EXTRACTION OF TONGUE CONTOUR
IN REAL-TIME MAGNETIC RESONANCE IMAGING SEQUENCES

Dawei Zhang, Minghao Yang, Jianhua Tao, Yang Wang, Bin Liu, Danish Bukhari

National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing, China
{dawei.zhang, mhyang, jhtao, yangwang, liubin, danishbukhari}@nlpr.ia.ac.cn

ABSTRACT

Real-time magnetic resonance imaging (rtMRI) is becoming a practical tool in speech production research and language pathology observation. It is still a challenge to extract the tongue contour accurately in rtMRI sequences, since tongue is a soft tissue and often touches other organs such as lips and upper mandible. This paper proposes a novel semi-automatic tongue contour extraction method from rtMRI sequences. The initial boundary image is obtained by combined multi-directional Sobel operators in tongue movement region; then a boundary intensity map is constructed to find the most probable tongue contour points by searching for the optimal boundary route with Viterbi algorithm; finally the tongue contour is obtained using B-Spline approximation. The proposed method could obtain accurate tongue contour from rtMRI sequences, even in the cases that some parts of tongue touch other organs. Experiments demonstrate the robustness of the proposed method.

Index Terms— Real-time magnetic resonance imaging sequences, tongue contour extraction, boundary intensity map

1. INTRODUCTION

As an important image observation tool nowadays, real-time magnetic resonance imaging (rtMRI) promises to be a viable tool for studying speech production and language pathology [1-3]. Many methods have been proposed for studying the speech production mechanism, including vocal tract segmentation from MRI sequences [4-6], discussing the tongue position for vowels [7-10], and the tongue motion patterns in speech production [11-14].

Researchers have proposed many methods of extracting tongue contours from MRI data [15-17]. Most of the methods need lots of manual operations in the early stage, and the boundaries of tongue in the images are usually distinct. A semi-automatic vocal tract segmentation method from rtMRI sequences was proposed in [18]. They provided a MATLAB tool to extract the vocal tract boundaries. This

software is an efficient tool for human speech production research [19-22]. However the extracted tongue contour is occasionally unreasonable especially when the tongue touches other organ boundaries such as lips and upper mandible. In [19-21], researchers studied the case that tongue touches other organs when pronouncing some special phonemes. However, their approaches are partly dependent on manual operation.

In this paper, a semi-automatic method is proposed to extract tongue contour. The tongue contour is defined as the boundary of the tongue in 2D magnetic resonance images. The proposed method consists of three steps:

1) register the tongue movement range and detect the initial boundary points using combined multi-directional Sobel operators;

2) construct a boundary intensity map to find the tongue contour points by searching for the optimal boundary route with Viterbi algorithm;

3) obtain the final tongue contour using B-Spline approximation [23].

The rest of this paper is organized as follows: the method of extracting accurate tongue contour is described in Section 2 and Section 3; experiments are discussed in Section 4; and the conclusion is given in Section 5.

2. INITIAL BOUNDARY DETECTION

2.1. Image acquisition

All data analyzed in this paper are from the USC-TIMIT database [24]. Subjects' upper airways were imaged in the midsagittal plane with a resolution of 68×68 pixels. The rtMRI sequences were reconstructed at the rate of 23.18 frames/second.

2.2. Registration of tongue movement range

A magnetic resonance image usually contains many kinds of tissue boundaries and noise information. A predefined range helps to find the initial boundary points quickly in the whole image. In [18], by marking four anatomical landmarks manually — the glottis, highest point on the palate, alveolar ridge and lips, a composite analysis grid is superimposed on each image frame as shown in Fig. 1 (a).
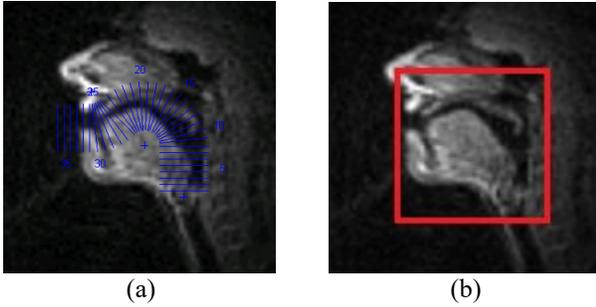
Fig. 1. Definition of the tongue's movements region: (a) a composite analysis grid on the frame [18]; (b) a rectangle on the frame in the proposed method.
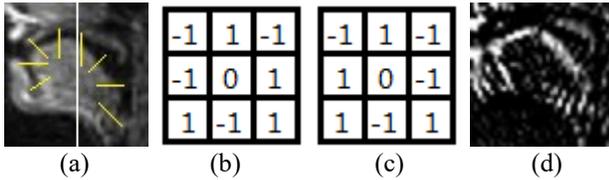
Fig. 2. Detection of initial boundary points: (a) tongue gradient directions in the front and back part; (b) Sobel operator for the front part; (c) Sobel operator for the back part; (d) the image of initial boundary points.

