Journal of Sound and Vibration (1997) **203**(4), 581–596

VOCAL TRACT SHAPE IDENTIFICATION FROM FORMANT FREQUENCY SPECTRA— A SIMULATION USING THREE-DIMENSIONAL BOUNDARY ELEMENT MODELS

Y. Kagawa, Y. Ohtani† and R. Shimoyama‡

Department of Electrical and Electronic Engineering, Okayama University,
Okayama 700, *Japan*

(*Received* 25 *July* 1994, *and in final form* 29 *November* 1996)

Identification of the three-dimensional vocal tract shape from the formant frequency spectra or transmission characteristic is discussed. The vocal tract wall is modelled by an assemblage of spline functions, which is deformable about the points of interest. Three-dimensional boundary elements are incorporated for evaluating the acoustic transmission characteristics. The identification is treated as an optimization process in such a way that the norm between some "measured" formant frequencies and those calculated for the assumed vocal tract shape is minimized by using the DFP algorithm.

1. INTRODUCTION

The vocal tract plays an important role in speech sound formation. The vocal tract can be considered to be an acoustic filter that discriminates among the frequency spectra of sounds produced by vocal cord vibration and transmits the spectra properly chosen for a particular vowel. The transmission characteristics depend on the the geometrical shape, for which the cross-sectional area distribution along the vocal tract is believed to be responsible. The investigation of the vocal tract as an acoustic transmission system is thus essential for speech analysis, synthesis and identification. To simulate the speech sound formation process in the vocal tract, realistic three-dimensional models are required. Three-dimensional vocal tract models have been proposed by using finite and boundary elements and their capability for evaluating their transmission characteristics have been discussed [1, 2]. The vocal tract is basically a hollow bent tube whose wall is made of muscles and whose shape is determined as a result of the movement of jaws, tongue and mouth together with partial tension in the muscles. The present paper presents an attempt to determine the three-dimensional vocal tract shape from the "measured" formant frequencies. There have been many investigations of this kind, but most of them are concerned about the determination of the cross-sectional area distribution along the vocal tract for which one-dimensional acoustic transmission models are entitled [3–6]. However, plane wave assumption corresponding to this one-dimensional modelling cannot be justified in the higher frequency range, as the measurement of the sound pressure distribution in the cast replica of the oral cavity shows that a plane wave cannot present in the oral cavity in higher frequency range [7]. The geometrical shape is therefore one of

† At present, with Sharp Co. Ltd., Nara 632, Japan.
‡ At present, with Tsuyama Technical College, Okayama 708, Japan.

0022–460X/97/240581 + 16 $25.00/0/sv960865

the important factors which determine the details of the acoustical characteristics of the vocal tract. The drawback of the one-dimensional models also lies in the fact that the cross-sectional area distribution can easily be obtained from the vocal tract shape given while, on the contrary, the corresponding three-dimensional geometrical shape cannot be recovered from the cross-sectional area distribution determined.

In the present paper, a deformable vocal tract expression is first discussed using a set of spline functions, for which the acoustic transmission field is modelled by the boundary elements as in the authors' previous work. Then one proceeds to the determination of the vocal tract shape from the formant frequencies. The inverse approach is here considered as an optimization process in which the norm between the ''measured'' formant frequencies and the calculated ones for an assumed vocal tract shape is minimized by using the Davidon–Fletcher–Powell (DFP) algorithm.

2. VOCAL TRACT MODELS AND THEIR ACOUSTIC CHARACTERISTICS

The vocal tract is considered as an acoustic filter and its transmission characteristics in steady state are numerically evaluated by using boundary element models [2]. The acoustic transmission system is illustrated as shown in Figure 1. The governing equation and the boundary conditions are

$$\nabla^2 p + k^2 p = 0 \qquad \text{in } \Omega, \tag{1}$$

$$\partial p/\partial n = -\mathrm{j}\omega\rho\hat{v}_g \qquad \text{on } \Gamma_1, \tag{2}$$

(for constant velocity excitation at the glottis)

$$\partial p/\partial n = -\mathrm{j}\omega\rho p/z_w \qquad \text{on } \Gamma_2, \tag{3}$$

$$\partial p/\partial n = -\mathrm{j}\omega\rho p/z_r \qquad \text{on } \Gamma_3. \tag{4}$$

Here p is the sound pressure, ρ the air density, k the wave number ($k = \omega/c$, ω being the angular frequency and c the sound speed), j the imaginary unit, $\partial/\partial n$ the normal derivative to the boundary, z_w the wall impedance, z_r the radiation impedance at the mouth and v_g the particle velocity at the glottis for excitation. ($\hat{\ }$) indicates the value prescribed. The mouth opening for acoustic radiation is terminated by the impedance equivalent to that of a rigid circular piston. The wall surface is divided into triangular surface patches over which linear interpolation functions are assumed for both sound pressure and particle velocity. The boundary element approach leads to discretized linear algebraic equations with respect to the nodal pressures and velocities of the form

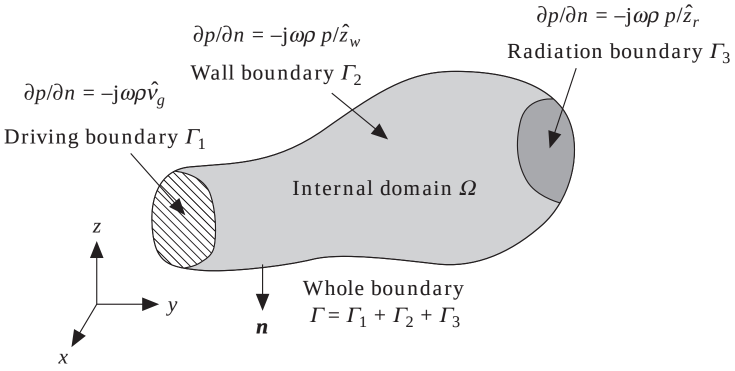
$\partial p/\partial n = -\mathrm{j}\omega\rho\, p/\hat{z}_r$
Radiation boundary Γ_3
$\partial p/\partial n = -\mathrm{j}\omega\rho\, p/\hat{z}_w$
Wall boundary Γ_2
$\partial p/\partial n = -\mathrm{j}\omega\rho\hat{v}_g$
Driving boundary Γ_1
Internal domain Ω
z
y
x
n
Whole boundary
$\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_3$

Figure 1. Acoustic transmission system in vocal tract.

