JSLHR

Supplement

Tissue-Point Motion Tracking in the Tongue From Cine MRI and Tagged MRI

Jonghye Woo,[a] Maureen Stone,[a] Yuanming Suo,[b] Emi Z. Murano,[c] and Jerry L. Prince[b]

Purpose: Accurate tissue motion tracking within the tongue can help professionals diagnose and treat vocal tract–related disorders, evaluate speech quality before and after surgery, and conduct various scientific studies. The authors compared tissue tracking results from 4 widely used deformable registration (DR) methods applied to cine magnetic resonance imaging (MRI) with harmonic phase (HARP)–based tracking applied to tagged MRI.
Method: Ten subjects repeated the phrase "a geese" multiple times while sagittal images of the head were collected at 26 Hz, first in a tagged MRI data set and then in a cine MRI data set. HARP tracked the motion of 8 specified tissue points in the tagged data set. Four DR methods including diffeomorphic demons and free-form deformations based on cubic B-spline with 3 different similarity measures were used to track the same 8 points in the cine MRI data set. Individual points were tracked and length changes of several muscles were calculated using the DR- and HARP-based tracking methods.
Results: The results showed that the DR tracking errors were nonsystematic and varied in direction, amount, and timing across speakers and within speakers. Comparison of HARP and DR tracking with manual tracking showed better tracking results for HARP except at the tongue surface, where mistracking caused greater errors in HARP than DR.
Conclusions: Tissue point tracking using DR tracking methods contains nonsystematic tracking errors within and across subjects, making it less successful than tagged MRI tracking within the tongue. However, HARP sometimes mistracks points at the tongue surface of tagged MRI because of its limited bandpass filter and tag pattern fading, so that DR has better success measuring surface tissue points on cine MRI than HARP does. Therefore, a hybrid method is being explored.

Key Words: deformable registration, tongue motion, MRI, muscle length, point tracking

Tongue motion is usually measured at the tongue surface. Imaging techniques such as cine magnetic resonance imaging (MRI) provide good brightness contrast, allowing extraction and tracking of tongue surface contour motion. Point tracking systems, such as electromagnetic articulography, track motion of pellets affixed to the tongue surface. These two types of data assess tongue surface shape, motion, and position in order to elucidate tongue and vocal tract function. Measurements within the tongue would enhance these data but are difficult to obtain. The only method currently available is tagged MRI (Parthasarathy, Prince, Stone, Murano, & Nessaiver, 2007), which is much less easily available than electromagnetic articulography or cine MRI. Therefore, in the present work we assessed the use of cine MRI movies to track internal tissue point motion in the tongue using deformable registration (DR) to see whether DR can track tissue points from cine MRI as accurately as they can be tracked from tagged MRI.

Tagged MRI has been used in previous work to track internal tissue motion of the tongue (Parthasarathy et al., 2007). Since Zerhouni and colleagues (Zerhouni, Parish, Rogers, Yang, & Shapiro, 1988) first developed tagged MRI to create visible myocardial markers, tagged MRI has been widely used in cardiac motion tracking (Axel, 2002; el Ibrahim, 2011). It has also been used for other motion-estimation tasks, including imaging motion of the tongue in speech (Parthasarathy et al., 2007). Tagged MRI changes the magnetization in tissue planes, causing image intensity changes whose motions are then tracked through time (Axel & Dougherty, 1989). More specifically, MRI records the density of hydrogen in tissue. Tagged MRI temporarily magnetizes planes of tissue so that tagging can be understood as the multiplication of the magnetization of the anatomy with a two-dimensional sinusoid (Parthasarathy et al., 2007). The

[a]University of Maryland School of Dentistry, Baltimore, MD
[b]Johns Hopkins University, Baltimore, MD
[c]Johns Hopkins Hospital, Baltimore, MD
Correspondence to Maureen Stone: mstone@umaryland.edu

Editor: Jody Kreiman
Associate Editor: Kate Bunton
Received June 29, 2012
Revision received February 4, 2013
Accepted September 26, 2013
DOI: 10.1044/2014_JSLHR-S-12-0208

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

tags are placed just before the subject begins speaking, and cine MRI images are captured throughout the speech task. As the tissue moves, the tags move with it, allowing the internal deformation of the tongue, lips, and velum to be observed and tracked. Having detailed information about soft tissue deformation makes it possible to compute many other function measurements, such as displacement, velocity, rotation, translation, elongation, strain, and local deformation.

Several tagging methods have been proposed, including spatial modulation of magnetization (SPAMM; Axel & Dougherty, 1989), complementary SPAMM (CSPAMM; Fischer, McKinnon, Maier, & Boesiger, 1993), and delay alternating with nutation for tailored excitation (Morris & Freeman, 2011). However, despite their abilities to offer different spatial tag patterns, quantifying the tissue motion itself still requires specialized postprocessing methods. Several methods to process tagged images have been proposed, including FindTags (Guttman, Prince, & McVeigh, 1994), harmonic phase (HARP; Osman, Kerwin, McVeigh, & Prince, 1999; Osman, McVeigh, & Prince, 2000), Gabor filters (Chen, Wang, Chung, Metaxas, & Axel, 2010), and SinMod (Arts et al., 2010). As well, both displacement encoding with simulated echoes (Aletras, Ding, Balaban, & Wen, 1999) and strain encoded imaging (Osman, Sampath, Atalar, & Prince, 2001) use stimulated echoes to establish more direct encoding of the displacement and strain, respectively, in the recovered signals. For an in-depth review of various tagged MRI and stimulated echo techniques and applications, see Axel, Montillo, and Kim (2005) and el Ibrahim (2011).

Cine MRI is a fast imaging technique that does not involve tags and that is frequently used in speech analysis because it allows noninvasive observation and measurement of tongue motion (Narayanan, Nayak, Lee, Sethy, & Byrd, 2004; Stone et al., 2001; Story, 2009; Winkler, Fuchs, Perrier, & Tiede, 2011). Cine MRI might be well suited to studies of tongue motion if it could map tongue muscle shortening to tongue surface deformation. However, cine MRI is weak at such mapping because it does not image internal muscle shortening or track tissue point motion from which to measure changes in position of the muscle origin or insertion. Despite the limitations of cine MRI, it could supplement tagged MRI, because tagged MRI is less successful at surface tracking than internal tracking. Therefore, this study was motivated by the idea that cine MRI could provide better tissue tracking at the tongue surface than tagged MRI. In a recent study, Woo, Stone, and Prince (2011) performed the fusion of cine MRI and high-resolution MRI in order to impose muscle anatomy from high-resolution MRI onto the time frames (TFs) of cine MRI. In the future, rapid development of MRI technology could also allow high-quality cine MRI to display textures within the tongue that reflect tissue characteristics, such as muscle bundles, fascia, and tendons. Textural information in cine MRI tongue images might allow the tracking of tissue points that represent points within or at the origins and insertions of muscles. In the present study, however, we used currently available cine MRI images that are representative of the current state of the art.