In the proposed method, the tongue's movement range is labeled manually (see the red rectangle in Fig. 1 (b)), where the image is the 60th frame in the rtMRI sequences of speaker F1 [24].

2.3. Detection of initial boundary points

As the magnetic resonance image contains noise information and tongue frequently touches other organs, it is difficult to obtain the boundary points with traditional Gaussian kernel edge-detector. According to the characteristic of tongue shape, the registered region is firstly divided into two parts - the front part and the back part of tongue, as shown in Fig. 2 (a). In the front part, gradient directions of the tongue boundary points are followed as Bottom to Up, Right-Bottom to Left-Top, Right to Left and Right-Top to Left-Bottom; while in the back part, the directions are mainly Bottom to Up, Left-Bottom to Right-Top, Left to Right and Left-Top to Right-Bottom. In this step, we use combined multi-directional Sobel operators to get equal gradient value on the four directions in each part, as shown in Fig. 2 (b) and (c). Fig. 2 (d) shows the image of initial boundary points extracted by the multi-directional Sobel operators above.

3. TONGUE COUNTOUR EXTRACTION

3.1. Construction of boundary intensity map

There are many outlier points in the image of initial boundary points. Then a boundary intensity map is constructed to remove the outliers and to find the most probable boundary points of the tongue.

Fig. 3. Central point and gridlines: (a) homogeneous gridlines; (b) inhomogeneous gridlines.
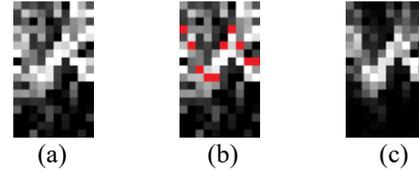
Fig. 4. Boundary intensity map: (a) the original boundary intensity map; (b) the previous tongue contour position in the original boundary intensity map; (c) the modified boundary intensity map.

3.1.1. Central point and gridlines

The center of the registered region is selected as the central point shown with the red point in Fig. 3 (a), which is a reference point to construct gridlines.

Gridlines are constructed around the central point and superimposed at a certain angle interval from the glottis to the glossodesmus as shown with yellow lines in Fig. 3 (a), which is different from [18] shown in Fig. 1 (a). Particularly, the glossodesmus is concerned in the proposed method, because it is expected to extract the entire contour of the tongue.

Considering the tongue movement characteristic, for instance, the tongue tip is more flexible than the rear part, the gridlines should be inhomogeneous in practice. Fig. 3 (b) shows a proper selection of the gridlines.

3.1.2. Boundary intensity map with gridlines

A boundary intensity map is constructed with the gridlines in this subsection, which is then used to find the optimal boundary points series of tongue contour.

Firstly, the sectors between every two adjacent gridlines are labeled successively from the glottis to the glossodesmus. Next, a boundary intensity map of $D \times N$ is used to find the most probable boundary points of tongue, where D is the maximum distance from the probable tongue boundary points to the central point and N is the total number of labeled sectors. In the boundary intensity map, the element in Row i and Colum j stands for the point which is i pixels away from the central point in the j-th sector of the image.

In order to enhance the robustness of extracting tongue contour in magnetic resonance images, along each arc of sectors from the central point to the distant, the maximum intensity of boundary points is selected as the corresponding element of the boundary intensity map shown in Fig. 4 (a).

3.2. Searching for the optimal boundary route

(a)   (b)
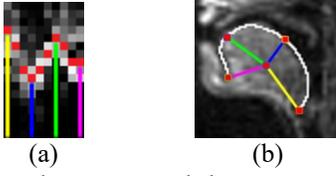
Fig. 5. The boundary route and the tongue contour: (a) the optimal boundary route of the intensity map; (b) the final extracted tongue contour.

3.2.1. Modifying the map with previous tongue contour

As the rtMRI sequences analyzed were reconstructed at a high rate [24], there are at most slight changes in tongue shape between two adjacent frames in general. Therefore the previous tongue contour is helpful in predicting that of the current frame, especially in the case that the tongue contour is incomplete in the image when tongue touches other organs.

