SEGMENTATION OF THE AIRWAY FROM THE SURROUNDING TISSUES ON MAGNETIC RESONANCE IMAGES: A COMPARATIVE STUDY

Alain Soquet, Véronique Lecuit*, Thierry Metens¨, Bruno Nazarian* and Didier Demolin°*

*Laboratoire de Phonétique Expérimentale, ¨Unité O.R.L. Hôpital Erasme, °Laboratoire de Phonologie Expérimentale, Université Libre de Bruxelles CP 110, 50 av. F.D. Roosevelt, 1050 Brussels Belgium

ABSTRACT

Magnetic Resonance Imaging techniques are uniquely attractive in their ability to provide an extensive body of information on the vocal tract geometry. Once the images are acquired, they must be further processed in order to segment the airway from the surrounding tissues, so as to locate the air passage. This problem has been addressed in several ways in the litterature.

In this paper, we carry out a comparative study of different approaches to the same body of data in order to assess the accuracy of the different methods. It is shown that the different methods present small average error and large error distribution.

1. INTRODUCTION

Advances in imaging techniques have allowed to renew the way the vocal tract geometry is studied. It has been recently confirmed that cross-sections can vary considerably along its length (Stone 1991 [7]; Demolin et al. 1996 [4]) and show a high degree of asymmetry (Stone 1991 [7]).

It is thus obvious that the vocal tract has to be apprehended as a three-dimensional structure in order to study both speech production and the link between the articulatory and the acoustic space.

Magnetic Resonance Imaging (MRI) techniques are uniquely attractive in their ability to provide an extensive body of information on the vocal tract geometry. Once the images are acquired, they must be further processed in order to segment the airway from the surrounding tissues, so as to locate the air passages. This problem has been addressed in several ways in the literature (Baer et al. 1991 [1]; Moore 1992 [5]; Demolin et al. 1996 [4]; Story et al. 1996 [8]).

Baer et al. 1991 [1], Moore 1992 [5] and Badin et al. 1998 [2] use semi-automatic threshold techniques. These approaches give systematic results based only on image properties. They are usually callibrated with regard to water filled tubes with known cross-sectionnal areas. Moore (1992) [5] reported underestimation of the vocal tract area by less than 12%, Crary and colleagues (1996) [3] suggest a maximal variation of less than 5% on their data.

Demolin et al. (1996) [4] outline the sections on a transparent sheet, the outlines are introduced in the computer by means of a digitization tablet, then each area is computed by a polygon surface computation algorithm.

The problem is complicated in some parts of the vocal tract by the presence of poorly imaged structures (e.g. teeth) and by the emergence of structures in the airspace (e.g. uvula, epiglottis).

The first problem cannot be handled automatically as the necessary information is not present in the image itself. Complementary data sources are thus necessary (e.g. X-rays or casts).

The second problem can be handled easily by subtracting the area of the emergent structure from the one of the air space.

In this paper, have focussed on the comparison of different measurement methods. We will therefore carry out measurements only in those regions of the vocal tract where the two problems mentionned above do not show up. All measurements have therefore been made in the pharynx between the epiglottis and the uvula.

2. METHODS

We have compared three segmentation methods. The first one (Demolin et al. 1996 [4]) requires the outline of the contour by hand on a transparent sheet. By means of a digitization tablet, the outlines are introduced in the computer and each area is computed by a polygon surface computation algorithm. This method will be refered to as the *manual method*. This digitization process however, might be biased by some human factors. A test of the accuracy and reproductibility of area measurements is displayed in table 1. Three different outlines corresponding to three different sections, a large, a medium and a small one where used as test samples. Each outline was measured 10 times. Mean and standard deviation are given for each section. Results show that standard deviation are similar in the three cases and is lower than 0.5 mm^2.

	Cross sectional areas [mm^2]		
	Large	Medium	Small
Mean	217.9	50.6	6.9
Std deviation	0.44	0.30	0.30

Table 1: Measurement accuracy of the manual method.

The second method consists in establishing a contour level (air-tissue contrast) for the images, then the contour can be obtained by a simple threshold technique. This method will be refered to as the *threshold method*. The same method has also been applied on the image zoomed four times and smoothed. This method will be refered to as the *threshold-zoom method*. The threshold for the air-tissue boundary was established for all images of one speaker. The area to be measured is selected by placing a mouse-controlled cursor inside the region of interest.

The third method is based on an elastic matching process: a small free-form curve is placed in the region of interest and extended inside the region until it reaches the air-tissue border. This method will be refered to as the *elastic method*. See Figure 1.