In this work, we applied four widely used DR methods to track tissue points from cine MRI so that we could compare the performance with HARP tracking from tagged MRI acquired in the same spatiotemporal coordinates as the cine MRI. In general, DR aims to find correspondences between source and target images by using a nonlinear transformation of the coordinate system. By using successive registrations of different TFs, it is possible to track motion of tissue points. In this work, we applied four intensity-based DR methods, including (a) diffeomorphic demons (Vercauteren, Pennec, Perchant, & Ayache, 2009) and free-from deformations based on parametric cubic B-splines (Rueckert et al., 1999) with three different similarity measures: (b) mutual information (MI), (c) sum of squared differences (SSD), and (d) normalized cross-correlation (NCC). We used these methods because they are state of the art and the different similarity measures and transformation models differentially affect their tracking performance. Each of the similarity measures has unique characteristics. MI is the most popular similarity measure for multimodal image registration (Pluim, Maintz, & Viergever, 2003). SSD and NCC are used when two images are acquired similarly with similar intensity range (Woo et al., 2010). In particular, NCC is well suited when intensity distributions between source and target images have an affine relationship (Hermosillo, Chefd'Hotel, & Faugeras, 2002). Diffeomorphic demons, like SSD, is based on the assumption that corresponding pixels have the same intensity values, but it uses spatial derivatives to form putative directions to warp the underlying coordinate system. Both transformation models are considered to be highly deformable (or so-called "free form"), but the B-spline model is parametric, whereas diffeomorphic demons uses a nonparametric model that enforces a diffeomorphism.

There exists literature on tracking the motion of the heart using cine MRI (Chandrashekara, Mohiaddin, & Rueckert, 2005; Gupta & Prince, 1995). Chandrasekhar and colleagues (2005) in particular used DR methods and found that cine measurements were well correlated to tagged measurements; however, cine MRI did not capture cardiac twisting well. Tongue motion differs from heart motion in that it is more deformable and its motions are faster and more varied. This could diminish the quality of tracking from DR. However, the tongue is less homogeneous than the heart, and therefore DR tracking might make use of the tissue heterogeneity.

In this work, midsagittal tissue points were specified to represent tissue points bounding intermediate segments of the superior longitudinal (SL) muscle and the origin and insertion of genioglossus muscles. Muscle length changes during speech were determined and validated against tagged MRI data for the same tasks to decide whether traditional cine MRI movies could provide reasonably accurate measurements of tissue point motion and muscle length change. To make these comparisons, both a cine MRI and a tagged MRI data set were collected in a single data recording session using identical spatial and temporal parameters. We used the tagged MRI data, analyzed by means of HARP, to assess the errors of the four DR estimations of tissue-point location

and muscle length. We used hand-tracked tissue points to compare tissue point tracks from HARP and DR.

Method

Subjects and Speech Task

The subjects were 10 normal native speakers of American English between ages 22 and 52 years (six women and four men). The speech task was "a geese." This task starts with a schwa, which positions the tongue in the center of the vowel space. The schwa is followed by upward tongue body motion (/g/), forward body motion into (/i/) and backing of the tongue body (/s/), while maintaining an elevated tip and blade. The jaw is minimally engaged, so that vocal tract shaping is heavily dependent on tongue deformation. Finally, the task can be repeated in 1 s, which was our MRI record time.

MRI Instrumentation and Data Collection

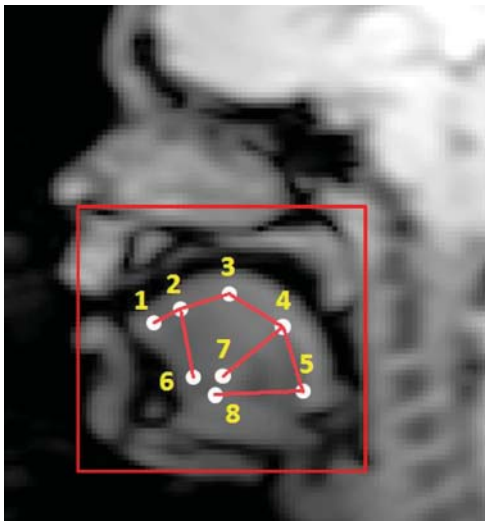
All MRI scanning was performed on a Siemens 3.0 T Tim Treo system (Siemens Medical Solutions, Malvern, PA) with a 16-channel head and neck coil. Two midsagittal MRI data sets were collected from each speaker in the same session: (a) a midsagittal cine MRI movie (cine) and (b) a midsagittal tagged MRI movie (tagged). The tagged MRI data were collected using magnitude-imaged CSPAMM reconstructed (MICSR) images (NessAiver & Prince, 2003; Parthasarathy et al., 2007). Both data sets had a 1-s record duration, 26 TFs per second, 6-mm slice thickness and tag separation (in the tagged data), no gap between slices, and a 1.875-mm in-plane resolution. Both data sets had identical parameters, including slice location, field of view, and so on. In-plane resolution was 1.875 mm × 1.875 mm, and seven sagittal slices were acquired. Other sequence parameters were repetition time (TR) 36 ms, echo time (TE) 1.47 ms, flip angle 6°, and turbo factor 11.

Both MRI methods produce a single "movie" for each slice by acquiring and summing multiple repetitions of the speech task. MICSR requires three repetitions per tissue slice acquired four times (12 repetitions), and the cine MRI algorithm we used requires five repetitions per tissue slice. The midsagittal data were extracted from a sagittal "stack" of five, seven, or nine slices depending on the subject's tongue size. Thus, this study focused on a midsagittal slice to analyze the motion. We optimized speaker precision during repetition by training each subject to speak to a metronome beat, which is also used in the scanner. Data from misaligned repetitions were discarded. The training method was based on the work of Masaki et al. (1999). Acoustic recordings of speech were made in the MRI scanner with a subtraction-type fiberoptic microphone (Optoacoustics Ltd., Israel) to corroborate phoneme locations in the MRI image sequences (Boersma & Weenink, 2010).

Tissue Point Selection and Tracking

During analysis, eight tissue points were selected manually within the tongue for each subject in the first TF (TF1) of the cine movie (see Figure 1). The points in Figure 1 represent the location of the SL muscle in the tongue—tip, blade, dorsum, pharynx, and root—as well as the origins of the genioglosus anterior (GGa), medial (GGm), and posterior (GGp).

Figure 1. Eight points and their affiliated muscles. Points 1 through 5 define segments within the superior longitudinal muscle. The rest indicate user-defined origins and insertions of fibers in the genioglosus anterior (GGa; 2–6), genioglosus medial (GGm; 4–7), and genioglosus posterior (GGp; 5–8) muscles. The red square defines the region of interest used in the analyses. The numbers represent the location of the superior longitudinal muscle in the tongue (1 = tip; 2 = blade; 3 = dorsum; 4 = pharynx; 5 = root) as well as the origins of the GGa (6), the GGm (7), and GGp (8).