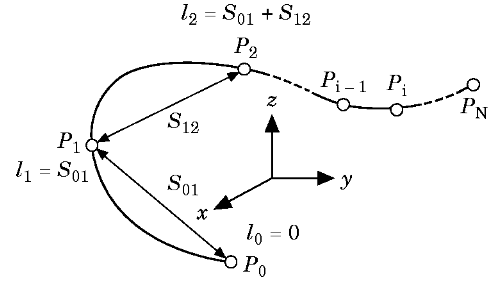

Figure 2. Nodes interpolated by spline functions in three-dimensional space.

$$[\mathbf{H}]\{\mathbf{p}\} = \mathrm{j}\omega\rho[\mathbf{G}]\{\mathbf{v}\}, \qquad \text{or} \qquad [\mathbf{Y}]\{\mathbf{p}\}=\{\mathbf{v}\}, \tag{5}$$

where $\{\mathbf{p}\}$ and $\{\mathbf{v}\}$ are the sound pressure and particle velocity vectors defined at the element nodes, and $[\mathbf{Y}] = -(1/\mathrm{j}\omega\rho)[\mathbf{G}]^{-1}[\mathbf{H}]$ is the acoustic admittance matrix, which includes the wall and radiation admittances. The transmission characteristics or transfer impedance is evaluated for the pressure at the center of the mouth opening against the constant velocity excitation over the glottis.

3. VOCAL TRACT EXPRESSION WITH A SET OF SPLINE FUNCTIONS

3.1. NODE CONNECTION WITH SPLINE FUNCTIONS

The vocal tract is a bent hollow tube with a variable cross-section. To express the tube by a smooth curved surface, nodes are taken around the circumference and toward the longitudinal direction, and are interpolated by piece-wise third-order spline functions for both directions [8]. This is not directly achieved but through transformation. The procedure is oulined as follows. As shown in Figure 2, one considers a set of nodes $P_i (i = 0, 1, 2, \dots, N)$ in the three-dimensional space (x, y, z). One first defines the total length of the base line as

$$l_i = \sum_{j=1}^{i} s_{j-1,j} \qquad (l_0 = 0), \tag{6}$$

where $s_{j-1,j} = |P_{j-1} - P_j|$ which depends on the nodal positions. The set of nodes in the x-, y-, z-co-ordinates are projected onto the l-x, l-y and l-z planes respectively. Spline interpolation is now applied to connect the nodes projected over these new planes. Figure 3 illustrates the nodes (l_i, x_i) $(i = 0, 1, 2, \dots, N)$ projected over the l-x plane. Here one then applies the third order splines to interpolate the nodes smoothly, which implies that

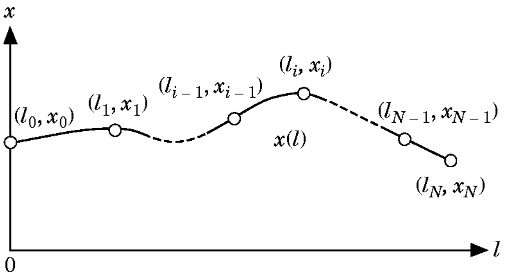

Figure 3. Projected expression to a l-x plane.

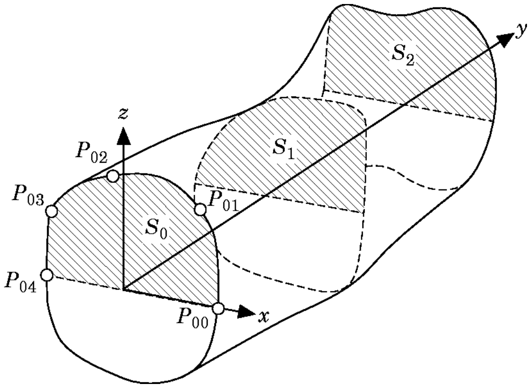

Figure 4. A tube as a succession of planes (symmetrical with respect to the x-y plane).

the function must be continuous to the second derivative at each node. The second derivative of the function $x(l)$ can be linear for the section $[l_{i-1}, l_i]$ as

$$x''(l) = M_{i-1}\frac{l_i - l}{h_i} + M_i\frac{l - l_{i-1}}{h_i}, \qquad i = 0, 1, 2, \ldots, N, \tag{7}$$

where $h_i = l_i - l_{i-1}$, and M_i is the second derivative of $x(l)$ at point P_i. Integrating twice with respect to l and noticing $x(l_i) = x_i$ and $x(l_{i-1}) = x_{i-1}$, one has

$$x(l) = M_{i-1}\frac{(l_i - l)^3}{6h_i} + M_i\frac{(l - l_i)^3}{6h_i} + \left(x_{i-1} - \frac{M_{i-1}h_i^2}{6}\right)\frac{l_i - l}{h_i} + \left(x_i - \frac{M_i h_i^2}{6}\right)\frac{l - l_{i-1}}{h_i}, \tag{8}$$

which forms $(N-1)$th order simultaneous equations with respect to M_i with $N+1$ unknowns ($i = 0, 1, 2, \ldots, N$). For the solution, both end conditions M_0 and M_N at $l = 0$ and l_N must be given, which are practically chosen to be zero. The third order spline functions are thus determined for arbitrary l_i. The same procedure is applied to a set of nodes projected to the l-y and l-z planes. Co-ordinate values at any position, not only the nodal position l_i but also the interpolated position l, are now obtained in the Cartesian co-ordinate space. Taking the nodes properly distributed over a tube surface, one can express a vocal tract of arbitrary shape. Figure 4 indicates the distribution of nodal positions P_{ji}. Subscript j corresponds to the cross-sectional planes S_j and subscript i indicates the order of the nodes to be connected between planes. The data are given for the nodes on successive planes S_j. Interpolation is first taken on planes S_j for the given data point P_{ji} (the figure illustrates the case $i = 0$–4) and the procedure is repeated for other planes. Interpolation is then made for the longitudinal direction connecting the nodes with common subscript number i. These nodes are again interpolated by spline functions as illustrated in Figure 5. The surface is thus complete, and the vocal tract of arbitrary shape is now expressed in terms of a set of spline functions. If necessary, new nodes can be created on the spline curves and again interpolated. Boundary elements of reasonable size are easily created by connecting the adjacent nodes over the surface. We employ simple triangular plain surface patches for present boundary element models, for which the acoustic transmission characteristics are calculated [2].

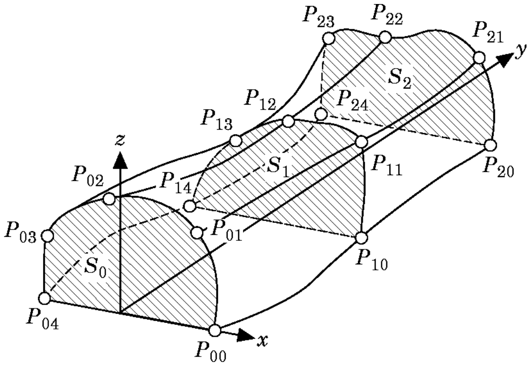

Figure 5. Connection of the planes (upper half).

