Acoust. Sci. & Tech. **25**, 6 (2004)

TECHNICAL REPORT

A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions

Hironori Takemoto[1,*], Tatsuya Kitamura[1,†], Hironori Nishimoto[2,‡] and Kiyoshi Honda[1,§]

[1]*ATR Human Information Science Laboratories, 2–2–2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619–0288 Japan*
[2]*School of Information Science, Japan Advanced Institute of Science and Technology, 1–1 Asahidai, Tatsunokuchi, Ishikawa, 923–1292 Japan*

(*Received 11 December 2003, Accepted for publication 24 May 2004*)

Abstract: Measurement of vocal tract area functions from MRI data requires a technique for tooth visualization because the teeth are as transparent as air in the images. In this article, a new method is proposed to accurately superimpose the teeth onto MRI volume data. Upper and lower tooth images with the surrounding bony structure are obtained by scanning a subject holding a contrast medium in the oral cavity. They are superimposed onto the target volume data via a three-dimensional transformation using landmarks sampled from the tooth images and target MRI data. The accuracy of the dental image superimposition is ensured by the minimization of the *error volume*, which is the mismatch volume of the dental image overlapping the surrounding soft tissue. The method is evaluated using five operators sampling the landmarks. Results show that the *error volume* is significantly reduced to a nearly constant value regardless of the operator's skill.

Keywords: MRI, Tooth superimposition, Vocal tract, Interdental space, *Error volume*

PACS number: 43.70.Jt [DOI: 10.1250/ast.25.468]

1. INTRODUCTION

The magnetic resonance imaging (MRI) technique has been used for measuring vocal tract shape in three dimensions and calculating the area functions [1–4] because MRI has no known harmful effects. The use of MRI has expanded from two-dimensional to three-dimensional imaging, and from static to dynamic imaging [5]. Thus MRI is becoming one of the most powerful tools for studying speech production, but it has several disadvantages with respect to measuring the vocal tract. The most serious one is the lack of the teeth; since calcified structures such as bones and teeth have little mobile hydrogen, they are almost invisible in MRI data. Consequently, the teeth and oral cavity have the same low pixel value (black) in an image, and are indistinguishable from each other. In such images, the oral cavity size is easily overestimated.

Various methods have been used to exclude the teeth volume from the oral cavity in MRI data. The use of dental stone casts is a commonly used procedure. Yang *et al.* [6] scanned casts in water using MRI to obtain the three-dimensional shape of the teeth for determining the boundaries of the oral cavity. Engwall [7] built polygon mesh data of the teeth and palate using a dental cast for the same purpose. Narayanan *et al.* [8,9] and Alwan *et al.* [10] measured casts using calipers, and sliced dental impressions to trace the outlines of the slices. This method was also adopted by Espy-Wilson *et al.* [11].

A few different approaches have also been attempted. Story *et al.* [3,4,12] and Tom *et al.* [13] used electron beam computed tomography (EBCT) to extract the dental images to obtain accurate volume data of the vocal tract. Matsumura *et al.* [14] and Niikawa *et al.* [15,16] used a dental crown plate made of vegetable oil to directly image the dental boundaries. Wakumoto *et al.* [17] used double plastic plates with an oil layer in between for the same purpose.

A few disadvantages can be noted among the previous methods described above. The methods using the dental cast [6–11] or dental crown plates [14–17] required image-by-image manual segmentation of the tooth boundaries, which takes time and effort. The methods using x-ray

*e-mail: takemoto@atr.jp
†e-mail: kitamura@atr.jp
‡e-mail: nhiro@jaist.ac.jp
§e-mail: honda@atr.jp

[3,4,12,13] are of limited use. The notable problem common across all methods is the lack of a criterion to evaluate accuracy. In this paper, we propose a new method to superimpose three-dimensional dental images on MRI volume data for accurate vocal tract measurement.

2. MATERIALS

The MRI data used in this study was obtained from a Japanese male subject (24 years old) using a clinical MRI scanner, the MAGNEX ECLIPSE 1.5T Power Drive 250 (Shimadzu-Marconi). The cervical part of the neck array coil was used to obtain dental images in Experiment 1, and the torso coil was used to record orofacial images in Experiment 2.

