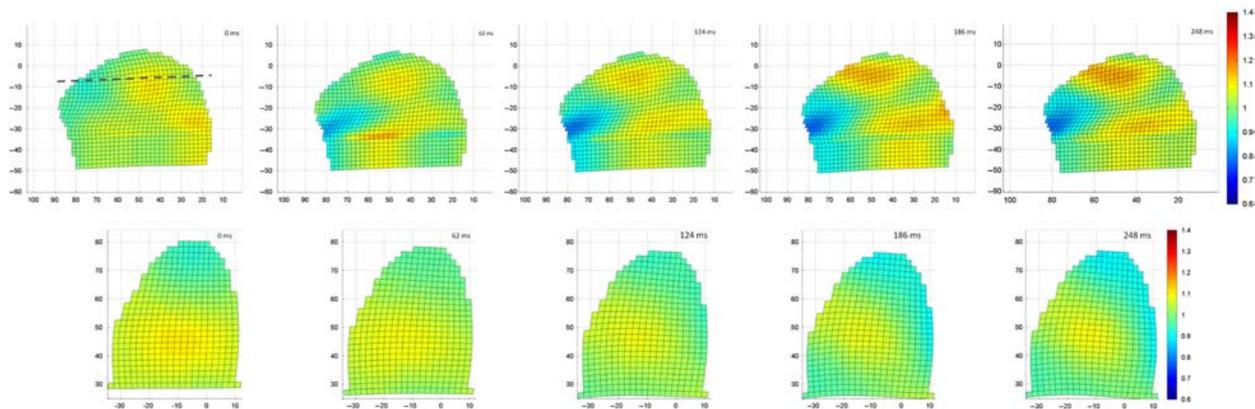


Figure 8. Colour-coded compression–expansion analysis image series for syllable /ba/ calculated from 4D analysis. Images in the top row belong to the midsagittal slice, and those below belong to the axial slice corresponding to the position marked by the dashed line on the first sagittal image. Note that the position of the dashed line (hence the position of the axial slice) deforms along with the tongue across each frame. Colour map of each image is included with relative extension values, where values higher than one signify expansion and those lower show compression. Therefore, red and blue colours represent expansion and compression, respectively (colour online).

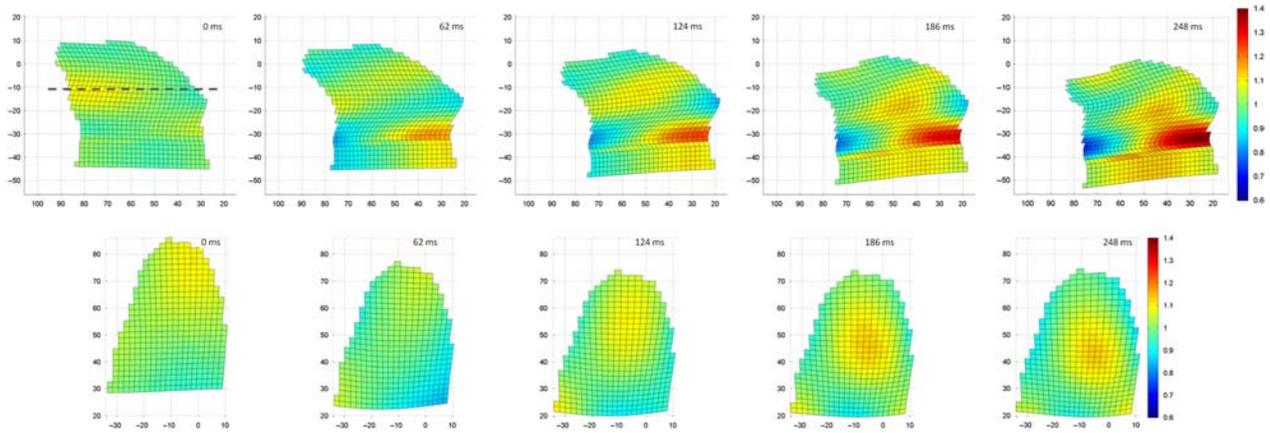


Figure 9. Colour-coded compression–expansion analysis image series for syllable /ta/ calculated from 4D analysis. Images in the top row belong to the midsagittal slice, and those below belong to the axial slice corresponding to the position marked by the dashed line on the first sagittal image. Red and blue colours represent expansion and compression, respectively (colour online).