3.2. VOCAL TRACT SHAPE EXPRESSED BY SPLINE FUNCTIONS

Figure 6 shows an example of the "observed" vocal tract profile in its middle plane for the Japanese vowel a: [9]. The depictions are based on X-ray tomographic pictures and its cross-sectional shapes cut at multiple planes and the corresponding ones modelled by using the present procedure are given in Figure 7. The column (b) of Figure 7 indicates the cross-section cut at the plane in the middle between planes 1′ and 2′, which looks reasonable. The numbers associated correspond to the planes shown in Figure 6. The central dotted curve in Figure 6 is drawn through the center of gravity taken at each cross-sectional plane. These planes are perpendicular to the center of the gravity line while the planes with primed numbers are inclined to those planes. The angle of each plane is given at the foot of the cross-sectional shapes in Figure 7. They require 6 to 25 input data points. The planes 10 and 11 are not given in reference [9], but are created by referring to other sources. Figure 8 shows the cross-sectional shape 3′ which is interpolated with 20 input data points given. Smooth and reasonable interpolation is possible with a relatively small amount of data except for sharp corners. The reproduction of the sharp corners may not be so essential acoustically so that fewer data points suffice in practice. Figure 9 is an example of the vocal tract produced with spline functions interpolating properly distributed data points. The surface is divided into triangular boundary element patches which are created by connecting the nodes generated over spline curves. This example has 40 divisions in the longitudinal direction and 12 divisions in the circumference, which results in 1224 elements and 665 nodes. The vocal tract ends with the mouth from

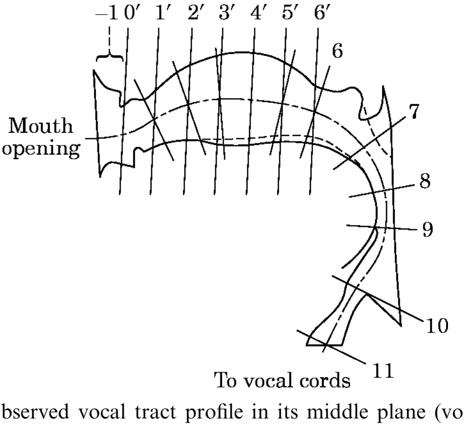

Figure 6. Observed vocal tract profile in its middle plane (vowel a:) [9].

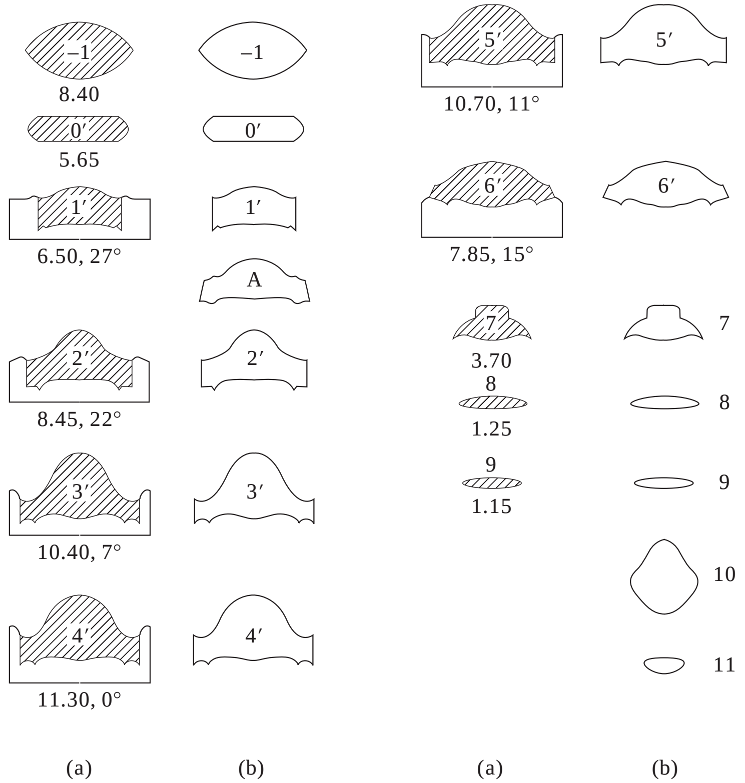

Figure 7. Cross-sectional shapes and their simulated counterparts (numbers in the figures correspond to the cross-sectional planes in Figure 6). (a) Observed [9]; (b) simulated with spline functions. Cross-sectional areas are given in cm^2.

which acoustic radiation takes place. With the present model, it is considered as a closed acoustic tube with a proper radiation impedance termination, for which the end surface is also divided into boundary elements. The termination is made with the acoustic impedance equivalent to the radiation impedance of a circular rigid piston of the same area in an infinite baffle, which is given as $z_r = \rho c\{(ka)^2/2 + j8ka/3\pi\}$ where a is the radius equivalent to the mouth opening. The vocal tract wall is made of muscle and assumed to provide the impedance $z_w = (14000 + j16\omega)$ (kg/m^2 s) [10]. Figure 10 compares the cross-sectional area distribution obtained from the ''observed'' data with the one based

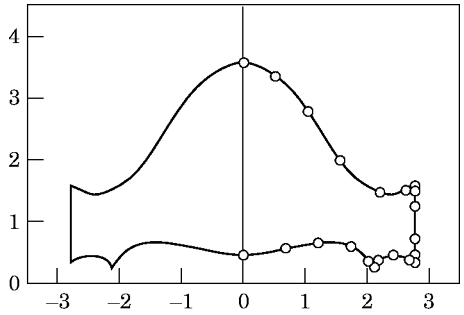

Figure 8. Input data points (○) and shape interpolated with spline functions (——) (plane 3^1). Scale values in cm.

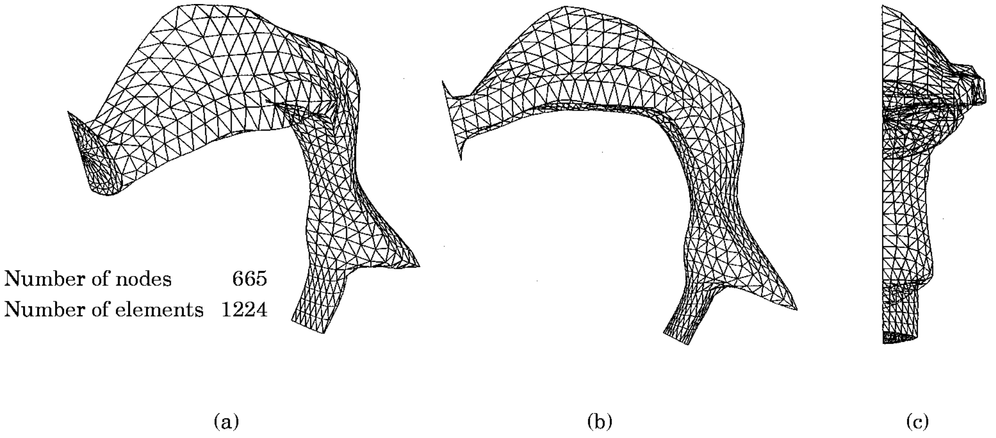

Figure 9. Vocal tract reproduced with spline functions and triangular boundary element patches (vowel a:) (not all shadowed lines are removed). (a) Bird's-eye review; (b) side view; (c) frontal view.

on the present spline model. The agreement is satisfactory though its accuracy depends on the number of data points given. The result shows that the present approach provides reasonable reproduction capability. It requires only a reasonable number of data points for the cross-section and the center of gravity curve to create the whole vocal tract. Figure 11 indicates the calculated transfer impedance of the present model, that is the sound pressure evaluated at the center of the mouth to the constant velocity excitation over the glottis, in which four resonances correspond to the formant frequencies.

