Ann.Bull.RILP(1986)No.20, 47-54

MEASUREMENTS OF THE THREE-DIMENSIONAL SHAPE OF THE VOCAL TRACT BASED ON THE MAGNETIC RESONANCE IMAGING TECHNIQUE

Mitsuhiro Rokkaku*, Kiyoshi Hashimoto*,
Satoshi Imaizumi, Seiji Niimi and Sigeru Kiritani

Purpose

Although it is possible to measure the three-dimensional shape of the tract using the X-ray projection technique, possible X-ray hazard prevents such data collection. Thus, the fine effects of the three-dimensional structure of the tract upon the speech production process have not been considered. For instance, analyses of the acoustic transfer characteristics of the vocal tract have been carried out based on an assumption that the vocal tract is an unbent acoustic tube of varying cross-sectional area, and, therefore, the effects of the curvature of the tract are usually not considered.

To resolve such problems, we have tried to measure the three-dimensional shape of the vocal tract using the magnetic resonance imaging (MRI)[1] technique. This technique poses no hazard to the human body[2], although it has several shortcomings such as insufficient space or time resolution.

Method

Data

Sagittal head images of a 24-year-old man pronouncing the Japanese vowel /a/ were measured by MRI.

Assuming that the tract is right-and-left symmetric to the central cross-sectional plane, four sagittal images of the left side were taken every 5mm shift in the direction perpendicular to central cross-sectional plane. The thickness of one slice was 10mm. The imaging time was about 25s per slice. The planar resolution was 1.2mm by 1.2mm. Fig. 1 shows some examples of these sagittal head images.

Procedure

The three-dimensional shape of the vocal tract was detected from the sagittal images using the following four steps.

* Univ. of Electro-Communications

Step I: First, for each sagittal image as shown in Fig. 2 (a), the mean direction of the gradients of the gray levels was calculated in a 5 by 5 matrix around pixel $I_{i,j}$. When the gray levels around pixel $I_{i,j}$ looked like Fig. 3, the direction of the gradient T was defined as

$$\begin{aligned} di &= (C+2E+H)-(A+2D+F) \\ dj &= (H+2G+F)-(C+2B+A) \\ T &= \mathrm{int}(8*\tan^{-1}(dj/di)/2\pi) \end{aligned} \qquad (1)$$

The average direction for the 5 by 5 pixels around $I_{i,j}$ were calculated. The local area was divided into two regions by a line perpendicular to the average direction as shown in Fig. 4 (c). Then, the mean gray level for each region was calculated. The gray level of pixel $I_{i,j}$ was replaced by the mean gray level which had a smaller absolute difference than the grey level of $I_{i,j}$. The result is shown in Fig. 2 (b). If linear averaging of the 5 by 5 neighborhood was applied, the gray level at pixel $I_{i,j}$ was replaced by 4.6. On the other hand, this method changed the gray level at pixel $I_{i,j}$ to 9. Thus, the discontinuity could be enlarged by this method.

Step II: After the process described above, spike noises were removed by a median filter on a 3 by 3 neighborhood. The result of this processing is shown in Fig. 2 (c).

Step III: In addition, this image was processed by an algorithm defined as an inter-class variance filter, and then the shape of the tract was detected.

First, 25 pixels in the neighborhood around $I_{i,j}$ were divided into 2 classes using a threshold so as to maximize the inter-class variance. Then, if the gray level at $I_{i,j}$ was greater than the threshold, the gray level at pixel $I_{i,j}$ was set to the inter-class variance. Otherwise, it was set to zero.

Step IV: The extracted vocal tract images from four slices were stacked in a direction perpendicular to the central cross-sectional plane and interpolated using a spline function to get a three-dimensional shape.

Results

Fig. 2 (a) shows an original sagittal image, (b) the output of Step I, and (c) the result of Step II. For reference purposes only, in Figs. 2 (d), (e), and (f) the differentiated images of Figs. 2 (a), (b), and (c) using a Sobel operator[3] are given. According to the processing Steps I and II, we can confirm the remaining portions. However, the edges of the vocal tract were smoothed to be dark or bright as shown in Figs. 2 (b) and (c). And as shown in Fig. 2 (c), the two-dimensional shape of the vocal tract can be clearly distinguished as the dark section.

Figs. 5 (a), (b), (c), and (d) show the shapes of the vocal tract extracted from the sliced images taken at planes which were 0, 5, 10, and 15mm distant from the central cross-sectional plane.

Fig. 6 (a) is the transversal shape on the plane perpendicular to the mid sagittal plane and is represented by the line connecting x and x in Fig. 6 (b).. This transversal shape was constructed by interpolating the cross points between the transversal plane and the edge of the tract on each sagittal image. This interpolation was carried out using a normalized B-spline function.

