Enhanced airway-tissue boundary segmentation for real-time magnetic resonance imaging data

Jangwon Kim, Naveen Kumar, Sungbok Lee, Shrikanth Narayanan

University of Southern California, Los Angeles, CA, U.S.A.

Abstract

This paper introduces an algorithm for robust segmentation of airway-tissue boundaries in the upper airway images recorded by the real-time magnetic resonance imaging. Compared to the previous method by Proctor et al. [1], the present algorithm performs image quality enhancement, including pixel sensitivity correction and grainy noise reduction, followed by robust estimation of airway path between the vocal tract walls. The airway path as well as the locations of the lips and the top of the larynx are found using the Viterbi algorithm. The tissue-airway boundaries are found for each grid line by searching the closest pixel of higher intensity than a threshold from the the estimated airway path. The accuracy of the tissue boundary segmentation was evaluated in terms of root-mean-squared-error as well as statistics (mean and standard deviation) of error for specific region in the vocal tract. Results suggest that the proposed algorithm shows significantly less estimation error than the previous method [1], especially for the front cavity and the lower boundary.

Keywords: real-time magnetic resonance imaging, image segmentation, automatic tracking, vocal tract analysis

1. Introduction

Real-time Magnetic Resonance Imaging (rtMRI) [2] is an important tool for studying human speech production. The rtMRI provides the entire mid-sagittal view of the upper airway of a subject. The airway-tissue boundary segmentation in the MR images is often required as a pre-processing for the analysis and modeling of the vocal tract movements [3] and of the morphological structure of the vocal tract [4]. Performing this segmentation automatically is essential for analyzing rtMRI data of speech production, that typically comprise hundreds or thousands of video frames; the complex structure of the vocal tract, non-uniform field sensitivity of the tissues in head and neck, grainy noise, magnetic resonance (MR) image artifact, and the rapidly varying irregular vocal tract shape, however, make this problem challenging. This paper presents an algorithm for more robust segmentation of the MR images, which includes (1) retrospective pixel intensity correction of the MR images, (2) detection of the front-most edge of the lips and the top of the larynx, (3) segmentation of airway-tissue boundary in the vocal tract, and (4) measurement of the distance between the upper and lower boundaries. The current method improves the robustness of the airway-tissue boundary estimation over the previous method [1] by using a combination of data-driven way of preprocessing of the MR images, robust airway path estimation, and model-based weighted linear curve fitting.

2. Methods

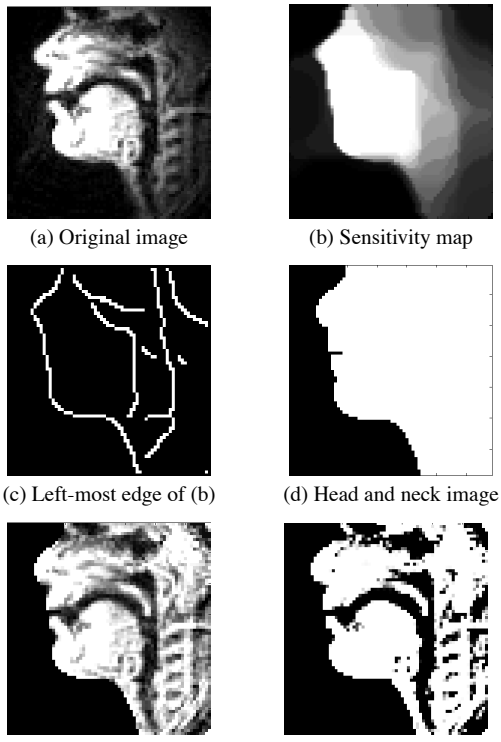
2.1. Pre-processing of MR images

The MR images in rtMRI data often suffer from grainy noise and non-uniform field sensitivity of the tissues, depending on recording configuration [2]. Figure 1 (a) shows an example of the MR images in the USC-EMO-MRI corpus [5], which was recorded at an image frame rate of 23.18 frames/sec and a spatial resolution of 68 $\times$ 68 pixels. The present algorithm uses a multi-resolution approach to minimizing the effects of the noise, artifacts, and non-uniform field sensitivity of the tissues. The details of the approach are as follows.