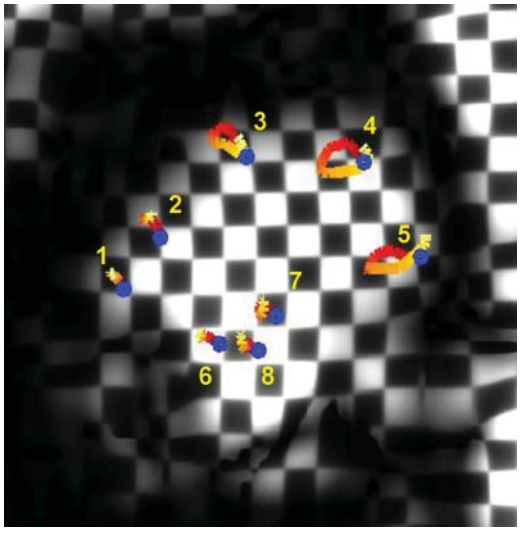
DR tracks, or registers, each pixel in TF1 of the cine MRI sequence with the closest match, based on image features, in each of the remaining 25 TFs. To track the motion of the tongue, we needed to find the coordinate system transformation describing how each point moves over time. To find this transformation, we used either (a) diffeomorphic demons (Vercauteren et al., 2009) or free-from deformations based on cubic B-spline (Rueckert et al., 1999) with similarity measures (b) MI, (c) NCC, or (d) SSD. These algorithms are available in the Insight Segmentation and Registration Toolkit library (Ibanez, Schroeder, Ng, & Cates, 2003). Whereas MI and NCC can deal with potential intensity differences in the registration process, SSD works under the assumption that the corresponding pixels have the same intensity values. Thus, prior to the SSD-based registration and diffeomorphic demons, we applied histogram matching. Histogram matching normalizes the intensity values of a source frame based on the intensity values of the target frame. Unlike SSD, which requires same-intensity values in the corresponding pixels, MI and NCC can compensate for intensity differences, so histogram matching was not used for these algorithms. All steps, including preprocessing and the registration algorithms themselves, were fully automated.

Let TF(x, t): $\Omega \subset \mathbb{R}^2 \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ denote the space-time acquisition of the frame, where Ω corresponds to an open and bounded domain and t denotes the TF (i.e., $t = 1, \ldots, 26$). For brevity, TF(x, 1): $\Omega \subset \mathbb{R}^2$ is the first TF, denoted as *TF1*. After DR, the locations of the eight tissue points chosen in TF1 space were tracked in the other 25 TFs using all four methods. We used a linear interpolation method to approximate the grid in the image.

For all four registration methods, we considered sequential versus individual tracking. Individual tracking takes a single tissue point in TF1 and independently deforms it to the optimal location in the other frames: $1 \rightarrow 2$, $1 \rightarrow 3$, $1 \rightarrow 4$, …, $1 \rightarrow 26$. Therefore, any one transformation is unaffected by the others. Sequential tracking deforms the single tissue point from TF1 to TF2, then from $2 \rightarrow 3$, $3 \rightarrow 4$, …, and so on, so that the path of that point is followed. Sequential tracking has the potential to propagate errors made in early frames to registration of later frames.

We used the tagged MRI data set, which is the best available measurement of tag motion, as the basis for calculating the error in the DR tracking estimates of point location and muscle length. A tagged MRI image is depicted with the motion path of eight tissue points for Subject 7 in Figure 2. The black-and-white grid depicts the intersections of horizontal and vertical tagged regions. It is used during motion to better visualize local tag deformation in the tongue. This is TF1 and no motion has occurred, so the grid is undeformed and just contains squares.

Figure 2. A tagged magnetic resonance image (MRI) illustrating the motion path of eight tissue points for Subject 7. The back of the tongue moves faster and farther than the front. Colors indicate time, with yellow occurring earlier than red. Note that the black-and-white grid is not an artifact; it is the tag grid filled in with black and white squares. The numbers represent the location of the superior longitudinal muscle in the tongue (1 = tip; 2 = blade; 3 = dorsum; 4 = pharynx; 5 = root) as well as the origins of the GGa (6), GGm (7), and GGp (8).



We analyzed the MICSR images using the HARP method, which tracks tags by determining the changes in harmonic phase over time (NessAiver & Prince, 2003). More specifically, tagged images have two harmonic peaks in the frequency domain (Xing et al., 2013). To isolate the spectral peaks, a bandpass filter is used, which reduces resolution of the reconstructed motion field and causes blurring. HARP uses phase information through time to track in every point of horizontal and vertical tagged images where a dense two-dimensional motion field is obtained in each direction, respectively. In addition, the refinement methods were used to address erroneous tracking due to large tongue motion (Liu & Prince, 2010) and tag jumping (Liu, Murano, Stone, & Prince, 2007; Liu & Prince, 2010), respectively. To date, HARP has provided fast, accurate assessment of the myocardial strains (Garot, Bluemke, Osman, Rochitte, McVeigh, et al., 2000) and regional function (Garot, Bluemke, Osman, Rochitte, Zerhouni, et al., 2000) of the heart from tagged MRI. It also has been recognized as a highly accurate program used by many in tagged MRI analysis. Therefore, we used it in the present study to compare the accuracy of the DR tracking methods (Cho, Chan, Leano, Strudwick, & Marwick, 2006) and cardiac torsional deformation (Notomi et al., 2005).

Because HARP has a less accurate tracking performance at tongue edges, the eight points were selected slightly below the surface in TF1 of the cine MR images. They were then superimposed onto TF1 of the tagged images. Recall that, absent head motion, the tissue points should be the same in both data sets. The locations of these tissue points were then tracked by HARP through the 26 TFs (cf. Parthasarathy et al., 2007). Examination of their motion in the tagged data sets allowed identification of possible tag jumping or mistracking. If a tracking or jumping error occurred in the tagged data, a neighboring point was selected in the tagged data set and tracked. Once a well-tracked point was identified, it also was used in the cine data set. The eight points were located, as much as possible, at or near the two ends of the muscle sections identified above.

Evaluation

In this section, we describe a series of experiments we conducted to assess and quantitatively compare the tracking performance of HARP and DR. This includes tracking individual points, measuring muscle length change, and using manual tracking in comparison to HARP and DR tracking.

Tracking motion of specific tissue points using DR and HARP. To test the accuracy of HARP and DR, we tracked the motion of eight tissue points. *Error* is defined as an absolute difference between HARP tracking and the DR method under consideration. We report global statistics on the tracking errors in all eight subjects and give special scrutiny to the two subjects whose tracking errors were the best and the worst.

Error in muscle length change using DR and HARP. We calculated muscle length using the Euclidean distance between tracked points. The distances between the circumferential points defined four regions of the SL muscle (1–2, 2–3, 3–4,

4–5) and three regions of the genioglossus: (a) GGa (2–6), (b) GGm (4–7), and (c) GGp (5–8; see Figure 1). Points 6 and 7 were above the tendonous origin of the genioglossus, and Point 8 was below. We computed the lengths using both DR- and HARP-based tracking.

Comparison between tissue point tracks using DR, HARP, and manual tracking. A test was performed to determine whether it might be advantageous to combine DR and HARP in tracking tissue points. The tongue is surrounded by air, and tissue points at the tongue surface are often poorly tracked by HARP (Parthasarathy et al., 2007). DR, on the other hand, excels at matching edges; thus, the combination of methods might provide excellent tissue point tracking both on and in the tongue. To test the idea of combining DR at the edges and HARP within the tongue, we selected three points from the subject with the best tracked points (Subject 7) to give HARP and NCC the best chance of tracking accurately. These points were also tracked by hand to compare the experimenter's judgment with the NCC and HARP results. In order for the experimenter to track the points, the points were chosen at intersecting gridlines in the MICSR images. The first was chosen at the gridline closest to Point 4. It was as superficial as the intersecting tag lines would allow. The other two points were chosen at the next two closest tag–line intersections that were successively deeper in the tongue. Measured distances were 6 mm between Points A and B and 8 mm between Points B and C. The three points were tracked manually by the second author to determine their position in each TF. These measurements were used to validate both HARP and the NCC point tracking trajectories.

Results

Tracking Motion of Specific Tissue Points Using DR and HARP

The results of four different tracking methods are presented in Tables 1 through 4. As shown in Tables 1 through 4, the NCC was the best by a small margin, so in this article we present the results of the NCC method. In addition, the MI method is the most commonly used in DR, so the MI results are also presented. The results indicated that the individual tracking method produced less error than sequential tracking for all subjects and methods. Therefore, we present only the individual tracking results.

