Extraction of Vocal Tract Area Function from Three-Dimensional Magnetic Resonance Images using Digital Waveguide Mesh

Kenji Inoue

January 2008

Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to express my gratitude to Associate Professor Hirotake Nakashima and Professor Mehdi N. Shirazi, my supervisors in the Faculty of Information Science and Technology at Osaka Institute of Technology. Their insightful advice and generous acceptance helped and encouraged me to conduct the research contained in this thesis.

My sincere gratitude also goes to Dr. Hironori Takemoto at ATR Human Information Science Laboratories. Without the proposal from and discussion with him, this research simply would not have been born. His enthusiasm and broad knowledge in the field was inalterable.

I would also like to appreciate Dr. Tatsuya Kitamura at Konan University and Dr. Shinobu Masaki at ATR Brain Activity Imaging Center. Their advice and support enlightened me to make this work more tangible and solid. My appreciation goes to my inspiring friends inside or outside of the school, and my every company in the Lab.

Finally, there should nothing exist without my parents, Takumi and Kazuko Inoue, and their love, to which this work is dedicated, first of all.

ATR MRI database of Japanese vowel production

The MRI data used in this work is part of the "ATR MRI database of Japanese vowel production", which was acquired and published by the ATR Human Information Science Laboratories based on the research "Enhancing human-machine communication technologies" commissioned by the National Institute of Information and Communications Technology. The use of the database and the publication of the results are under the license agreement

with ATR-Promotions.

Abstract

A method is proposed in this thesis to extract the vocal tract area function from the three-dimensional magnetic resonance images. The proposed method uses the digital waveguide mesh, an implementation of the finite-difference time-domain (FDTD) method, to simulate wave propagation in the vocal tract from the glottis to the lips. The dimensions of the vocal tract areas are then calculated along the traveling wavefront that emerges from the simulation.

Formant analysis has been conducted for Japanese five vowels /aeiou/ to show the validity of the proposed method. The calculated formant frequencies of the area functions obtained by the proposed method and other existing methods have been compared to those measured from the solid models of the imaged vocal tract shapes.

Chapter 1

Introduction

The articulatory synthesis is a speech synthesis technique based on the human vocal tract and articulation models. It digitally simulates the glottal excitation and the wave propagation in the vocal tract which is controlled by the articulators such as tongue and lips.

In the recent years, improvements are made on the techniques to produce the 3D shapes of the human vocal tract during the speech production using the magnetic resonance imaging (MRI) [5, 8].

Some industrial applications, such as the simulation of the speech synthesis based on the 3D vocal tract shape, are, however, difficult in the modern computers to be carried out in real-time due to the computational complexity introduced by the 3D model. One of the solutions to the problem is to use an approximated 1D model estimated from the 3D vocal tract shape.

The vocal tract area function is a well developed model for such a 1D vocal tract model. It is represented as a series of area values (typically expressed in the magnitude of cm^2), each value representing a dimension of a cylindrical tube. The cylindrical tubes are assumed to be connected in cascade to consist an approximated human vocal tract. No bend and other fine components of the vocal tract shape are included in the model. In the simulation of the speech production or the computation of the transfer function using the area function, only the planar wave is considered to be propagated through the tube.

There have been many methods proposed to estimate the 1D vocal tract area function from the 3D vocal tract shape. Story [16, 17] and Story *et al.* [18, 19, 20] proposed the iterative bisection method to calculate the geometrical center points of the centerline of the vocal tract. Kröger *et al.* [7] proposed another method which the rectilinear grid system is first specified by the user and then the centerline points of the vocal tract are calculated from the geometrical center points of the intersection area of the grid lines and the vocal tract.

Although these methods try to find the geometrical centerline of the 3D vocal tract shape, the path of the actual acoustic wave propagated in the vocal tract are found not to go through the geometrical center points above the supraglottis [13, 21]. Nakai *et al.* [14] proposed a technique which exploits the finite element method (FEM) to estimate the area function from the sound intensity at near the first formant frequency.

The accuracy and validity of these methods are, however, not well investigated comparing with each other. No standard method is therefore established in the literature. The present study is intended to address the issue by proposing a method which computes the path of the acoustic wave solving the wave equation in the time domain and estimates the vocal tract area function from the centerline along the propagation path, and then by comparing the proposed method with the reference model and other existing methods. To solve the partial differential equation (PDE) of the wave propagation in computer, digital waveguide mesh (DWM) [15, 4, 11, 12, 10] is exploited to discretize the equation. Lower four formants have been obtained from the estimated vocal tract area function by applying the proposed method to the vowels contained in the "ATR MRI database of Japanese vowel production" [2] and the comparison of the results has been made against the formant frequencies of the solid models [6] created based upon the applied dataset, as well as with those produced from the other existing methods.

Chapter 2

Methods

In this study, as the algorithms that calculate the centerline of the region of the vocal tract shape, the proposed method using DWM and other three methods, namely a method using the Manhattan distance proposed by Takemoto *et al.* [22], the iterative bisection method adopted by Story *et al.* [18], and a method using the nearest points on the opposite edge (as cited as "conventional method" in [9]), were implemented and investigated.

In the whole process of the estimation of the vocal tract ara function, the same procedure is taken for the processes except the calculation of the centerline. In this study, the whole process of the extraction of the vocal tract area function from the 3D MRI data was conducted in the following manner and order.

1. Application of the noise reduction filter.
2. Segmentation of the airway from the surrounding tissue by thresholding.
3. Manual specification of the glottis and lip positions.
4. Segmentation of the main vocal tract region from the branches (e.g. nasal tract).
5. Calculation of the centerline of the vocal tract.
6. Smoothing of the centerline applying the spline interpolation.

7. Determination of the grid lines and their resliced oblique sections locally perpendicular to the centerline.

8. Derivation of the vocal tract area function by counting the number of airway voxels in the oblique sections, starting just above the glottis.

In the following sections, each process of the extraction procedure is detailed.

2.1 Preprocessing

The 3D MRI data processed in this study is assumed to be an 8-bit grayscaled volumetric image. An example of the used volumetric MRI data is shown in Fig. 2.1.

2.1.1 Noise reduction

First, noise reduction filters may be applied to the 3D MRI data. MR images tend to be affected by a significant level of noise which is approximately assumed to be an additive white Gaussian noise (AWGN). The desirable filter needed to be applied would depend on the image to be processed.

In the software the author created for this study, median filter and moving average filter were implemented. Median filter picks the neighbor voxels and take the median value of them (using the sorting) as the value of the center voxel. Moving average filter sums up the values of neighbor voxels and center voxel its own and take the average value of them as the value of the center voxel.

Since the separation of the airway and tissue regions is the fundamental role in the preprocessing state, the median filter which preserves the edge information more than the moving average filter would be preferable.

Other noise reduction filters such as the ones exploit the statistical characteristics of AWGN may need to be applied or developed for better results.

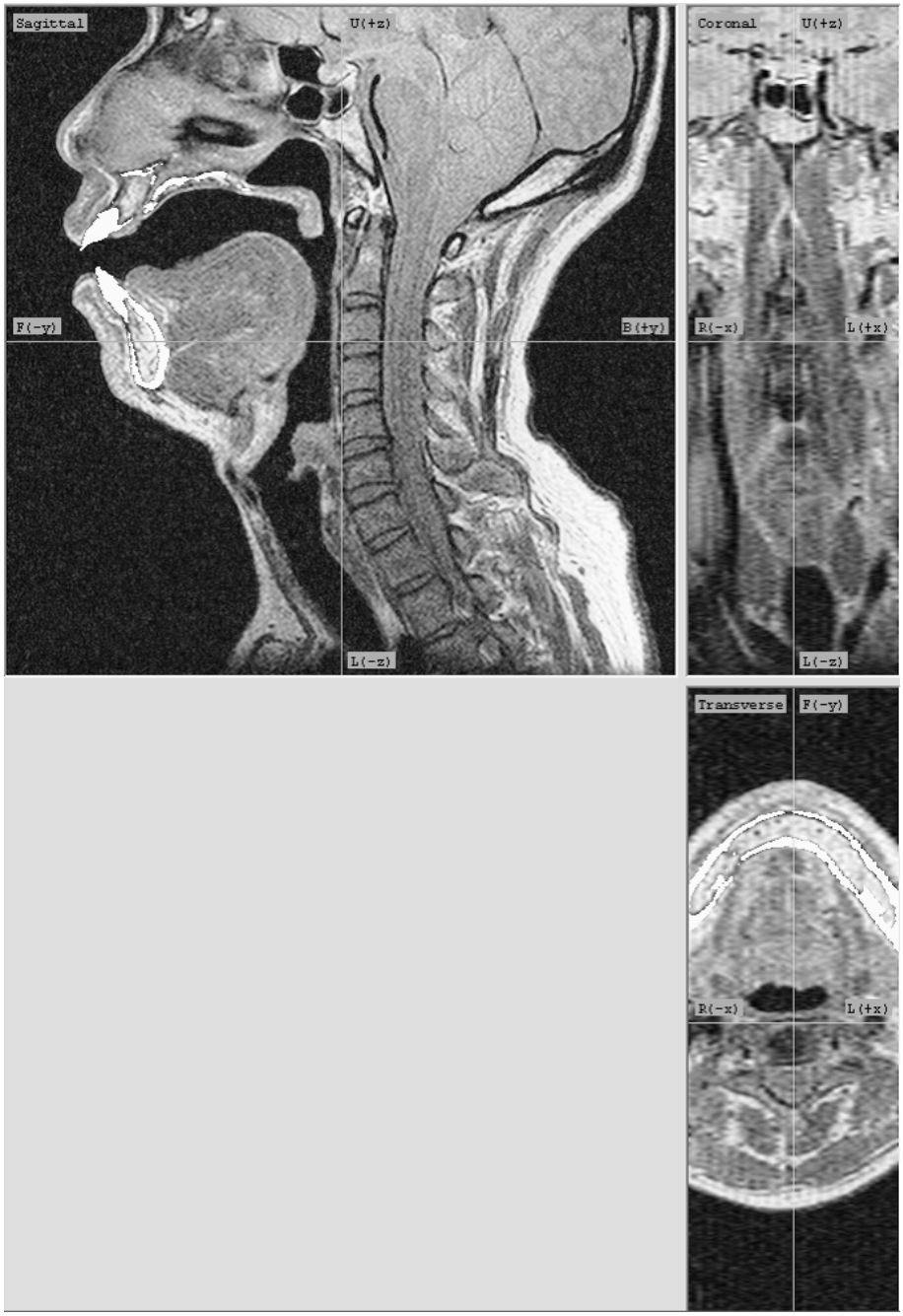

Figure 2.1: An example of the volumetric MRI data used in this study. (Upper left) sagittal view (mid-sagittal plane is shown). (Upper right) coronal view. (Lower right) transverse view.

2.1.2 Airway identification

Second, airway region is segmented from the surrounding tissue by thresholding, or binarization, filter. Binarization filter takes a threshold and all the voxels whose value is lower than the threshold are blackened. In this stage, black ($R = G = B = 0$) voxels are identified as the airway after the binarization.

An appropriate thresholding value also depends on the image to be processed. In this study, the threshold is determined manually by the author image by image.

Some isolated voxels within the airway may be misidentified as a tissue in this process. Those isolated noise voxels are left as tissue in this process but determined as airway at the later stage.

2.1.3 Glottis and lip positions

Glottis and lip locations can then be specified manually (Fig. 2.2).

Figure 2.2: Specification of the glottis and lip lines

The two end points of the glottis line are given by the operator person.

Given an assumption that the glottis region is ellipse-like 3D shape, the glottis region is then computed as follows: in the sagittal plane at which the two end points located, a linear airway region is first determined from the median point of the specified glottis line by extending the point to the direction and its opposite direction of the specified line (treating it as a vector from an end point to the other end) and collecting the airway points. For all the points consisting the linear region, airway region is next determined by extending the point to the direction and its opposite direction of the line perpendicular to the sagittal plane and collecting the airway points. The joint region of the computed airway regions result to the glottis region.

The two end points of the lip line are given by the operator person. The lip region is determined as the airway region of a circle whose center is at the median point of the given lip line, whose radius is the half length of the given lip line, and whose two axes are the direction of the line and the direction perpendicular to the sagittal plane.

The lip line has not to be the actual lip location; it, in fact, specifies the location of the end position of the centerline calculation algorithms. It may be better to set the lip location in the air some millimeters away from the top and bottom lips.

2.1.4 Detection of isolated noise voxels

The isolated noise voxels within the airway is determined using the connectivity of region.

All the non-black (i.e. tissue) voxels four-neighboring in 2D (or six-neighboring in 3D) to the glottis region are chosen as the starting points of the non-airway set. Until no more voxel can be added into the non-airway set, add the non-black voxels four- or six-neighboring to one of the points in the non-airway set into the non-airway set used in the next iteration.

For each of the four-neighbor voxels to one of the voxels of the airway region, add it to the airway region if it is not in the non-airway set. This addition to the airway region is iterated until no more voxel is added.

2.2 Segmentation of the main vocal tract from the branches

The algorithms used in this study require that the branch regions such as the nasal tract, piriform fossa, and epiglottic vallecula are also segmented from the airway region beforehand. Since it is usually obtained manually and no such segmentation algorithm is found in the literature, an automatic branch region detection algorithm is developed in this study as follows.

First, a distance map from the glottis is calculated using the method of Manhattan distance, whose detail is explained in the later section. A distance map $dm(p)$ denotes a correspondent step distance at position p from the glottis. Next, since the main vocal tract region is considered as the region such that the sound wave originated from the glottis to the lips can reach there without backward propagation, the main vocal tract region is extracted by finding the backward propagation path from the lips to the glottis with the following algorithm.

1. Assign $t = 0$. Define $R(t) = R(0)$ as a set of voxel positions consisting the lip region.

2. For all the positions $p_i \in R(t)$ $(0 \leq i < |R(t)|)$, calculate the set $Q_i = \{q \mid q \in FN(p), q \notin R(t), dm(q) \leq dm(p_i), dm(q) \geq 0\}$, and find the set $R(t+1) = R(t) \cup Q_1 \cup \cdots \cup Q_n$, where $FN(p)$ is the four neighbors in 2D (or six neighbors in 3D) of position p.

3. Iterate procedure 2. until $R(t+1) = R(t)$ satisfies. If it satisfies, the algorithm ends and the $R(t)$ is the voxel positions consisting the main vocal tract region.

This algorithm runs in $O(NM)$ time for N voxels and M neighbors using a queue as its data structure.

An example of the result of the branch segmentation is depicted in Fig. 2.3.

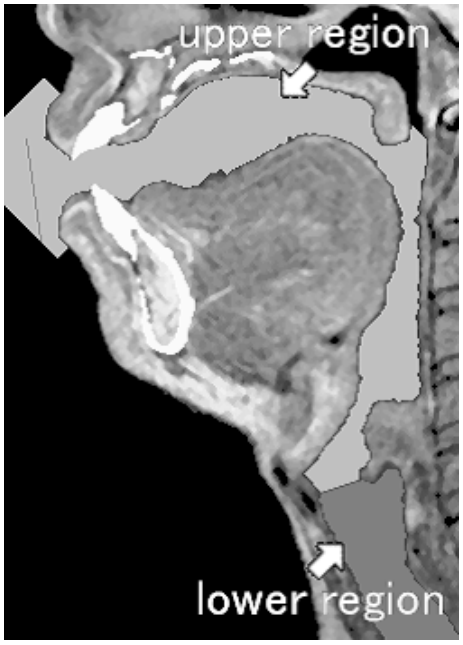

Figure 2.3: Upper (light gray) and lower (dark gray) main vocal tract regions after branch segmentation

2.3 Centerline calculation

In this study, two methods are used as the algorithms which compute the centerline along the propagation path of the acoustic wave; one is a method using the DWM to solve the wave equation of the wave propagation in time domain, and the other is a method using the Manhattan distance as an approximation of the wave propagation solution with lower computational complexity.

Besides them, another two methods are used as the algorithms which compute the centerline from the geometrical center points of the vocal tract region; one is the iterative bisection method, and the other is a method using the nearest points on the opposite edge.

Methods using the DWM, Manhattan distance, and the iterative bisection method have their 2D and 3D implementations while the method using the nearest points on the opposite edge solely have its 2D implementation as it cannot be naively extended to 3D.

2.3.1 Method using digital waveguide mesh

In the 1D acoustic tube model of the vocal tract, sound wave propagated in the 3D space is considered as the propagation of planar wave. Thus, vocal tract area functions whose acoustical characteristics are similar to those of 3D shapes may be able to be obtained by calculating how the acoustic wave propagates in the 3D vocal tract shape and computing the cross-sectional areas following the path and direction of wave propagation.