3.3. DEFORMABLE VOCAL TRACT MODEL

The shape of the vocal tract depends on the position of the jaws and tongue, and the shape of the mouth. Here one first assumes a basic shape or a shape in its mean position, from which a deformation is made to create the shape corresponding to a certain vowel. One would like to realize this by shifting the wall at the least possible number of points. Figure 12 indicates the case when the shift of displacement d_{pq} is made at point (x_p, y_q), with which its vicinity also deforms as much as $d_{pq}\omega_{xi}$, where ω_{xi} is a distribution function arbitrarily chosen. One possible answer for proper distribution could be the solution of Poisson's equation for an elastic membrane with the shape of the vocal tract to which a concentrated force is applied. The vocal tract muscle is not a simple thin membrane so that we here employ the rather empirical approach of choosing a Gaussian distribution

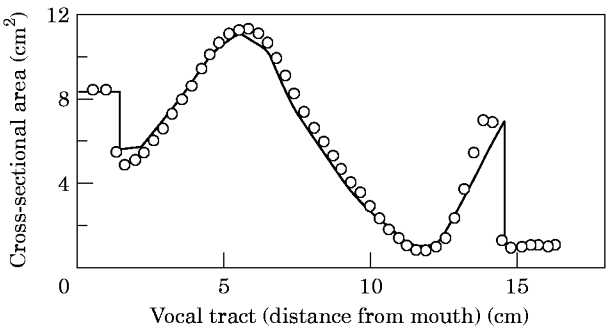

Figure 10. Cross-sectional area distribution (vowel a:). ——, Based on present spline method; –○–, based on "observed" data.

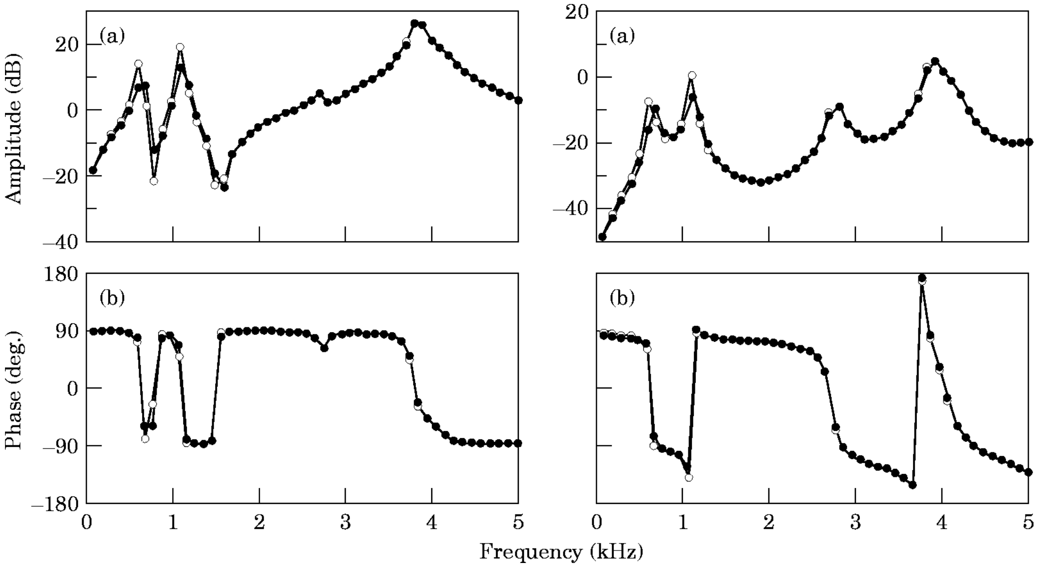

Figure 11. Input (left) and transfer (right) impedances normalized with respect to *ρc* (vowel a:). (a) Amplitude; (b) phase, –○–, rigid wall; –●–, muscle wall.

$$\omega_{xi} = \exp[-\{A_x(l_i - l_p)\}^2], \quad i = 0, 1, 2, \ldots, N, \tag{9}$$

for the x direction, where l_i is defined in equation (6) for $s_{k-1,k} = |x_{k-1} - x_k| \cdot A_x$ gives a factor indicating how far the shifting extends. ω_{xi} is of unit value for $i = p$. This is illustrated in Figure 12. Similarly, in the y direction, one has

$$\omega_{yj} = \exp[-\{A_y(l_j - l_q)\}^2], \quad i = 0, 1, 2, \ldots, N. \tag{10}$$

Therefore the displacement distribution d_{ij} for the surface about the center (x_i, y_j) where the shifting is made is

$$d_{ij} = d_{qp}\omega_{xi}\omega_{yi}. \tag{11}$$

Extension to the curved surface is straightforward, for which the definition is applied to the orthogonal curved lines taken over the curved surface. Multiple shifting can also be made by taking a linear combination of each independent shift. The place and the number of the positions when the shift is taken and other factors should properly be chosen to express the vocal tract shape of interest. Figure 13 shows an example when a part of the wall is shifted. In Figure 13(a), the original vocal tract to which displacement is applied

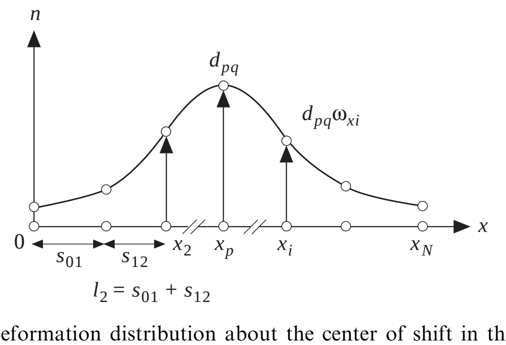

Figure 12. Deformation distribution about the center of shift in the *x*-*n* plane.