The tissue point trajectories of two subjects, tracked by HARP, MI, and NCC, are displayed in Figure 3. The circles indicate the location of the first (red) and last (blue) tissue point in the motion of "a geese." The green line shows the trajectory through the 26 TFs. Subject 7, the best tracked subject, had mean errors of 2.2 mm and 1.7 mm for the MI and NCC tracking, respectively. Subject 3 had large mean errors in both methods—4.4 mm (MI) and 3.2 mm (NCC)—and had particularly poor tracking of the tongue tip and dorsum (Points 1 and 3).

The four DR methods all generated considerable tracking error. Errors for the best and worst methods are shown in Tables 1 through 4 and Figure 4. Tables 1 through 4 show the mean error over the 26 TFs for each point by subject for MI (worst) and NCC (best). NCC was considerably better than MI, but in most cases both methods had poor performance within the tongue (see Figure 4). Points 6, 7, and 8 had smaller errors because they were near the mandible and tended to move very little. The other points had fairly large errors, in particular Points 1 through 4, which had large motions. In Figure 4 are graphed the errors of the two methods relative to HARP for Points 1 and 3 for Subjects 3 and 7. There were no errors in TF1 because the starting points were the

Table 1. Error (in millimeters) for the mutual information–tracked tissue points, each averaged over 26 time frames.

	Subject									
Point	1	2	3	4	5	6	7	8	9	10
1	3.9	3.1	7.7	6.4	3.5	6.3	1.1	4.1	4.3	4.3
2	6.8	4.2	3.7	3.2	7.2	4.3	3.8	3.3	8.1	5.9
3	7.0	5.3	6.5	4.3	7.4	4.5	2.5	7.1	3.6	6.1
4	2.5	3.0	4.0	2.3	3.5	2.5	2.8	5.0	5.2	4.8
5	2.0	1.9	2.7	2.9	2.4	2.0	2.7	4.5	4.5	6.8
6	1.6	3.8	2.5	3.0	1.3	4.3	1.3	1.8	4.7	1.9
7	3.4	5.7	4.1	2.9	2.6	6.2	1.7	2.0	5.6	2.0
8	2.4	5.5	4.0	2.5	2.4	4.6	1.9	1.9	4.0	2.0
M	3.7	4.1	4.4	3.4	3.8	4.3	2.2	3.7	5.0	4.2
SD	2.1	1.4	1.8	1.3	2.3	1.5	0.9	1.9	1.4	2.0

Table 2. Error (in millimeters) for the normalized cross-correlation–tracked tissue points, each averaged over 26 time frames.

	Subject									
Point	1	2	3	4	5	6	7	8	9	10
1	2.7	3.4	7.6	5.5	2.4	6.7	1.3	3.1	2.8	3.9
2	3.7	3.3	3.3	3.8	5.2	4.7	2.2	2.7	4.1	3.1
3	4.7	2.6	2.7	3.9	4.2	2.6	2.2	6.5	2.7	3.2
4	4.5	2.9	2.8	2.1	2.7	2.4	2.2	5.1	3.7	3.6
5	2.7	2.1	2.1	2.4	1.6	2.0	2.0	4.1	3.1	4.5
6	2.9	3.8	2.6	3.3	2.1	3.2	0.8	1.8	3.0	2.6
7	2.8	5.2	2.2	2.7	2.1	3.5	1.1	2.0	3.7	2.6
8	1.8	5.0	2.4	1.9	1.6	1.6	1.8	1.6	2.9	2.2
M	3.2	3.5	3.2	3.2	2.7	3.3	1.7	3.4	3.2	3.2
SD	1.0	1.1	1.8	1.2	1.3	1.7	0.6	1.7	0.5	0.8

Table 3. Error (in millimeters) for the sum of squared differences–tracked tissue points, each averaged over 26 time frames.

	Subject									
Point	1	2	3	4	5	6	7	8	9	10
1	2.7	3.6	7.4	6.0	2.6	6.6	1.5	2.8	4.1	5.0
2	6.0	5.1	4.0	5.0	7.8	4.6	3.0	3.8	7.0	4.9
3	4.9	2.4	2.8	3.8	4.4	3.5	2.0	7.8	3.8	3.7
4	5.0	3.1	3.7	2.0	3.2	1.7	2.7	5.8	4.7	4.1
5	2.2	2.2	2.5	2.9	2.3	2.0	1.9	4.1	5.4	5.1
6	2.9	5.1	2.3	4.7	2.8	4.0	1.2	1.8	4.8	2.5
7	3.0	7.7	2.7	4.2	4.0	6.2	1.2	2.1	6.8	2.1
8	2.7	7.3	3.1	3.0	3.4	2.5	3.3	1.3	5.0	2.0
M	3.7	4.6	3.6	4.0	3.8	3.9	2.1	3.7	5.2	3.7
SD	1.9	2.3	1.8	2.2	2.0	1.8	1.3	1.8	2.5	1.6

Table 4. Error (in millimeters) for the diffeomorphic demons–tracked tissue points, each averaged over 26 time frames.

	Subject									
Point	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	3.6	2.1	7.7	5.4	0.8	6.5	1.1	4.5	3.3	3.6
2	4.8	2.4	2.6	3.2	1.7	4.9	2.0	3.3	4.3	4.4
3	6.4	3.3	3.7	2.7	2.8	4.0	1.3	4.9	5.1	4.6
4	5.0	3.2	2.6	1.0	1.9	2.6	1.5	3.8	6.9	4.7
5	2.8	2.5	2.2	2.0	1.3	1.4	2.4	3.6	7.6	4.5
6	2.3	2.0	1.9	3.9	1.4	4.4	0.6	1.1	3.8	2.3
7	2.9	2.9	1.7	2.5	2.4	4.3	1.1	1.6	6.2	1.4
8	1.8	2.8	2.2	2.5	2.2	4.2	1.1	1.3	5.1	1.9
M	3.7	2.6	3.1	2.9	1.8	4.1	1.4	3.0	5.3	3.4
SD	1.7	1.4	1.4	1.4	0.9	1.7	0.9	1.4	2.5	1.3

same in both data sets. For both subjects, NCC (bottom panels) was tracked more accurately than MI, although Point 1 for Subject 3 had nearly identical error using both DR methods.

Error in Muscle Length Change Using DR and HARP

The SL muscle had small length changes for this task, possibly because the segments were short. In addition, errors for all SL muscle segments were highly variable. Because GGp length was derived from Points 5 and 8 (see Figure 1), both of which were tracked fairly well for all subjects (Tables 1–4), we use only the DR data for the GGp, which were among the best, to demonstrate the error patterns.