DWM is an implementation of the finite-difference time-domain (FDTD) method [23] which can efficiently compute the numerical solution of the wave equation of the acoustic wave propagation in time domain.

For sound pressure $p(t, x, y, z)$ in the Cartesian coordinate, 2D or 3D wave equation

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \tag{2.1}$$

is digitized and solved with rectilinear mesh (where c is the speed of sound).

In this study, a junction is set to each voxel of the image (Fig. 2.4). Each junction has a waveguide for each voxel of its four- or six-neighbors.

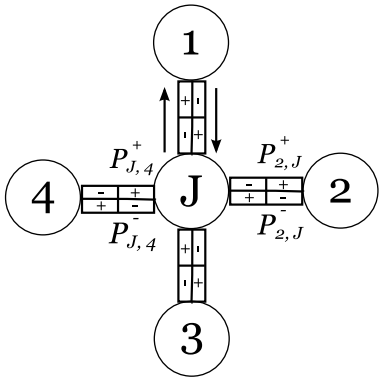

Figure 2.4: A waveguide junction with four connections

For a junction J with M waveguide connections, an outgoing pressure to the k-th connection p_k^- is formulized as

$$p_k^- = r_k p_k^+ + \sum_{i=1, i\neq k}^{M} (1 + r_i) p_{J,i}^+ \tag{2.2}$$

where p_k^+ is an incoming pressure from k-th connection, $p_{J,i}^+$ is an incoming pressure from i-th connection a unit time step before, and r_k is a reflection coefficient for k-th connection defined as

$$r_k = \frac{2Y_k - \sum_{i=1}^{M} Y_i}{\sum_{i=1}^{M} Y_i} \tag{2.3}$$

where Y_k is an admittance of k-th waveguide connection. Impedance Z_k is defined as the inverse number of an admittance Y_k as

$$Z_k = \frac{1}{Y_k} \tag{2.4}$$

The input pressure of the simulation is given in the initial time step $t = 0$ such that excited pressure 1.0 is set to every voxel consisting the glottis and it is released just after the pressure is given, provided that the total excited pressure 1.0 is equally divided by the number of its waveguide connections to be set to each waveguide (i.e. 0.25 is set for each waveguide at a four-neighbor junction).

In the simulation steps, for each voxel of the main vocal tract region value of distance map $dm(p)$ at voxel position p is set to the step time when the sum of the energy passed through the voxel by the step time first exceeds a certain threshold T_{dm}, regarding this step time as a distance from the glottis. If the threshold T_{dm} is properly configured, contour planes of sound pressure distribution at each step distance from the glottis can be obtained by this method.

In this study, Euclidean norm is used for the summation energy. The value of the distance map $dm(p)$ at voxel position p is thus the step time $t_s = 0, 1, 2, ...$ that first satisfies the condition

$$\sum_{t=0}^{t_s} |p_t(p)| > T_{dm} \tag{2.5}$$

where $|p_t(p)|$ is an absolute value of the real-valued pressure at point p at step time t.

Centerline is obtained by calculating the centroids for every step distance

in the computed distance map. The centerline, however, contains small but sharp fluctuations, so that the smoothing process explained in a later section is applied.

2.3.2 Method using Manhattan distance

The method that utilizes the Manhattan distance, proposed by Takemoto *et al.* [22], calculates the distance map from the glottis using the Manhattan, or L_1, metric. The distance map computed in this way can be considered as a result of roughly approximated solution of the wave propagation (see Fig. 4.1 and compare the difference of contour distance maps).

The distance map for this method using the Manhattan distance, $dm_M(p)$, is given by

$$dm_M(p) = \min_{q \in FN(p)} \{dm_M(q)\} + 1 \tag{2.6}$$

where p is a voxel position, $FN(p)$ is the four neighbors in 2D (or six neighbors in 3D) of position p, and the value of $dm_M(p)$ is set to 0 for voxels consisting the glottis, and ∞ for voxels not in the main vocal tract region.

dm_M can be calculated in $O(NM)$ time for N voxels and M neighbors using a queue as its data structure.

The processes of smoothing and centerline extraction is same as the way used for the method using DWM.

2.3.3 Iterative bisection method

The iterative bisection method, adopted by Story [16, 17] and Story *et al.* [18, 19, 20], calculates the centerline from geometrical center points of the vocal tract shape.

Firstly, the algorithm connects the glottis and lip points with a line segment and computes a plane which passes through the median of the line segment and perpendicular to the line segment. It then calculates the centroid of the voxel positions contained in both the perpendicular plane and the vocal tract region and adds the centroid to the point set of centerline to be constructed. Secondly, it bisections the vocal tract region by the perpen-

dicular plane and recursively applies the same procedure to the two centroid-to-glottis and centroid-to-lip line segments.

2.3.4 Method using the nearest points on the opposite edge

The method that finds the nearest points on the opposite edge, conventionally used in the literature as in [9], also calculates the centerline from geometrical center points of the vocal tract shape. Since this algorithm cannot be naively extended to its 3D version, the centerline is obtained using the 2D mid-sagittal plane of the MRI data in this study.

Given that the two sets of the voxel positions on the edge of the non-airway region obtained by following the edge contour from the two end points of the glottis are E_1 and E_2, the algorithm finds a correspondent point $e_2 \in E_2$ for each point $e_1 \in E_1$ such that the Euclidean distance between e_1 and e_2 is minimized, as formulated as

$$e_2 = e_2(p) = \arg\min_{q \in E_2} \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} \tag{2.7}$$

and adds the median point between e_1 and e_2 to the point set of centerline to be constructed.

2.4 Smoothing

In the two methods using the DWM and Manhattan distance which are based on the computed distance map, centerline is computed by calculating the centroid for each step distance. The centerline computed in this way, however, contains a little variations which results to produce an unstably discontiguous vocal tract area function.

The centerline is, therefore, smoothed by first recalculating the centroids by each predefined step distance D_{smooth} to cut off the number of centerline points to be its $1/D_{smooth}$ points and then applying the spline interpolation.

2.5 Derivation of the vocal tract area function

After the centerline points are calculated, it is resampled at the given slice interval, which is set to 2.5 mm in this study. Reslice grid planes locally perpendicular to the resampled points are then obtained (Fig. 2.5).

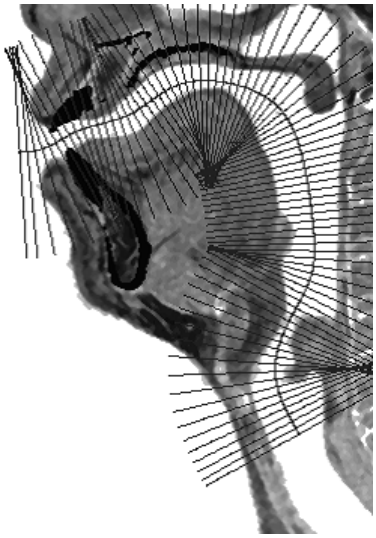

Figure 2.5: A computed centerline and its grid planes at slice interval 2.5 mm projected on the sagittal plane, calculated and superimposed for vowel /e/

The dimension of each grid plane is determined by counting the number of voxels of main vocal tract region intersecting to the grid plane. The dimension d_k at k-th cross-sectional area from the glottis to the lips can thus be calculated by

$$d_k = n_k l^2 \tag{2.8}$$

where n_k is the number of voxels of main vocal tract region on the k-th grid plane and l is the cubic voxel side-length of the 3D MRI image.

Chapter 3

Experiments

The 3D MRI data of the Japanese five vowels /aeiou/, spoken by an adult male Japanese speaker, contained in the "ATR MRI database of Japanese vowel production" [2] is used for the experiments and evaluations of the centerline calculation algorithms.

There are solid models formed by a stereo-lithographic technique [6] for those five vowels and their acoustical characteristics are well investigated. By comparing to the formant frequencies of the solid models, as opposed to those of the recorded utterance which many disturbance factors may have involved, the acoustical difference between the 3D MRI data and its 1D derivation, the vocal tract area function, are precisely evaluable.

Firstly, an investigation was conducted to show which values can be used for the parameters of the proposed method using DWM to precisely estimate the vocal tract area function. There are three major parameters for the proposed method: a threshold value T_{dm} to calculate the distance map, waveguide impedances of the air inside vocal tract Z_{air} and the vocal tract wall Z_{wall}, and a step distance D_{smooth} of the smoothing process. In this study, threshold value T_{dm} to calculate the distance map was configured in all cases as

$$T_{dm}(t) = 0.01/N_g(t^2 + 1) \tag{3.1}$$

for 2D simulation and

$$T_{dm}(t) = 0.1/N_g(t^3 + 1) \tag{3.2}$$

for 3D simulation, where t was the step time in the simulation. The impedance of the air Z_{air} was always set to a constant 1 and the impedance of the vocal tract wall Z_{wall} was changed to 1, 10, 100, 1000, or 10000 to see that the calculated contour distance points give proper contour sound pressure planes. Note that all the pressure that goes to the wall junction was set to be lost (no energy back from the wall) in the simulation. The step distance D_{smooth} of the smoothing process was changed to 10, 20, 30, 40, 50, 60, 70, or 80.

In the preprocessing stage of all the experiments conducted in this study, two consecutive eight-neighbor (2D, 3×3) median filters were applied and then thresholding with threshold value 64 was applied to identify the vocal tract region. The lip-side end point of the centerline drawn from the glottis was set 1 cm away from around the lips. When calculating the vocal tract area function, therefore, its areas within 1 cm from the lip-side end were unconditionally removed, and the areas whose dimension is greater than the manually preset threshold value for each vowel, as shown in Table 3.1, were further removed from there due to the lip opening.

Table 3.1: Threshold values to cut off the lip-side areas due to the lip opening

Vowel	Threshold (cm^2)
a	7.0
e	2.5
i	1.5
o	1.5
u	1.0

From the vocal tract area functions obtained changing the parameter values, transfer functions were calculated in frequency domain using a transmission line model which included energy losses due to viscosity, heat conduction, and radiation [1, 3]. The yielding wall effect was not adopted (i.e.

the wall is rigid) to make the simulation condition same as the reference solid models by Kitamura *et al.* The lower four formants were determined by finding the peaks in the transfer function and compared to those of solid models to decide the most precisely matched parameter values.

Secondly, formants frequencies were derived by the same procedure for the vocal tract area functions obtained from the method using the DWM with the most precisely matched parameter values and using other three existing methods. The obtained formants frequencies were compared to those measured from the solid models.

Chapter 4

Results and discussion

4.1 Contour maps

The contour distance maps obtained by the methods using the DWM and Manhattan distance are shown in Fig. 4.1 through Fig. 4.10.

For some conditions for vowel /e/ and /i/, namely $Z_{wall} = 1000, 10000$ in Fig. 4.3 (2D, /e/), $Z_{wall} = 10, 100, 1000, 10000$ in Fig. 4.5 (2D, /i/), and $Z_{wall} = 1000, 10000$ in Fig. 4.6 (3D, /i/), 2D and 3D DWM methods failed to compute adequate contour sound pressure maps that show the continuous wave propagation. In those conditions, scattered dots are depicted just after the places where narrow constrictions occur in the vocal tract (note that /e/ does not actually have so much narrow constriction; it is, to some extent, introduced by the unremoved noise voxels). Considering that the same conditions tend to work for their 3D cases, it can be said that the shortage of the number of waveguide junctions in the constrictions, which is directly derived from the mesh size, introduced a numerical instability in the simulation that resulted to the false contour maps. The square or cubic mesh size was 0.5 mm in the simulation, and the narrowest constrictions were about 4 pixels wide for /e/ and 3 pixels wide for /i/ in the mid-sagittal planes on which the 2D simulations were performed.

The most numerically stable condition over all the vowels in this experiment was the condition $Z_{air} = 1$ and $Z_{wall} = 1$. It also drew the most natural

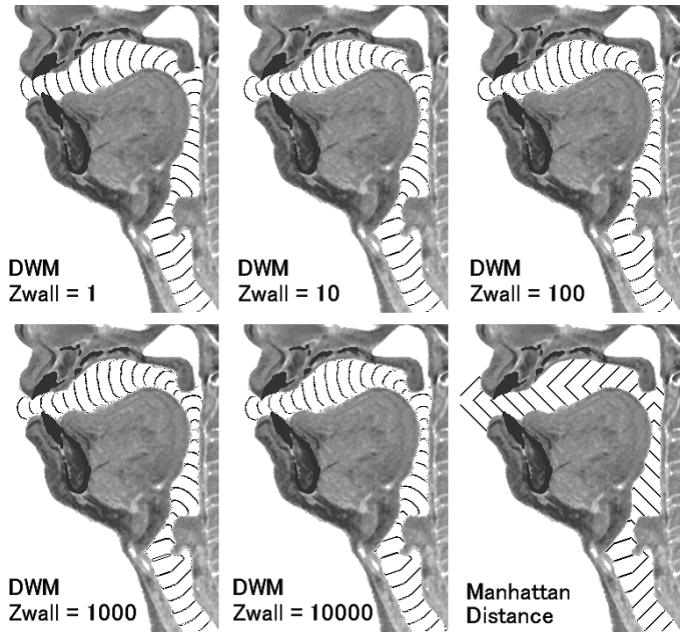

Figure 4.1: Contour distance maps of 2D versions of the DWM and Manhattan distance methods for vowel /a/

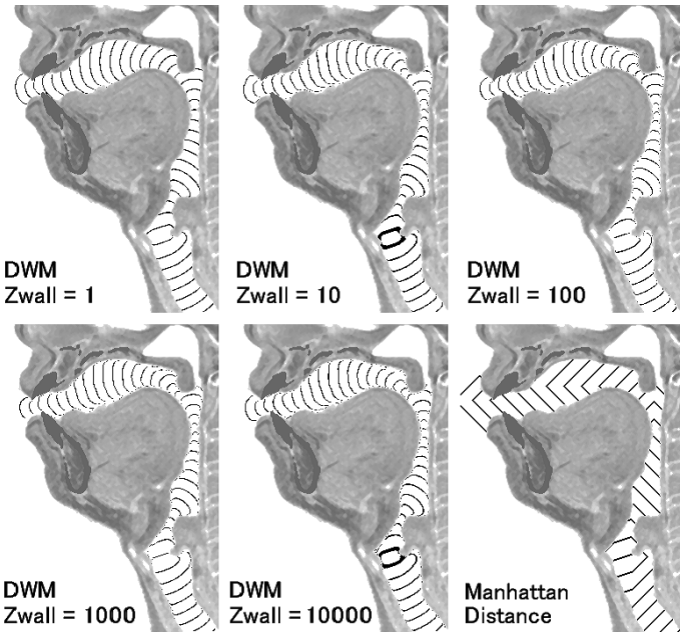

Figure 4.2: Contour distance maps of 3D versions of the DWM and Manhattan distance methods for vowel /a/

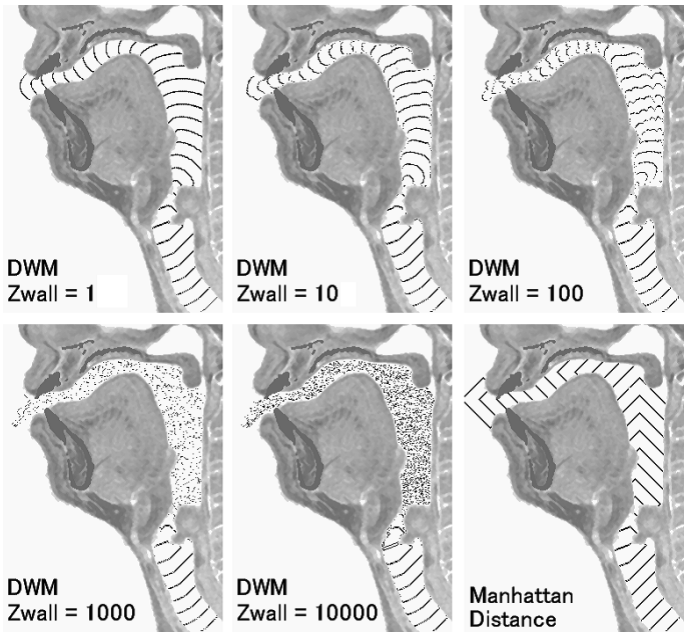

Figure 4.3: Contour distance maps of 2D versions of the DWM and Manhattan distance methods for vowel /e/