Discussion

The results discussed above suggest the following.

Some parts of the vocal tract, such as the lips, tongue, soft palate and pharynx, can be clearly detected. Other parts, such as the teeth or teethridge, are not easy to detect by this method. This difficulty comes from the fact that MRI can image only tissues consisting of many protons. In other words, hard tissues, such as teeth, teethridge and their vicinity, can not be imaged. Such portions must be observed using some other technique.

Furthermore, it was difficult to detect the shape of the vocal tract wall as the slice was shifted far from the central cross-section plane. This may be due to the fact that the thickness of one slice was large, and, thus, the edge of the vocal tract was smeared to some extent. The thickness can be reduced, but this may introduce some imaging noise.

Since the imaging time per slice was about 20s, the measured shape of the tract was the one averaged within the duration. Therefore, the dynamic characteristics of the articulation could not be measured.

Due to these limitations, the vocal tract shape was not always detectable as a continuous curve.

Conclusion

We confirmed that the cross-sectional shape of any point of the vocal tract can be measured. In spite of X-ray projection, it is possible to measure the three-dimensional shape without any hazard to the human body. Since the imaging speed is as slow as 25s, the MRI method is not suitable for measuring active movements of the articulators.

However, three-dimensional measurements of the dynamic movements of articulation will be possible by utilizing both of the static measurements given by this method, and dynamic measurements provided by some other techniques, such as ultrasonic tomography.

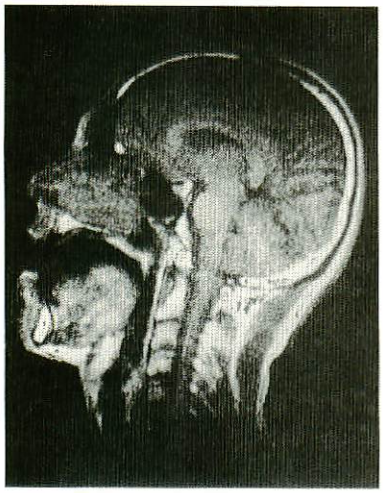
Acknowledgement

We wish to express our sincere gratitude to associate Prof. Y. Itai, Dr. S. Aoki, and Mr. K. Hurukawa, Department of Radiology, Faculty of Medicine, University of Tokyo, for their technical help and valuable comments. We would like to appreciate Mr. Hiroshi Nakayama, Waseda University, for his kind cooperation.

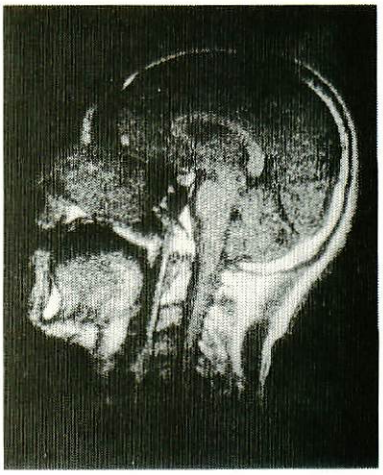
References

1) L. E. Crooks (1985); An introduction to magnetic resonance imaging, IEEE in Medicine and Biology Magazine, 4, 3 , 8-15.
2) T. F. Budinger (1985); Health effects of in vivo nuclear magnetic resonance, IEEE in medicine and biology magazine, 4, 3, 31-38.
3) H. Ozaki, K. Taniguchi, and H. Ogawa; Image Processing -- from its basis to its applications, (Kyoritsu Shuppan, Tokyo, 1983), 152-153, (in Japanese).

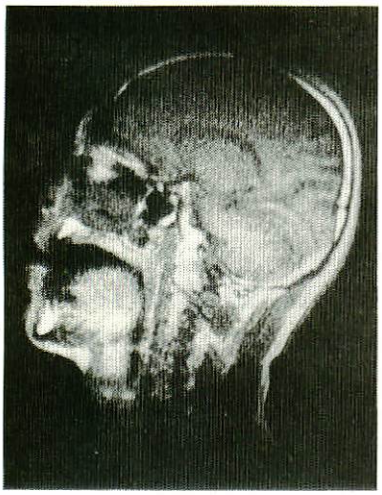
(a)



(b)



(c)



(d)

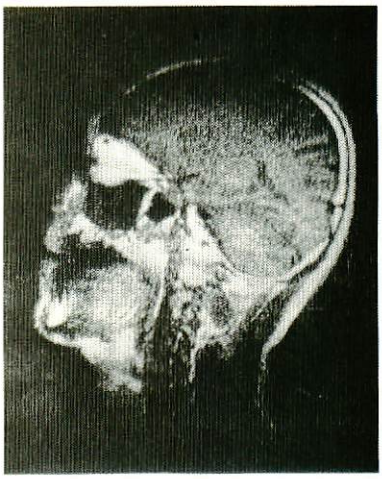


Fig. 1. MRI images of the head in planes, the distances of which from the center are (a) 0mm, (b) 5mm, (c) 10mm and (d) 15mm.