The changes in muscle length for GGp, calculated using HARP, are shown in Figure 5, Panel A, along with the lengths calculated using MI and NCC. In addition, Panel B of Figure 5 illustrates the GGp muscle length change as a percentage with respect to the HARP tracking results. The contact points /g/, /i/, and /s/ were derived using the visual inspection of the vocal tract of the motion pattern in the midsagittal MR images. HARP tracking of the GGp (solid lines) shows that it started to shorten before tongue–palate contact was made for /g/ (first vertical line) in all subjects, with the most rapid shortening occurring before /i/ contact (second vertical line). Thus, shortening of the GGp was associated most strongly with /i/ production. The GGp remained shortened during the /i/ and began to lengthen just after /s/ contact (third vertical line) in all subjects except Subject 6, who continued shortening the GGp through the /s/. The results of the MI (dashed lines) and NCC (dotted lines) registration were sometimes similar in pattern and timing to the HARP results (solid lines), as seen for Subject 5. The errors, however, were not systematic within or across subjects. DR overestimated length in Subject 1 and underestimated length in Subjects 3, 4, and 8. For Subjects 2, 6, and 10, DR did both at different times, and the DR was quite noisy for Subjects 7 and 9. Timing errors also varied from subject to subject. The frame at which the GGp began to shorten or lengthen was fairly accurate for Subject 1, but not in most of the others. As with the direction and extent of

Figure 3. Motion patterns of eight points for Subjects 7 and 3 using harmonic phase (HARP) tracking in the tagged MRI data set and using mutual information (MI) and normalized cross-correlation (NCC) in the cine MRI data set. The red circle is the first time frame, the blue circle is the last time frame, and the path is green.

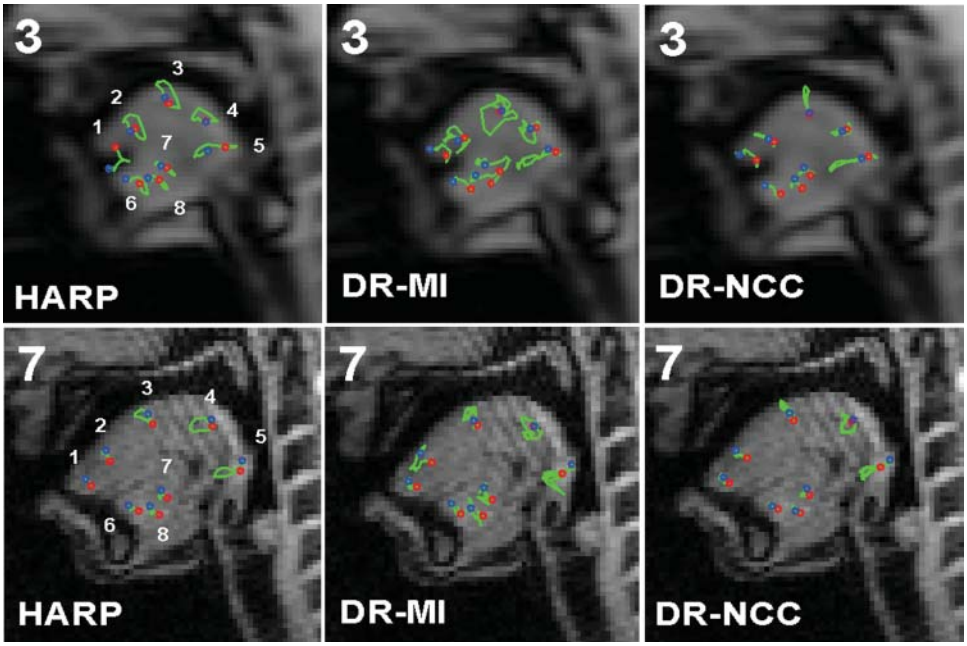
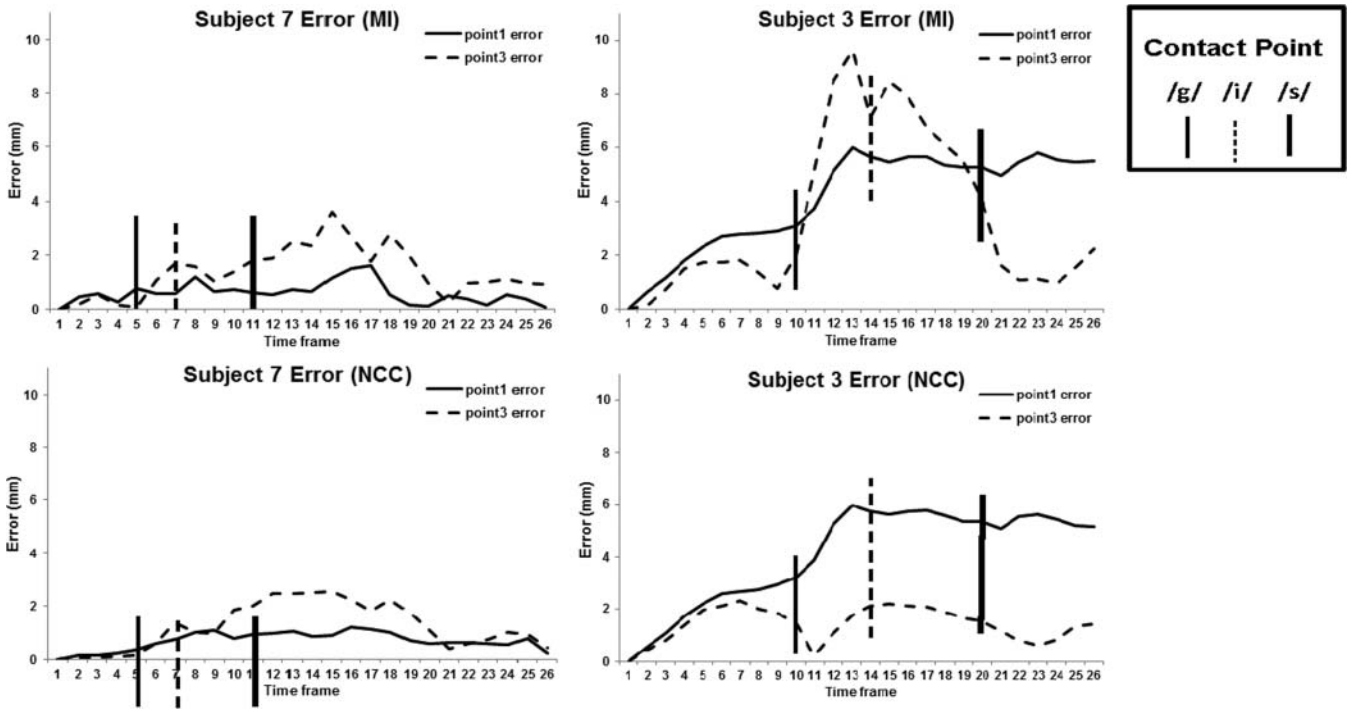


Figure 4. Error compared to HARP tracking of Points 1 and 3 for subjects with small (Subject 7) and large (Subject 3) DR tracking errors.



motion, timing errors were not predictable. Similar error patterns were seen for the GGa, GGm, and SL muscles (results not shown).