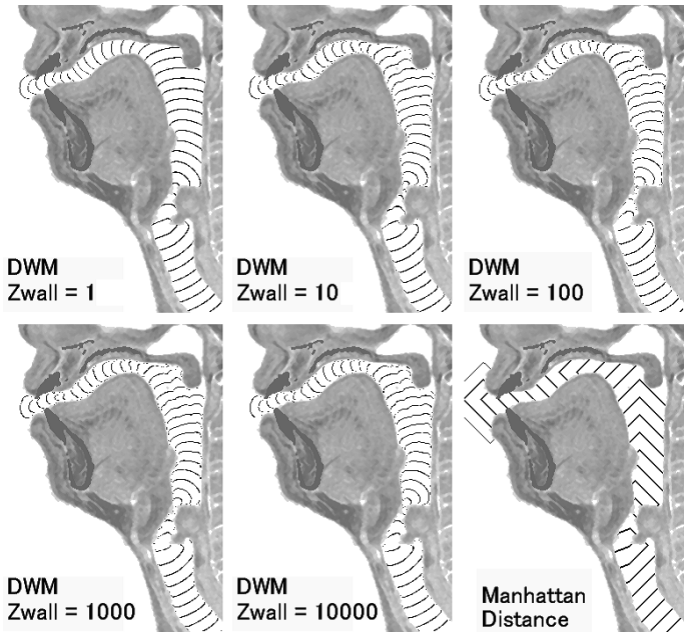

Figure 4.4: Contour distance maps of 3D versions of the DWM and Manhattan distance methods for vowel /e/

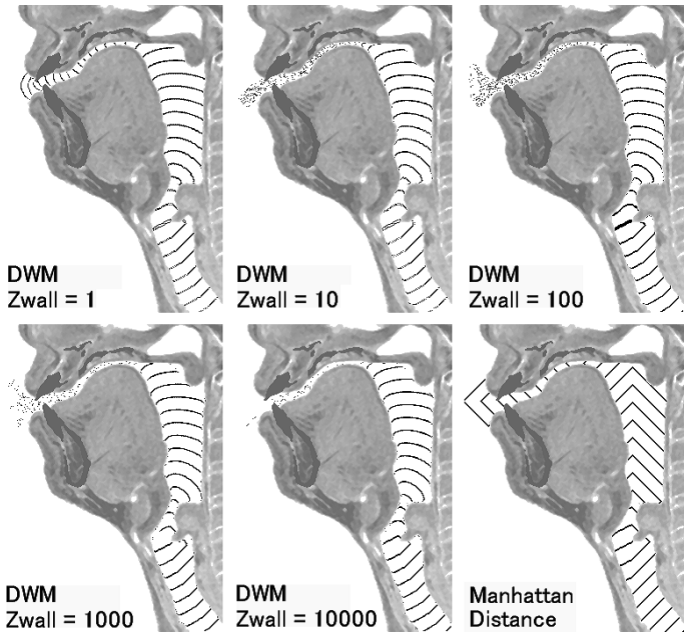

Figure 4.5: Contour distance maps of 2D versions of the DWM and Manhattan distance methods for vowel /i/

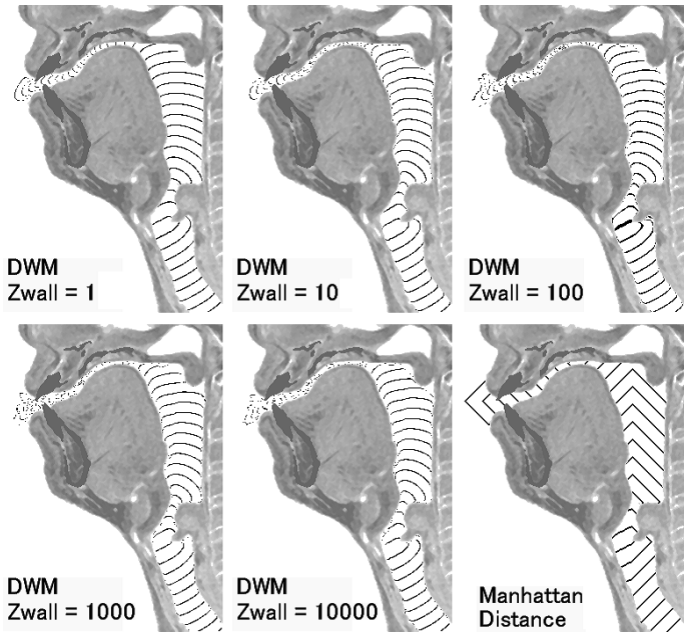

Figure 4.6: Contour distance maps of 3D versions of the DWM and Manhattan distance methods for vowel /i/

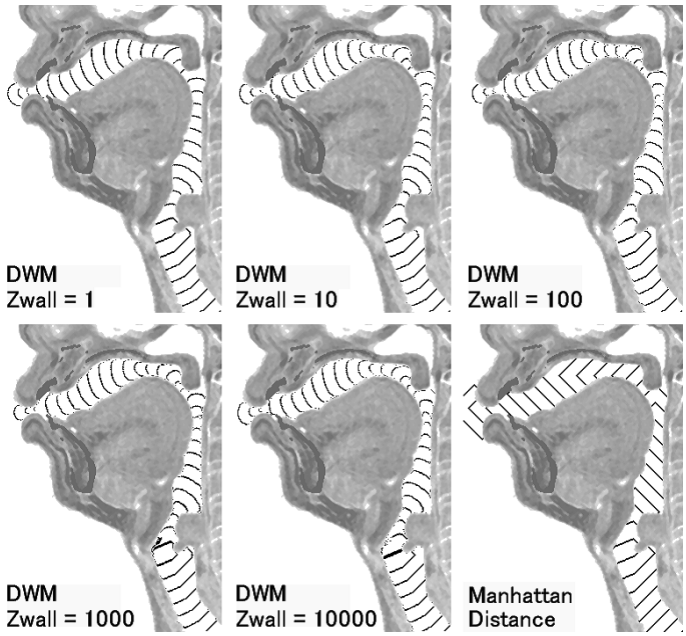

Figure 4.7: Contour distance maps of 2D versions of the DWM and Manhattan distance methods for vowel /o/

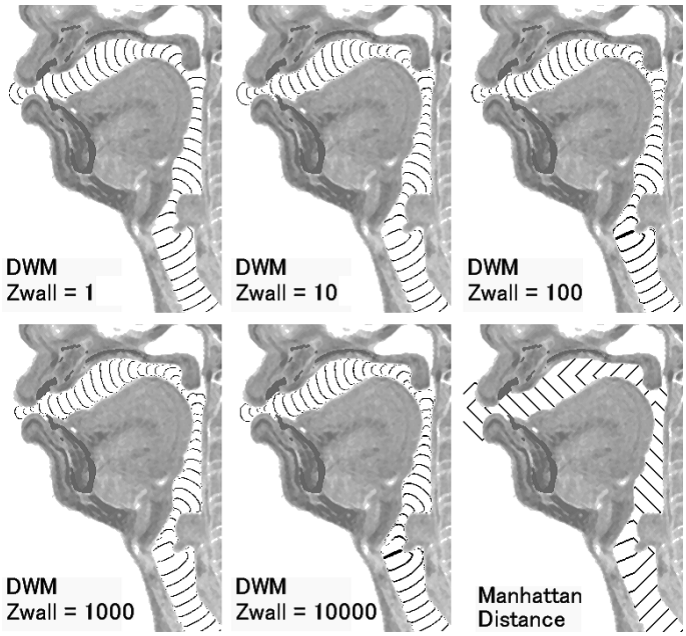

Figure 4.8: Contour distance maps of 3D versions of the DWM and Manhattan distance methods for vowel /o/

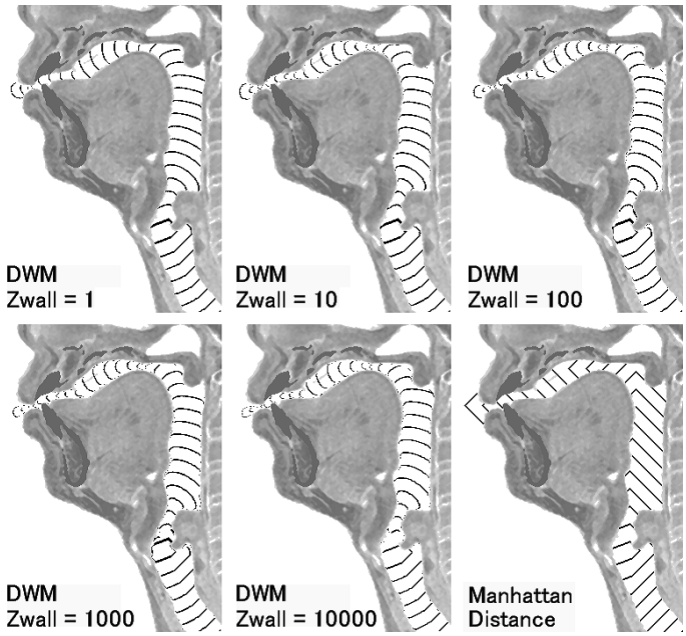

Figure 4.9: Contour distance maps of 2D versions of the DWM and Manhattan distance methods for vowel /u/

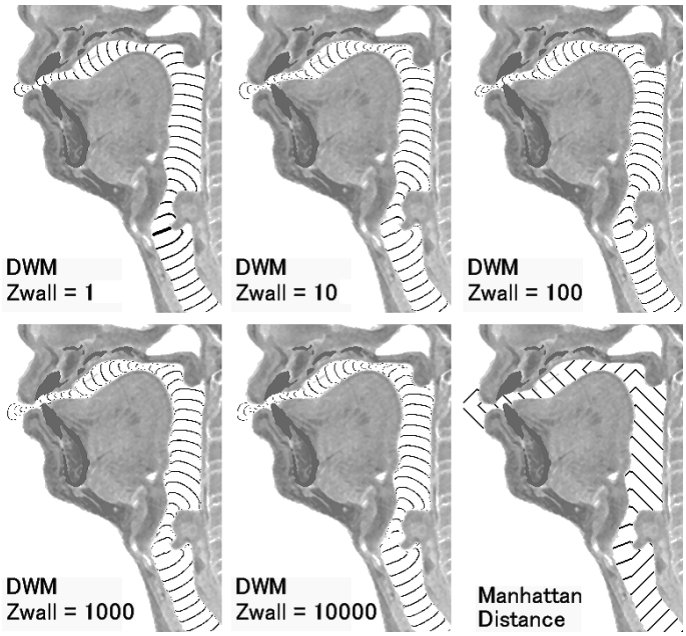

Figure 4.10: Contour distance maps of 3D versions of the DWM and Manhattan distance methods for vowel /u/

contour lines of wave propagation at the first glance, but the condition that says the air and wall impedances are same, which also says no reflection occurs at the wall boundary, is unnatural. Note that the simulation configured in this study gives no energy reflection from the medium of vocal tract wall to the airway, (to reduce the amount of computing time for over 100 DWM simulations conducted in this study), and the wall of the reference solid model is considered to be rigid which can be virtually simulated by setting the wall impedance Z_{wall} higher enough. As the higher wall impedance showed the higher possibility of the numerical instability for the introduced threshold value $T_{dm}(t)$, the effect of changing the wall impedance value Z_{wall} to the resulted vocal tract area functions and its relation to the mesh size and threshold value $T_{dm}(t)$ should need to be investigated in detail in future studies.

4.2 Vocal tract area functions and their transfer functions

For each pair of vocal tract wall impedance $Z_{wall} = 1, 10, 100, 1000, 10000$ and step distance of the smoothing process $D_{smooth} = 10, 20, 30, 40, 50, 60,$ vocal tract area function and its transfer function were computed for each vowel. The frequencies of the lower four formants, F1-F4, of the calculated transfer functions were found by peak peaking and their four relative percent errors (RPEs), Δ1-Δ4, to those of the reference solid models were computed. The sum of the four RPEs, $\sum \Delta$, was also computed for each area function. The calculated lower four formants of all the vocal tract area functions are shown in Appendix A. Some area functions and their transfer functions could have not been properly computed due to the numerical instability explained the former section.

In this study, the sum of the four RPEs was used as a general index for the evaluation of the validity of the vocal tract area function. Considering the fact that some showed better F1 RPE than F2 and some showed vice versa, the use of this index may need to be discussed. The fifth and higher

formants were not considered since the higher frequency region involve many factors such as the effects of the piriform fossae and transverse mode.

The sum of the four RPEs were compared by vocal tract wall impedance $Z_{wall} = 1, 10, 100, 1000, 10000$ and step distance of the smoothing process $D_{smooth} = 10, 20, 30, 40, 50, 60$ for each vowel (Fig. 4.11 through Fig. 4.20).

In the figures Fig. 4.11, Fig. 4.13, Fig. 4.15, Fig. 4.17, and Fig. 4.19, the smallest RPE sum $\sum \Delta$ was chosen as the best matched representative among the group of same simulation dimension (i.e. 2D or 3D), same smoothing factor D_{smooth}, but different vocal tract wall impedances Z_{wall}. In a similar fashion, the smallest RPE sum was chosen among the group of same simulation dimension and same wall impedance Z_{wall} but different smoothing factors in the figures Fig. 4.12, Fig. 4.14, Fig. 4.16, Fig. 4.18, and Fig. 4.20.

For the comparisons with the wall impedance variations, no obvious trend is visible. Lower impedances tend to get better result for both 2D and 3D to vowel /i/ (Fig. 4.15) and 3D to /o/ (Fig. 4.17), midst impedances better for 2D to /e/ (Fig. 4.13) and both 2D and 3D to /u/ (Fig. 4.19), and no explicit trend for others.

On the other hand, the comparisons with the smoothing factor shows a trend that the lower smoothing step distances tend to get better RPE sums for most cases except for the 2D to vowel /i/ (Fig. 4.16) and 3D to /u/ (Fig. 4.20). This result could be considered as a direct reflection of the smoothing effect that shortened the total vocal tract length which resulted the worse RPEs, or it also could be arisen from the modeling inconsistency, such as the addition of the piriform fossa, between the solid model and 1D area function, which might have shifted the formant frequencies of the 1D area function from the actual 3D model.

This result indicates that the effect of the smoothing process to the formant frequencies tends to be greater than that of the vocal tract wall impedance.