1. Create a field sensitivity map, denoted by S, of an original MR image using a morphological closing operation, followed by 2-dimensional median filtering. Figure 1 (b) shows the sensitivity map of the image in Figure 1 (a). The morphological closing operation selectively exclude low-intensity pixels (of grainy noise or artifacts in general) in the airway region when creating S.
2. Create the set of edge points, as in Figure 1 (c), of the sensitivity map using the Canny edge detector [6] implemented in MATLAB. Likewise, create the set of edge points of the original image. Let E_O and E_{SM} denote the sets of edge points of the sensitivity map and the original image, respectively.
3. Create the head and neck boundary line E_H by finding the left-most points of E_O and E_{SM}. Then, create a binary image, denoted by B, of the head-neck region by setting the pixel intensity to be 1 for pixels in the right side of E_H in each row and setting the pixel intensity to be 0 otherwise, as in Figure 1 (d).
4. Multiply the pixel intensity of the original image and the inverse of the pixel intensity of S for non-zero elements in B, while setting the non-tissue pixel intensity to be zero. Figure 1 (e) shows the result image, denoted by C.
5. Perform a sigmoid warping of the pixel intensity in C for suppressing grainy noise as well as highlighting tissue. Figure 1 (f) shows the final image.

3. Construction of grid lines

In order to detect the lips, the larynx, and the airway-tissue boundaries, the present algorithm constructs grid lines, adopting from the previous method [1]. The previous method is motivated by the analysis of the upper airway image by Ohman [7]. The grid construction method requires four manually chosen anatomical landmarks near the larynx, the highest point on the palate, the alveolar ridge, and the center of the lips, in one of the MR images. See [1] for the details of grid construction that we follow. The differences from the previous method are (i) that



(a) Original image (b) Sensitivity map

(c) Left-most edge of (b) (d) Head and neck image

(e) Intensity corrected (f) Final image

Figure 1: *The MR image after each pre-processing step*

a user chooses the distance between the center of adjacent grid lines in the present algorithm, not by the number of grid lines as in the previous method, and (ii) that the origin of the reverse polar grid lines is placed at the top of the image on the horizontal coordinate of the labial landmark point. This point offers more smooth transition from the forward polar grid lines to the reverse polar grid lines. Note that such method of the grid line construction assumes that the head and neck are aligned such that the subject faces the left side of the image and the neck is vertically straight.

4. Lips and Larynx detection

For each frame, the initial and the final grid lines correspond to the locations of the top of the larynx and the front-most edge of the lips, respectively. Since these articulatory positions vary slowly and smoothly over time, the present algorithm finds each of their optimal positions by constraining rapid change of the estimated locations of them.

Assume q_t is a state at instance t. N denotes the number of states. $S^T_{q_i,q_j}$ denotes the transition score from q_i to q_j. $S^L_{q_i}$ is the likelihood score (of the observation) for q_i. P_i is the prior score of q_i. K is the number of instances. Q denotes a sequence of states $q_1, q_2, \cdots, q_K$, one state for each instance. The objective score $\mathcal{J}$ of Q is defined as follows:

$$\mathcal{J} = \left(P_1 S^L_{q_1} + w S^T_{q_2,q_1}\right) + \left(\sum_{u=2}^{K-1} S^L_{q_u} + w S^T_{q_{u+1},q_u}\right) \quad (1)$$

where w is a weighting factor for $S^T_{q_i,q_j}$. The optimal sequence $Q^\star$ is obtained by finding Q associated with the minimum $\mathcal{J}$:

$$Q^\star = \underset{[q_1,q_2,\cdots,q_K]}{\arg\min}\ \mathcal{J} \quad (2)$$

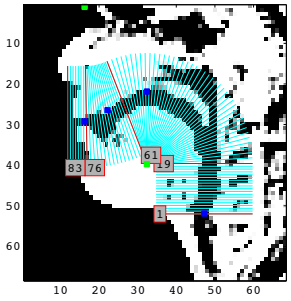


Figure 2: *The grid lines (cyan color) superimposed on an MR image. Four blue dots are the manually selected landmarks. The origins (green color) of the forward polar grid lines (19 ∼ 61) and the reverse polar grid line (61 ∼ 76) are determined based on the landmarks.*