Fig. 4 (b) shows the previous tongue contour superimposed on the boundary intensity map shown with small red squares. After that the original map is modified in every column of the map according to Eq. (1):

$$M_{ij} = K_{ij} \times \exp\left(-\left(\frac{i - i_j}{\sigma_M}\right)^2\right) \quad (1)$$

where K_{ij} is the element in Row i and Column j of the original map, M_{ij} is the element in Row i and Column j of the modified map, i_j is the row number of the previous tongue contour in Column j, and σ_M is the modifying standard deviation. The modified boundary intensity map is shown in Fig. 4 (c). Specially, the original map of the first frame has not to be modified.

3.2.2. The optimal boundary route in modified map

The maximum in each column of the boundary intensity map is usually regarded as the boundary points. However, there are exceptional cases: (i) some outlier points of maximum intensity in the boundary image are error contour points, which bring difficulties in extracting the true boundary points; (ii) there are no obvious boundary points when tongue touches other organs.

To solve these problems, it is beneficial to use the transition probabilities in the process of searching for the optimal boundary route in the boundary intensity map. The optimal boundary route contains a sequence of probable tongue boundary points, and it is defined as the Euclidean distance in pixels successively in every column from the most probable boundary point to the central point. As adjacent boundary points are close to each other, the transition probabilities are given in Eq. (2):

$$T_{hi} = \exp\left(-\left(\frac{h - i}{\sigma_T}\right)^2\right) \quad (2)$$

where T_{hi} is the transition probability between the Row h of the previous column and the Row i of the next column, and σ_T is the transition standard deviation.

Table 1. A typical configuration of parameters.

Parameter	Value	Description
D	17	Row number of boundary intensity map — the maximum distance in pixels away from the central point
N	10	Column number of boundary intensity map — number of labeled sectors
σ_M	5	Modifying standard deviation
σ_T	6	Transition standard deviation

The optimal boundary route is found by solving Eq. (3):

$$r^* = \arg\max_{r} P(r|M, \sigma_T)$$
$$= \arg\max_{r} \prod_{j=1}^{N} T_{r(j-1)r(j)} M_{r(j)j} \quad (3)$$

where r^* denotes the optimal boundary route, $P(\cdot)$ is the corresponding probability, r is a possible boundary route, $r(j)$ is the j-th element of r, and specially $T_{r(0)r(1)}$ is defined to be 1 for simplicity.

Finally, according to the modified boundary intensity map and the transition probabilities, the optimal boundary route is obtained using the Viterbi algorithm. Fig. 5 (a) shows the optimal boundary route of the modified boundary intensity map.

3.3. Obtaining tongue contour

The last step is to map the optimal boundary route of the boundary intensity map to the tongue contour points. According to the optimal boundary route, the probable boundary point in every column of the boundary intensity map is mapped to the tongue contour point in the image, with the angle along the angular bisector of the corresponding labeled sector (the column number of the map), and the distance (the row number of the map) away from the central point. After mapping the boundary points, the tongue contour is obtained using B-spline approximation [23]. Fig. 5 (b) shows the finally extracted tongue contour.

4. EXPERIMENTS

4.1. Parameters configuration

A typical parameters configuration is shown in Table 1. Meanwhile the parameters of the latest version MATLAB tool [18] released in 2014 are all recommended default values, and the anatomical landmarks are marked manually according to the prompted messages. All MRI data analyzed in the next subsection are from Speaker F1 [24].

4.2. Results and discussion

4.2.1. Extraction results of tongue contour

In Fig. 6, images in the first row are results of the baseline method [18], and the second row are the proposed method.
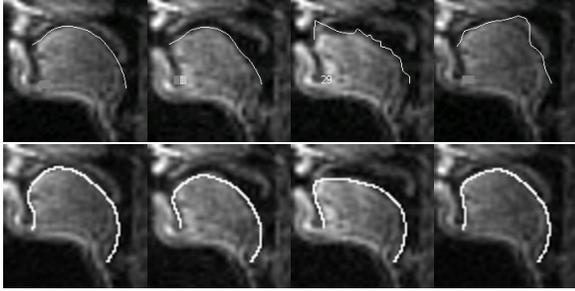
Fig. 6. Tongue contour extraction results. The first row: the baseline method; the second row: the proposed method.

Fig. 6 demonstrates that the baseline method and the proposed method are both effective in most cases. However, the baseline method may extract the unreasonable tongue contour when tongue touches other organ boundaries. While the proposed method works well in this case. By constructing the boundary intensity map and searching for the optimal boundary route with transition probabilities, the proposed method shows improved robustness in more different situations.