The difference between the RPE sums of 2D and 3D versions is not obvious. The 2D version shows better results in some cases and 3D version in other cases. It suggests that the shape information necessary to extract the centerline of the vocal tract shape is amply supplied by the mid-sagittal

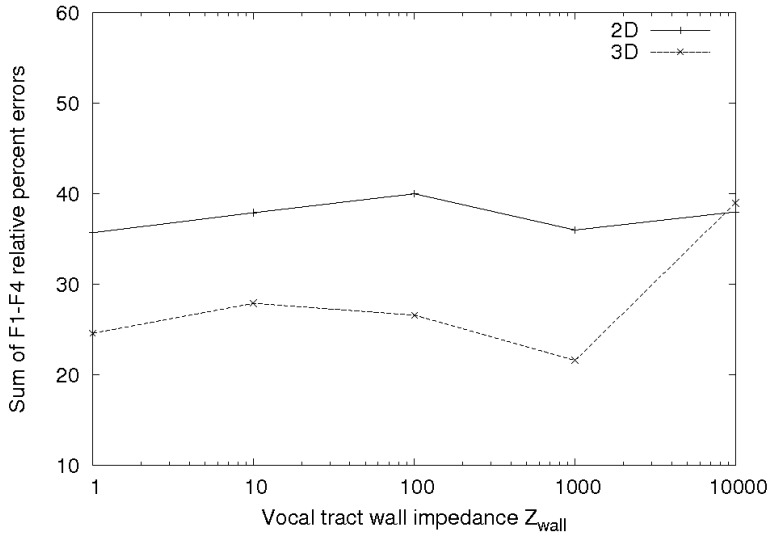

Figure 4.11: Comparison of the DWM conditions by the vocal tract wall impedance Z_{wall} for vowel /a/

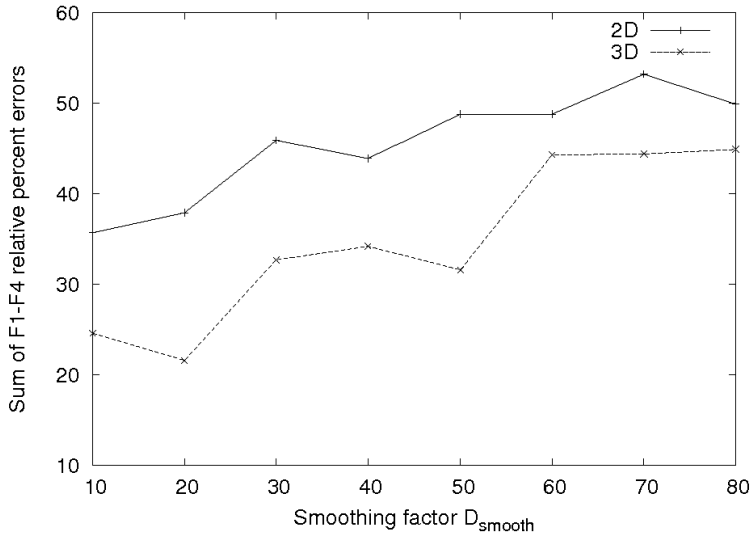

Figure 4.12: Comparison of the DWM conditions by the step distance of the smoothing process D_{smooth} for vowel /a/

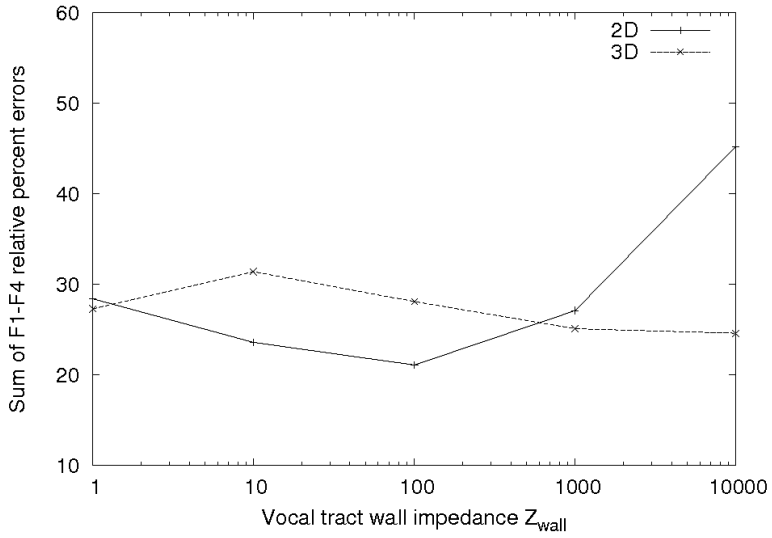

Figure 4.13: Comparison of the DWM conditions by the vocal tract wall impedance Z_{wall} for vowel /e/

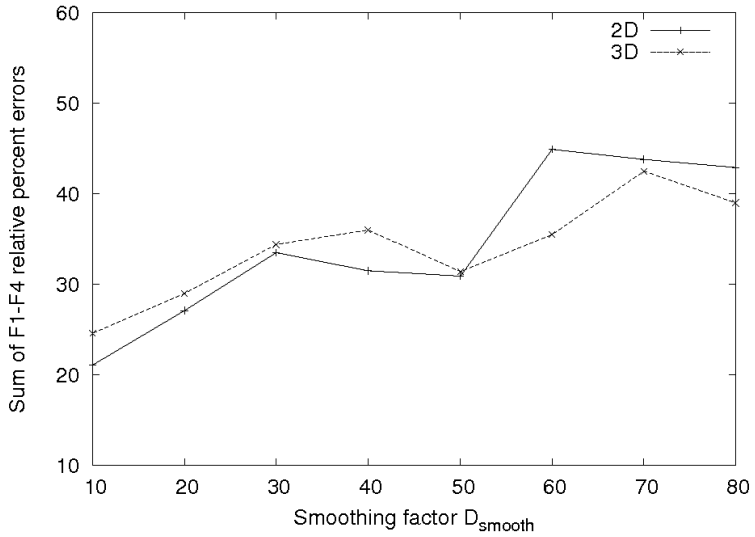

Figure 4.14: Comparison of the DWM conditions by the step distance of the smoothing process D_{smooth} for vowel /e/

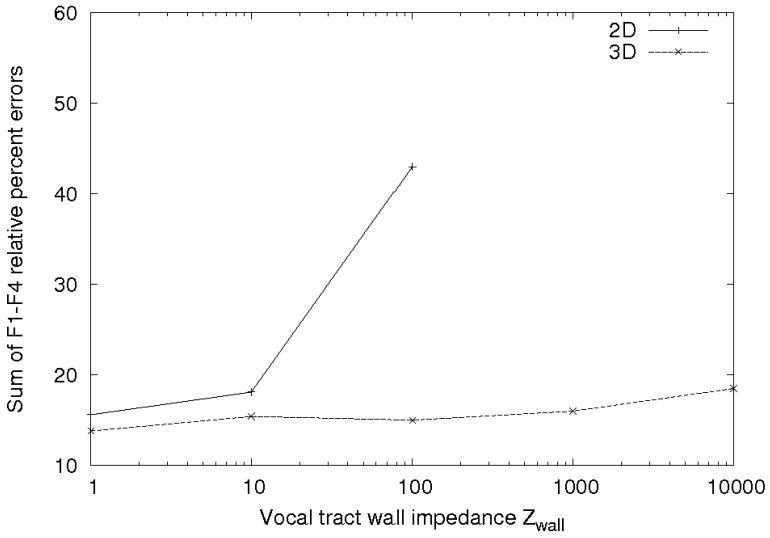

Figure 4.15: Comparison of the DWM conditions by the vocal tract wall impedance Z_{wall} for vowel /i/

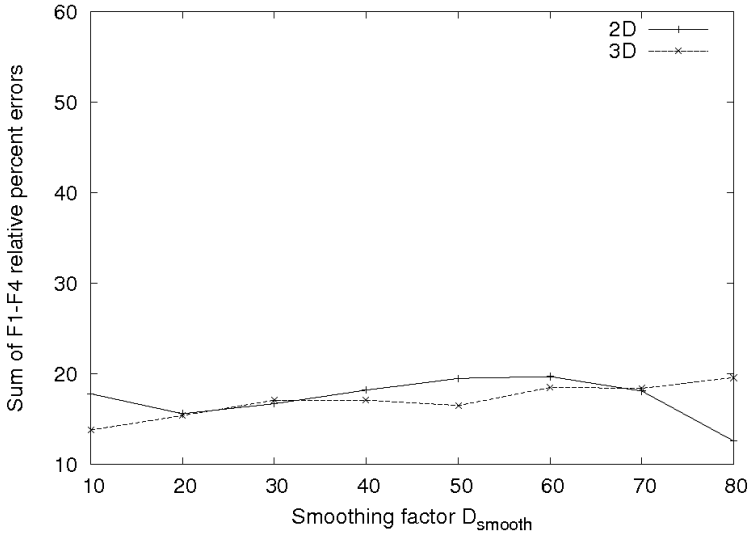

Figure 4.16: Comparison of the DWM conditions by the step distance of the smoothing process D_{smooth} for vowel /i/

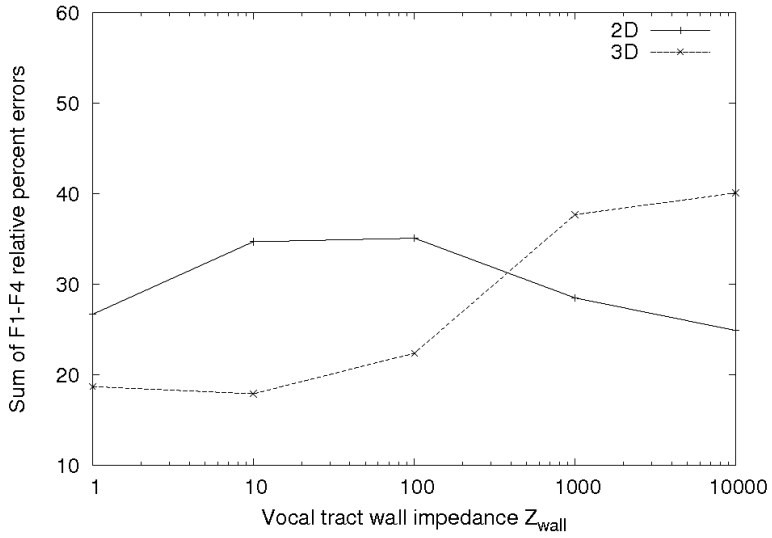

Figure 4.17: Comparison of the DWM conditions by the vocal tract wall impedance Z_{wall} for vowel /o/

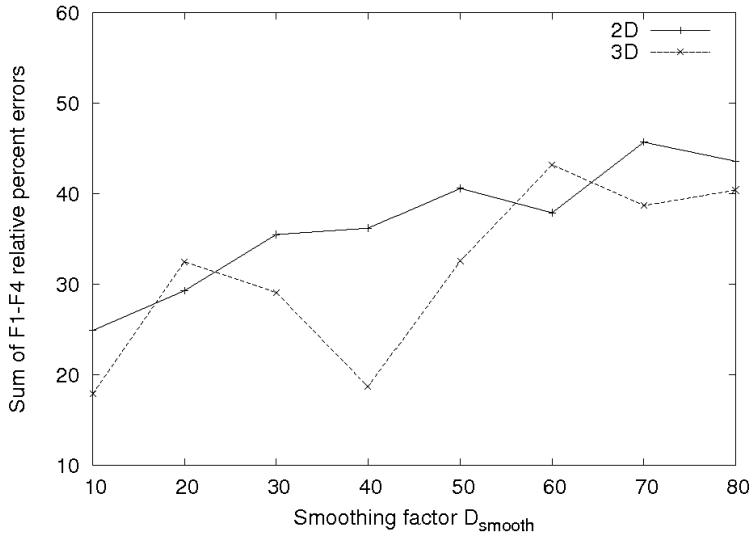

Figure 4.18: Comparison of the DWM conditions by the step distance of the smoothing process D_{smooth} for vowel /o/

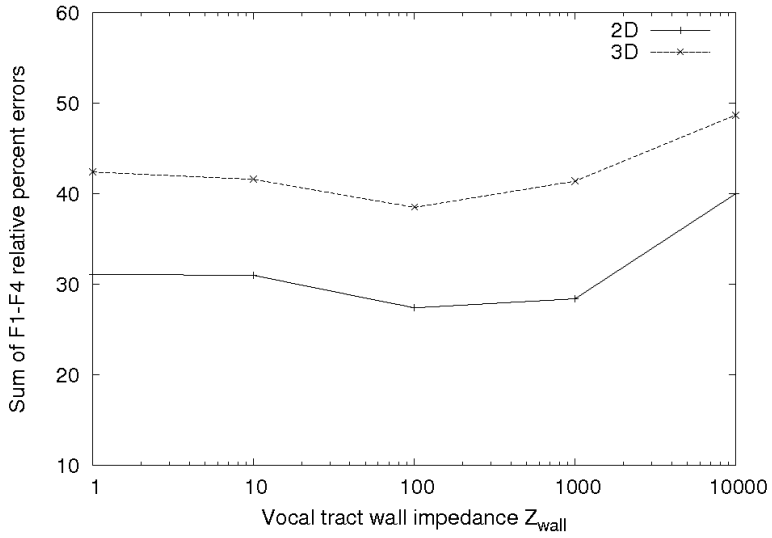

Figure 4.19: Comparison of the DWM conditions by the vocal tract wall impedance Z_{wall} for vowel /u/

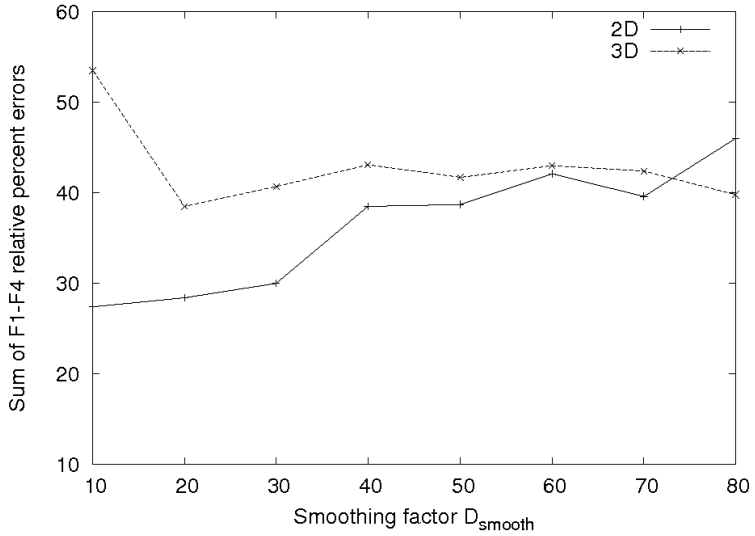

Figure 4.20: Comparison of the DWM conditions by the step distance of the smoothing process D_{smooth} for vowel /u/

plane. Considering the amount of the resources needed to compute the 3D simulation, 2D version of the DWM method in the mid-sagittal plane would be sufficient for this use.

To investigate the relationship between the smoothing factor and the consequent vocal tract length, vocal tract length is plotted for each smoothing factor $D_{smooth} = 10, 20, 30, 40, 50, 60, 70, 80$ for the most stable condition $Z_{wall} = 1$ in Fig. 4.21 through Fig. 4.25.

In the 2D versions, when the smoothing factor increases the vocal tract length monotonically decreases in almost all cases. The sole exception is the case of vowel /o/ (Fig. 4.24) where the $D_{smooth} = 40$ and 50. In the 3D versions, the vocal tract length tends to decreases when the smoothing factor increases, but large variation can be seen in figures Fig. 4.21 and Fig. 4.24.

Compared to the reported vocal tract lengths such as in [22], where the lengths are reported 17.750 for /a/, 17.250 for /e/, 17.500 for /i/, 18.500 for /o/, and 18.000 for /u/, the lower smoothing factors may be calculating the area functions too long for vowels /a/, /o/, and /u/ in this study. Considering the personal difference and the fact that no clear criteria for cutting the lip-side end areas is provided, nothing concrete can be said for small vocal tract length differences.

Since lower smoothing factors may be variable and tend to report longer vocal tract lengths, and higher smoothing factors tend to report vocal tract area functions which have worse RPEs, the author selected the medium value of $D_{smooth} = 60$ for comparison with other existing methods.

Combining with that the condition $Z_{wall} = 1$ was shown most stable, the condition $Z_{wall} = 1, D_{smooth} = 60$ was eventually chosen for the comparison in the next experiment. The calculated vocal tract area functions for the 2D and 3D DWM methods with that condition are depicted in Fig. 4.26 through Fig. 4.35 and their values are listed in Table 4.1 through Table 4.3.