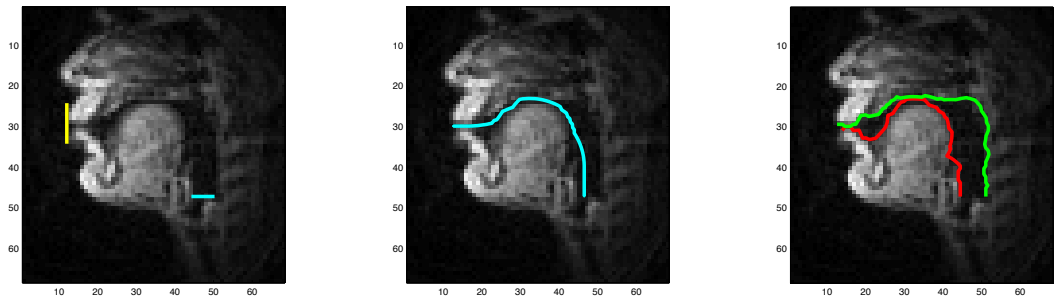
For detection problem of the edge of the lips, q_i corresponds to the i-th grid line, where $q_{N/2}$ is placed on the grid line of the labial landmark (the 77-th grid line in Figure 2). Also, $S^T_{x,y}$ is the Euclidean distance between the centers of grid lines x and y. S^L_x is the maximum pixel intensity of all pixels in the grid line x. Note that the length and width of searching region for the lip detection are specified by users.

For the top of the larynx, q_i corresponds to the i-th grid line where $q_{N/2}$ is placed on the grid line of the larynx landmark (the first grid line in Fig. 2). $S^T_{x,y}$ is the same as defined for lips detection. Let D^L_x be the mean of the first-order derivatives of pixel intensities of x, computed along the grid lines. Then, $S^L_x = D^L_x \times W_x$, where W_x is an optional weighting term which gives more weight on higher grid line. W_x often helps for better estimation, especially when the low part of the larynx in MR images protrudes. This algorithm detects the point where the pixel intensity increases the most, searching from the top grid line.

The length and the width of searching regions (grid lines) for the lip detection and the larynx detection are specified by users. w is set to be 1 for these problems. One example of lips and larynx detection results is shown in Figure 3 (a).

5. Airway-path detection

The key idea behind improving airway-tissue boundary segmentation is to find an accurate and possibly approximate airway path in the upper airway first, from which the optimal airway-tissue boundaries can be determined easily and more robustly. The optimal airway paths passing through all grid lines in an MR image are determined by finding the paths of the minimum score, using the Viterbi algorithm. For this problem, each possible path in a grid line corresponds to a state, while each grid line corresponds to an instance. q_i corresponds to the i-th bin, where $q_{N/2}$ is located at the center of the grid line; $S^T_{x,y}$ is the Euclidean distance between bins x and y, where the bins are located in the adjacent grid lines, one bin for each grid line. S^L_x is the pixel intensity (observation) of the bin x, determined for each instance. Then, the optimal airway path is found by minimizing the score of possible bins as in the equation 2. The reason of using all bins, not only local minima as in the previous method [1] is that all local minima are sometimes found outside the upper airway when some regions in the vocal tract is fully closed. The estimated airway path in our method can still stay within the region of interest during full contact in the upper airway, restricted by the transition costs between states.



(a) Lips and larynx (b) Airway path (c) Tissue Boundaries

Figure 3: *Estimated vocal tract parameters: (a) estimated locations of forward-most edge of the lips (yellow color) and top of the larynx (cyan color), (b) airway path (cyan color), (c) airway-tissue boundaries (red line for lower boundary, green line for upper boundary).*

Optionally, our algorithm performs a smoothing of the pixel intensity matrix (observaions) using the mean of the 25% and 75% quantiles of the intensity values of neighboring pixels. We found that this smoothing is effective for reducing the estimation error caused by the low-intensity pixels outside the vocal tract walls, because this smoothing tends to increase their intensity values. Also, the smoothing assists the airway path to stay inside the upper airway when a part of vocal tract is fully closed, by forcing the intensity of the present pixels of fully closed region to be low (because the past and future pixel intensities are low). Neighbors in the range of four instances, eight grids and four bins were used for the estimated airway path in Figure 3 (b). w in eqn. 1 was set to be 3.

6. Airway-tissue boundary segmentation

Two airway-tissue boundaries, i.e., the upper and lower boundaries of the vocal tract walls, are determined at the first bins whose pixel intensity is over a certain threshold in the upper direction and lower direction, respectively. The threshold was set to be 0.5, where the maximum pixel intensity of each MR image is 1.