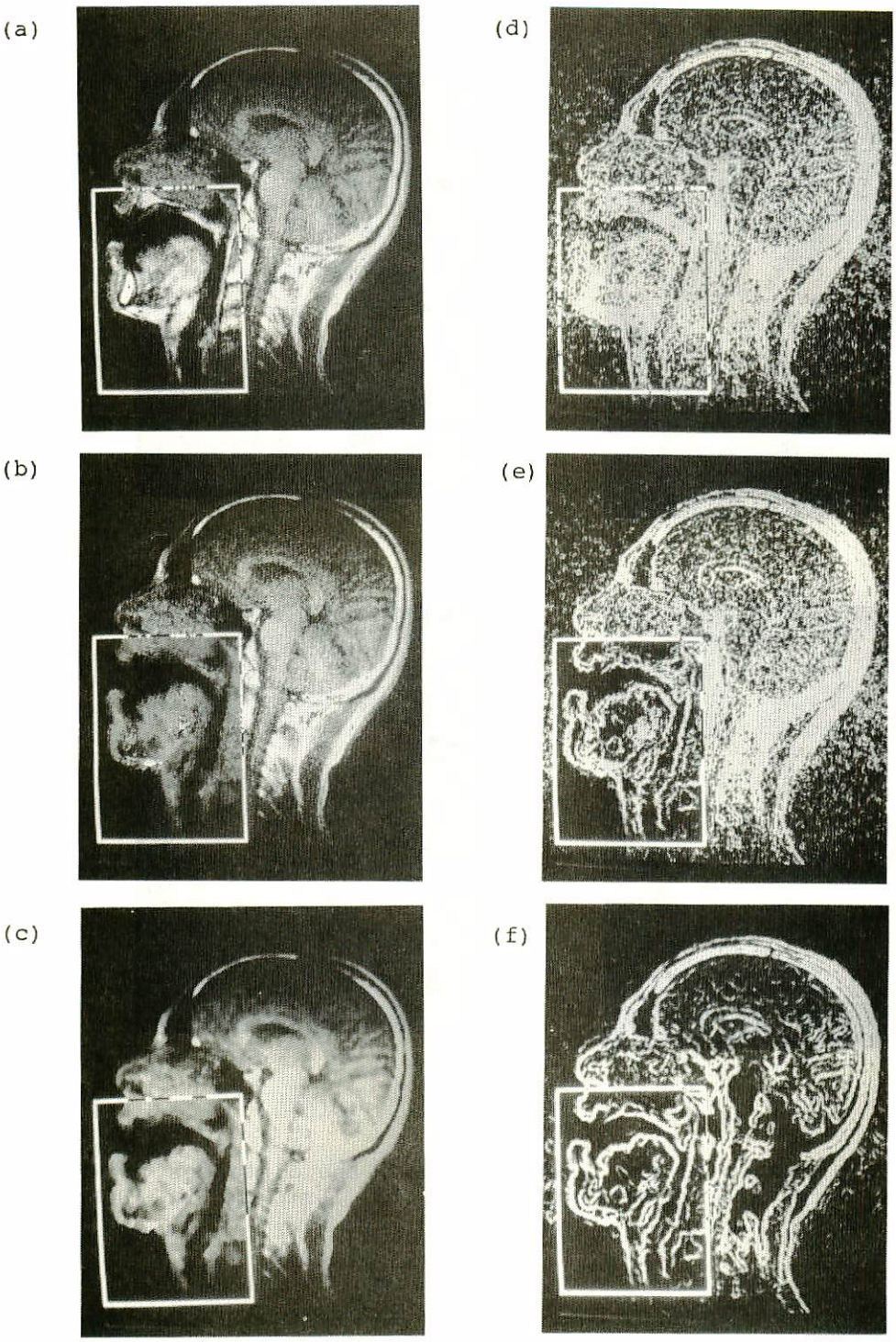


Fig. 2. Processed images and reference images: (a) Original, (b) smoothing of (a), (c) median filtering of (b), (d) differentiation of (a), (e) differentiation of (b) and (d) differentiation of (c). The processed area is indicated by the square.