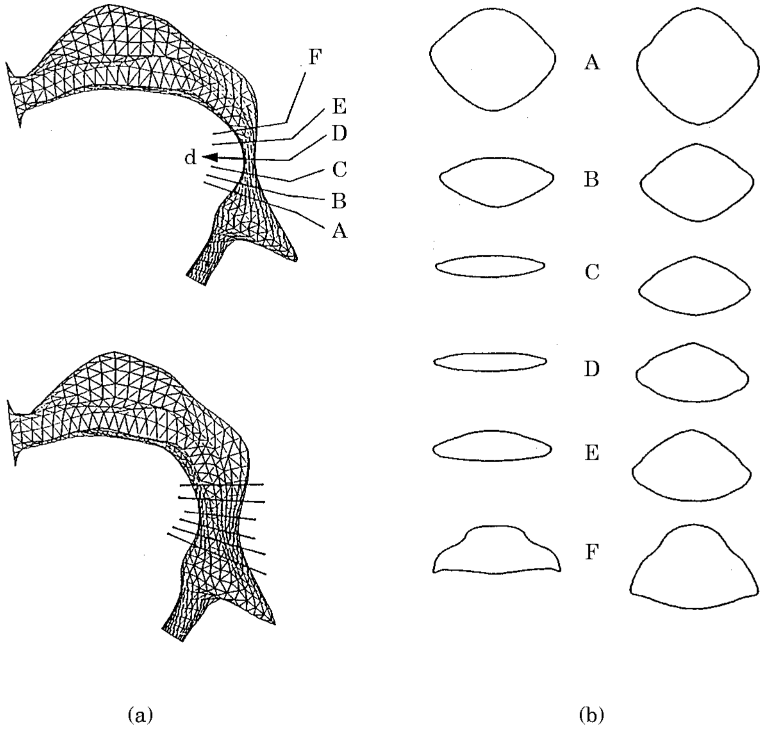

Figure 13. Deformed example. (a) Side review; (b) cross-section, left before deformation, right after deformation.

as indicated and the deformed shape are shown. In Figure 13(b), the cross-sectional shapes cut at various planes are shown. In the present case the nodes after deformation are interpolated with splines and triangular element division is again made.

3.4. VOCAL TRACT SHAPE IDENTIFICATION

Vocal tract shape identification from the observed formant frequencies is a typical inverse problem [11]. As no direct inversion is possible, an optimization technique must be employed, for which a step-by-step correction approach is used. The transmission characteristics are first calculated for the vocal tract of an assumed shape and its

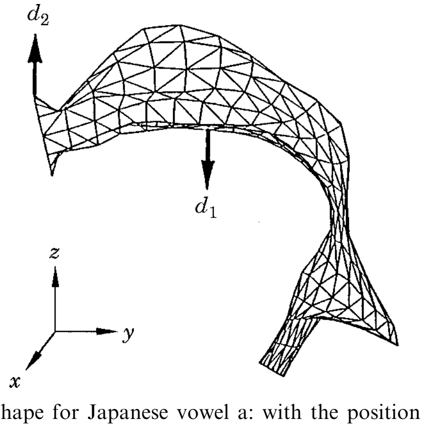

Figure 14. Vocal tract shape for Japanese vowel a: with the position and directions altered.

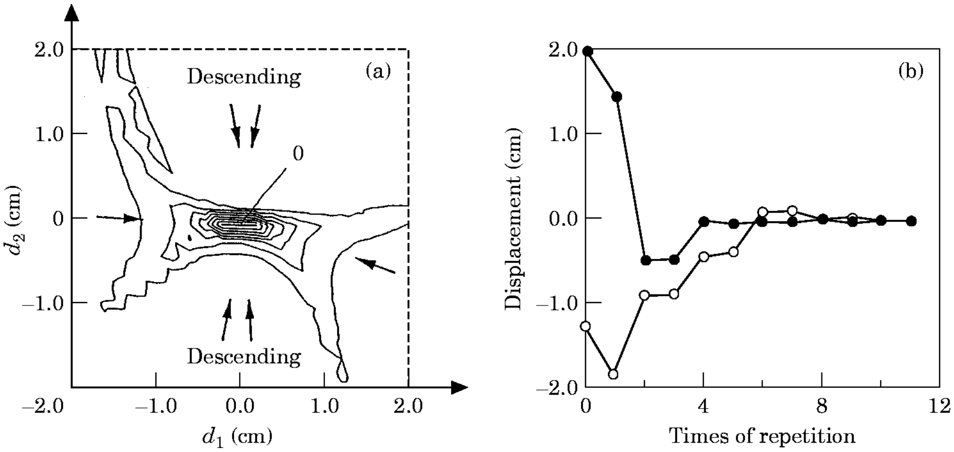

Figure 15. Convergence characteristic. (a) Distribution of the objective function (equi-amplitude lines) with respect to the displacements d_1 and d_2; (b) displacement convergence to the original, –○–, d_1 (jaw and tongue), –●–, d_2 (mouth and upper lip).

formant frequency distribution is compared with the "observed" one. The wall surface is then shifted until the calculated frequency distribution meets the observed case. One needs a measure for correction as this is considered as an optimization problem, for which one utilizes the DFP method [12].

3.4.1. *Davidon–Fletcher–Powell method*

The Davidon–Fletcher–Powell method is a modification of the Newton method, which is well-known for obtaining approximate solutions of non-linear problems progressively. Here the approach is outlined for convenience in the following. Consider the problem of minimizing a function $W(\mathbf{d})$ for the variation of the independent variable $\mathbf{d} = \{d_1, d_2, \ldots, d_n\}$, and assume that $W(\mathbf{d})$ can be expanded in terms of the quadrature of $\mathbf{d}$ in the vicinity of its minimum. The variable that minimizes $W(\mathbf{d})$ is given as

$$\mathbf{d}_{\min}^{(k+1)} = \mathbf{d}^{(k)} - (\mathbf{B}^{(k)})^{-1}\mathbf{b}^{(k)}, \tag{12}$$

where the components of $\mathbf{B}^{(k)}$, $\mathbf{b}^{(k)}$ are

$$\mathbf{B}_{ij}^{(k)} = \frac{\partial^2}{\partial d_i^{(k)}\,\partial d_j^{(k)}} W(\mathbf{d}^{(k)}), \quad i = 0, 1, 2, \ldots, N, \quad \mathrm{j} = 1, 2, \ldots, N, \tag{13}$$

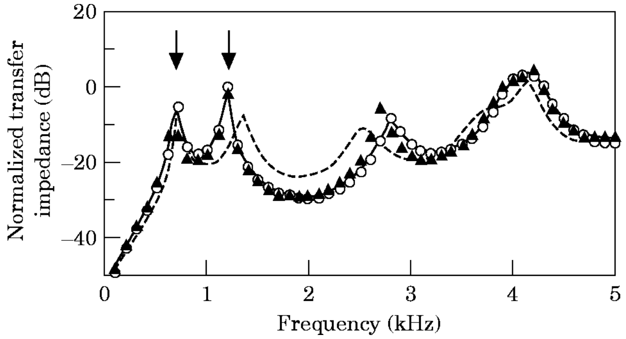

Figure 16. Variation of transmission characteristic (vowel a:). - - -, Initially deformed ($d_1 = -1{\cdot}3$ cm, $d_2 = -2{\cdot}0$ cm); ——, objective (original); –○–, converged; –▲–, after three repetitions.