The estimated airway-tissue boundary points are smoothed by the robust local regression using weighted linear least squares and a 1-st degree polynomial model, implemented in MATLAB, for each image frame. Figure 3 (c) illustrates the smoothed airway-tissue boundaries. Finally, a distance function for the airway-tissue boundaries is obtained by computing the Euclidean distance (in pixel unit) between the upper and lower boundaries or between the upper boundary point and the closest lower boundary point regardless of their grid line. It was observed that the later (green line in Fig. 4) is less erroneous, in particular near the lip region, than the former (blue line in Fig. 4). The initial boundary point for computing the distance function is in the grid line of the estimated larynx. The final boundary point is in the grid line of the first local minimum distance from the final grid line. Figure 4 illustrates a distance function in the upper airway. The software package which contains the MATLAB codes for the present algorithm and the subsets of data for demonstration is freely available at http://sail.usc.edu/old/software/rtmri_seg.

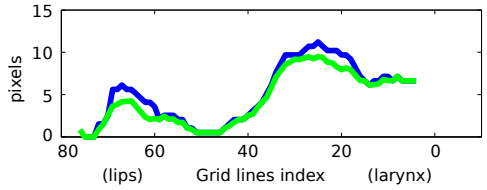


Figure 4: *Distance function from the larynx to the lips for Fig. 3 (c). Green line is the shortest distance from the estimated upper boundary point for each grid line to the closest point in the lower boundary to it. Blue line is the distance between lower and upper boundary points for each grid line.*

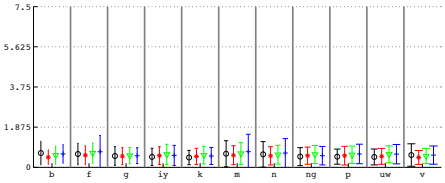
7. Evaluation of estimated airway-tissue boundaries

The estimated airway-tissue boundaries are evaluated against manually annotated tissue boundaries. For this purpose the annotators were instructed to sketch the lower and upper vocal tract walls using a continuous curve. For each of lower and upper boundaries, the Euclidean distance between each estimated boundary point and the closest point in the reference boundary for the estimated point is measured.

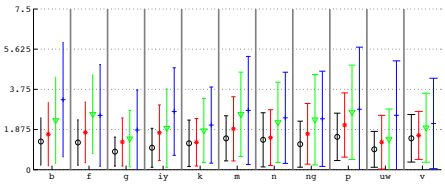
The statistics (mean and standard deviation) of the distance values are computed for each sub-region in the vocal tract and each phone. The sub-regions of the present algorithm are (1) grid lines $1 \sim 19$ for pharyngeal region, (2) grid lines $20 \sim 52$ for velar and dorsal constriction region, (3) grid lines $53 \sim 67$ (alveolar ridge landmark) for the hard palate region, and (4) grid lines $68 \sim 77$ for labial constriction region. The sub-regions of the previous algorithm are also determined in a similar way. The previous algorithm does not include the lip detection, thus large estimation error is observed in the grids after the lip landmark. For a fair comparison, the final grid line for analysis is fixed to the lip landmark point.

The palatal and dental corrections, and the mean pharyngeal wall were used as pre-processing for the baseline system [1]. See [1] for more details. For the present algorithm, the mean of the estimated boundary in the palatal region and the vertical position of the palate landmark is used in the final upper boundary. The reason for the palatal corrections in both argorithms is that the soft tissue in the hard palate region often shows significantly lower pixel intensity than other tissues, thus not sufficiently contrasted to the airway.

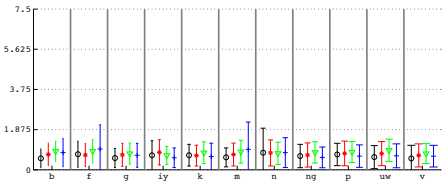
The list of phones used for evaluation is [B, F, G, IY, K, M, N, NG, P, UW, V]. Producing speech sound for these phones involves highly constricted or fully closed articulatory gestures, where the error of the estimated airway-tissue segmentation tends to be high. For each phone, 10 phone instances were randomly selected in a male subjects' data in the USC-EMO-MRI corpus. The acoustic phone boundary of each phone instance is obtained using an adaptive speech-text alignment tool, SailAlign [8]. The image frames within the starting and final times with one marginal frame in each side were selected. In



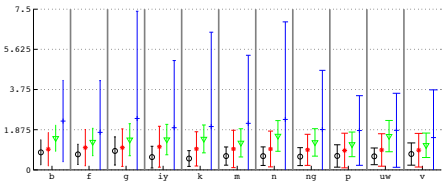
(a) Inner boundary; proposed



(b) Inner boundary; baseline



(c) Outer boundary; proposed



(d) Outer boundary; baseline

Figure 5: *Errorbar of the distance (in pixel unit) between manual airway-tissue boundary and estimated airway-tissue boundary. From left to right in each phone, each errorbar is for pharyngeal region (black color), velar and dorsal region (red color), palatal region (green color), and labial region (blue color).*

total, 492 image frames were used for evaluation.