tongue, consistent with recoil from the released tongue dorsum and stiffening of the posterior tongue to control the rate of backward motion. Later frames reflect continued inward motion at the tip and the anterior lower tongue which may be directing the expansion up and back into /a/. The slight tongue tip and blade expansion in time frame 1 is consistent with tongue lowering just prior to the release from the palate for the /sh/ (0 ms). The axial images capture the 3D component reflecting the asymmetrical release; both tongue palate contact for consonants and motion into the vowels are often asymmetrical.

In the other two repetitions (not shown) of /sha/, sagittal expansion also originates in the upper-central region, expands to the upper-back region of the tongue and is accompanied by compression in the region under the tip. For the first repetition, there is expansion (yellow) in the upper-central and lower-front regions of the tongue, and

compression under the tip and in the root. Repetitions 2 and 3 (not shown) show compression either below the tip (2) or in the root (3), indicating a trade-off between these two regions. Muscle trade-offs may allow different muscle combinations to produce the same surface shape, facilitating coordination of complex speech gestures. However, variation between repetitions could also result from other sources including, different surface shapes consistent with the different muscle patterns and also from errors in the tag tracking given that tagged data are always collected with a low resolution. For all the axial data, the centre of the tongue expands first and then extends backwards asymmetrically. The outer regions show compression. The variation between repetitions may reflect fine tuning of a single basic gesture in which anterior, posterior and lateral regions trade-off their contributions by compressing to greater or lesser extents. Since the tongue is volume

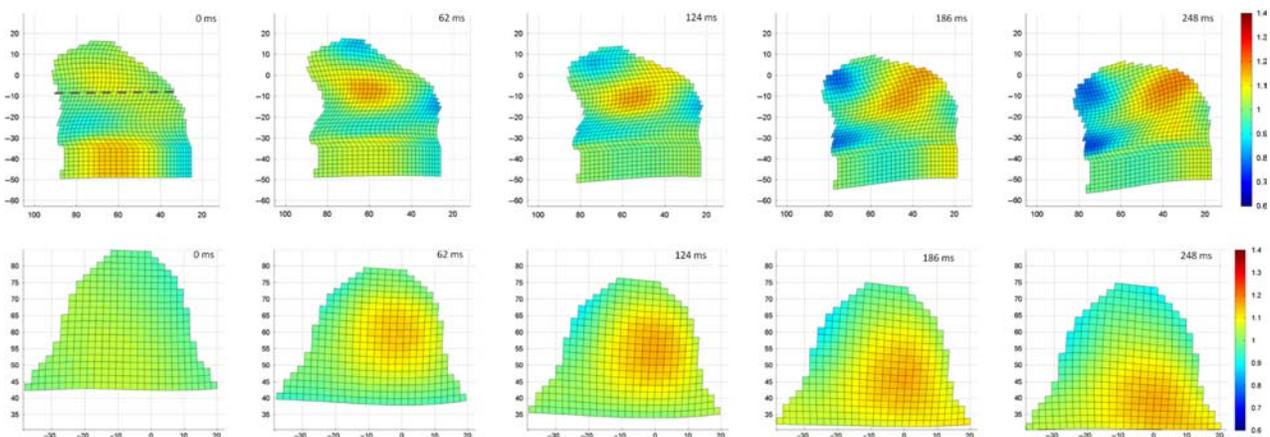


Figure 10. Colour-coded compression–expansion analysis image series for syllable /sha/ calculated from 4D analysis. Images in the top row belong to the midsagittal slice, and those below belong to the axial slice corresponding to the position marked by the dashed line on the first sagittal image. Red and blue colours represent expansion and compression, respectively (colour online).

preserving, the overall compression correlates with the backward expansion in the tongue body, despite muscle trade-offs.