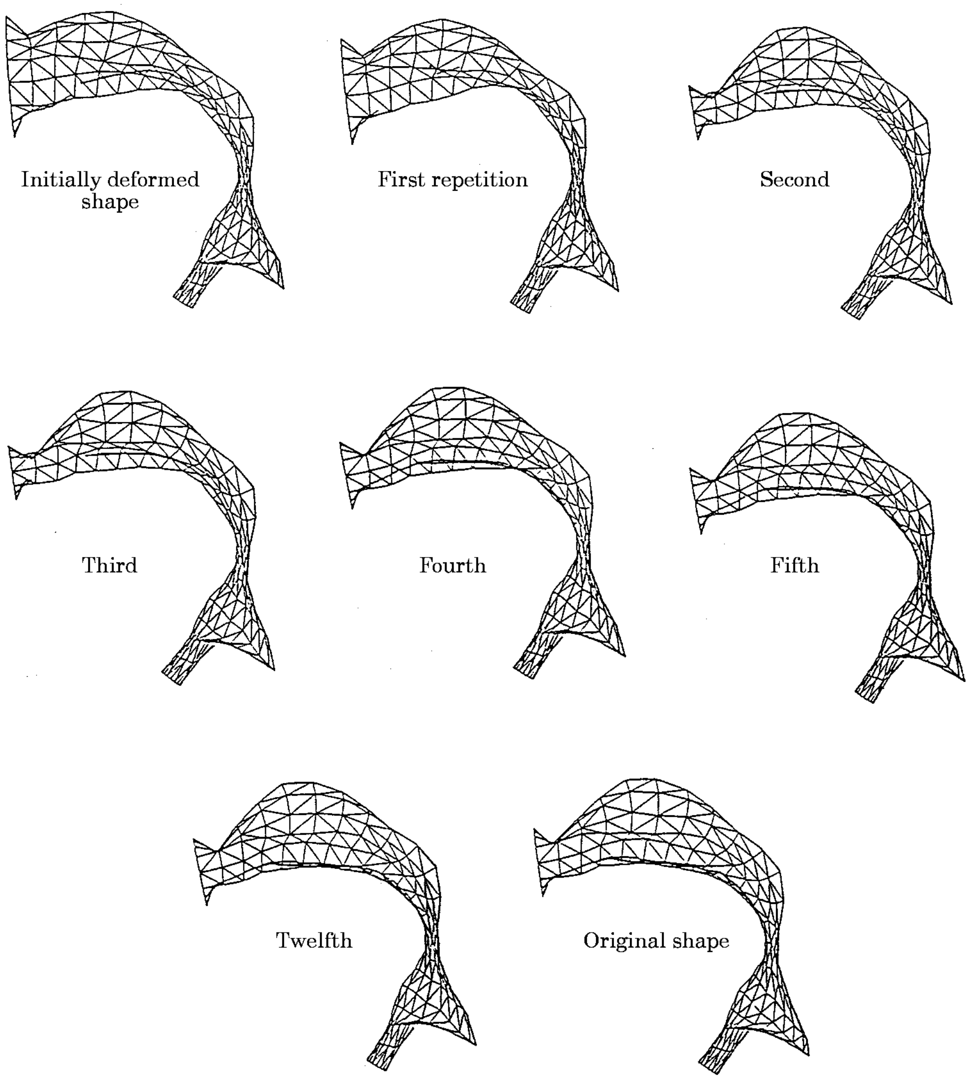

Figure 17. Shape transformation process, from the deformed to the original shape.

and

$$\mathbf{b}_i^{(k)} = (\partial/\partial d_i^{(k)})\, W(\mathbf{d}^{(k)}), \quad i = 0, 1, 2, \dots, N, \tag{14}$$

where $(\mathbf{d})^{(k)}$ is an assumed initial value or the value at step k. This approach is known as the Newton method. Finding $\mathbf{d}_{\min}^{(k+1)}$ in a single step is possible only when the initial value $\mathbf{d}^{(k)}$ is properly chosen very close to $\mathbf{d}_{\min}^{(k+1)}$, and repetition is generally required for convergence. Inversion of the Hessian matrix $(\mathbf{B}^{(k)})^{-1}$ must be evaluated at each step. The solution procedure is not always practical because the evaluation must be made at each operation. The DFP method avoids the direct evaluation of $(\mathbf{B}^{(k)})^{-1}$ and instead evaluates the matrix $\mathbf{S}^{(k)}$ which is a matrix that converges to $(\mathbf{B}^{(k)})^{-1}$ by choosing the gradient direction for successive evaluation of the $(\mathbf{B}^{(k)})^{-1}$ value.

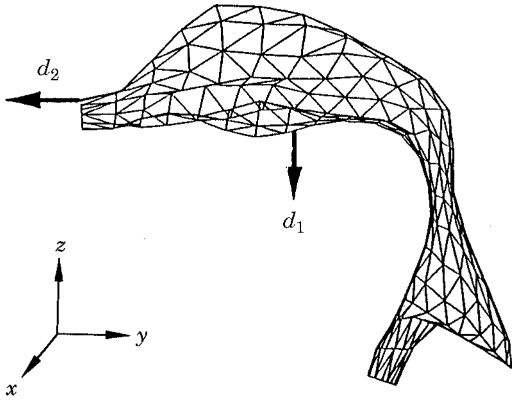

Figure 18. Vocal tract shape for Japanese vowel o: with the positions and directions to be shifted.

The procedure is as follows.

Step 1. Find to give minimum $W(\mathbf{d}^{(k)} - \lambda^{(k)}\mathbf{S}^{(k)}\mathbf{b}^{(k)})$ by the linear searching method in the direction of $-\mathbf{S}^{(k)}\,\mathbf{b}^{(k)}$. Upon introducing the coefficient $\lambda^{(k)}$, the displacement to be shifted in the next step will be

$$\mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} - \lambda^{(k)}\mathbf{S}^{(k)}\mathbf{b}^{(k)}. \tag{15}$$

Step 2. The components of $\mathbf{b}^{(k+1)}$ and $\mathbf{S}^{(k+1)}$ are evaluated as

$$b_i^{(k+1)} = \frac{\partial}{\partial d_i^{(k+1)}}\, W(\mathbf{d}^{(k+1)})$$

and

$$\mathbf{S}^{(k+1)} = \mathbf{S}^{(k)} + \frac{\Delta\mathbf{d}^{(k)}(\Delta\mathbf{d}^{(k)})^{\mathrm{T}}}{(\Delta\mathbf{d}^{(k)})^{\mathrm{T}}\Delta\mathbf{b}^{(k)}} - \frac{\mathbf{S}^{(k)}\Delta\mathbf{b}^{(k)}(\Delta\mathbf{b}^{(k)})^{\mathrm{T}}\mathbf{S}^{(k)}}{(\Delta\mathbf{b}^{(k)})^{\mathrm{T}}\mathbf{S}^{(k)}\Delta\mathbf{b}^{(k)}}, \tag{16}$$

where

$$\Delta\mathbf{b}^{(k)} = \mathbf{b}^{(k+1)} - \mathbf{b}^{(k)}, \qquad \Delta\mathbf{d}^{(k)} = \mathbf{d}^{(k+1)} - \mathbf{d}^{(k)} = -\lambda^{(k)}\mathbf{S}^{(k)}\mathbf{b}^{(k)}. \tag{17, 18}$$