4.2.2. Extraction accuracy of tongue contour key points

Four key points on tongue tip (TT), tongue dorsum (TD), rear part of tongue body (TB) and tongue root (TR) are given as shown with the red points in Fig. 7, which are usually used for studying tongue's continuous motion.

Table 2 shows the extraction errors of the key points in 50 continuous frames respectively from Speaker M1 and F1. The extraction errors are defined as the Root Mean Square Errors (RMSE) of Euclidean distance in pixels between the automatically extracted points and the true key points which are marked in these frames by three volunteers (one volunteer is a professor studying speech production and other two volunteers are medical college students).

The data in Table 2 show that the extraction error of tongue tip is obviously larger than those of other key points. The reason may be that tongue tip moves very flexibly, and it touches other organs frequently. It is still a challenge to extract tongue tip accurately.

4.2.3. Extraction accuracy of different speakers

To examine the extraction accuracy of tongue contour, 50 image frames are respectively taken from rtMRI sequences of ten different speakers [24]. The errors of tongue contour extraction results using the baseline method and the proposed method are shown in Table 3, which are the mean RMSE of the tongue contour key points presented in the previous subsection. Meanwhile the key points in these frames are also marked manually.

Table 3 indicates that the proposed method is effective in automatically extracting the tongue contour with reduced errors, demonstrating that the proposed method is more robust than that in [18] statistically. The experiment results show that the proposed method is speaker-independent.
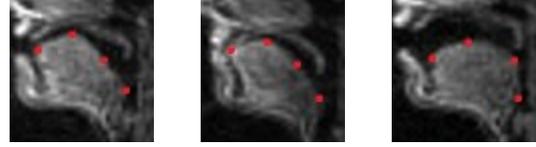
Fig. 7. Key points of tongue contour: TT, TD, TB and TR.

Table 2. RMSE in pixels of key points.

Key points	Speaker M1	Speaker F1	Mean
TT	1.88	1.57	1.73
TD	0.41	0.51	0.46
TB	0.86	0.69	0.78
TR	0.07	0.13	0.10

Table 3. Mean RMSE in pixels of ten speakers.

Speaker	The baseline method	The proposed method
M1	1.06	0.81
M2	0.86	0.70
M3	0.57	0.63
M4	0.65	0.46
M5	1.16	0.92
F1	0.91	0.73
F2	1.02	0.69
F3	0.74	0.58
F4	0.80	0.55
F5	0.54	0.61
Mean	0.83	0.67

However, there may be a big difference of mean RMSE among different speakers, such as Speaker M4 and M5, so the robustness needs to be further improved in the future.

5. CONCLUSIONS

A novel method of semi-automatic tongue contour extraction in rtMRI sequences is proposed. The initial boundary image is obtained with combined multi-directional Sobel operators in registered region; then a boundary intensity map is constructed to find the most probable tongue contour points by searching for the optimal route with Viterbi algorithm; finally the tongue contour is obtained using B-Spline approximation. Experiments show that the proposed method is of higher accuracy in tongue contour extraction, especially when some parts of tongue touch other organs. The proposed method could be extended to extract contours of other organs such as joints and backbones in MRI or X-ray images.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] J.M. Unger, "The oral cavity and tongue: magnetic resonance imaging," *Radiology*, vol. 155, no. 1, pp. 151-153, 1985.

[2] A.M. Sulter, D.G. Milker, R.F. Wolf, H.K. Schutte, H.P. Wit, and E.L. Mooyaart, "On the relation between the dimensions and resonance characteristics of the vocal tract: a study with MRI," *Magnetic Resonance Imaging*, vol. 10, no.3, pp. 365–373, 1992.

[3] S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd, "An approach to real-time magnetic resonance imaging for speech production," *JASA*, vol. 115, no. 4, pp. 1771–1776, 2004.

[4] E. Bresch, Y.C. Kim, K. Nayak, D. Byrd, and S. Narayanan, "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 123-132, 2008.

[5] E. Bresch, S. Narayanan, "Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 28, no. 3, pp. 323–338, 2009.

[6] A. Martins, N. Mascarenhas, and C. Suazo, "Spatio-temporal resolution enhancement of vocal tract MRI sequences based on image registration," *Integrated Computer-Aided Engineering*, vol. 18, no. 2, pp. 143-155, 2011.

[7] M. Stone, E.P. Davis, A.S. Douglas, M.N. Aiver, R. Gullapalli, W.S. Levine, and A.J. Lundberg, "Modeling tongue surface contours from Cine-MRI images," *Journal of Speech and Hearing Research*, vol. 44, no. 5, pp. 1026–1040, 2001.