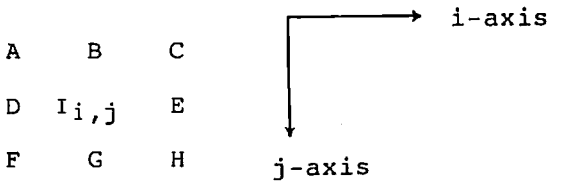


Fig. 3. Configuration of the 3 by 3 neighborhood of the pixel $I_{i,j}$.

(a)

0	0	5	9	9
0	0	5	9	9
0	0	5	9	9
0	0	5	9	9
0	0	5	9	9

(b)

×	→	→	→	×
×	→	→	→	×
×	→	→	→	×
×	→	→	→	×
×	→	→	→	×

× : undefined

(c)

0	0	9	9
0	0	9	9
0	0	9	9
0	0	9	9
0	0	9	9

Fig. 4. Schematic representation of Step I.
(a) Configuration of a 5 by 5 neighborhood, (b) directions of the gradient and (c) Division of the region

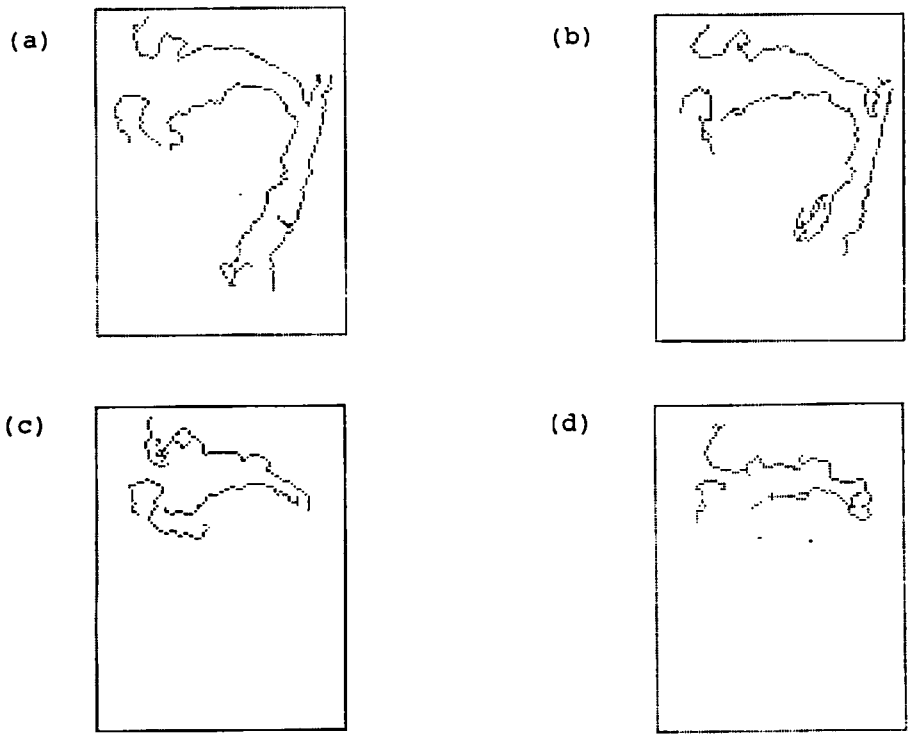


Fig. 5. Detected shapes of the vocal tract in the planes: (a) at 0mm, (b) at 5mm, (c) at 10mm and (d) at 15mm from the center.

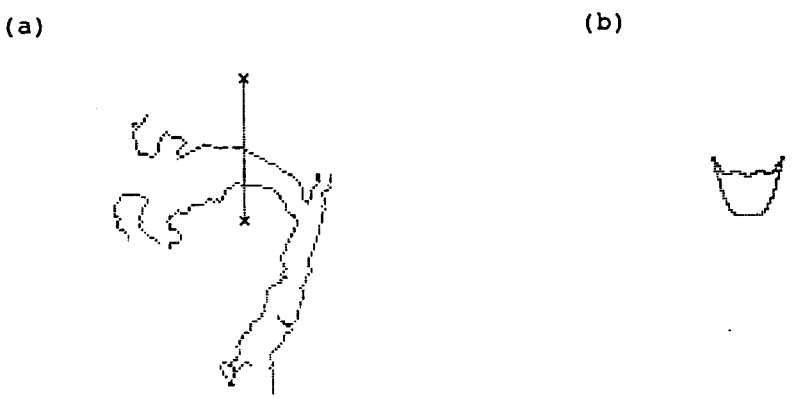


Fig. 6. Transversal shape of the vocal tract (a) in a plane perpendicular to the mid sagital plane represented by a line connecting x and x (b).