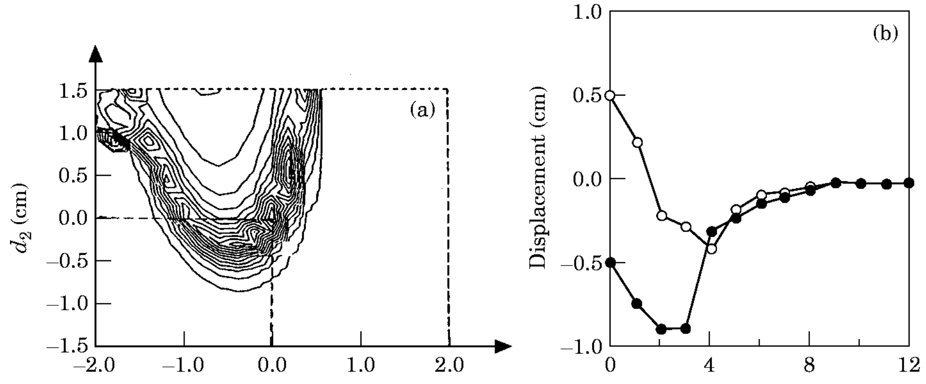

Figure 19. Convergence characteristics. (a) Distribution of the objective function with respect to the displacements d_1 and d_2; (b) displacement convergence to the original –○–, d_1 (jaw and tongue), –●–, d_2 (mouth and upper lip).

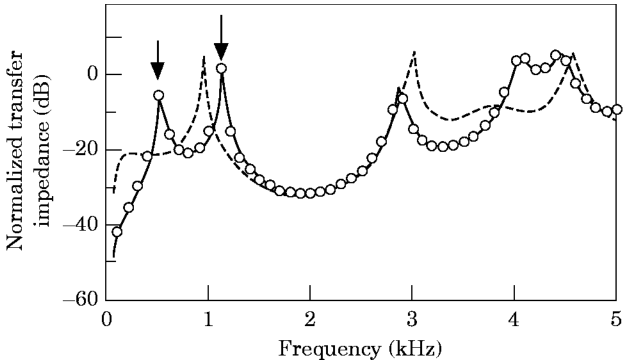

Figure 20. Variation of transmission characteristic (vowel o:). - - -, Initially deformed ($d_1 = 0{\cdot}5$ cm, $d_2 = -0{\cdot}5$ cm); ——, objective (original); –○–, converged.

Step 3. Go back to step 1.

Upon repeating step 1 and step 2, $\mathbf{S}^{(k+1)}$ arrives at $(\mathbf{B}^{(k+1)})^{-1}$ and $\mathbf{d}^{(k+1)}$ goes to $\mathbf{d}_{\min}$. For the initial value of $\mathbf{S}^{(0)}$ the unit matrix is chosen.

3.4.2. *Objective function*

The frequency spectra observed at the mouth for a particular vowel are the spectra that are produced at the glottis as the result of vocal cord vibration, weighted by the transmission frequency characteristics of the vocal tract. For simplicity, in the present simulation one takes the transmission frequency characteristic instead of the formant frequency spectra as they are equivalent when the spectra driven at the glottis are known. Thus the norm between the transfer impedance evaluated at the resonances for assumed vocal tact shape and the "measured" one is chosen as the objective function W, which is to be minimized. The assumed shape that minimizes the objective function must be one very close to the true shape. The transfer impedance Z_f, as defined earlier as the solution of equation (5), is the ratio of the pressure evaluated at the center of the mouth opening to the velocity excitation v_g over the glottis at frequency f. The objective function is now defined as

$$W(\mathbf{d}) = \frac{1}{N}\sum_{f=1}^{N} |Z_f(\mathbf{d}) - \hat{Z}_{f0}|^2 + \alpha(\mathbf{d}), \tag{19}$$

where $Z_f(\mathbf{d})$ is the transfer impedance (normalized) for the assumed shape at frequency f and Z_{f0} is the "measured" or observed impedance for the vocal tract of interest. $\mathbf{d}$ is the shift displacement at the wall surface towards the direction indicated by the arrows in Figure 14. $\alpha(\mathbf{d})$ is a penalty term for the limit as the shape is physically bounded within a certain range, within which a local minimum of the objective function is sought. The convergence is achieved when the following criterion is satisfied for a small value ε_w:

$$|W(\mathbf{d}^{(i+1)}) - W(\mathbf{d}^{(i)})| < \varepsilon_w. \tag{20}$$

4. SOME SIMULATED EXAMPLES

Here one examines the recovery of the vocal tract shape for a particular vowel from the spectral difference between the "measured" transfer impedance for that vowel and the spectra calculated for an assumed vocal tract shape. In the present simulation, the transfer impedance calculated from the vocal tract shape of that particular vowel is used instead of the "measured" spectra. From the authors' work [2], as one knows the true reasonable vocal tract shape, one chooses a shape slightly deformed at several points from the known shape as the assumed shape. Figure 14 shows the vocal tract corresponding to Japanese a:, which consists of 496 elements with 281 nodes, providing double nodes for sharp corners. Two positions are movable to simulate the mouth and lower jaw motion. The variation of the vocal tract shape is limited within the extension of the motions of mouth, jaws and tongue. The movable range is assumed to be within ±2·0 cm, which is set in the penalty term in equation (19). The distribution factor is chosen so that the movable area extends to 2·5 cm in radius at which the displacement is only 10% of that of the center. The objective function evaluated at two resonance frequencies corresponding to the first and second formant frequencies of the transfer impedance characteristic and the convergence characteristic when the DFP algorithm is used are shown in Figure 15. The criterion of the convergence ε_w is chosen to be of the order of 10^{-7}. The objective function monotonically descends for the variation of the displacements as shown in Figure 15(a),