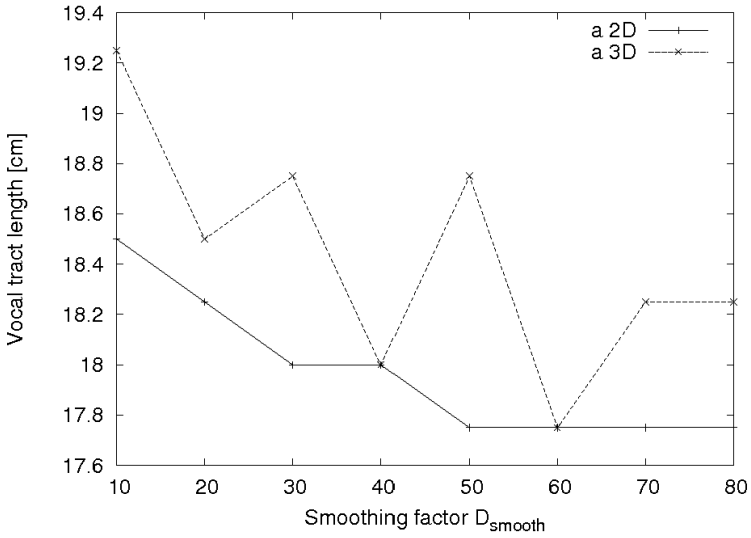

Figure 4.21: Calculated vocal tract length for each smoothing factor D_{smooth} for vowel /a/

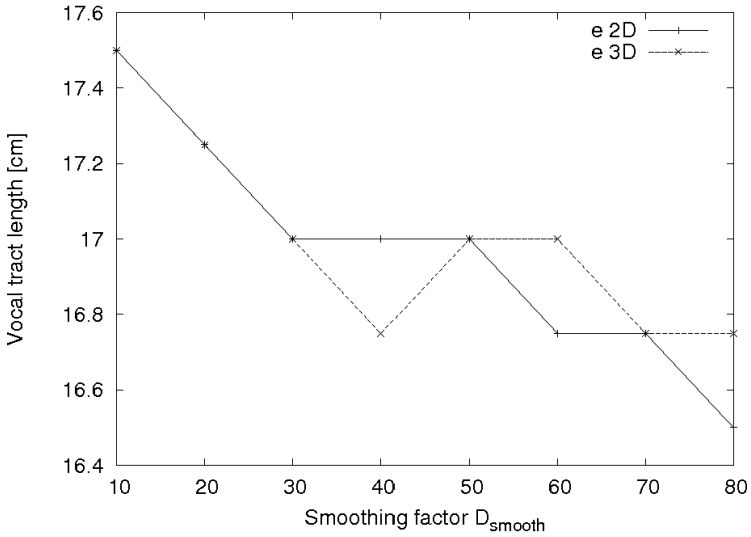

Figure 4.22: Calculated vocal tract length for each smoothing factor D_{smooth} for vowel /e/

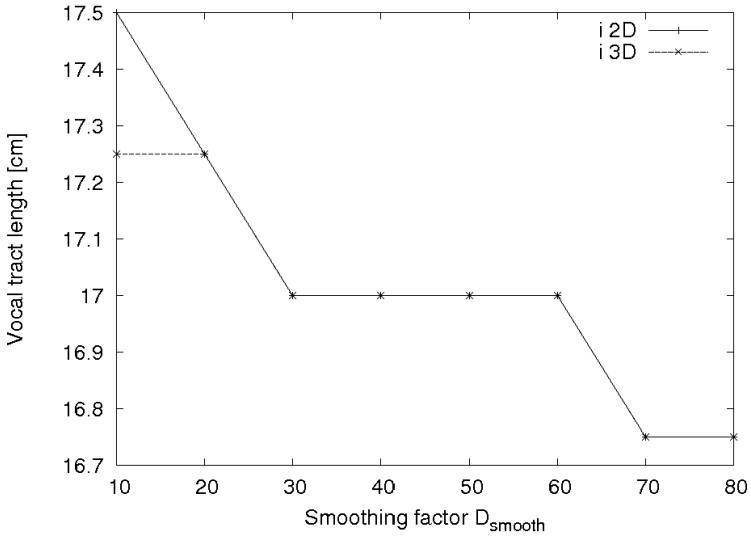

Figure 4.23: Calculated vocal tract length for each smoothing factor D_{smooth} for vowel /i/

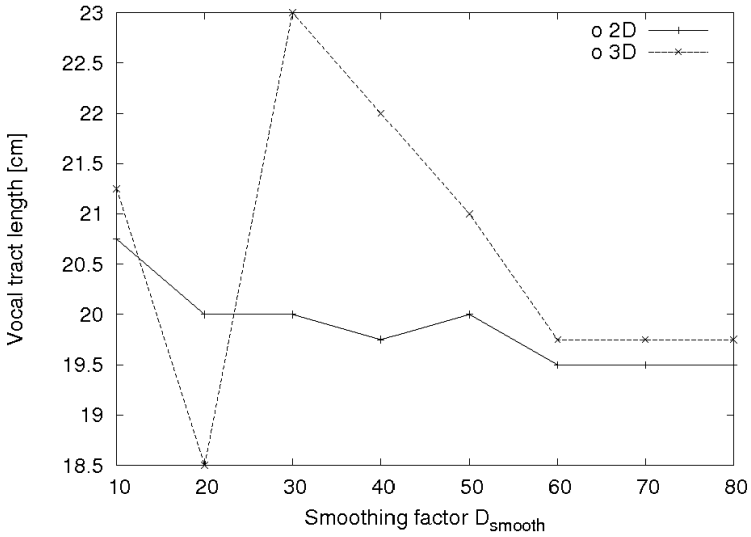

Figure 4.24: Calculated vocal tract length for each smoothing factor D_{smooth} for vowel /o/

Table 4.1: Calculated vocal tract area function from section 1–30 using 2D and 3D DWM methods with condition $Z_{wall} = 10, D_{smooth} = 60$ for five vowels. Units are cm^2. Each section length is 0.25 cm.

Section	a 2D	a 3D	e 2D	e 3D	i 2D	i 3D	o 2D	o 3D	u 2D	u 3D
1	0.55	0.61	0.53	0.56	0.85	0.56	0.56	0.56	0.42	0.54
2	1.02	1.15	0.54	0.57	0.66	0.66	0.95	0.96	1.26	1.21
3	0.90	0.77	0.48	0.46	1.11	1.08	0.57	0.55	0.77	0.72
4	0.60	0.56	0.40	0.37	0.67	0.57	0.49	0.49	0.56	0.54
5	0.58	0.51	0.36	0.36	0.49	0.45	0.48	0.47	0.54	0.53
6	0.43	0.47	0.32	0.31	0.41	0.38	0.45	0.41	0.57	0.51
7	0.36	0.35	0.23	0.24	0.39	0.37	0.43	0.41	0.55	0.56
8	0.36	0.40	0.15	0.16	0.35	0.33	0.46	0.44	0.57	0.61
9	0.41	0.42	0.17	0.17	0.39	0.38	0.47	0.53	0.68	0.70
10	0.46	0.45	0.35	0.33	0.47	0.47	0.61	0.58	0.82	0.83
11	0.68	0.70	2.19	0.91	0.62	0.64	1.12	0.99	2.67	1.32
12	2.73	1.39	4.26	3.82	2.55	1.14	4.01	2.69	5.20	4.75
13	4.49	4.31	4.39	4.91	4.75	4.48	4.55	4.69	5.50	6.22
14	4.53	5.17	4.18	4.86	5.16	5.74	4.37	5.06	5.45	6.40
15	4.29	4.87	4.11	4.47	5.19	5.97	4.00	4.59	5.52	6.14
16	3.95	4.22	3.90	4.07	5.21	5.81	3.81	4.06	5.36	5.57
17	3.42	3.63	3.50	3.61	5.06	5.58	3.25	3.41	5.26	5.24
18	2.94	3.06	3.18	3.35	5.01	5.17	2.72	2.88	4.87	4.88
19	2.42	2.60	3.11	3.19	4.85	4.91	2.71	2.46	4.45	4.67
20	2.78	2.33	3.81	3.32	4.75	4.85	3.09	2.67	5.21	4.51
21	2.62	2.72	3.90	3.98	5.22	4.99	2.96	3.10	5.38	5.12
22	2.63	2.63	4.13	3.98	5.71	5.28	2.92	2.96	5.51	5.62
23	2.37	2.55	4.35	4.27	5.79	5.79	2.79	2.92	5.51	5.50
24	2.21	2.36	4.57	4.48	6.12	5.92	2.48	2.74	5.62	5.57
25	1.86	2.10	4.25	4.57	6.12	6.08	2.06	2.37	5.30	5.77
26	1.60	1.74	3.97	4.17	5.87	5.81	1.74	1.93	4.67	4.95
27	1.23	1.44	3.83	3.92	5.56	5.73	1.31	1.43	4.11	4.34
28	1.03	1.14	3.60	3.72	5.28	5.49	1.12	1.19	3.66	3.84
29	0.85	0.97	3.37	3.46	4.89	5.07	0.84	0.92	3.19	3.38
30	0.78	0.80	3.23	3.22	4.76	4.75	0.66	0.76	2.97	3.02

Table 4.2: Calculated vocal tract area function from section 31–60.

Section	a 2D	a 3D	e 2D	e 3D	i 2D	i 3D	o 2D	o 3D	u 2D	u 3D
31	0.64	0.66	3.10	3.09	4.48	4.51	0.54	0.58	2.63	2.73
32	0.60	0.59	3.27	3.01	4.46	4.46	0.52	0.51	2.48	2.53
33	0.63	0.61	3.44	3.27	4.54	4.46	0.46	0.49	2.38	2.40
34	0.74	0.73	3.59	3.48	4.44	4.45	0.49	0.50	2.50	2.45
35	0.87	0.84	3.66	3.58	4.32	4.42	0.63	0.56	2.77	2.74
36	1.10	1.00	3.31	3.32	4.11	4.16	0.86	0.78	2.88	2.79
37	1.47	1.35	3.18	3.16	3.88	3.88	0.99	1.01	2.87	2.82
38	1.69	1.62	2.99	3.15	3.73	3.85	1.05	1.01	2.39	2.54
39	1.74	1.79	2.94	3.06	3.44	3.58	0.77	0.93	2.10	2.21
40	1.27	1.27	2.68	2.90	3.12	3.41	0.71	0.71	2.02	2.04
41	1.29	1.26	2.17	2.51	2.64	2.91	0.70	0.67	1.92	2.00
42	1.34	1.27	1.77	2.00	2.06	2.34	0.81	0.71	1.65	1.74
43	1.71	1.65	1.45	1.61	1.58	1.78	1.03	0.98	1.23	1.35
44	1.73	1.79	1.31	1.36	1.22	1.32	0.95	1.01	0.86	0.92
45	1.70	1.70	1.06	1.12	0.93	1.02	0.73	0.82	0.55	0.63
46	1.79	1.75	0.97	1.06	0.70	0.74	0.58	0.68	0.34	0.38
47	1.84	1.78	0.97	0.99	0.52	0.59	0.52	0.51	0.28	0.30
48	2.12	2.08	1.19	1.08	0.41	0.46	0.59	0.53	0.28	0.27
49	2.77	2.74	1.34	1.29	0.32	0.38	0.73	0.68	0.32	0.27
50	3.27	3.16	1.44	1.40	0.24	0.26	0.99	0.88	0.54	0.49
51	3.84	3.93	1.55	1.51	0.20	0.22	1.44	1.24	0.93	0.86
52	4.71	4.79	1.71	1.59	0.15	0.17	2.10	1.88	1.24	1.16
53	5.67	5.58	1.75	1.73	0.16	0.16	2.74	2.58	1.54	1.49
54	6.09	6.13	2.01	1.82	0.15	0.15	3.37	3.16	1.77	1.76
55	6.51	6.75	2.07	1.99	0.19	0.18	3.82	3.66	2.05	1.96
56	7.05	7.42	2.02	2.07	0.26	0.20	4.21	4.66	2.27	2.24
57	7.67	7.61	2.07	2.07	0.18	0.21	5.11	5.07	2.42	2.34
58	7.89	7.71	1.96	1.96	0.19	0.20	5.68	5.40	2.38	2.39
59	7.94	7.87	1.91	1.92	0.20	0.20	5.97	5.75	2.29	2.32
60	7.70	7.75	1.79	1.79	0.18	0.18	6.26	5.95	2.21	2.28

Table 4.3: Calculated vocal tract area function from section 61 to lip end. The bottom line denotes the total vocal tract length (VTL) in cm.

Section	a 2D	a 3D	e 2D	e 3D	i 2D	i 3D	o 2D	o 3D	u 2D	u 3D
61	7.64	7.65	1.77	1.79	0.20	0.16	6.32	6.25	2.01	2.10
62	7.14	7.29	1.57	1.69	0.27	0.24	6.42	6.50	1.92	1.98
63	6.74	6.77	1.36	1.53	0.28	0.26	6.15	6.22	1.66	1.68
64	6.39	6.55	1.17	1.15	0.32	0.28	6.06	6.02	1.41	1.49
65	6.31	6.28	1.20	1.13	0.32	0.27	6.15	6.13	0.99	1.06
66	5.99	5.68	1.67	1.54	0.43	0.41	5.85	5.89	0.81	0.79
67	5.25	5.56	2.25	1.81	0.50	0.47	5.60	5.64	0.60	0.55
68	5.10	5.33		2.46	0.67	0.50	5.18	5.00	0.50	0.56
69	5.37	5.27					5.37	5.24	0.53	0.46
70	5.45	5.61					5.54	5.35	0.76	0.64
71	5.77	5.67					5.55	5.58	1.05	1.06
72							4.89	5.58	1.17	1.16
73							3.65	4.64	1.05	1.05
74							2.56	3.55	0.64	0.64
75							1.78	2.85	0.24	0.36
76							1.04	2.40	0.20	0.17
77							0.40	1.68		
78							0.30	0.68		
79								0.35		
VTL	17.75	17.75	16.75	17.0	17.0	17.0	19.5	19.75	19.0	19.0

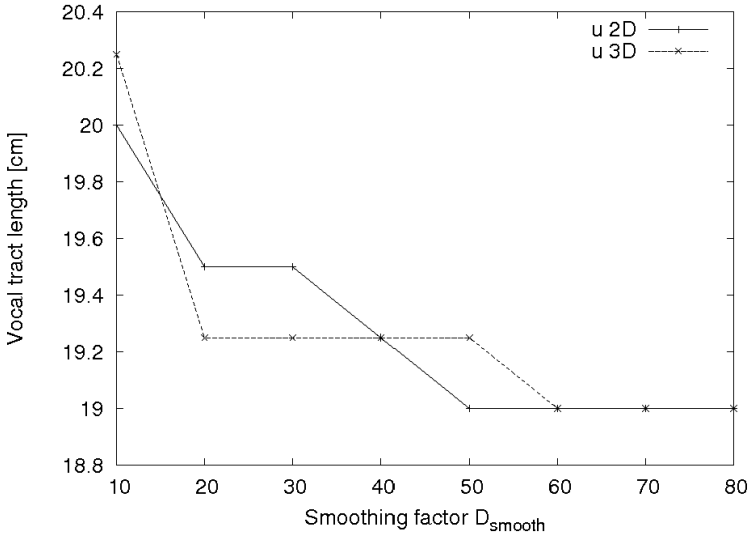

Figure 4.25: Calculated vocal tract length for each smoothing factor D_{smooth} for vowel /u/

4.3 Comparison against the solid model with existing methods

From the aforementioned discussion in the former sections, $Z_{wall} = 1, D_{smooth} = 60$ was found to be the most stable condition. With that condition, a comparison with other existing methods agaist the solid model was conducted and the result is listed in Table 4.4 through Table 4.8. DWM, MD, BS, and NE represent methods using digital waveguide mesh, Manhattan distance, iterative bisection, and nearest edge point, respectively. The VTL column denotes the total vocal tract length in cm. Δ represents the relative percent error (RPE) to the solid model. The RPEs for first formant (F1) for vowel /i/ were not calculated due to the lack of measurement for the solid model (see Kitamura *et al.* [6] for details).

The first and second formant frequencies for the solid model are notably lower than those of the typical male utterances. According to Kitamura *et al.* [6], this is because of the effect of rigid vocal tract. For all vowels except /i/,

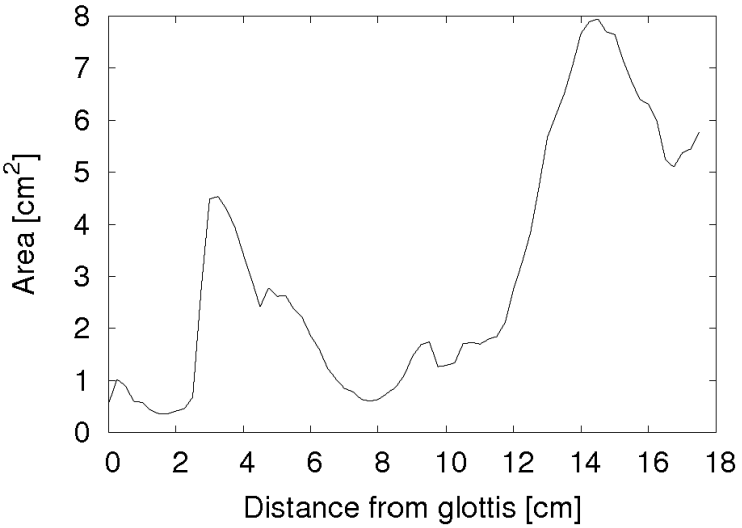

Figure 4.26: Calculated vocal tract area function using 2D DWM method with condition $Z_{wall} = 10, D_{smooth} = 60$ for vowel /a/

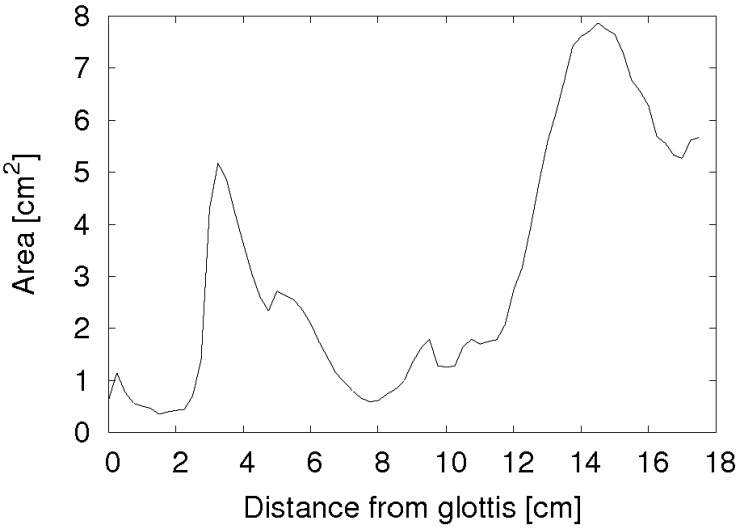

Figure 4.27: Calculated vocal tract area function using 3D DWM method with condition $Z_{wall} = 10, D_{smooth} = 60$ for vowel /a/