Comparison Between Tissue Point Tracks Using DR, HARP, and Manual Tracking

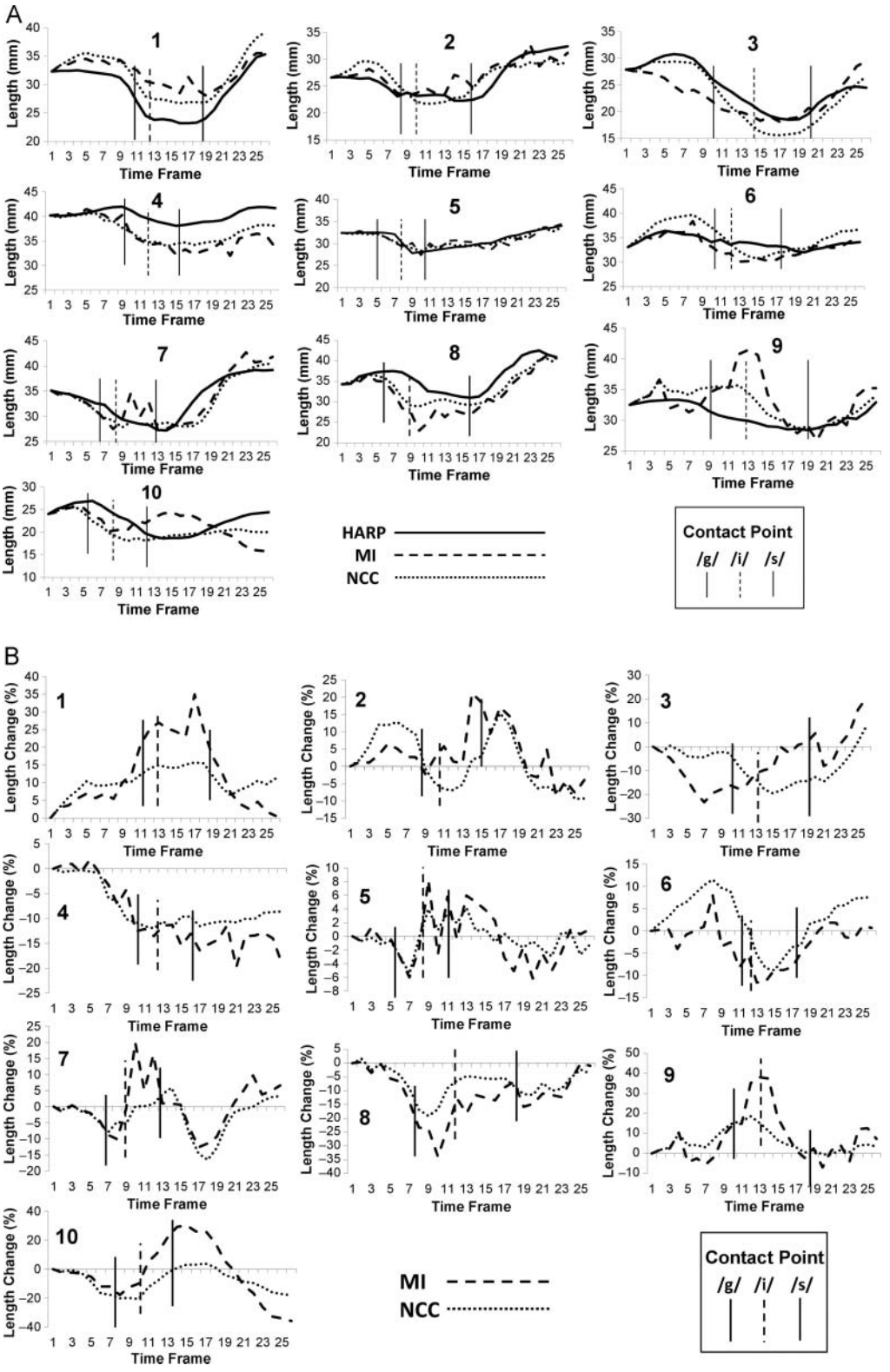
Three additional tissue points were manually tracked (see Figure 6, Panel A) and compared to automatic tracking by NCC and HARP tracking methods. Errors between both automatic methods and the manual tracks are shown for the three points in Figure 6. Point A was the most superficial point. In the HARP errors (solid lines in Figure 6, Panel B), well-tracked TFs usually had subpixel errors (recall that one pixel = 1.875 mm in both data sets). The large errors seen in Point A starting at TF17 were caused by HARP mistracking at the tongue surface. The tag pattern faded significantly in the image by that time and the algorithm tracked a point occurring in the airway. The error was then propagated through the rest of the TFs. The NCC errors (dashed lines in Figure 6, Panel B) were consistently larger than the HARP errors, except when the mistracking occurred. Although the DR errors were greater overall, they were more consistent.

Discussion

Tissue point tracks calculated by HARP from tagged MRI data revealed varied amounts of motion across subjects and points (see Figure 3). DR tracking was poorer in tracking internal points, especially when large motions were seen. This is partly because the DR we used in the present work finds pixel correspondences based on appearance measures such as intensity. Thus, a distinctive intensity change, such as edges, is captured well, whereas the internal tissue is not because of lack of the distinctive internal features. The internal tissue patterning of the cine MRI did not sufficiently aid DR, and the resultant point tracking values were not close enough to HARP to be useful. The registration task is often cast as an optimization problem in which data fidelity (i.e., similarity measure) and regularization are used to find the best transformation that aligns two images. Because the registration is an ill-posed problem, regularization is used (Woo et al., 2010). In the present work, in the homogeneous regions where intensity values do not change, regularization plays an important role using B-spline (Modersitzki, 2004). However, although the B-spline carries an intrinsic regularization (i.e., a type of model on the deformations), the effect of regularization is limited. This is because the basis functions of cubic B-splines have a local support in the neighborhood of control points (Rueckert et al., 1999) and the regularization may not reflect how the tongue deforms accurately. This is why tags have an advantage; the tags are features that give a regular indication of motion on a local basis without any prior assumption about the deformation. In addition, they are based on an entirely different method that tracks actual material points whose phases differ due to magnetic tagging.

The individual tracking method performed better than the sequential tracking method. As stated before, this is

Figure 5. Panel A: comparison of GGp muscle length changes for 10 subjects. MI (dashed lines) and NCC (dotted lines) estimates differ from HARP tracks (solid lines) by overestimation, underestimation, or both. Panel B: comparison of GGp muscle length changes relative to the HARP tracking as a percentage (%).