The syllable /ga/ shows very large expansion in the tongue root and upper tongue surface immediately after the release of /g/, combined with compression in the anterior tongue below the tip. This is likely due to the rapid release of the tongue from the palate for the start of the /g/-to-/a/ motion. The tongue is compressed against the palate in time frame 1 and is released in time frame 2. The posterior tongue root was pulled forward during the /g/ and some backward motion positioning is observed at 248 ms. The tongue body was already high for /g/, however, and so no upward motion is seen in the posterior tongue. There is probably lateral expansion as well in this region. The extrinsic tongue muscles, such as styloglossus (not seen in these scans) may also contribute to the motion of the posterior tongue into /a/.

For the other two repetitions of /ga/ (not shown), we observe a similar expansion pattern in the tongue root. Regarding the upper tongue surface, expansion in repetition 2 is similar to that observed in repetition 1. However, for repetition 3 the expansion pattern is more diffuse in the upper tongue surface, and larger in the tongue body and in the root, indicating a trade-off between these regions.

In Table 1, we observe the maximum extension percentage values of each repetition of the syllables. The initial positions of the rectangular grids are taken as the reference. Considering all the syllables and repetitions, the most compression and expansion is observed for the syllable /ga/ in repetition 1 with 29.1% and 42.0%, respectively. The differences between compression and expansion percentages of each repetition can vary highly like in syllable /ga/ (11.8–29.1% compression and 27.0–42.0% expansion). But, this is not a drawback because these values are just the maximums, not the overall extensions. The syllables /sha/ and /ba/ show consistency within the repetitions, whereas we see high deviations for the repetitions of syllables /ta/ and /ga/.

Table 1. Maximum compression and expansion percentages inside the tongue for different syllables and repetitions.

Syllable	Compression (%)			Expansion (%)		
	rep. 1	rep. 2	rep. 3	rep. 1	rep. 2	rep. 3
/ba/	26.3	21.4	–	22.4	23.1	–
/ta/	14.9	25.9	17.6	19.8	30.0	30.9
/sha/	21.8	23.4	20.8	21.5	21.2	18.2
/ga/	29.1	11.8	23.0	42.0	27.0	33.1

Notes: Each row represents extension percentages of one syllable and each column belongs to one repetition, which is abbreviated by ‘rep.’. Columns 2–4 show the percentages of maximum compression observed for one syllable, whereas last three columns show those of expansion. Therefore, during the first repetition of syllable /sha/, we observe maximum compression of 21.8% and maximum expansion of 21.5% in the tongue.

5. Discussion

Segmented k-space imaging has been previously implemented in tongue imaging, but reliable synchronisation of imaging with auditory cues becomes a major limitation. In this paper, we took the approach of real-time imaging with the addition of MR tagging. This approach is to acquire consecutive images as rapidly as possible, using as short a time to repeat (TR) as possible. Images acquired in this fashion generally have limited spatial and temporal resolution. Improvements can be achieved by using echo-planar techniques (i.e. collecting several k-space lines per TR by modulating the gradients, the approach which is used here) or by parallel imaging via the use of multiple receiver coils (Pettigrew et al. 1999; Reeder and Faranesh 2000).

Error analysis of motion-tracking for the syllables as well as their repetitions showed that the errors were mostly no more than one pixel. There was only one error exceeding this bound, which is probably due to the motion-tracking algorithm originally developed for the heart.

The 2D displacement analysis included only the distances of selected material points throughout time relative to the initial time frame. Observations indicated consistency between the experimental results and theoretical knowledge on how the tongue moves during the four consonant-to-vowel syllables.