[8] V.J. Wedeen, T.G. Reese, V.J. Napadow, and R.J. Gilbert, "Demonstration of primary and secondary muscle fiber architecture of the bovine tongue by diffusion tensor magnetic resonance imaging," *Biophysical Journal*, vol. 80, no. 2, pp. 1024-1028, 2001.

[9] T. Sayoko, K. Honda, "An MRI analysis of the extrinsic tongue muscles during vowel production," *Speech Communication*, vol. 49, no. 1, pp. 49–58, 2007.

[10] M. Stone, E.P. Davis, A.S. Douglas, M.N. Aiver, R. Gullapalli, W.S. Levine, and A.J. Lundberg, "Multi-subject atlas built from structural tongue magnetic resonance images," *Proceedings of Meetings on Acoustics Acoustical Society of America*, vol. 19, pp. 1-4, 2013.

[11] R. Lauder, Z.F. Muhl, "Estimation of tongue volume from magnetic resonance imaging," *Angle Orthodontist*, vol. 61, no.3, pp. 175-184, 1991.

[12] M. Stone, E.P. Davis, A.S. Douglas, M. NessAiver, R. Gullapalli, W.S. Levine, and A. Lundberg, "Modeling the motion of the internal tongue from tagged Cine-MRI images." *Journal of the Acoustical Society of America1*, Vol. 9, no. 6, pp. 2974-2982, 2001.

[13] R.E. Martin, B.J. MacIntosh, R.C. Smith, A.M. Barr, T.K. Stevens, J.S. Gati, and R.S. Menon, "Cerebral areas processing swallowing and tongue movement are overlapping but distinct: A functional magnetic resonance imaging study," *Journal of Neurophysiology*, vol. 92, no. 4, pp. 2428-2493, 2004.

[14] H. Li, J.H. Tao, M.H. Yang, and B. Liu, "Estimate articulatory MRI series from acoustic signal using deep architecture," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), pp. 4854-4858, 2015.

[15] T. Peng, E. Kerrien, and M.O. Berger, "A shape base framework to segmentation of tongue contours from MRI data," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), pp. 662 - 665, 2010.

[16] C. Song, J. Wei, Q. Fang, S. Liu, Y. Wang, and J. Dang, "Tongue shape synthesis based on Active Shape Model," *International Symposium on Chinese Spoken Language Processing* (*ISCSLP*), pp. 383-386, 2012.

[17] J. Lee, J. Woo, F. Xing, E.Z. Murano, M. Stone, and J.L. Prince, "Semi-automatic segmentation for 3D motion analysis of the tongue with dynamic MRI," *Computerized Medical Imaging and Graphics*, vol. 38, no. 8, pp. 714-724, 2014.

[18] M.I. Proctor, D. Bone, N. Katsamanis, and S. Narayanan, "Rapid semi-automatic segmentation of real-time magnetic resonance images for parametric vocal tract analysis," *Proceedings of the IEEE International Speech Communication Association* (*Interspeech*), pp. 1576-1579, 2010.

[19] M.I. Proctor, A. Lammert, A. Katsamanis, L. Goldstein, C. Hagedorn, and S. Narayanan, "Direct estimation of articulatory kinematics from real-time magnetic resonance image sequences," *Proceedings of the IEEE International Speech Communication Association* (*Interspeech*), pp. 281-284, 2011.

[20] M.I. Proctor, R. Walker, "Articulatory bases of sonority in English liquids," *Berlin De Gruyter*, pp.289-316, 2012.

[21] C. Smith, M. Proctor, K. Iskarous, L. Goldstein, and S. Narayanan, "Stable articulatory tasks and their variable formation: Tamil retroflex consonants," *Proceedings of the IEEE International Speech Communication Association* (*Interspeech*), pp. 2006-2009, 2013.

[22] F.Y. Hsieh, L. Goldstein, D. Byrd, and S. Narayanan, "Truncation of pharyngeal gesture in English diphthong [aɪ]," *Proceedings of the IEEE International Speech Communication Association* (*Interspeech*), pp. 968-972, 2013.

[23] M.H. Yang, J.H. Tao, and D.W. Zhang, "Extraction of tongue contour in X-ray videos," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), pp. 1094-1098, 2013.

[24] S. Narayanan, A. Toutios, V. Ramanarayanan, A. Lammert, J. Kim, and S. Lee, "Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC)," *The Journal of the Acoustical Society of America*, vol. 136, pp. 1307-1311, 2014.