2.1. MRI Experiment 1: Dental Scans for Extracting "Digital Jaw Casts"

The purpose of the dental scan is to obtain volume data for the upper and lower jaws with teeth. MRI data were obtained to produce a set of "digital jaw casts" for superimposing the dental images onto the target volume images. The subject took a prone position during the scan, holding blueberry juice inside the mouth as an oral contrast medium. It is known that blueberry juice can serve as a contrast medium for MRI [18]. In addition, the subject bit on a small wooden chip between the upper and lower teeth so that the interdental space could be visualized. The scan parameters were as follows: fast spin echo (FSE) scan sequence, sagittal slice plane, 1.5 mm slice thickness, no slice gap, no averaging, 256×256 mm field of view (FOV), 512×512 pixel image size, 51 slices, 11 ms TE, and 3,000 ms TR. The total acquisition time was approximately 180 s.

2.2. MRI Experiment 2: Orofacial Scans for Vocal Tract Imaging While Producing Five Japanese Vowels

Five sets of volume data were acquired to extract the vocal tract during sustained productions of Japanese vowels /a/, /i/, /u/, /e/, and /o/. The subject took a supine position during the scans, and produced each vowel while maintaining the speech organs in a steady position. All of the scan parameters were the same as those in the dental scan, except the slice thickness (2.0 mm). The acquisition time was 180 s for each scan.

3. METHOD

The procedures in our method are as follows: (1) Extract the upper and lower jaws with teeth from the dental scan data to reconstruct a set of upper and lower "digital jaw casts"; (2) sample the coordinates of four anatomical landmarks on the upper and lower jaws from all the volume data; (3) generate initial affine transformation matrices from four landmarks of the "digital jaw casts" and those of the target volume data so that the former can be superimposed on the latter; and (4) optimize the affine transformation matrices to minimize the *error volume* using a simulated annealing (SA) method [19].

3.1. Image Processing and Volume Reconstruction

The original images in a 16-bit DICOM format exported from the MRI system were converted into 8-bit images using a threshold selection algorithm based on a mean adjacent-pixel number [20]. This method can automatically determine the threshold value at which the body tissue is appropriately segmented from the airway in a 16-bit image. After the threshold value was calculated, a linear conversion was performed so that the value in a 16-bit image was mapped to 40 in an 8-bit one. This process was necessary to eliminate any variation in image histograms among different image sets. After the conversion, the air-tissue threshold value was 40 in all 8-bit image sets.

The dental scan data shows that the upper and lower jaws with teeth can be extracted by manually tracing the boundaries to reconstruct the "digital jaw casts." Anatomically the upper jaw represents the maxilla and palatal bone inferior to the base of the zygomatic process, and the lower jaw is the mandible except for the right and left rami. Because the contrast medium in the oral cavity had high pixel values (white) but the teeth had low pixel values (black), the dental boundaries can be identified and the upper and lower jaws can be segmented in each image.

The images of the upper and lower jaws and those of orofacial scans were interpolated using a tri-linear algorithm to reconstruct volume data sets consisting of a cubic voxel of $0.5 \times 0.5 \times 0.5$ mm. Figure 1 represents the visualized "digital jaw casts."

3.2. Anatomical Landmarks and Sampling Method

Four anatomical landmarks were defined on the upper jaw and four on the lower jaw, as shown in Fig. 2.

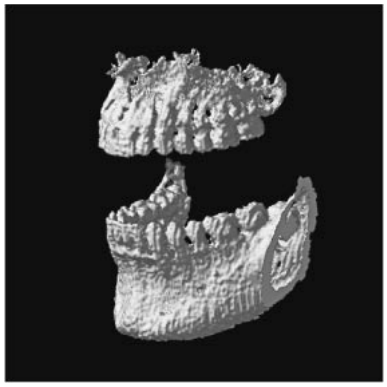
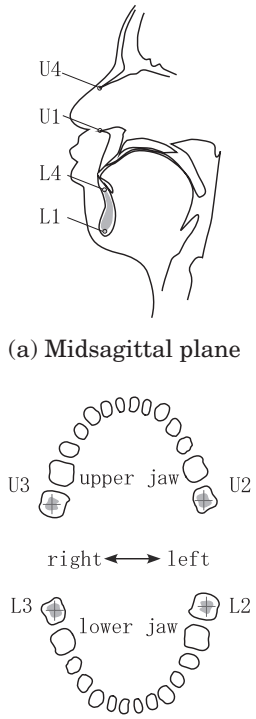


Fig. 1 Upper and lower "digital jaw casts." They were extracted from the data in MRI Experiment 1.