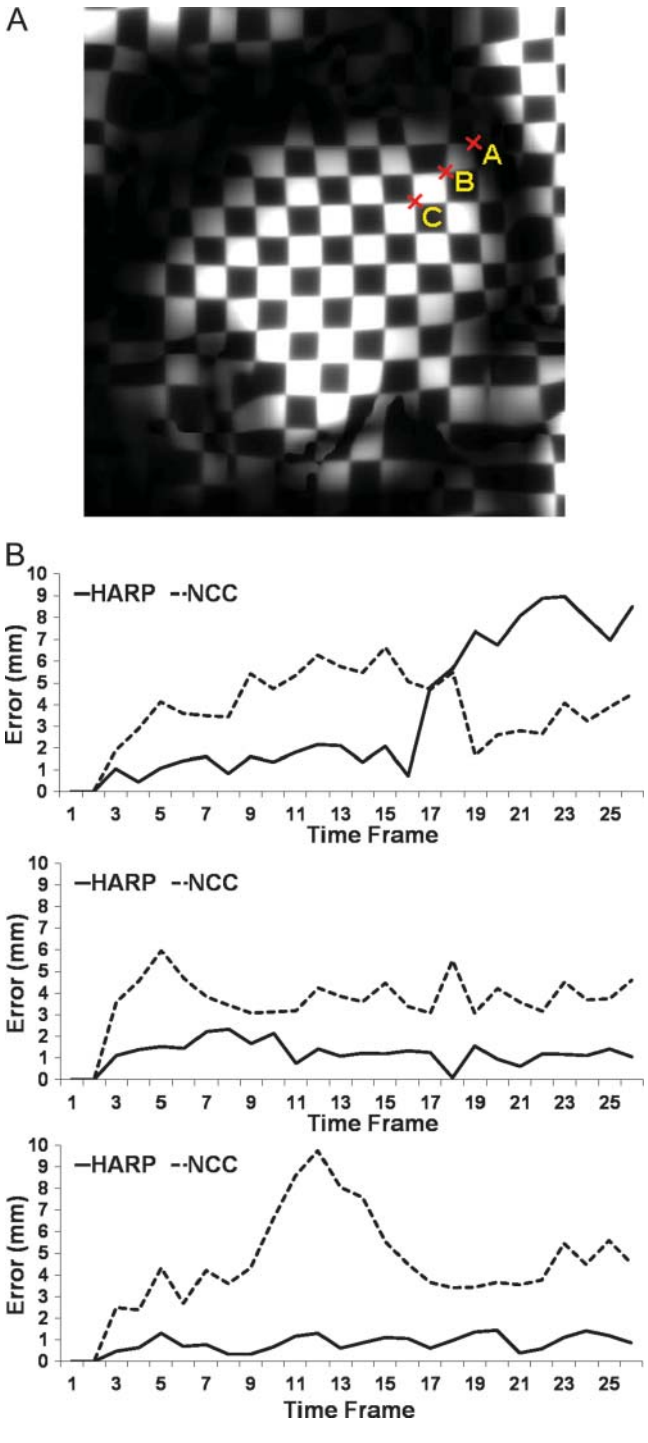


Figure 6. Panel A: three tissue points, including: A (surface point), B (point 6 mm deep to Point A), and C (point 8 mm deep to Point B). Panel B: tongue tracking errors of HARP (solid lines) and NCC (dashed lines) relative to manual tracking of three tissue points.

because the sequential tracking propagates errors, whereas the individual tracking method compares the target frame only to TF1. In addition, only the DR method using the B-spline with NCC and MI similarity measures were used for the computations, because the B-spline with SSD and diffeomorphic demons and some of the other methods failed to track large deformations.

Muscle shortening patterns across subjects were generally linked to phoneme identity in HARP (see Figure 5, Panel A). During the upward and forward tongue body movement into /g/ and /i/, the GGp muscle shortened, and after the start of /s/ it began to lengthen, consistent with the local deformations that occur during tongue motion. Of course, some intersubject differences may have been due to inadvertently choosing nonidentical points across subjects. Better methods of muscle identification will be helpful in preventing of that type error. Muscles tracked with DR sometimes had similar patterns of shortening and timing to the HARP tracks; however, the errors were not systematic (see Figure 5). Therefore, with current image quality and DR methods it is difficult to reliably predict the true shortening pattern and reduce the errors of overestimation, underestimation, extreme enhancement, or poor tracking. However, with improvement in image quality to contain detailed muscle information, DR methods could produce results that are comparable to the HARP tracking method as demonstrated in other applications (Chandrashekara et al., 2005; Ouyang, Li, & El Fakhri, 2013).

The DR tracking method has some limitations that can be improved and others that are inherent in the task. One inherent limitation is that cine MRI images are not designed to capture tissue points. The cine MR images used in this study had low resolution (1.875 mm × 1.875 mm × 6 mm). Higher resolution images would contain tissue features that could be exploited better with a DR algorithm. This would be true of ultrasound or X-ray images as well; the better the resolution of different tissue types, the better a DR algorithm would perform. A second limitation is that DR focuses on edges and image features that are distinctive and does a poorer job of tracking pixels in homogeneous regions. Several modifications could improve the DR estimation of tissue point motion in cine MRI. A motion model, such as the average tongue motion (i.e., motion atlas) derived from a large corpus of tagged MRI, could be used as statistical prior information for regularization to constrain and direct the tracking algorithm, which is a topic of ongoing research. A refinement technique, which starts at a well-tracked set of pixels such as the tongue surface, could constrain tracking of neighboring pixels, continuing until the entire tongue is tracked. Regularization methods such as these may be more limited in their use with certain client populations, however, because many clients, such as those with neurological impairment, are more likely to be variable and unpredictable in their utterances.

One of the limitations of HARP is that near edges, such as the tongue surface, it is more likely to track tissue points erroneously because of blurring of the tag pattern near the edges caused by the HARP bandpass filter. Another limitation

is that HARP's ability to track degrades as the tags fade. Although retagging techniques or data interleaving techniques are available to address tags fading, speech applications require many repetitions, hampering the use of such techniques. Cine MRI and DR have no such limitations; thus, it is possible that, with the proper constraints, DR could be used at the tongue surface especially at later TFs to supplement HARP measurements. Such a hybrid algorithm has been reported in a limited data set (Xing et al., 2013). Finally, the comparisons in the present work were based on two-dimensional midsagittal slices. This can be improved by comparing DR point tracking using full three-dimensional super-resolution volumes (Woo, Murano, Stone, & Prince, 2012) with an incompressible deformation estimation algorithm (Liu et al., 2012), which also is a subject of ongoing research.

Conclusion

DR applied to cine MRI has the potential to allow calculation of tissue point and muscle properties if the image is sufficiently detailed. However, at present this approach cannot track at subpixel resolution, as does HARP. However, where HARP cannot be used or has limitations itself, such as at the tissue–air interface, DR might be a useful strategy for some studies, provided that its limitations are well recognized and accounted for in the analysis.

Acknowledgments

This research was supported in part by Grants R01 CA133015 (PI: M. Stone) and K99 DC012575 (PI: J. Woo) from the National Institutes of Health. Parts of this article were presented at the 2011 International Seminar on Speech Production, Montreal, Quebec, Canada.

References

Aletras, A. H., Ding, S., Balaban, R. S., & Wen, H. (1999). DENSE: Displacement encoding with stimulated echoes in cardiac functional MRI. *Journal of Magnetic Resonance, 137,* 247–252.

Arts, T., Prinzen, F. W., Delhaas, T., Milles, J., Rossi, A. C., & Clarysse, P. (2010). Mapping displacement and deformation of the heart with local sine-wave modeling. *IEEE Transactions on Medical Imaging, 29,* 1114–1123.

Axel, L. (2002). Biomechanical dynamics of the heart with MRI. *Annual Review of Biomedical Engineering, 4,* 321–347.

Axel, L., & Dougherty, L. (1989). MR imaging of motion with spatial modulation of magnetization. *Radiology, 171,* 841–845.

Axel, L., Montillo, A., & Kim, D. (2005). Tagged magnetic resonance imaging of the heart: A survey. *Medical Image Analysis, 9,* 376–393.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer [Computer software]. Retrieved from www.praat.org

Chandrashekara, R., Mohiaddin, R. H., & Rueckert, D. (2005). Comparison of cardiac motion fields from tagged and untagged MR images using nonrigid registration. In A. F. Frangi, P. I. Radeva, A. Santos, & M. Hernandez (Eds.), *Functional imaging and modeling of the heart: Vol. 3504. Third International Workshop, FIMH 2005, Barcelona, Spain, June 2–4, 2005* (pp. 425–433). Heidelberg, Germany: Springer-Verlag.

Chen, T., Wang, X., Chung, S., Metaxas, D., & Axel, L. (2010). Automated 3D motion tracking using Gabor filter bank, robust point matching, and deformable models. *IEEE Transactions on Medical Imaging, 29,* 1–11.

Cho, G. Y., Chan, J., Leano, R., Strudwick, M., & Marwick, T. H. (2006). Comparison of two-dimensional speckle and tissue velocity based strain and validation with harmonic phase magnetic resonance imaging. *American Journal of Cardiology, 97,* 1661–1666.