(a) (b) (c) (d)

Figure 1: The elastic matching process. For each image, we initialize a closed free-form curve in the interest area (a). This curve is defined by a set of control points. A set of criteria is computed at this point to caracterize this area : pixel level, density gradiant or texture analysis. Then, this curve is growing by a global moving of its set of control points (b, c, and d). Each moving is controlled by the local properties of the curve and the image properties.

2. DATA

MRI data has been acquired for two female speakers (subject 1 and 2) and two male speakers (subject 3 and 4), all of them native French speakers living in Brussels. The task of the subjects was to pronounce and to sustain nasal and oral French vowels. The reference was a word containing the vowel to be pronounced. This reference word was given a few seconds before the recording by one of the experimenters.

Subjects 1 and 4 have pronounced 4 nasal vowels and subjects 2 and 3 four nasal and four oral vowels.

The magnetic resonance images have been acquired at the Magnetic Resonance Unit of the Hôpital Erasme, Université Libre de Bruxelles on a 1.5 T MRI system with a quadrature Head-Neck coil (Philips Gyroscan NT ACS, Best The Netherlands). The images consisted of 1 stack of 18 transversal slices. This proton density weighted acauisition took 13.8 seconds with the following parameters: TR: 1716 ms, TE: 9ms, gradient profile low-high, ETL = 11, Partial Fourier encoding: 60%, Field of view 250 x 200 mm, Matrix 161 x 256. The stack of tranversal slices was defined with slice thickness 4 mm and no gap between adjacent slices of the stack. The positioning of the stack was planned on a sagittal survey image realized during similar phonation. The stacks have been positioned so as to be above the epiglottis with slices perpendicular to the oro-pharynx wall. Acquisition was launched immediately after the beginning of phonation.

3. RESULTS

Measures have been carried out with the four methods described above in the pharynx region (above the epiglottis and below the uvula). The number of measured sections is not the same for the different subjects due to variation in vocal tract length. The number of measurements is summarized in table 2.

Subject	Measurements
1	44
2	78
3	128
4	65
All	315

Table 2: Measurement number for the different subjects.

The manual method has been taken has reference for two main reasons. First, it has shown to be accurate and reproducible, and second, the manual outlining of the boundary allows the operator to take into account knowledge about the vocal tract geometry.

Table 3 shows the correlation between the errors of the different methods. It can be seen that the correlation between the elastic and the threshold methods is low. The correlation is of course higher between the threshold and the threshold-zoom methods. This shows that the manual method does not give rise to a systematic bias in comparison with the three other methods.

	Threshold	Threshold-Zoom
Elastic	0.58418	0.523507
Threshold		0.879809

Table 3: Correlation between the errors of the different methods.

Tables 4 and 5 show respectively the average and the standard deviation of the measurement errors between the different methods and the manual. Tables 6 and 7 the average and the standard deviation of the relative errors between the different methods and the manual.

Subject	Threshold	Threshold-Zoom	Elastic
1	2.63	1.65	11.98
2	-0.17	-1.59	3.76
3	4.89	3.50	0.96
4	-2.22	-3.88	-3.67
All	1.85	0.46	2.23

Table 4: Average measurement errors between the different methods and the manual method in mm^2.

Subject	Threshold	Threshold-Zoom	Elastic
1	37.18	28.81	28.49
2	11.04	20.19	21.43
3	58.47	49.22	61.95
4	13.15	17.14	16.96
All	40.73	35.52	42.94

Table 5: Standard deviation of measurement errors between the different methods and the manual method in mm^2.

Subject	Threshold	Threshold-Zoom	Elastic
1	1.36	0.26	8.49
2	1.90	0.31	7.11
3	2.42	-0.16	6.10
4	-4.08	-5.70	-2.66
All	0.80	-1.13	4.87

Table 6: Average relative measurement errors between the different methods and the manual method in %.

Subject	Threshold	Threshold-Zoom	Elastic
1	20.26	20.06	17.34
2	26.56	27.77	24.784
3	21.67	23.03	61.20
4	15.50	15.75	11.11
All	21.78	22.69	41.80

Table 7: Standard deviation of relative measurement errors between the different methods and the manual method.

From these results it can be seen that even if the mean error is relatively small for the different methods, the distribution of the errors is very widespread.

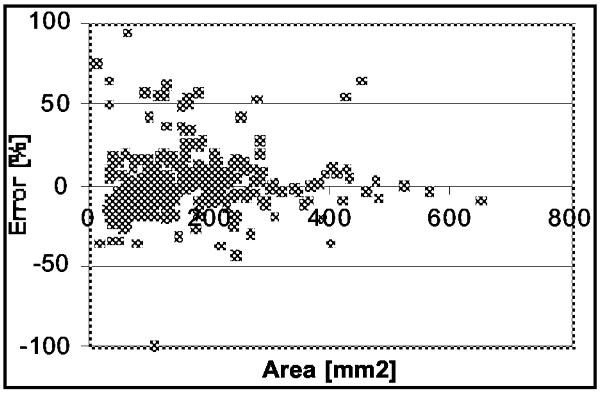

Figure 1: Relative measurement errors between the threshold method and the manual method.