(a) Midsagittal plane

(b) Upper and lower jaws

Fig. 2 Anatomical landmarks. Four landmarks were defined on the upper jaw and four on the lower jaw. U1: anterior nasal spine (ANS), U2: center of the dental pulp of the maxillary left second molar, U3: center of the dental pulp of the maxillary right second molar, U4: rhinion (the midline point at the inferior free end of the internasal suture), L1: superior point of the cancellous bone of the mandible on the midsagittal plane L2: center of the dental pulp of the mandibular left second molar, L3: center of the dental pulp of the mandibular right second molar, and L4: inferior point of the cancellous bone of the mandible in the midsagittal plane.

The landmarks do not necessarily need to be the above-mentioned ones; any anatomical landmark is useful if it can be identified both in the "digital jaw casts" and target volume data. In the present study, these points were selected because they are located apart, thus being clearly observable in all data sets.

Three operators (A1, A2, and A3) sampled the three-dimensional coordinates of the above landmarks using a Multi Planar Reconstruction (MPR) program. This program can display three orthogonal planes of volume data passing through the position of a mouse pointer and can record the pointer's position. To reduce sampling errors, they picked each point five times for each of the volume data sets. The averaged coordinates were used in the following procedure. In addition, using the same data sets, two operators (B1 and B2) sampled the landmarks only once so that these roughly sampled data could serve in examining the validity of SA optimization.

3.3. Initial Affine Transformation Matrix

The initial affine transformation matrices used for superimposition were computed from the coordinates of the anatomical landmarks. Two matrices were generated for each target volume data: one for the upper jaw, and the other for the lower jaw. Each matrix was calculated as follows. The four anatomical landmarks on the upper or lower "digital jaw cast" were defined as $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$, $P_3(x_3, y_3, z_3)$, and $P_4(x_4, y_4, z_4)$. The corresponding four points on the target volume were defined as $P'_1(x'_1, y'_1, z'_1)$, $P'_2(x'_2, y'_2, z'_2)$, $P'_3(x'_3, y'_3, z'_3)$, and $P'_4(x'_4, y'_4, z'_4)$. Using an affine transformation matrix M, P_1, P_2, P_3, and P_4 were mapped to $P''_1(x''_1, y''_1, z''_1)$, $P''_2(x''_2, y''_2, z''_2)$, $P''_3(x''_3, y''_3, z''_3)$, and $P''_4(x''_4, y''_4, z''_4)$, using the following equation:

$$\begin{pmatrix} x''_1 & y''_1 & z''_1 & 1 \\ x''_2 & y''_2 & z''_2 & 1 \\ x''_3 & y''_3 & z''_3 & 1 \\ x''_4 & y''_4 & z''_4 & 1 \end{pmatrix} = \begin{pmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{pmatrix} M, \quad (1)$$

where M is a 4×4 matrix, including two submatrices R and T. R is a 3×3 matrix for rotation, and T is a 1×3 matrix for translation.

The parameters for rotation and translation in the initial transformation matrix are obtained by minimizing the total distance D calculated for the corresponding points in the matrices:

$$D = \sum_{k=1}^{4} \overline{P'_k P''_k}. \quad (2)$$

3.4. *Error Volume* and Matrix Optimization

The *error volume* was introduced as a parameter to evaluate the accuracy of the superimposition of the "digital jaw casts" on the target volume data. Its value was defined as the number of voxels in the "digital jaw casts" that overlapped the soft tissue of the target volume data, which was represented in cubic centimeters. If the "digital jaw casts" were superimposed onto the target volume data in the appropriate position, most of the voxels would be placed on dark regions in the target volume (i.e., the jaws and teeth), and the *error volume* would be small. On the contrary, if the cast were moved to an inappropriate position, the *error volume* would increase. The *error volume* was calculated using the following algorithm.