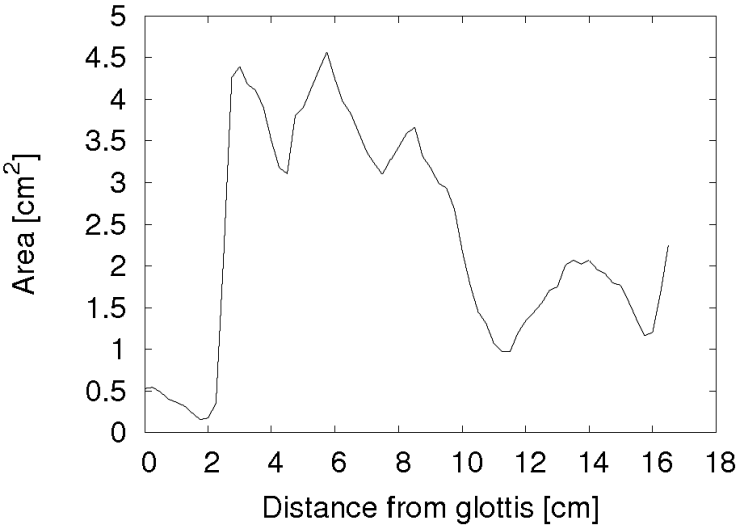

Figure 4.28: Calculated vocal tract area function using 2D DWM method with condition $Z_{wall} = 10, D_{smooth} = 60$ for vowel /e/

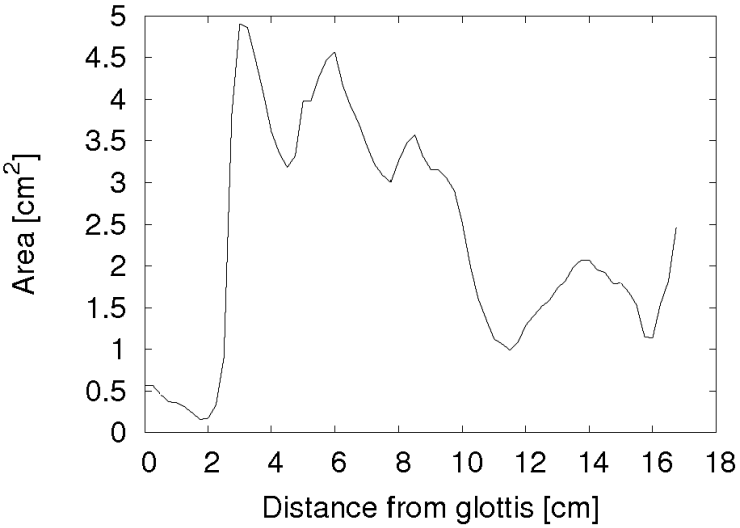

Figure 4.29: Calculated vocal tract area function using 3D DWM method with condition $Z_{wall} = 10, D_{smooth} = 60$ for vowel /e/

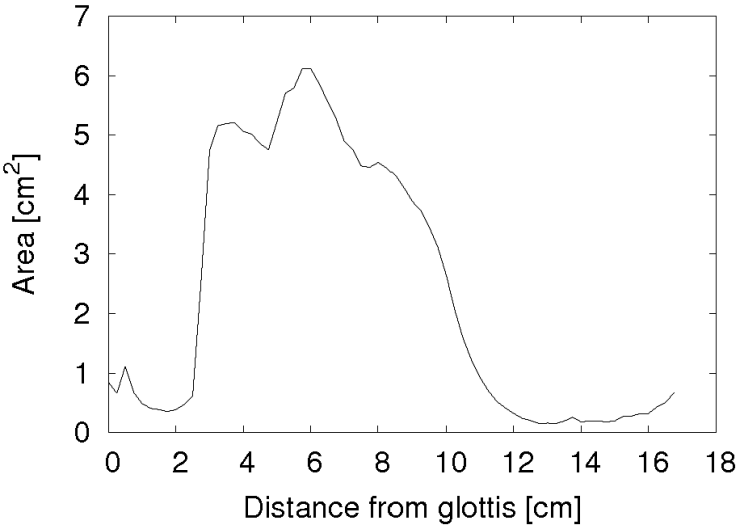

Figure 4.30: Calculated vocal tract area function using 2D DWM method with condition $Z_{wall} = 10, D_{smooth} = 60$ for vowel /i/

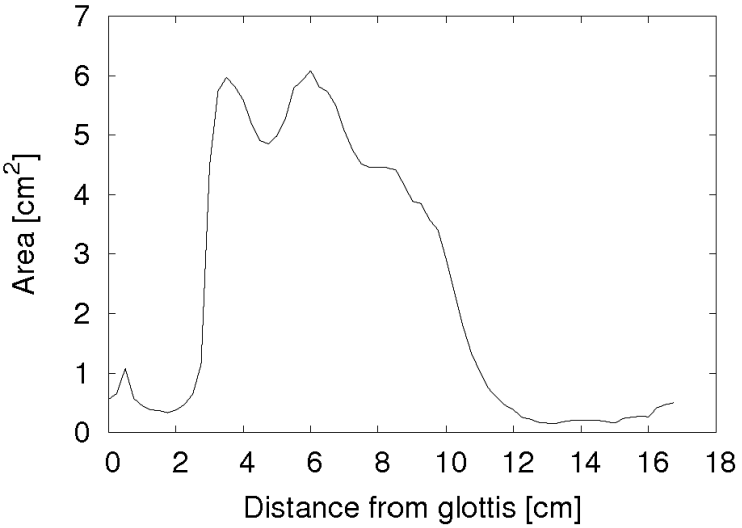

Figure 4.31: Calculated vocal tract area function using 3D DWM method with condition $Z_{wall} = 10, D_{smooth} = 60$ for vowel /i/

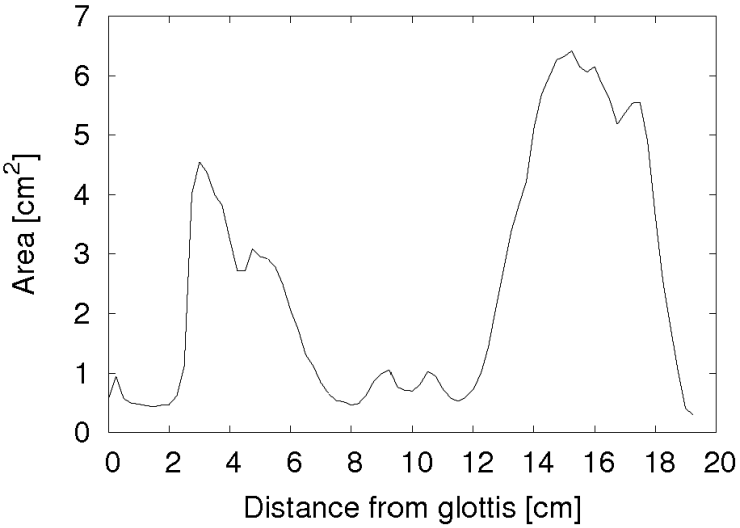

Figure 4.32: Calculated vocal tract area function using 2D DWM method with condition $Z_{wall} = 10, D_{smooth} = 60$ for vowel /o/

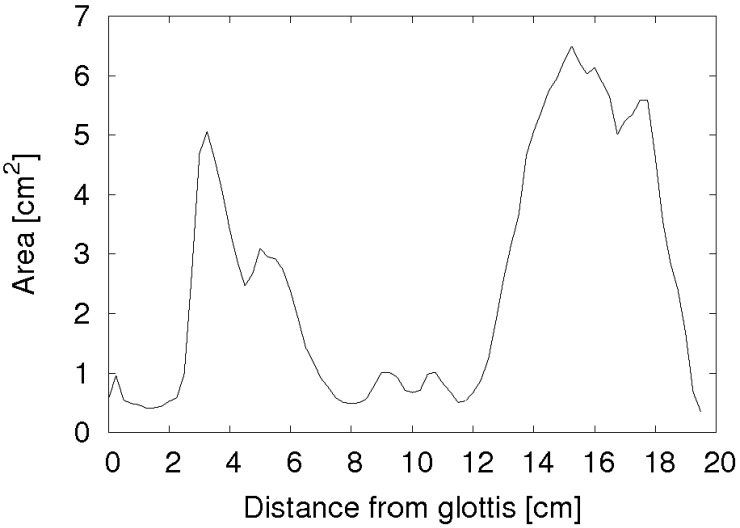

Figure 4.33: Calculated vocal tract area function using 3D DWM method with condition $Z_{wall} = 10, D_{smooth} = 60$ for vowel /o/

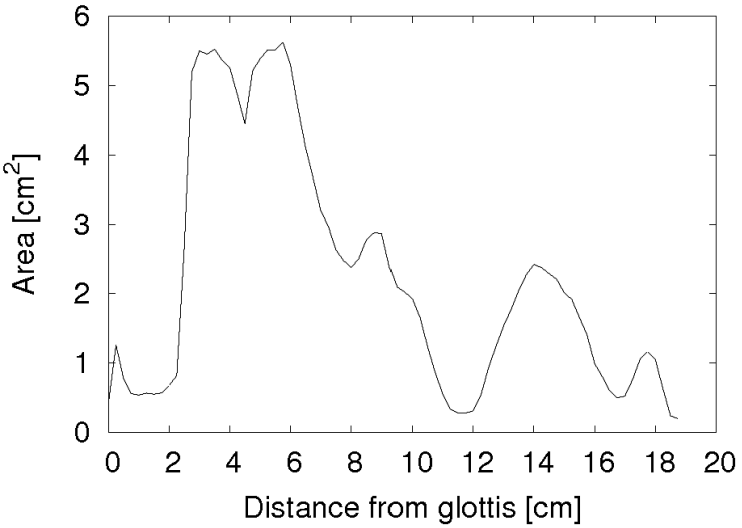

Figure 4.34: Calculated vocal tract area function using 2D DWM method with condition $Z_{wall} = 10, D_{smooth} = 60$ for vowel /u/

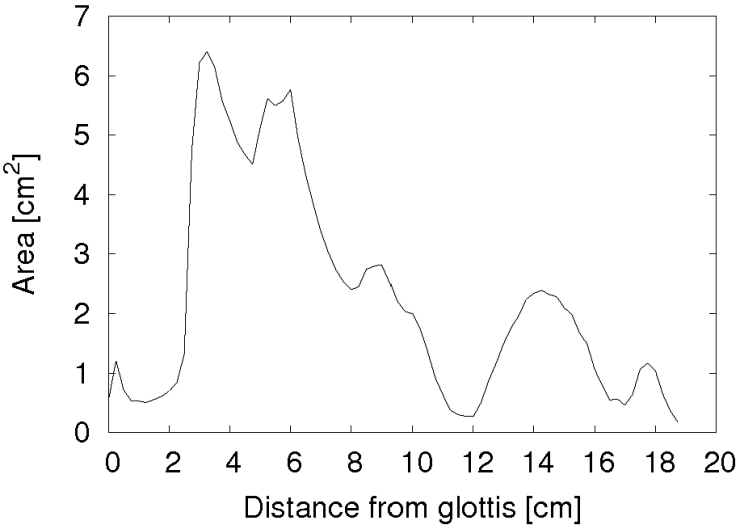

Figure 4.35: Calculated vocal tract area function using 3D DWM method with condition $Z_{wall} = 10, D_{smooth} = 60$ for vowel /u/

almost all RPEs for the first (F1) and second (F2) formants were reported greater than 10 percent, some scoring greater than 20 percent. The cause of these significant differences have not been addressed yet, but one possible reason is that the absence of the piriform fossa in the 1D area function model. The dimensions above the glottis region may also be involved.

In this result, the DWM methods were third-best for 2D and second for 3D for vowel /a/, fourth for 2D and second for 3D for vowel /e/, second for 2D and first for 3D for vowel /i/, fourth for 2D and fifth (worst) for 3D for vowel /o/, and second for 2D and first for 3D for vowel /u/. The result showed that the DWM methods had a slightly better result than other methods in general.

There were no significant difference between the 2D and 3D verions as discussed in the previous section.

4.4 Future work

Further research should consider the following issues.

Modeling inconsistency

The modeling inconsistency between the 3D vocal tract shape and 1D area function, the presence or absence of the piriform fossa, should be addressed in the future studies. A possible solution is adding the piriform fossa to the 1D area function, and another one is removing them from the 3D shape and analyzing the acoustic characteristics of the 3D shape. The DWM method can also be used for the acoustic analysis.

Lip radiation modeling

Using the 3D DWM method, the vocal tract area function can directly be computed from the contour pressure planes, without calculating the center-line of the 3D shape. The area functions derived in this manner contain the lip radiation characteristic in its model as the wave sound propagation will

Table 4.4: Comparison of lower four formants on each method against the solid model for vowel /a/

	VTL	F1	F2	F3	F4	Δ1	Δ2	Δ3	Δ4	$\sum\Delta$
Solid model		450	1070	2407	2696					
DWM 2D	17.75	539	1318	2531	2865	19.8	23.2	5.2	6.3	54.4
DWM 3D	17.75	533	1318	2514	2889	18.4	23.2	4.4	7.2	53.2
MD	17.75	533	1318	2496	2877	18.4	23.2	3.7	6.7	52.0
BS	17.75	527	1301	2092	2865	17.1	21.6	13.1	6.3	58.1
NE	17.75	527	1301	2625	2941	17.1	21.6	9.1	9.1	56.8

Table 4.5: Comparison of lower four formants on each method against the solid model for vowel /e/

	VTL	F1	F2	F3	F4	Δ1	Δ2	Δ3	Δ4	$\sum\Delta$
Solid model		370	1629	2193	2684					
DWM 2D	16.75	422	1852	2420	2854	14.1	13.7	10.4	6.3	44.4
DWM 3D	17.0	416	1828	2391	2813	12.4	12.2	9.0	4.8	38.5
MD	16.5	428	1869	2414	2865	15.7	14.7	10.1	6.7	47.2
BS	16.75	422	1529	1981	2783	14.1	6.1	9.7	3.7	33.5
NE	16.75	422	1846	2385	2848	14.1	13.3	8.8	6.1	42.2

Table 4.6: Comparison of lower four formants on each method against the solid model for vowel /i/

	VTL	F1	F2	F3	F4	Δ1	Δ2	Δ3	Δ4	$\sum\Delta$
Solid model		-	2038	2853	3134					
DWM 2D	17.0	170	2104	2690	3492	-	3.2	5.7	11.4	20.4
DWM 3D	17.0	164	2109	2707	3481	-	3.5	5.1	11.1	19.7
MD	16.5	170	2191	2906	3609	-	7.5	1.9	15.2	24.5
BS	17.0	170	1746	2467	3475	-	14.3	13.5	10.9	38.7
NE	16.75	170	2109	2695	3539	-	3.5	5.5	12.9	21.9

Table 4.7: Comparison of lower four formants on each method against the solid model for vowel /o/

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		293	616	2175	2924					
DWM 2D	19.5	352	721	2385	2988	20.1	17.0	9.7	2.2	49.0
DWM 3D	19.75	363	762	2379	2947	23.9	23.7	9.4	0.8	57.8
MD	19.75	346	703	2385	2936	18.1	14.1	9.7	0.4	42.3
BS	19.75	334	703	2027	2531	14.0	14.1	6.8	13.4	48.4
NE	19.75	346	709	2350	2912	18.1	15.1	8.0	0.4	41.6

Table 4.8: Comparison of lower four formants on each method against the solid model for vowel /u/

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		196	934	2136	3091					
DWM 2D	19.0	234	1078	2291	3182	19.4	15.4	7.3	2.9	45.0
DWM 3D	19.0	234	1084	2279	3135	19.4	16.1	6.7	1.4	43.6
MD	19.0	234	1084	2315	3176	19.4	16.1	8.4	2.7	46.6
BS	19.25	229	1061	2063	2654	16.8	13.6	3.4	14.1	48.0
NE	19.0	234	1066	2285	3258	19.4	14.1	7.0	5.4	45.9

be calculated after the wave reaches the lip end. As it is not obvious that the boundary between the vocal tract and air, it is currently problematic to decide the lip-side end of the vocal tract. The 3D DWM method may possibly solve this problem by calculating the wave propagation to the place where the sound wave pressure would be measured in the speech synthesis simulation.