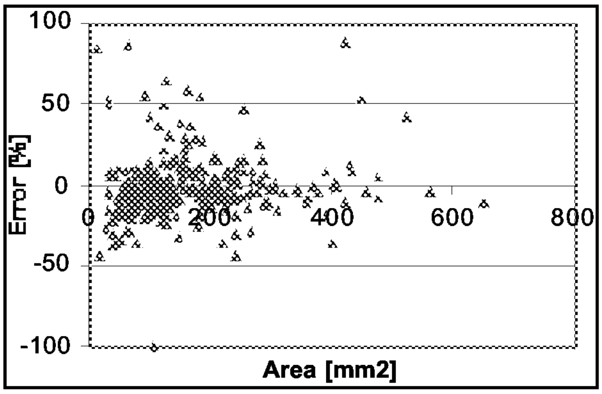

Figure 2: Relative measurement errors between the threshold-zoom method and the manual method.

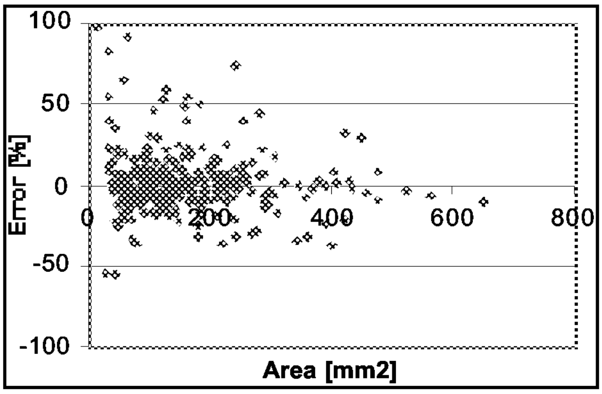

Figure 3: Relative measurement errors between the elastic method and the manual method.

Figures 1, 2 and 3 present the relative errors for the threshold method, the threshold-zoom method and the elastic method respectively.

It can be seen from figure 1 and 2 that the error distribution is similar for the two threshold methods, which is not surprising, and that figure 3 shows a different pattern.

These results illustrate the behavior of the different method in real world situation. The calibration of the segmentation method with water filled tubes with known cross-sectionnal areas. (Moore 1992 [5]; Crary et al. 1996 [3]) seems to give only an optimistic evaluation of the segmentation error. Indeed the contrast between the air and the surrounding tissues is lower than the contrast between air and water.

4. CONCLUSIONS

Results show that (i) the three methods give comparable results (small average error and large error distribution) with a somewhat lower dispersion for the threshold methods, and (ii) the settings of the parameters of each method (contrast, threshold level, free-form curve) have an impact on the resulting area.

5. REFERENCES

1 Baer T., Gore J.C., Gracco L. C. and Nye P.W. “Analysis of vocal tract shape and di;ensions using magnetic resonance imaging: Vowels,” ,” *J. Acoust. Soc. Am.* 90 (2): 799-828, 1991.

2 Badin P., Pouchoy L., Bailly G., Raybaudi M., Segebarth C., Lebas J.F., Tiede M., Vatikiotis-Bateson E. and Tohkura Y. “Un modèle articulatoire tridimensionnel du conduit vocal basé sur des données I.R.M.,” *Actes des XX1èmes JEP*, Martigny, Switzerland, 1998.

3 Crary M.A., Kotzur I. M., Gauger J., Gorham M. and Burton S. “Dynamic magnetic resonace imaging in the study of vocal tract configuration,” *Journal of Voice*, 10 (4):378-388, 1996.

4 Demolin D., Metens T. and Soquet A. “Three_dimensional measurement of the vocal tract by MRI,”*International Conference on Spoken Language Processing*. Philadelphie. USA, 1996.

5 Moore C.A. “The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images,” *J. Speech Hear. Res.*: 35. 1009-1023, 1992.

6 Nazarian B. “Reconstruction automatique de structures anatomiques tubulaires à l’aide de surfaces de forme libre. Applications en milieu Hospitalier,” PhD Thesis. Université de la Méditerrannée Aix-Marseille II. Marseille, France, 1996.

7 Stone M. “Toward a model of three-dimensional tongue movement,” *J. of Phonetics*: 19. 309-320, 1991.

8 Story B.H., Titze I.R. and Hoffman E.A. “Vocal tract area functions from magnetic resonance imaging,” *J. Acoust. Soc. Am.* 100: 537-554, 1996.