The coordinate of each voxel, (x, y, z), of the "digital jaw cast" is mapped to the destination coordinate, (x', y', z'), using the equation:

$$(x', y', z', 1) = (x, y, z, 1)\, M, \quad (3)$$

where M is the affine transformation matrix for the "digital jaw cast" to be superimposed onto the target volume. If the voxel value of the target volume at the coordinate (x', y', z') is greater than the threshold value for extracting the teeth

and jaws from the soft tissue, the voxel is counted as *error volume*. In the present study, the threshold value was set at 90. While the teeth and thick bony regions of bones consisting of compact substance are adequately segmented from soft tissue with the threshold at 40, the thin bony regions, such as dental sockets and the palatal bone, are not because these regions are influenced by the surrounding soft tissues and blueberry juice, which have high pixel values. The optimization of the initial affine transformation matrices to minimize the *error volume* was based on the SA method [19].

The SA method is an efficient algorithm used for optimization in nonlinear and stochastic systems, and it is also known to be resistant to the problem of local minima. In the present study, affine transformation matrices were optimized through an iterative process in which the rotation or translation parameters in the matrix were minutely changed and the change was judged to be accepted or not by referring to the *error volume*. The change was accepted when it successfully decreased the error volume. When the *error volume* was not reduced, it was accepted with a probability given by $\exp(-\Delta E/T)$, where ΔE was the increase of the *error volume* and T was the "temperature" parameter that gradually decreased in the iteration. The change was rejected when a large increase of the *error volume* resulted. The total iteration number was set at 1,000.

Because it takes considerable time to calculate the *error volume* on entire images, the calculation was performed only in the three orthogonal planes passing through the centroid of the "digital jaw cast." The total calculation time was approximately 70 s (FreeBSD operating system, 1-GHz CPU, and 1-GB RAM).

4. RESULTS AND DISCUSSION

4.1. Sampling Error (Standard Deviation) of Anatomical Landmarks

Table 1 shows the sampling errors for the landmark coordinates of the upper and lower jaws sampled by the three operators (A1, A2, and A3) from the volume data for the five vowels and dental scan. These values index the standard deviations of the landmarks in units of $0.5 \times 0.5 \times 0.5$ mm voxels.

4.2. Results of Superimposition Using Initial Affine Transformation Matrix

Table 2 shows the values of *error volume* in cm^3, when the initial affine transformation matrices were used for superimposition. The total volume of the upper jaw was 12.44 cm^3, and that of the lower jaw was 21.53 cm^3. The values show a large deviation, ranging from 1.20 to 6.20 cm^3. Moreover, the values of A1, A2, and A3 were significantly smaller than those of B1 and B2 (t-test: $p < 0.01$): i.e., the superimpositions by A1, A2, and A3 were more accurate than those by B1 and B2.

Figure 3 shows three orthogonal planes (sagittal, coronal, and transverse) of the volume for vowel /u/. The images in Fig. 3(a) are the results of A2 and those in (b) are those of B1. The slice locations are common between Figs. 3(a) and (b). Figure 3 clearly shows that the result of superimposition by A2 is superior to that by B1.

Table 1 Sampling error (standard deviation) of anatomical landmarks. Units are voxels ($0.5 \times 0.5 \times 0.5$ mm).

(a) Upper jaw

Operator	/a/	/i/	/u/	/e/	/o/	Dental scan
A1	0.87	0.48	0.68	0.44	0.35	0.34
A2	0.95	0.75	1.30	0.88	0.83	1.61
A3	1.13	0.99	1.03	1.18	1.13	1.07

(b) Lower jaw

Operator	/a/	/i/	/u/	/e/	/o/	Dental scan
A1	0.50	0.36	0.34	0.56	0.63	0.29
A2	1.36	1.15	1.53	1.40	1.73	1.44
A3	1.16	1.26	1.23	0.72	1.50	1.12

Table 2 *Error volume* before optimization in cm^3. The total volume of the upper jaw is 12.44 cm^3, and that of the lower jaw is 21.53 cm^3. The average of the values in the upper jaw is 2.10 cm^3 and that in the lower jaw is 3.14 cm^3. The total standard deviation of the values in the upper jaw is 1.01, and that in the lower jaw is 1.45.