Vocal tract with stops and branches

The current proposed method cannot be applied to certain stop consonants such as /k/, /t/ or /m/ as the wave cannot reach the opposite side of the stops caused by the tongue. It also cannot be directly used for the consonants which make two separate airways in the vocal tract such as /l/. Those separated sideways can be modeled as branches and the extraction method needs to deal with them.

Chapter 5

Conclusions

It was shown in this study that the proposed method using the DWM can be used to estimate the vocal tract area function from 3D MRI data. The parameters of the DWM simulation need to be set adequately to conduct the estimation. The smoothing factor has a greater effect in the formants of the resulted vocal tract area function than the parameters of the DWM simulation.

The comparison to the reference solid model, with other existing methods, showed that the proposed method can produce the vocal tract area functions more precisely matched to the reference model than existing methods in many cases.

Appendix A

Calculated formants

The following lists are the calculated lower four formants of all the vocal tract area functions obtained by DWM methods. The VTL column denotes the total vocal tract length in cm. Δ represents the relative percent error (RPE) to the solid model. The first column is the condition of the simulation, the dimension, smoothing factor D_{smooth}, and vocal tract wall impedance Z_{wall} from left to right.

The list lacks some conditions which the lower four formants could have not been properly calculated or failed to be picked up the peaks.

Table A.1: Lower four formants obtained by DWM method compared against the solid model for vowel /a/ (page 1)

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		450	1070	2407	2696					
2D 10 1	18.5	527	1272	2402	2701	17.1	18.9	0.2	0.2	36.4
2D 10 10	18.25	527	1289	2438	2748	17.1	20.5	1.3	1.9	40.8
2D 10 100	18.5	527	1283	2332	2707	17.1	19.9	3.1	0.4	40.5
2D 10 1000	18.75	522	1266	2356	2719	16.0	18.3	2.1	0.9	37.3
2D 10 10000	18.75	516	1272	2338	2660	14.7	18.9	2.9	1.3	37.7
2D 20 1	18.25	527	1283	2367	2731	17.1	19.9	1.7	1.3	40.0
2D 20 10	18.5	527	1272	2373	2731	17.1	18.9	1.4	1.3	38.7
2D 20 100	18.25	527	1295	2379	2742	17.1	21.0	1.2	1.7	41.0
2D 20 1000	18.25	527	1289	2397	2754	17.1	20.5	0.4	2.2	40.1
2D 20 10000	18.25	527	1289	2402	2748	17.1	20.5	0.2	1.9	39.7
2D 30 1	18.0	539	1307	2479	2836	19.8	22.1	3.0	5.2	50.1
2D 30 10	18.25	533	1289	2443	2801	18.4	20.5	1.5	3.9	44.3
2D 30 100	18.25	533	1283	2455	2801	18.4	19.9	2.0	3.9	44.2
2D 30 1000	18.25	533	1283	2455	2813	18.4	19.9	2.0	4.3	44.7
2D 30 10000	18.0	539	1295	2484	2824	19.8	21.0	3.2	4.7	48.8
2D 40 1	18.0	527	1301	2479	2818	17.1	21.6	3.0	4.5	46.2
2D 40 10	18.0	527	1289	2467	2818	17.1	20.5	2.5	4.5	44.6
2D 40 100	18.0	527	1289	2467	2813	17.1	20.5	2.5	4.3	44.4
2D 40 1000	18.0	527	1283	2473	2818	17.1	19.9	2.7	4.5	44.3
2D 40 10000	18.0	533	1289	2502	2842	18.4	20.5	3.9	5.4	48.3
2D 50 1	17.75	533	1324	2543	2859	18.4	23.7	5.7	6.0	53.9
2D 50 10	18.0	533	1295	2537	2859	18.4	21.0	5.4	6.0	50.9
2D 50 100	18.0	533	1289	2537	2854	18.4	20.5	5.4	5.9	50.2
2D 50 1000	18.0	533	1289	2549	2865	18.4	20.5	5.9	6.3	51.1
2D 50 10000	17.75	539	1307	2555	2883	19.8	22.1	6.1	6.9	55.0
2D 60 1	17.75	539	1318	2531	2865	19.8	23.2	5.2	6.3	54.4
2D 60 10	18.0	533	1295	2531	2859	18.4	21.0	5.2	6.0	50.7
2D 60 100	18.0	533	1295	2531	2865	18.4	21.0	5.2	6.3	50.9
2D 60 1000	18.0	533	1295	2531	2859	18.4	21.0	5.2	6.0	50.7
2D 60 10000	17.75	539	1307	2549	2889	19.8	22.1	5.9	7.2	55.0

Table A.2: Lower four formants obtained by DWM method compared against the solid model for vowel /a/ (page 2)

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		450	1070	2407	2696					
3D 10 1	19.25	486	1248	2426	2719	8.0	16.6	0.8	0.9	26.3
3D 10 10000	18.5	492	1377	2438	2848	9.3	28.7	1.3	5.6	45.0
3D 20 1	18.5	504	1313	2432	2783	12.0	22.7	1.0	3.2	39.0
3D 20 10	19.0	492	1236	2420	2754	9.3	15.5	0.5	2.2	27.5
3D 20 100	19.0	492	1225	2432	2760	9.3	14.5	1.0	2.4	27.2
3D 20 1000	19.5	486	1166	2443	2748	8.0	9.0	1.5	1.9	20.4
3D 20 10000	18.0	498	1371	2479	2871	10.7	28.1	3.0	6.5	48.3
3D 30 1	18.75	516	1242	2432	2807	14.7	16.1	1.0	4.1	35.9
3D 30 10	18.75	498	1242	2461	2795	10.7	16.1	2.2	3.7	32.7
3D 30 100	18.75	504	1236	2461	2795	12.0	15.5	2.2	3.7	33.4
3D 30 1000	18.25	510	1307	2479	2848	13.3	22.1	3.0	5.6	44.1
3D 30 10000	18.25	516	1289	2490	2854	14.7	20.5	3.4	5.9	44.4
3D 40 1	18.0	533	1348	2467	2859	18.4	26.0	2.5	6.0	53.0
3D 40 10	18.75	510	1225	2467	2813	13.3	14.5	2.5	4.3	34.7
3D 40 100	18.75	516	1219	2449	2807	14.7	13.9	1.7	4.1	34.5
3D 40 1000	18.25	522	1277	2461	2842	16.0	19.3	2.2	5.4	43.0
3D 40 10000	18.25	522	1266	2455	2854	16.0	18.3	2.0	5.9	42.2
3D 50 1	18.75	522	1207	2402	2801	16.0	12.8	0.2	3.9	32.9
3D 50 10	18.75	516	1219	2414	2807	14.7	13.9	0.3	4.1	33.0
3D 50 100	18.75	516	1213	2408	2795	14.7	13.4	0.0	3.7	31.7
3D 50 1000	18.75	522	1207	2420	2801	16.0	12.8	0.5	3.9	33.2
3D 50 10000	18.5	522	1236	2484	2859	16.0	15.5	3.2	6.0	40.8
3D 60 1	17.75	533	1318	2514	2889	18.4	23.2	4.4	7.2	53.2
3D 60 10	18.25	527	1266	2484	2848	17.1	18.3	3.2	5.6	44.3
3D 60 100	18.25	527	1260	2502	2859	17.1	17.8	3.9	6.0	44.9
3D 60 1000	18.0	533	1289	2520	2889	18.4	20.5	4.7	7.2	50.8
3D 60 10000	18.0	533	1283	2525	2900	18.4	19.9	4.9	7.6	50.8

Table A.3: Lower four formants obtained by DWM method compared against the solid model for vowel /e/ (page 1)

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		370	1629	2193	2684					
2D 10 1	17.5	410	1793	2326	2736	10.8	10.1	6.1	1.9	28.9
2D 10 10	17.5	404	1787	2250	2701	9.2	9.7	2.6	0.6	22.1
2D 10 100	17.75	404	1740	2262	2690	9.2	6.8	3.1	0.2	19.4
2D 20 1	17.25	410	1793	2361	2783	10.8	10.1	7.7	3.7	32.2
2D 20 10	17.25	410	1799	2338	2772	10.8	10.4	6.6	3.3	31.1
2D 20 100	17.25	410	1799	2332	2772	10.8	10.4	6.3	3.3	30.9
2D 20 1000	17.5	410	1758	2326	2754	10.8	7.9	6.1	2.6	27.4
2D 30 1	17.0	416	1828	2379	2813	12.4	12.2	8.5	4.8	37.9
2D 30 10	17.0	410	1816	2350	2795	10.8	11.5	7.2	4.1	33.6
2D 30 100	17.0	422	1828	2373	2830	14.1	12.2	8.2	5.4	39.9
2D 30 1000	17.0	422	1811	2344	2813	14.1	11.2	6.9	4.8	36.9
2D 40 1	17.0	416	1822	2320	2783	12.4	11.8	5.8	3.7	33.8
2D 40 10	17.0	416	1822	2332	2795	12.4	11.8	6.3	4.1	34.8
2D 40 100	17.0	422	1834	2379	2836	14.1	12.6	8.5	5.7	40.8
2D 40 1000	16.75	422	1846	2373	2830	14.1	13.3	8.2	5.4	41.0
2D 50 1	17.0	416	1822	2303	2789	12.4	11.8	5.0	3.9	33.2
2D 50 10	17.0	416	1822	2291	2783	12.4	11.8	4.5	3.7	32.4
2D 50 100	16.75	422	1852	2350	2830	14.1	13.7	7.2	5.4	40.3
2D 50 1000	16.75	422	1863	2361	2848	14.1	14.4	7.7	6.1	42.2
2D 60 1	16.75	422	1852	2420	2854	14.1	13.7	10.4	6.3	44.4
2D 60 10	16.75	422	1852	2438	2859	14.1	13.7	11.2	6.5	45.4
2D 60 100	16.5	428	1869	2397	2865	15.7	14.7	9.3	6.7	46.5
2D 60 1000	16.5	422	1875	2461	2889	14.1	15.1	12.2	7.6	49.0

Table A.4: Lower four formants obtained by DWM method compared against the solid model for vowel /e/ (page 2)

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		370	1629	2193	2684					
3D 10 1	17.5	404	1793	2361	2731	9.2	10.1	7.7	1.8	28.7
3D 10 10	17.0	416	1846	2338	2772	12.4	13.3	6.6	3.3	35.6
3D 10 100	17.25	410	1799	2291	2731	10.8	10.4	4.5	1.8	27.5
3D 10 1000	17.5	404	1787	2279	2719	9.2	9.7	3.9	1.3	24.1
3D 10 10000	17.5	404	1787	2273	2707	9.2	9.7	3.6	0.9	23.4
3D 20 1	17.25	416	1822	2326	2766	12.4	11.8	6.1	3.1	33.4
3D 20 10	17.0	416	1846	2408	2807	12.4	13.3	9.8	4.6	40.1
3D 20 100	16.75	416	1840	2367	2783	12.4	13.0	7.9	3.7	37.0
3D 20 1000	17.25	410	1805	2291	2754	10.8	10.8	4.5	2.6	28.7
3D 20 10000	17.25	416	1811	2326	2777	12.4	11.2	6.1	3.5	33.1
3D 30 1	17.0	416	1828	2361	2801	12.4	12.2	7.7	4.4	36.7
3D 30 10	17.0	416	1846	2367	2795	12.4	13.3	7.9	4.1	37.8
3D 30 100	17.0	416	1834	2361	2789	12.4	12.6	7.7	3.9	36.6
3D 30 1000	17.0	416	1828	2361	2783	12.4	12.2	7.7	3.7	36.0
3D 30 10000	17.0	416	1828	2373	2801	12.4	12.2	8.2	4.4	37.2
3D 40 1	16.75	422	1863	2338	2824	14.1	14.4	6.6	5.2	40.2
3D 40 10	16.75	422	1869	2379	2818	14.1	14.7	8.5	5.0	42.3
3D 40 100	17.0	416	1828	2320	2783	12.4	12.2	5.8	3.7	34.1
3D 40 1000	17.0	416	1828	2356	2801	12.4	12.2	7.4	4.4	36.4
3D 40 10000	16.75	422	1846	2367	2818	14.1	13.3	7.9	5.0	40.3
3D 50 1	17.0	416	1828	2326	2795	12.4	12.2	6.1	4.1	34.8
3D 50 10	17.0	416	1822	2309	2772	12.4	11.8	5.3	3.3	32.8
3D 50 100	17.0	416	1828	2338	2795	12.4	12.2	6.6	4.1	35.4
3D 50 1000	17.0	416	1834	2391	2818	12.4	12.6	9.0	5.0	39.0
3D 50 10000	16.75	422	1852	2367	2830	14.1	13.7	7.9	5.4	41.1
3D 60 1	17.0	416	1828	2391	2813	12.4	12.2	9.0	4.8	38.5
3D 60 10	17.0	416	1834	2315	2789	12.4	12.6	5.6	3.9	34.5
3D 60 100	17.0	422	1840	2356	2813	14.1	13.0	7.4	4.8	39.2
3D 60 1000	17.0	422	1840	2344	2818	14.1	13.0	6.9	5.0	38.9
3D 60 10000	17.0	422	1846	2356	2830	14.1	13.3	7.4	5.4	40.2

Table A.5: Lower four formants obtained by DWM method compared against the solid model for vowel /i/ (page 1)

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		-	2038	2853	3134					
2D 10 1	17.5	164	2086	2695	3451	-	2.4	5.5	10.1	18.0
2D 10 10	19.25	152	1828	2156	2766	-	10.3	24.4	11.7	46.5
2D 20 1	17.25	170	2115	2754	3387	-	3.8	3.5	8.1	15.3
2D 20 10	17.75	164	1875	2127	2748	-	8.0	25.4	12.3	45.8
2D 30 1	17.0	164	2150	2830	3457	-	5.5	0.8	10.3	16.6
2D 30 10	17.0	164	2139	2789	3498	-	5.0	2.2	11.6	18.8
2D 40 1	17.0	164	2133	2760	3498	-	4.7	3.3	11.6	19.5
2D 40 10	17.0	170	2127	2748	3469	-	4.4	3.7	10.7	18.7
2D 50 1	17.0	164	2086	2648	3451	-	2.4	7.2	10.1	19.7
2D 50 10	17.0	170	2104	2678	3492	-	3.2	6.1	11.4	20.8
2D 60 1	17.0	170	2104	2690	3492	-	3.2	5.7	11.4	20.4
2D 60 10	17.0	170	2104	2695	3463	-	3.2	5.5	10.5	19.3

Table A.6: Lower four formants obtained by DWM method compared against the solid model for vowel /i/ (page 2)

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		-	2038	2853	3134					
3D 10 1	17.25	158	2086	2772	3416	-	2.4	2.8	9.0	14.2
3D 10 10	17.5	158	2080	2701	3393	-	2.1	5.3	8.3	15.7
3D 10 100	17.75	158	2080	2701	3381	-	2.1	5.3	7.9	15.3
3D 10 1000	20.0	398	2203	2801	4313	-	8.1	1.8	37.6	47.5
3D 10 10000	23.5	527	2273	2842	4377	-	11.5	0.4	39.7	51.6
3D 20 1	17.25	164	2092	2748	3434	-	2.6	3.7	9.6	15.9
3D 20 10	17.25	164	2109	2789	3475	-	3.5	2.2	10.9	16.6
3D 20 100	17.25	164	2080	2666	3498	-	2.1	6.6	11.6	20.2
3D 20 1000	17.5	164	2080	2684	3375	-	2.1	5.9	7.7	15.7
3D 20 10000	18.25	158	1494	2414	2941	-	26.7	15.4	6.2	48.2
3D 30 1	17.0	164	2098	2748	3504	-	2.9	3.7	11.8	18.4
3D 30 10	17.0	164	2109	2766	3492	-	3.5	3.0	11.4	18.0
3D 30 100	17.0	164	2115	2789	3481	-	3.8	2.2	11.1	17.1
3D 30 1000	17.0	164	2133	2813	3481	-	4.7	1.4	11.1	17.1
3D 30 10000	17.75	164	1799	2637	3199	-	11.7	7.6	2.1	21.4
3D 40 1	17.0	164	2121	2742	3481	-	4.1	3.9	11.1	19.0
3D 40 10	17.0	164	2121	2754	3434	-	4.1	3.5	9.6	17.1
3D 40 100	17.0	164	2121	2760	3475	-	4.1	3.3	10.9	18.2
3D 40 1000	17.25	164	2080	2654	3451	-	2.1	7.0	10.1	19.2
3D 40 10000	17.25	170	2057	2631	3469	-	0.9	7.8	10.7	19.4
3D 50 1	17.0	164	2127	2760	3486	-	4.4	3.3	11.2	18.9
3D 50 10	17.0	170	2133	2789	3434	-	4.7	2.2	9.6	16.5
3D 50 100	17.0	170	2133	2801	3510	-	4.7	1.8	12.0	18.5
3D 50 1000	17.0	170	2133	2777	3457	-	4.7	2.7	10.3	17.6
3D 50 10000	17.0	170	2115	2766	3492	-	3.8	3.0	11.4	18.3
3D 60 1	17.0	164	2109	2707	3481	-	3.5	5.1	11.1	19.7
3D 60 10	17.25	164	2115	2748	3533	-	3.8	3.7	12.7	20.2
3D 60 100	17.25	170	2127	2766	3486	-	4.4	3.0	11.2	18.6
3D 60 1000	17.0	170	2133	2777	3539	-	4.7	2.7	12.9	20.2
3D 60 10000	17.0	170	2133	2772	3539	-	4.7	2.8	12.9	20.4