Fischer, S. E., McKinnon, G. C., Maier, S. E., & Boesiger, P. (1993). Improved myocardial tagging contrast. *Magnetic Resonance in Medicine, 30,* 191–200.

Garot, J., Bluemke, D. A., Osman, N. F., Rochitte, C. E., McVeigh, E. R., Zerhouni, E. A., … Lima, J. A. (2000). Fast determination of regional myocardial strain fields from tagged cardiac images using harmonic phase MRI. *Circulation, 101,* 981–988.

Garot, J., Bluemke, D. A., Osman, N. F., Rochitte, C. E., Zerhouni, E. A., Prince, J. L., & Lima, J. A. (2000). Transmural contractile reserve after reperfused myocardial infarction in dogs. *Journal of the American College of Cardiology, 36,* 2339–2346.

Gupta, S. N., & Prince, J. L. (1995). On variable brightness optical flow for tagged MRI. *Information Processing in Medical Imaging, 3,* 323–334.

Guttman, M. A., Prince, J. L., & McVeigh, E. R. (1994). Tag and contour detection in tagged MR images of the left ventricle. *IEEE Transactions on Medical Imaging, 13,* 74–88.

Hermosillo, G., Chefd'Hotel, C., & Faugeras, O. (2002). Variational methods for multimodal image matching. *International Journal of Computer Vision, 50,* 329–343.

Ibanez, L., Schroeder, W., Ng, L., & Cates, J. (2003). *The ITK software guide*. Albany, NY: Kitware Inc.

el Ibrahim, S. H. (2011). Myocardial tagging by cardiovascular magnetic resonance: Evolution of techniques—Pulse sequences, analysis algorithms, and applications. *Journal of Cardiovascular Magnetic Resonance, 13,* 1–40.

Liu, X., Abd-Elmoniem, K. Z., Stone, M., Murano, E. Z., Zhuo, J., Gullapalli, R. P., & Prince, J. L. (2012). Incompressible deformation estimation algorithm (IDEA) from tagged MR images. *IEEE Transactions on Medical Imaging, 31,* 326–340.

Liu, X., Murano, E., Stone, M., & Prince, J. L. (2007). HARP tracking refinement using seeded region growing. In *Proceedings of the 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 372–375). Piscataway, NJ: IEEE.

Liu, X., & Prince, J. L. (2010). Shortest path refinement for motion estimation from tagged MR images. *IEEE Transactions on Medical Imaging, 29,* 1560–1572.

Masaki, S., Tiede, M., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., & Ninomiya, N. (1999). MRI-based speech production study using a synchronized sampling method. *The Journal of the Acoustical Society of Japan, 20,* 375–379.

Modersitzki, J. (2004). *Numerical methods for image registration*. Oxford, UK: Oxford University Press.

Morris, G. A., & Freeman, R. (2011). Selective excitation in Fourier transform nuclear magnetic resonance. *Journal of Magnetic Resonance, 213,* 214–243.

Narayanan, S., Nayak, K., Lee, S., Sethy, A., & Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. *The Journal of the Acoustical Society of America, 115,* 1771–1776.

NessAiver, M., & Prince, J. L. (2003). Magnitude image CSPAMM reconstruction (MICSR). *Magnetic Resonance in Medicine, 50,* 331–342.

Notomi, Y., Setser, R. M., Shiota, T., Martin-Miklovic, M. G., Weaver, J. A., Popovic, Z. B., … Thomas, J. D. (2005). Assessment of left ventricular torsional deformation by Doppler

tissue imaging: Validation study with tagged magnetic resonance imaging. *Circulation, 111,* 1141–1147.

Osman, N. F., Kerwin, W. S., McVeigh, E. R., & Prince, J. L. (1999). Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. *Magnetic Resonance in Medicine, 42,* 1048–1060.

Osman, N. F., McVeigh, E. R., & Prince, J. L. (2000). Imaging heart motion using harmonic phase MRI. *IEEE Transactions on Medical Imaging, 19,* 186–202.

Osman, N. F., Sampath, S., Atalar, E., & Prince, J. L. (2001). Imaging longitudinal cardiac strain on short-axis images using strain-encoded MRI. *Magnetic Resonance in Medicine, 46,* 324–334.

Ouyang, J., Li, Q., & El Fakhri, G. (2013). Magnetic resonance-based motion correction for positron emission tomography imaging. *Seminars in Nuclear Medicine, 43,* 60–67.

Parthasarathy, V., Prince, J. L., Stone, M., Murano, E. Z., & Nessaiver, M. (2007). Measuring tongue motion from tagged cine MRI using harmonic phase (HARP) processing. *The Journal of the Acoustical Society of America, 121,* 491–504.

Pluim, J. P., Maintz, J. B., & Viergever, M. A. (2003). Mutual-information–based registration of medical images: A survey. *IEEE Transactions on Medical Imaging, 22,* 986–1004.

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., & Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging, 18,* 712–721.

Stone, M., Davis, E. P., Douglas, A. S., Aiver, M. N., Gullapalli, R., Levine, W. S., & Lundberg, A. J. (2001). Modeling tongue surface contours from Cine-MRI images. *Journal of Speech, Language, and Hearing Research, 44,* 1026–1040.

Story, B. H. (2009). Vowel and consonant contributions to vocal tract shape. *The Journal of the Acoustical Society of America, 126,* 825–836.

Vercauteren, T., Pennec, X., Perchant, A., & Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage, 45*(Suppl. 1), S61–S72.

Winkler, R., Fuchs, S., Perrier, P., & Tiede, M. (2011). Biomechanical tongue models: An approach to studying inter-speaker variability. In *Proceedings of INTERSPEECH 2011: 12th Annual Conference of the International Speech Communication Association, Florence, Italy, August 27–31, 2011* (pp. 273–276) Retrieved from http://hal.archives-ouvertes.fr/docs/00/61/92/38/PDF/winkler_et_al_interspeech_2011_Hal.pdf

Woo, J., Dey, D., Cheng, V. Y., Hong, B. W., Ramesh, A., Sundaramoorthi, G., … Slomka, P. J. (2010). Nonlinear registration of serial Coronary CT Angiography (CCTA) for assessment of changes in atherosclerotic plaque. *Medical Physics, 37,* 885–896.

Woo, J., Lee, J., Bogovic, J., Murano, E. Z., Xing, F., Stone, M., & Prince, J. L. (2013). Multi-subject atlas built from structural tongue magnetic resonance images. *Proceedings of Meetings on Acoustics, 19*(060194), 1–4. doi:10.1121/1.4799736

Woo, J., Murano, E. Z., Stone, M., & Prince, J. L. (2012). Reconstruction of high-resolution tongue volumes from MRI. *IEEE Transactions on Biomedical Engineering, 59,* 3511–3524.

Woo, J., Stone, M., & Prince, J. L. (2011). Deformable registration of high-resolution and cine MR tongue images. *Medical Image Computing and Computer Assisted Intervention, 14*(Pt. 1), 556–563.

Xing, F., Woo, J., Murano, E., Lee, J., Stone, M., & Prince, J. L. (2013). Estimating 3D tongue motion with MR images. *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 41–48). Nagoya, Japan.

Zerhouni, E. A., Parish, D. M., Rogers, W. J., Yang, A., & Shapiro, E. P. (1988). Human heart: Tagging with MR imaging—A method for noninvasive assessment of myocardial motion. *Radiology, 169,* 59–63.