(a) Upper jaw

Operator	/a/	/i/	/u/	/e/	/o/	Avg.	Std.
A1	1.73	1.20	1.31	1.34	1.47	1.41	0.20
A2	1.67	1.37	1.74	1.42	1.59	1.56	0.16
A3	1.52	1.22	1.46	1.37	1.56	1.43	0.14
B1	3.84	3.54	4.07	3.50	4.00	3.79	0.26
B2	3.66	1.65	2.56	1.33	2.29	2.30	0.90
Avg.	2.48	1.80	2.23	1.79	2.18		
Std.	1.16	0.99	1.14	0.96	1.07		

(b) Lower jaw

Operator	/a/	/i/	/u/	/e/	/o/	Avg.	Std.
A1	2.36	1.29	1.64	1.23	2.38	1.78	0.56
A2	2.49	1.56	1.75	2.85	2.49	2.23	0.55
A3	3.04	1.69	2.35	2.53	3.29	2.58	0.62
B1	3.51	3.77	5.85	3.12	6.20	4.49	1.43
B2	4.90	5.57	5.21	4.13	3.41	4.64	0.87
Avg.	3.26	2.78	3.36	2.77	3.55		
Std.	1.03	1.84	2.01	1.05	1.55		

(a) Results of operator A2

(b) Results of operator B1

Fig. 3 Three orthogonal planes (sagittal, coronal, and transverse) of the volume for vowel /u/, which were superimposed onto the "digital jaw casts" according to the initial transformation matrices.

(a) Results of operator A2

(b) Results of operator B1

Fig. 4 Three orthogonal planes (sagittal, coronal, and transverse) of the volume for vowel /u/, which were superimposed onto the "digital jaw casts" according to the optimized transformation matrices.

Table 3 *Error volume* after optimization in cm^3. The average of the values in the upper jaw is 1.19 cm^3 and that in the lower jaw is 1.56 cm^3. The total standard deviation of the values in the upper jaw is 0.11, and that in the lower jaw is 0.26.

(a) Upper jaw

Operator	/a/	/i/	/u/	/e/	/o/	Avg.	Std.
A1	1.35	1.09	1.22	1.05	1.21	1.18	0.12
A2	1.35	1.03	1.28	1.09	1.18	1.18	0.13
A3	1.23	1.06	1.22	1.15	1.23	1.18	0.07
B1	1.31	1.10	1.35	1.11	1.34	1.24	0.13
B2	1.22	1.01	1.26	1.07	1.24	1.16	0.11
Avg.	1.29	1.06	1.26	1.09	1.24		
Std.	0.06	0.04	0.06	0.04	0.06		

(b) Lower jaw

Operator	/a/	/i/	/u/	/e/	/o/	Avg.	Std.
A1	1.63	1.25	1.64	1.20	1.85	1.51	0.28
A2	1.72	1.30	1.54	1.32	1.85	1.54	0.24
A3	1.88	1.28	1.52	1.28	1.94	1.58	0.32
B1	1.82	1.30	1.51	1.27	1.91	1.56	0.30
B2	1.89	1.46	1.63	1.19	1.88	1.61	0.29
Avg.	1.79	1.32	1.57	1.25	1.89		
Std.	0.11	0.08	0.06	0.05	0.04		