The 4D strain analyses show that the strain patterns of the tongue during the utterances of the four syllables are unique for each consonant, and that these patterns do not settle but propagate within the tongue through time. These sagittal and axial images only hint at the causes of this stretch. For example, Figure 9 shows that the back of the tongue expands posteriorly and vertically (248 ms). The yellow oblique region could be passive expansion, but also it is consistent with pull on the tissue from the styloglossus muscle. Similarly, the compression in the lower anterior tongue is consistent with muscle activation in genioglossus anterior, or activation of the jaw muscles, which would also contribute to tongue body elevation.

Figures 8–11 are consistent with trade-offs in compression and expansion used to produce the motion of muscular hydrostats (Smith and Kier 1985). Moreover, the differences between repetitions of the same syllables also support motor equivalence in the tongue, i.e. similar surface shapes can be produced with different combinations of muscles. For example, more variability in strain pattern across repetitions was seen for /ta/ and /ga/ than for /sha/ and /ba/. The /sh/ requires greater precision than the three stop consonants because it requires a narrow airstream to be directed precisely towards the palate. The stops /b/, /t/ and /g/ only require forceful closure and rapid release. Multiple repetitions of the stops may be produced with different amounts of closure force and release precision, thus increasing the variability of the strains in the tongue. Variability in /b/ closure would not be visible

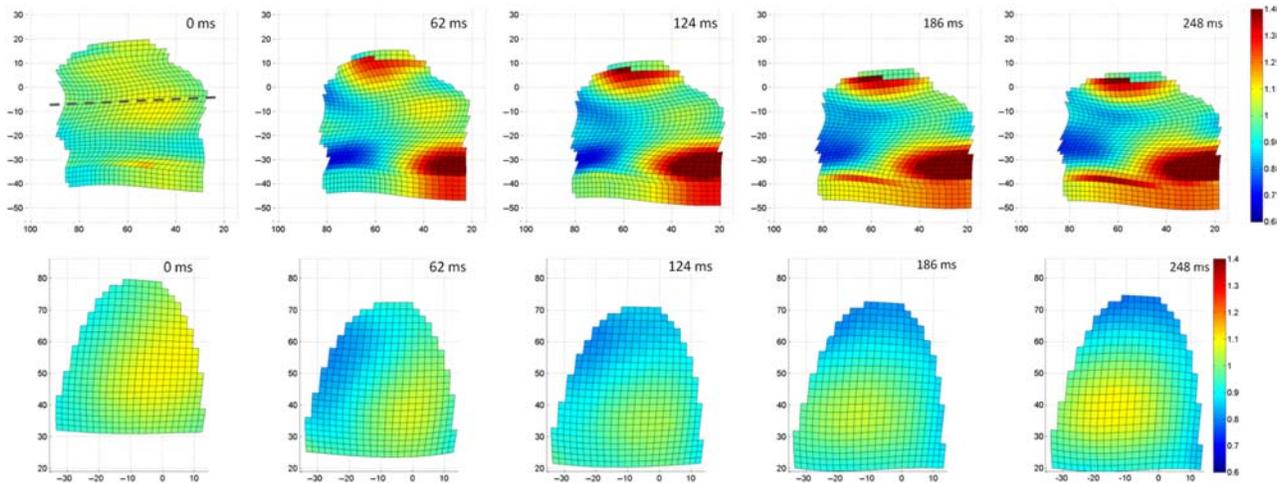


Figure 11. Colour-coded compression–expansion analysis image series for syllable /ga/ calculated from 4D analysis. Images in the top row belong to the midsagittal slice, and those below belong to the axial slice corresponding to the position marked by the dashed line on the first sagittal image. Red and blue colours represent expansion and compression, respectively (colour online).

in tongue motion, but this could account for the greater variability in tongue strain patterns seen in repeated motions from /t/ and /g/.