Table A.7: Lower four formants obtained by DWM method compared against the solid model for vowel /o/ (page 1)

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		293	616	2175	2924					
2D 10 1	20.75	328	668	2203	2777	11.9	8.4	1.3	5.0	26.7
2D 10 10	20.5	334	686	2244	2830	14.0	11.4	3.2	3.2	31.7
2D 10 100	20.5	340	686	2250	2824	16.0	11.4	3.4	3.4	34.3
2D 10 1000	20.75	328	674	2221	2807	11.9	9.4	2.1	4.0	27.5
2D 10 10000	21.0	334	668	2186	2813	14.0	8.4	0.5	3.8	26.7
2D 20 1	20.0	346	709	2291	2824	18.1	15.1	5.3	3.4	41.9
2D 20 10	20.25	340	691	2273	2813	16.0	12.2	4.5	3.8	36.5
2D 20 100	20.25	340	691	2279	2830	16.0	12.2	4.8	3.2	36.2
2D 20 1000	20.25	334	680	2297	2889	14.0	10.4	5.6	1.2	31.2
2D 20 10000	20.0	340	697	2303	2912	16.0	13.1	5.9	0.4	35.5
2D 30 1	20.0	340	691	2315	2854	16.0	12.2	6.4	2.4	37.0
2D 30 10	20.0	340	691	2326	2854	16.0	12.2	6.9	2.4	37.6
2D 30 100	20.0	340	686	2315	2877	16.0	11.4	6.4	1.6	35.4
2D 30 1000	20.0	334	686	2320	2906	14.0	11.4	6.7	0.6	32.6
2D 30 10000	19.75	346	697	2350	2947	18.1	13.1	8.0	0.8	40.1
2D 40 1	19.75	340	691	2356	2959	16.0	12.2	8.3	1.2	37.7
2D 40 10	19.75	340	697	2356	2947	16.0	13.1	8.3	0.8	38.3
2D 40 100	19.75	340	697	2338	2936	16.0	13.1	7.5	0.4	37.1
2D 40 1000	19.75	340	691	2356	2959	16.0	12.2	8.3	1.2	37.7
2D 40 10000	19.75	346	703	2350	2865	18.1	14.1	8.0	2.0	42.3
2D 50 1	20.0	340	691	2309	2672	16.0	12.2	6.2	8.6	43.0
2D 50 10	20.0	340	691	2320	2690	16.0	12.2	6.7	8.0	42.9
2D 50 100	20.0	340	697	2344	2783	16.0	13.1	7.8	4.8	41.8
2D 50 1000	20.0	340	697	2350	2807	16.0	13.1	8.0	4.0	41.2
2D 50 10000	19.75	346	709	2361	2865	18.1	15.1	8.6	2.0	43.8
2D 60 1	19.5	352	721	2385	2988	20.1	17.0	9.7	2.2	49.0
2D 60 10	19.75	340	697	2350	2965	16.0	13.1	8.0	1.4	38.6
2D 60 100	19.75	340	697	2361	2971	16.0	13.1	8.6	1.6	39.3
2D 60 1000	19.75	340	697	2361	2982	16.0	13.1	8.6	2.0	39.7
2D 60 10000	19.5	346	709	2385	3012	18.1	15.1	9.7	3.0	45.9

Table A.8: Lower four formants obtained by DWM method compared against the solid model for vowel /o/ (page 2)

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		293	616	2175	2924					
3D 10 1	21.25	117	486	2127	2625	60.1	21.1	2.2	10.2	93.6
3D 10 10	21.25	311	639	2250	2807	6.1	3.7	3.4	4.0	17.3
3D 10 100	21.75	322	639	2232	2713	9.9	3.7	2.6	7.2	23.5
3D 20 10	20.5	334	686	2309	2865	14.0	11.4	6.2	2.0	33.5
3D 20 1000	20.5	334	691	2332	2865	14.0	12.2	7.2	2.0	35.4
3D 30 1	23.0	322	598	2086	2555	9.9	2.9	4.1	12.6	29.5
3D 30 10	20.25	346	715	2338	2848	18.1	16.1	7.5	2.6	44.3
3D 30 100	20.75	334	674	2315	2842	14.0	9.4	6.4	2.8	32.6
3D 30 1000	20.0	352	750	2356	2865	20.1	21.8	8.3	2.0	52.2
3D 30 10000	20.25	352	721	2367	2871	20.1	17.0	8.8	1.8	47.8
3D 40 1	22.0	322	609	2180	2643	9.9	1.1	0.2	9.6	20.9
3D 40 10	20.25	340	691	2332	2818	16.0	12.2	7.2	3.6	39.1
3D 40 100	20.0	340	709	2350	2836	16.0	15.1	8.0	3.0	42.2
3D 40 1000	20.0	346	715	2367	2854	18.1	16.1	8.8	2.4	45.4
3D 40 10000	20.0	340	697	2379	2854	16.0	13.1	9.4	2.4	41.0
3D 50 1	21.0	340	650	2279	2783	16.0	5.5	4.8	4.8	31.2
3D 50 10	20.0	346	709	2356	2848	18.1	15.1	8.3	2.6	44.1
3D 50 100	20.0	352	732	2361	2848	20.1	18.8	8.6	2.6	50.1
3D 50 1000	20.0	346	709	2379	2865	18.1	15.1	9.4	2.0	44.6
3D 50 10000	20.0	340	697	2397	2877	16.0	13.1	10.2	1.6	41.0
3D 60 1	19.75	363	762	2379	2947	23.9	23.7	9.4	0.8	57.8
3D 60 10	19.75	357	744	2367	2930	21.8	20.8	8.8	0.2	51.7
3D 60 100	20.0	357	732	2350	2777	21.8	18.8	8.0	5.0	53.7
3D 60 1000	20.0	340	703	2367	2801	16.0	14.1	8.8	4.2	43.2
3D 60 10000	19.75	346	715	2408	2842	18.1	16.1	10.7	2.8	47.7

Table A.9: Lower four formants obtained by DWM method compared against the solid model for vowel /u/ (page 1)

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		196	934	2136	3091					
2D 10 1	20.0	223	1025	2174	2941	13.8	9.7	1.8	4.9	30.2
2D 10 10	20.0	229	1043	2174	2959	16.8	11.7	1.8	4.3	34.6
2D 10 100	20.0	223	1014	2197	3041	13.8	8.6	2.9	1.6	26.8
2D 10 1000	21.5	217	920	1951	2209	10.7	1.5	8.7	28.5	49.4
2D 10 10000	24.25	199	832	1588	2057	1.5	10.9	25.7	33.5	71.6
2D 20 1	19.5	229	1049	2227	3012	16.8	12.3	4.3	2.6	36.0
2D 20 10	19.75	229	1037	2232	3035	16.8	11.0	4.5	1.8	34.2
2D 20 100	19.5	234	1061	2238	3053	19.4	13.6	4.8	1.2	39.0
2D 20 1000	19.75	223	1025	2238	3065	13.8	9.7	4.8	0.8	29.1
2D 20 10000	20.75	223	920	1957	2291	13.8	1.5	8.4	25.9	49.5
2D 30 1	19.5	229	1043	2244	3076	16.8	11.7	5.1	0.5	34.0
2D 30 10	19.5	229	1043	2232	3059	16.8	11.7	4.5	1.0	34.0
2D 30 100	19.5	229	1037	2244	3059	16.8	11.0	5.1	1.0	34.0
2D 30 1000	19.5	234	1061	2256	3088	19.4	13.6	5.6	0.1	38.7
2D 30 10000	19.25	234	1072	2273	3094	19.4	14.8	6.4	0.1	40.7
2D 40 1	19.25	229	1049	2262	3088	16.8	12.3	5.9	0.1	35.1
2D 40 10	19.25	234	1061	2256	3082	19.4	13.6	5.6	0.3	38.9
2D 40 100	19.25	234	1055	2256	3065	19.4	13.0	5.6	0.8	38.8
2D 40 1000	19.25	229	1055	2256	3123	16.8	13.0	5.6	1.0	36.4
2D 40 10000	19.0	234	1078	2279	3129	19.4	15.4	6.7	1.2	42.7
2D 50 1	19.0	234	1066	2279	3182	19.4	14.1	6.7	2.9	43.2
2D 50 10	19.25	229	1037	2273	3129	16.8	11.0	6.4	1.2	35.5
2D 50 100	19.25	234	1061	2279	3129	19.4	13.6	6.7	1.2	40.9
2D 50 1000	19.0	234	1055	2279	3188	19.4	13.0	6.7	3.1	42.2
2D 50 10000	19.25	234	1078	2273	3070	19.4	15.4	6.4	0.7	41.9
2D 60 1	19.0	234	1078	2291	3182	19.4	15.4	7.3	2.9	45.0
2D 60 10	19.0	234	1066	2291	3152	19.4	14.1	7.3	2.0	42.8
2D 60 100	19.0	234	1072	2291	3176	19.4	14.8	7.3	2.7	44.2
2D 60 1000	19.0	234	1055	2291	3170	19.4	13.0	7.3	2.6	42.2
2D 60 10000	18.75	240	1119	2309	3193	22.4	19.8	8.1	3.3	53.7

Table A.10: Lower four formants obtained by DWM method compared against the solid model for vowel /u/ (page 2)

	VTL	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$	$\sum \Delta$
Solid model		196	934	2136	3091					
3D 10 1	20.25	240	1148	2215	3053	22.4	22.9	3.7	1.2	50.3
3D 10 10	20.5	264	1295	2232	3047	34.7	38.7	4.5	1.4	79.3
3D 10 100	20.75	264	1254	2238	3070	34.7	34.3	4.8	0.7	74.4
3D 20 1	19.25	252	1213	2232	3018	28.6	29.9	4.5	2.4	65.3
3D 20 10	19.75	234	1078	2227	3012	19.4	15.4	4.3	2.6	41.6
3D 20 100	19.75	229	1055	2238	3041	16.8	13.0	4.8	1.6	36.2
3D 30 1	19.25	252	1172	2244	3006	28.6	25.5	5.1	2.7	61.9
3D 30 10	19.25	258	1201	2250	3023	31.6	28.6	5.3	2.2	67.8
3D 30 100	19.5	234	1066	2244	3035	19.4	14.1	5.1	1.8	40.4
3D 40 1	19.25	234	1084	2256	3006	19.4	16.1	5.6	2.7	43.8
3D 40 10	19.25	246	1125	2262	3018	25.5	20.4	5.9	2.4	54.2
3D 40 100	19.25	240	1113	2268	3076	22.4	19.2	6.2	0.5	48.3
3D 50 1	19.25	234	1072	2262	3029	19.4	14.8	5.9	2.0	42.1
3D 50 10	19.25	240	1102	2268	3029	22.4	18.0	6.2	2.0	48.6
3D 50 100	19.25	234	1072	2268	3059	19.4	14.8	6.2	1.0	41.4
3D 60 1	19.0	234	1084	2279	3135	19.4	16.1	6.7	1.4	43.6
3D 60 10	19.0	234	1090	2279	3129	19.4	16.7	6.7	1.2	44.0
3D 60 100	19.0	246	1148	2279	3106	25.5	22.9	6.7	0.5	55.6

Bibliography

[1] S. Adachi and M. Yamada. An acoustical study of sound production in biphonic singing, Xöömij. *The Journal of the Acoustical Society of America*, 105(5):2920–2932, 1999.

[2] ATR-Promotions. *Manual of the ATR MRI database of Japanese vowel production*, fourth edition, Apr. 2007.

[3] R. Caussé, J. Kergomard, and X. Lurton. Input impedance of brass musical instruments—Comparison between experiments and numerical models. *The Journal of the Acoustical Society of America*, 75(1):241–254, 1984.

[4] S. A. V. Duyne and J. O. Smith III. Physical modeling with the 2-D digital waveguide mesh. In *Proc. Int. Computer Music Conf. (ICMC'93)*, pages 40–47, Tokyo, Japan, Sept. 1993.

[5] T. Kitamura and S. Masaki. Progress of studies on the human speech production system based on MRI observations. *The Journal of the Acoustical Society of Japan*, 62(5):385–390, 2006.

[6] T. Kitamura, H. Takemoto, and K. Honda. Acoustic characteristics of solid models based on vowel production MRI data. *Technical Report of IEICE. EA*, EA2007-89:19–24, 2007.

[7] B. J. Kröger, R. Winkler, C. Mooshammer, and B. Pompino-Marschall. Estimation of vocal tract area function from magnetic resonance imaging: Preliminary results. In *Proceedings of 5th Seminar on Speech Production: Models and Data*, pages 333–336, Bavaria, Germany, May 2000.

[8] S. Masaki, T. Kitamura, and H. Takemoto. A breakthrough in speech research with MRI technology. In *Proceedings of the 2007 Autumn Meeting of the Acoustical Society of Japan*, pages 453–456, Kofu, Japan, Sept. 2007.

[9] K. Mochizuki and T. Nakai. Estimation of area function from 3-D magnetic resonance images of vocal tract using finite element method. *Acoustical Science and Technology*, 28(5):346–348, 2007.

[10] J. Mullen. *Physical Modelling of the Vocal Tract with the 2D Digital Waveguide Mesh*. PhD thesis, The University of York, Apr. 2006.

[11] J. Mullen, D. M. Howard, and D. T. Murphy. Digital waveguide mesh modeling of the vocal tract acoustics. *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 119–122, 2003.

[12] J. Mullen, D. M. Howard, and D. T. Murphy. Acoustical simulations of the human vocal tract using the 1D and 2D digital waveguide software model. In *Proceedings of the 7th International Conference on Digital Audio Effects (DAFX-04)*, pages 311–314, Naples, Italy, Oct. 2004.

[13] T. Nakai, K. Satoh, and Y. Suzuki. Sound pressure distribution and propagation path in the vocal tract with the supraglottis and the pyriform fossa. *Technical Report of IEICE. HIP*, 98(504):31–38, 1999.

[14] T. Nakai, M. Shota, and S. Tetsuya. Estimation of cross sectional area of 2-d vocal tract by analysis of finite element method. *Technical Report of IEICE. SP*, 102(248):1–4, 2002.

[15] J. O. Smith III. Physical modeling using digital waveguides. *Computer Music Journal*, 16(4):74–91, 1992.

[16] B. H. Story. Synergistic modes of vocal tract articulation for american english vowels. *The Journal of the Acoustical Society of America*, 118(6):3834–3859, 2005.

[17] B. H. Story. Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002. *The Journal of the Acoustical Society of America*, 123(1):327–335, 2008.

[18] B. H. Story, I. R. Titze, and E. A. Hoffman. Vocal tract area functions from magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 100(1):537–554, 1996.

[19] B. H. Story, I. R. Titze, and E. A. Hoffman. Vocal tract area functions for an adult female speaker based on volumetric imaging. *The Journal of the Acoustical Society of America*, 104(1):471–487, 1998.

[20] B. H. Story, I. R. Titze, and E. A. Hoffman. The relationship of vocal tract shape to three voice qualities. *The Journal of the Acoustical Society of America*, 109(4):1651–1667, 2001.

[21] Y. Suzuki, T. Nakagawa, T. Nakai, J. Dang, and K. Honda. Propagation paths in the vocal tract of vowels by finite element method. In *Proceedings of the 2001 Spring Meeting of the Acoustical Society of Japan*, pages 247–248, Tsukuba, Japan, Mar. 2001.

[22] H. Takemoto, K. Honda, S. Masaki, Y. Shimada, and I. Fujimoto. Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. *The Journal of the Acoustical Society of America*, 119(2):1037–1049, 2006.

[23] K. Yee. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *Antennas and Propagation, IEEE Transactions on*, 14(3):302–307, 1966.