4.3. Results of Superimposition Using Optimized Affine Transformation Matrix

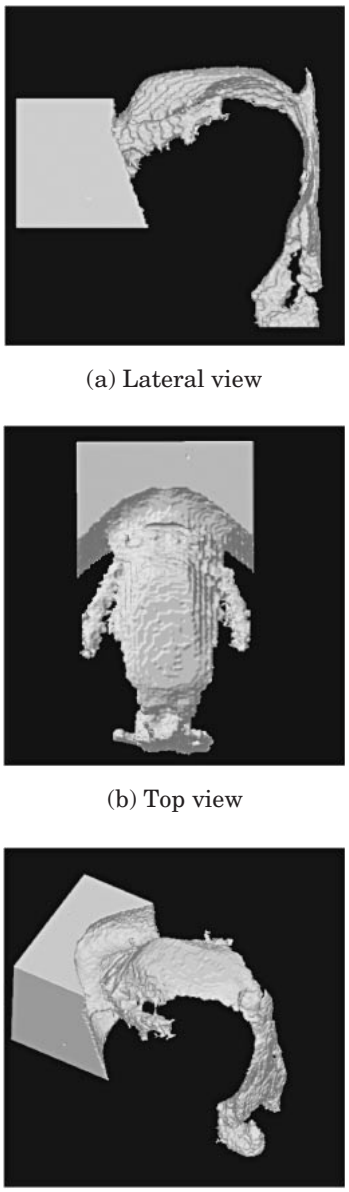
Table 3 shows the values of *error volume* in cm^3, when optimized affine transformation matrices were used for superimposition. The values for the same data fall into a similar small range across the five operators, in contrast to Table 2. Furthermore, the standard deviations of the values are significantly reduced from 1.01 (before optimization) to 0.11 (after optimization) in the upper jaw and from 1.45 to 0.26 in the lower jaw because the optimization process significantly reduced the difference in the non-optimized values of the *error volume* between the operators of A and those of B. This result indicates that the values of *error volume* in all of the data sets were successfully converged into a similar value by applying the SA optimization, regardless of the initial values or operators.

The values of *error volume* for the back vowels (/a/, /u/, and /o/) are significantly larger than those for the front vowels (/i/ and /e/) in both the upper and lower jaws (*t*-test: $p < 0.01$). This seems to be due to the motion artifacts in the image, which differed between the front and back vowels. In this subject, the speech organs tended to be more stable in front vowels than in back vowels during image acquisition.

Figure 4 shows three orthogonal planes of the volume for the vowel /u/. The slice locations are the same as in Fig. 3. Figure 4(a) shows the results of A2, while (b) shows those of B1. The difference between the images of (a) and those of (b) is very small in contrast to Fig. 3, suggesting that the optimization is successful.

4.4. Extracted 3D Vocal Tract Shape

Figure 5 shows lateral, top and oblique views of the three-dimensional vocal tract shape extracted from MRI volume data based on the proposed method. The dental images were superimposed onto the volume data for the vowel /a/ using the optimized transformation matrices. The vocal tract was extracted by a region-growing



(a) Lateral view

(b) Top view

(C) Oblique view

Fig. 5 Visualization of the 3D vocal tract shape extracted from the volume data for vowel /a/, which was superimposed onto the "digital jaw casts."

algorithm from a seed point set in the oral cavity, then visualized using a ray-tracing algorithm. This example shows that the teeth compose a part of the vocal tract wall that continues to the lip wall. It also shows that the space between the upper and lower teeth (interdental space) is included in the vocal tract. Since these walls move with speech articulation and form a part of vocal tract sidebranches, dental image superimposition onto MRI volume data is always a necessary procedure for MRI-based studies of the vocal tract.

5. CONCLUSION

A method was proposed to combine 3D MRI data with dental images for vocal tract visualization. Optimization of affine transformation matrices successfully reduces possible error in landmark sampling and ensures a certain level of accuracy regardless of the operator's skill. The evaluation result confirmed that this method provides an accurate and automatic procedure to solve the MRI-specific problem of the lack of dental images.

ACKNOWLEDGEMENT

This research was conducted as part of 'Research on Human Communication' with funding from the Telecommunications Advancement Organization of Japan.

REFERENCES

[1] S. Adachi and M. Yamada, "An acoustical study of sound production in biphonic singing, Xöömij," *J. Acoust. Soc. Am.*, **105**, 2920–2932 (1999).

[2] J. Dang and K. Honda, "Acoustic characteristics of the piriform fossa in models and humans," *J. Acoust. Soc. Am.*, **101**, 456–465 (1997).