6. Open problems and future work

This paper provides the framework for future studies of multi-dimensional motion analysis of the tongue during different utterances. It is clear that understanding the mechanics of the internal muscles of the tongue is as important as revealing the tongue surface motion.

Tags are very useful in analysing the movement of the tongue, but for limited durations. However, our technique can be used to analyse the tongue movements for different utterances like vowel-to-consonants, with different choices of rounded–unrounded, tense-lax vowels and fricative-stop-sonorant consonants. Most of the work in this specific field remains to be done. Hopefully, with the improvement of our technique to a piecewise acquisition scheme, detailed motion analysis of the tongue during free speech could become a reality in the near future.

Acknowledgements

We wish to thank Dr Danielle Dick and the anonymous reviewers for their valuable contributions to this study.

References

- Akgul Y, Kambhamettu C, Stone M. 1999. Automatic extraction and tracking of the tongue contours. *IEEE Trans Med Imaging*. 18(10):1035–1045.
- Amini A, Chen Y, Curwen R, Mani V, Sun J. 1998. Coupled B-snake grids and constrained thin-plate splines for analysis of

2-D tissue deformations from tagged MRI. *IEEE Trans Med Imaging*. 17(3):344–356.

- Axel L, Dougherty L. 1989. MR imaging of motion with spatial modulation of magnetization. *Radiology*. 171(3):841–845.
- Badin P, Henrich N, Lamalle L. 2008. Articulatory comparison of spoken and sung vowels based on MRI. *J Acoust Soc Am*. 123(5):3730.
- Baer T, Gore J, Boyce S, Nye P. 1987. Application of MRI to the analysis of speech production. *Magn Reson Imaging*. 5(1):1–7.
- Buchaillard S, Perrier P, Payan Y. 2008. To what extent does tagged-MRI technique allow to infer tongue muscles' activation pattern? A modelling study. In: *INTERSPEECH*. Brisbane, Australia: ISCA. p. 2839–2842.
- Demolin D, Metens T, Soquet A. 1998. Real time MRI and articulator coordinations in vowels. In: *Proceedings of International Conference on Spoken Language Processing (ICSLP)*. Sydney, Australia p. 272–275.
- Dick D. 2000. Three-dimensional tracking of tongue motion during speech motion using tagged cine-MRI [master's thesis]. Baltimore, MD: The Johns Hopkins University.
- Fujita S, Dang J, Suzuki N, Honda K. 2007. A computational tongue model and its clinical application. *Oral Sci Int*. 4(2):97–109.
- Guttman R, Zerhouni E, McVeigh E. 1997. Analysis of cardiac function from MR images. *Comput Graph Appl IEEE*. 17(1):30–38.
- Honda K, Kusakawa N. 1997. Compatibility between auditory and articulatory representations of vowels. *Acta Otolaryngol Suppl*. 532:103–105.
- Kerwin WS, Prince JL. 2000. A k-space analysis of MR tagging. *J Magn Reson*. 142(2):313–322.
- Kiritani S, Itoh K, Fujimura O. 1975. Tongue-pellet tracking by a computer-controlled X-ray microbeam system. *Acoust Soc Am*. 57(6):1516–1520.
- MacNeilage PF, Sholes GN. 1964. An electromyographic study of the tongue during vowel production. *J Speech Hear Res*. 7(3):209–232.
- Marieb EN, Hoehn K. 2006. *Human anatomy and physiology*. 7th ed. San Francisco, CA: Benjamin Cummings.