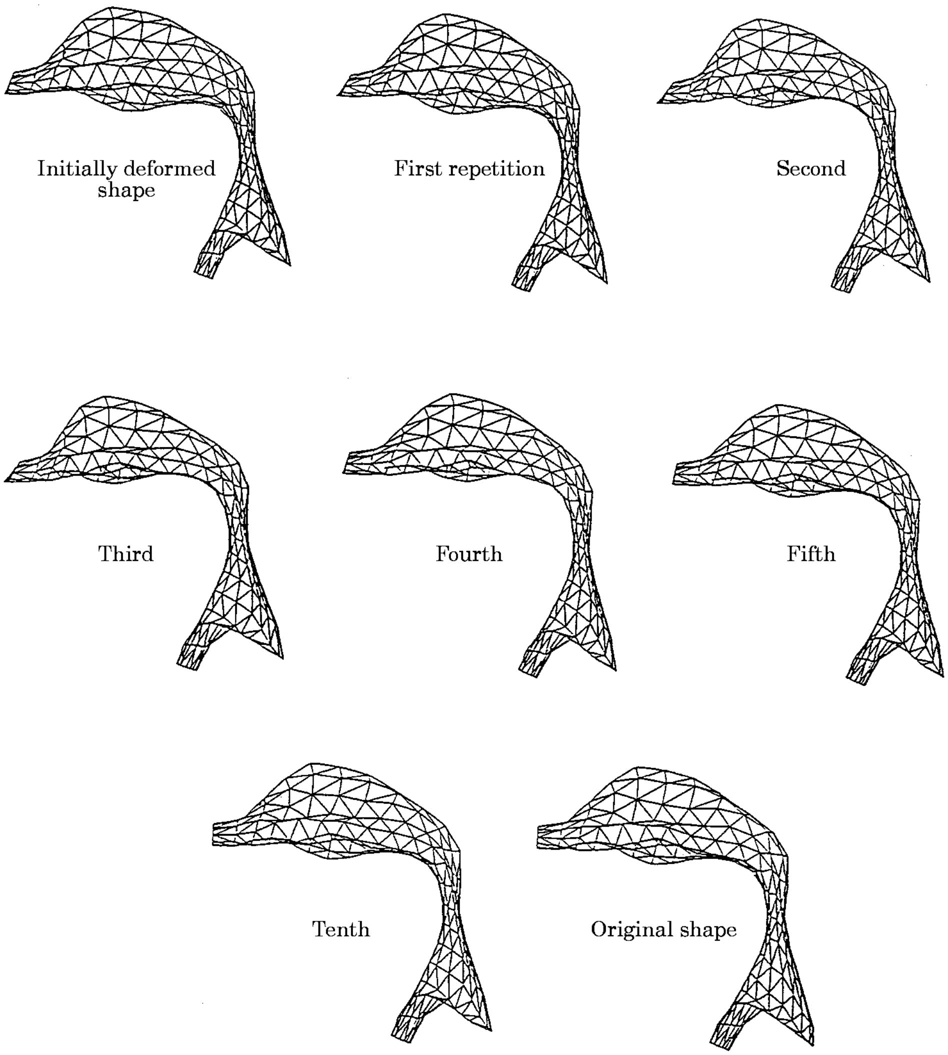

Figure 21. Shape transformation process, from the deformed to the original (Japanese vowel o:).

which shows the recovery of the original positions after several operations as shown in Figure 15(b). Figure 16 shows an example of the transfer impedance characteristic for vowel a:, in which the vocal tract is deformed initially as much as $d_1 = -1{\cdot}3$ cm and $d_2 = 2{\cdot}0$ cm. Figure 17 shows the sequence of the transformation process of the shape from the deformed to the original. Another initially deformed case when $d_1 = 1{\cdot}8$ cm and $d_2 = 2{\cdot}0$ cm is examined. The original shapes are again recovered after nine repetitive calculations. Figure 18 shows the vocal tract shape corresponding to the Japanese vowel o:. Two positions and directions of the displacements are chosen where d_1 creates the jaw movement and d_2 models the push of the lips. Figure 19 shows the convergence characteristic both for the objective function and displacements. There are many local minima in the objective function as shown in Figure 19(a), but the well corresponding to the positions $d_1 = 0$ and $d_2 = 0$ is the deepest. For recovery simulation, the initial deformation, $d_1 = 0{\cdot}5$ cm and $d_2 = -0{\cdot}5$ cm, is assumed and convergence is shown in Figure 19(b). Reasonable convergence is also possible. Figure 20 indicates the change of the transfer impedance as the displacements recover. It is interesting to note that for both cases the third and fourth formant frequencies also come close to the original frequencies as the first and second recover the original frequencies or the "measured" ones. Figure 21 indicates the successive process of transformation.

5. CONCLUDING REMARKS

Here a technique has been proposed to identify the vocal tract shape from the formant frequency spectra observed in front of the mouth. Its validity and capability were examined through numerical simulation. Identification was shown to be possible if the initial shapes were properly assumed. The examples shown are limited. Examination should be extended to identifying the vocal tract shapes corresponding to all vowels starting from a certain common shape or mean shape as the initial shape. It would also be interesting to investigate the transformation of the vocal tract shape when changing from one vowel to another.

The work was partly presented at The 14th Computational Electromagnetics and Electronics Symposium, Japan Society of Simulation Technology [14].

REFERENCES

1. Y. Kagawa 1981 *Finite Element Method for Acoustical Engineers—Fundamental and Applications*. Tokyo: Baifukan Press (in Japanese).
2. Y. Kagawa, R. Shimoyama, T. Yamabuchi, T. Murai and K. Takarada 1992 *Journal of Sound and Vibration* **157**, 385–403. Boundary element models of the vocal tract and radiation field and their response characteristics.
3. See, for example, M. R. Schroeder 1967 *Journal of Acoustical Society of America* **41**, 1002–1010. Determination of geometry of the human vocal tract by acoustic measurements.
4. H. Wakita 1967 *IEEE Transactions Audio Electroacoustics* **AU-21**, 417–427. Direct estimation of the vocal-tract shape by inverse filtering of acoustic speech waveforms.
5. M. M. Sondi 1974 *Journal of the Acoustical Society of America* **55**, 1070–1075. Model for wave propagation in a lossy vocal tract.
6. K. Shirai and M. Honda 1978 *IECJ Transactions* **J61-A**, 409–416. Estimation of articulatory parameters from speech waves (in Japanese).
7. K. Motoki, N. Miki and N. Nagai 1990 *Proceedings of ICSLP-90*, 433–436. Measurement of sound wave characteristics in the vocal tract.
8. J. H. Ahlberg, E. H. Nelson and J. L. Walsh 1969 *The Theory of Splines and Their Applications*. New York: Academic Press.
9. T. Chiba and M. Kajiyama 1958 *The Vowel—Its Nature and Structure*. Tokyo: Kaiseikan Press.

10. J. Suzuki 1978 *Journal of the Acoustical Society of Japan* **34**, 149–156. Discussion on vocal tract wall impedance.
11. G. M. L. Gladwell 1986 *Inverse Problems in Vibration*. Dordrecht: Martinus Nijhoff.
12. S. L. S. Jacogy, J. S. Kowalik and J. T. Pizzo 1972 *Iterative Methods for Nonlinear Optimization Problems*. Englewood Cliffs, NJ: Prentice-Hall.
13. M. J. Box, D. Davies and W. H. Swann 1969 *Nonlinear Optimization Techniques*. London: Oliver and Boyd.
14. Y. Ohtani, R. Shimoyama and Y. Kagawa 1993 *Proceedings of the* 14*th Computational Electromagnetics and Electronics*, *Japan Society of Simulation Technology*, Session V-7, 271–276. The shape identification of a vocal tract from frequency spectra with boundary element model (in Japanese).