Figure 5 shows the errorbar (as standard deviation) of the distance for each region and each phone for each estimated boundary. For lower boundary, the mean and standard deviation of the proposed algorithm is significantly smaller in all four regions than those of the baseline algorithm. Especially, the larger error in the front cavity (the palatal and labial regions) is significantly suppressed in the proposed algorithm. For the upper boundary, the baseline algorithm performs significantly better in the regions from the pharynx to the palate regions than for the lower boundary, presumably partially by the dental and palatal correction. However, the labial region still shows significantly large error. The amount of error in the labial region is significantly suppressed in the proposed algorithm. Table 1 shows the root-mean-squared-error (RMSE) for all estimated boundary points in each of the lower and upper boundaries. The proposed algorithm shows significantly lower RMSE for both lower and upper trajectories than the baseline algorithm. These results suggest that the proposed algorithm generates significantly more accurate airway-tissue boundaries than the baseline algorithm. In sum, the proposed algorithm generates more robust airway-tissue boundaries regardless of the phone and the vocal tract region than the baseline algorithm.

Table 1: *RMSE between the estimated and manually-labeled boundaries in pixel unit.*

Baseline		Proposed	
lower	upper	lower	upper
2.56	2.13	0.71	0.93

8. Conclusion and future work

The present algorithm estimates the airway-tissue boundaries from a robustly estimated airway path in each enhanced MR image. According to the quantitative evaluation on the estimated boundaries, the estimation error is significantly reduced by the present algorithm than the previous method [1] in terms of RMSE (2.56 to 0.71 for the lower boundary; 2.13 to 0.93 for the upper boundary). A major advantage of the proposed method over the baseline is robustness across different regions in the vocal tract . The proposed algorithm also extracts the positions of the front-most edge of the lips and the top of the larynx automatically. This helps constrain the search space of the airway-tissue boundaries, resulting more robust boundary estimation. In addition, with the algorithm one can estimate the length of the vocal tract above the larynx.

Automatic head movement correction for each MR image is an on-going work that we would like to use for more robust and convenient tissue boundary estimation. In addition, this approach also calls for a preprocessing technique that is better suited to this imaging modality.

9. Acknowledgements

This work is supported by NSF IIS-1116076 and NIH DC007124.

10. References

[1] Michael Proctor, Danny Bone, Nassos Katsamanis, and Shrikanth S Narayanan, "Rapid semi-automatic segmentation of real-time magnetic resonance images for parametric vocal tract analysis," in *Proceedings in Interspeech*. 2010, pp. 1576–1579, ISCA.

[2] Shrikanth Narayanan, Krishna Nayak, Sungbok Lee, Abhinav Sethy, and Dani Byrd, "An approach to real-time magnetic resonance imaging for speech production," *Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1771 – 1776, 2004.

[3] Vikram Ramanarayanan, Louis Goldstein, Dani Byrd, and Shrikanth S. Narayanan, "An investigation of articulatory setting using real-time magnetic resonance imaging," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 510–519, 2013.

[4] Adam Lammert, Michael I. Proctor, and Shrikanth S. Narayanan, "Interspeaker variability in hard palate morphology and vowel production," *Journal of Speech, Language, and Hearing Research*, 2013, (in press).

[5] Jangwon Kim, Asterios Toutios, Yoon-Chul Kim, Yinghua Zhu, Sungbok Lee, and Shrikanth Narayanan, "USC-EMO-MRI corpus: An emotional speech production database recorded by real-time magnetic resonance imaging," in *Proceedings of the 10th International Seminar on Speech Production (Accepted)*, 2014.

[6] J Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, 1986.

[7] S. E. G. Öhman, "Numerical model of coarticulation," *Journal of the Acoustical Society of America*, vol. 41, no. 2, pp. 310 – 320, 1967.

[8] Athanasios Katsamanis, Matthew Black, Panayiotis G. Georgiou, Louis Goldstein, and Shrikanth S. Narayanan, "SailAlign: Robust long speech-text alignment," in *Workshop on New Tools and Methods for Very-Large Scale Phonetics Research*, Philadelphia, PA, Jan 2011.