- McVeigh ER. 1996. MRI of myocardial function: motion tracking techniques. *Magn Reson Imaging*. 14(2):137–150.
- McVeigh ER, Atalar E. 1992. Cardiac tagging with breath-hold cine MRI. *Magn Reson Med*. 28(2):318–327.
- Mermelstein P. 1973. Articulatory model for the study of speech production. *J Acoust Soc Am*. 53(4):1070–1082.
- Moulton MJ, Creswell LL, Downing SW, Actis RL, Szabo BA, Vannier MW, Pasque MK. 1996. Spline surface interpolation for calculating 3-D ventricular strains from MRI tissue tagging. *Am J Physiol Heart Circ Physiol*. 270(1):H281–H297.
- Ozturk C, McVeigh ER. 1999. Four-dimensional B-spline-based motion analysis of tagged cardiac MR images. In: Chen CT, Clough AV, editors. *Proc. SPIE medical imaging*. Vol. 3660. San Diego, CA: SPIE. p. 46–56.
- Ozturk C, McVeigh ER. 2000. Four-dimensional B-spline based motion analysis of tagged MR images: introduction and *in vivo* validation. *Phys Med Biol*. 45(6):1683–1702.
- Palmer J. 1973. Dynamic palatography. *Phonetica*. 28:76–85.
- Parthasarathy V, Prince JL, Stone M, Murano EZ. 2007. Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing. *J Acoust Soc Am*. 121(1):491–504.
- Pettigrew RI, Oshinski JN, Chatzimavroudis G, Dixon WT. 1999. MRI techniques for cardiovascular imaging. *J Magn Reson Imaging*. 10(5):590–601.
- Radeva P, Amini AA, Huang J. 1997. Deformable B-solids and implicit snakes for 3D localization and tracking of SPAMM MRI data. *Comput Vis Image Underst*. 66(2):163–178.
- Reeder SB, Faranesh AZ. 2000. Ultrafast pulse sequence techniques for cardiac magnetic resonance imaging. *Top Magn Reson Imaging*. 11(6):312–330.
- Shechter G, Declerck J, Ozturk C, McVeigh ER. 1999. Fast template based segmentation of cine cardiac MR. *Proc Int Soc Magn Reson Med*. 7(1):480.
- Smith KK, Kier WM. 1985. Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats. *Zool J Linnean Soc*. 83(4):307–324.
- Smith KK, Kier WM. 1989. Trunks, tongues, and tentacles: moving with skeletons of muscle. *Am Sci*. 77:28–35.
- Stark DD, Bradley WG. 1999. *Magnetic resonance imaging*. 3rd sub ed. St. Louis, MO: C.V. Mosby.
- Stone M. 1990. A three-dimensional model of tongue movement based on ultrasound and X-ray microbeam data. *J Acoust Soc Am*. 87(5):2207–2217.
- Stone M. 1991. Toward a three dimensional model of tongue movement. *J Phon*. 19:309–320.
- Stone M, Lundberg A. 1996. Three-dimensional tongue surface shapes of English consonants and vowels. *J Acoust Soc Am*. 99(6):1–10.
- Stone M, Faber A, Raphael L, Shawker T. 1992. Cross-sectional tongue shape and linguopalatal contact patterns in [s], [sh], and [l]. *J Phon*. 20:253–270.
- Stone M, Davis E, Douglas A, NessAiver M, Gullapalli R, Levine W, Lundberg A. 2001. Modeling the motion of the internal tongue from tagged cine-MRI images. *J Acoust Soc Am*. 109(6):2974–2982.
- Stone M, Liu X, Chen H, Prince JL. 2010. A preliminary application of principal components and cluster analysis to internal tongue deformation patterns. *Comput Methods Biomech Biomed Eng*. 13(4):493–503.
- Takano S, Honda K. 2006. Measurement of tissue deformation in the tongue during a vowel sequence /ei/ using tagged-cine MRI. In: 7th International Seminar on Speech Production. Ubatuba-SP, Brazil. p. 63–71.
- Unay D. 2001. Analysis of tongue motion using tagged cine-MRI [master's thesis]. Istanbul: Bogazici University.
- Zerhouni EA, Parish DM, Rogers WJ, Yang A, Shapiro EP. 1988. Human heart: tagging with MR imaging – a method for noninvasive assessment of myocardial motion. *Radiology*. 169(1):59–63.