[3] B. H. Story, I. R. Titze and E. A. Hoffman, "Vocal tract area functions for an adult female speaker based on volumetric imaging," *J. Acoust. Soc. Am.*, **104**, 471–487 (1998).

[4] B. H. Story, I. R. Titze and E. A. Hoffman, "The relationship of vocal tract shape to three voice qualities," *J. Acoust. Soc. Am.*, **109**, 1651–1667 (2001).

[5] S. Masaki, M. K. Tiede, K. Honda, Y. Shimada, I. Fujimoto, Y. Nakamura and N. Ninomiya, "MRI-based speech production study using a synchronized sampling method," *J. Acoust. Soc. Jpn.* (*E*), **20**, 375–379 (1999).

[6] C. Yang, H. Kasuya, S. Kanou and S. Satou, "An accurate method to measure the shape and length of the vocal tract for the five Japanese vowels by MRI," *Jpn. J. Logop. Phoniatr.*, **35**, 317–321 (1994).

[7] O. Engwall, "Using linguopalatal contact patterns to tune a 3D tongue model," *Proc. Eurospeech 2001*, Vol. 2, pp. 1475–1478 (2001).

[8] S. S. Narayanan, A. A. Alwan and K. Haker, "An articulatory study of fricative consonants using magnetic resonance imaging," *J. Acoust. Soc. Am.*, **98**, 1325–1347 (1995).

[9] S. S. Narayanan, A. A. Alwan and K. Haker, "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals," *J. Acoust. Soc. Am.*, **101**, 1064–1077 (1997).

[10] A. Alwan, S. Narayanan and K. Haker, "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics," *J. Acoust. Soc. Am.*, **101**, 1078–1089 (1997).

[11] C. Y. Espy-Wilson, S. E. Boyce, M. Jackson, S. Narayanan and A. Alwan, "Acoustic modeling of American English /r/," *J. Acoust. Soc. Am.*, **108**, 343–356 (2000).

[12] B. H. Story, I. R. Titze and E. A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.*, **100**, 537–554 (1996).

[13] K. Tom, I. R. Titze, E. A. Hoffman and B. H. Story, "Three-dimensional vocal tract imaging and formant structure: Varying vocal register, pitch, and loudness," *J. Acoust. Soc. Am.*, **109**, 742–747 (2001).

[14] M. Matsumura, T. Niikawa, Y. Matsushige, K. Shimizu, Y. Hashimoto and T. Morita, "Measurement of 3D shapes of vocal tract, dental crown, and nasal cavity using MRI," *Tech. Rep. IEICE*, MBE93-131, pp. 41–48 (1994).

[15] T. Niikawa, M. Matsumura, K. Shimizu, Y. Hashimoto, T.

Tachimura and T. Wada, "Accurate measurement of three-dimensional shapes of vocal tract using magnetic resonance imaging: Japanese fricative consonants," *Tech. Rep. IEICE*, MBE97-86, pp. 1–7 (1997).

[16] T. Niikawa, M. Matsumura, A. Okuno, T. Tachimura and T. Wada, "Estimation of sound pressure distribution and transfer characteristics of three-dimensional vocal tract measured by MRI," *Tech. Rep. IEICE*, SP98-8, pp. 1–4 (1998).

[17] M. Wakumoto, S. Masaki, J. Dang, K. Honda, Y. Shimada, I. Fujimoto and Y. Nakamura, "Visualization of dental crown shape in an MRI-based speech production study," *Int. J. Oral Maxillofac. Surg.*, **26**, 189–190 (1997).

[18] K. Hiraishi, I. Narabayashi, O. Fujita, K. Yamamoto, A. Sagami, Y. Hisada, Y. Saika, I. Adachi and H. Hasegawa, "Blueberry juice: preliminary evaluation as an oral contrast agent in gastrointestinal MR imaging," *Radiology*, **194**, 119–123 (1995).

[19] S. Kirkpatrick, C. D. Gelatt Jr. and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, **220**, 671–680 (1983).

[20] K. Sasakawa, S. Kuroda and S. Ikebata, "A threshold selection method based on the mean of adjacent number in binary images," *Tech. Rep. IEICE*, PRU89-25, pp. 49–56 (1989).