Two-Dimensional Symmetric Digital Waveguide Model for the Vocal Tract

Tahir MUSHTAQ *
Department of Mathematics, COMSATS University Islamabad, Vehari Campus, Vehari (61100), Pakistan, tahirmushtaq@cuivehari.edu.pk

Naeem SARWAR
Department of Mathematics, COMSATS University Islamabad, Vehari Campus, Vehari (61100), Pakistan, naeem555sarwar7736@gmail.com

Muhammad Zubair Akbar QURESHI
Department of Mathematics, Air University, Islamabad, Multan 66000, Pakistan, zubair@aumc.edu.pk

* Author to whom correspondence should be addressed

Abstract: - The existing waveguide modeling of the vocal tract has significant limitations concerning acoustic accuracy and simulation cost. The multi-dimensional waveguide modeling gives precise acoustic wave propagation in the vocal tract at the expense of high computational cost. The efficiency of multi-dimensional digital waveguide modeling has remained a big concern. In this work, we evaluate a new implementation that intends to overcome these limitations. In this paper, we examine a new implementation that aims to address this issue. We introduce the symmetric two-dimensional digital waveguide model by taking half of the mesh with the symmetric line. The symmetric conditions are implemented on the symmetric line on the meshing of the digital waveguide model of the vocal tract. For the evaluation of the current model, the standard two-dimensional non-symmetric digital waveguide model is taken as a benchmark model. The comparison of our symmetric model with the benchmark model has been presented in terms of frequency profiles, formant frequencies, and efficiency. The phonetic vowels /æ/, /ʌ/, /I/, /ɜ-/, and /o/ have been chosen for the demonstration of the proposed work. In the simulation of the current work, the symmetric model is found to be about identical to that of the benchmark model for frequency profiles and formant frequencies. In all cases taken in the proposed work, the symmetric digital waveguide model is more than 90% more efficient than the benchmark model.

Keywords: symmetric, digital waveguide, vocal tract, delay lines, rectilinear uniform grid.

1. INTRODUCTION

Speech modeling is very popular as it enables machines to comprehend and generate human-like speech. This technology supports transcription services, language translation, and virtual assistants, enhancing communication between humans and machines. Advancements in speech modeling contribute to more natural and effective human-computer interactions, making it an interesting area of research. These days, many researchers are working on speech synthesizer modelling with the goal of reproducing speech that is as natural as possible. The air volume between the vocal folds and the lips is referred to as the vocal tract. The various vowels of speech are determined by the resonances of this volume. Vocal tract resonances play a role in voice quality in addition to encoding intelligible sounds, and their control is an important aspect of singing techniques.

In the process of speech production in the human being, the lungs push the air, which goes through the glottis and is modified by the vocal tract. In the end, the airflow is expelled from the lips. The human speech framework is partitioned into two fundamental subsystems such as vocal folds and the vocal tract. The vocal folds produce the train of pulses under the pressure of the lungs, which is why the vocal folds are known to be the source of the sound. The vocal tract imposes the frequency pattern according to its configuration in the train of pulses. Many research studies have been centered around the working of the vocal folds, with many varieties [1-5].

The configuration of the vocal tract has an impact on the quality of sound because of variations in its shape and dimensions, affecting the resonance and timbre of the produced sounds during speech. The vocal tract is assumed to be the concatenation of multiple cylindrical tubes, each with varying cross-sectional areas. There have been a lot of studies done on the vocal tract [6-12]. In a digital waveguide

model of the vocal tract, a regular grid is employed. Each node on this grid is treated as a scattering junction, interconnected by unit waveguide elements [13-17]. This approach allows for a detailed simulation of acoustic interactions within the vocal tract, enhancing the model's accuracy. Many vocal tract methodologies exist, including tube-shaped models [8, 18] and cone-shaped models [19, 20], which are being utilized effectively for wave propagation.

One-dimensional and multi-dimensional digital waveguide models are being used in the modeling of the vocal tract. The advantages of the one-dimensional waveguide model are simple, efficient, and capture basic characteristics of wave behavior. However, the disadvantages are the Oversimplification of complex vocal tract geometry and inaccuracy. The advantages of the multi-dimensional waveguide model are its accurate representation of the size and shape of the vocal tract and its ability to handle the coupling between the vocal tract and the nasal cavity, which enhances the quality of speech. The disadvantages of the multi-dimensional waveguide model are its requirement of a large amount of computation and memory, its tendency to introduce numerical error and artifacts, and its inability to account for the nonlinear effects and turbulence.

The most recent waveguide models are an expansion of the previous waveguide models, and they are well-known for their high-quality and realistic voice production. These models have recently been used in literature works [16, 18, 20-23].

The standard two-dimensional waveguide model is an acceptable choice for demonstrating the vocal tract when sound quality is a primary consideration [17, 24, 25]. The one-dimensional waveguide model efficiently computes wave propagation in the vocal tract, using less time and memory as compared to the two-dimensional waveguide model. This computational efficiency makes the one-dimensional model advantageous for faster and more resource-friendly simulations of acoustic wave behavior in the vocal tract. The two-dimensional waveguide has a limitation due to its high computational cost [26-28].

In this work, we introduce the symmetric two-dimensional digital waveguide model by taking half of the mesh with the symmetric line. The symmetric conditions are implemented on the symmetric line on the meshing of the digital waveguide model of the vocal tract. This approach enables the proposed model to be computationally efficient and realistic in comparison with the multi-dimensional waveguide model. For the evaluation of the current model, the standard two-dimensional complete digital waveguide model is assumed as a benchmark model. The comparison of our symmetric model with the benchmark model has been presented in terms of frequency profiles, formant frequencies, and efficiency. For the simulations of wave propagation in the vocal tract, we also consider the five selected phonetic vowels /æ/, /ʌ/, /I/, /ɜ/, and /o/ [29] for the evaluation of the current work.

In the methodology of the current work, a rectilinear mesh is generated to match the geometrical shape of the symmetric vocal tract for each vowel. The symmetric condition is imposed on the central line of the vocal tract. Input samples are provided at the inlet, while output results are computed at the outlet. Each node of the mesh is considered a junction.

In the first stage, we measure the incoming pressure at each junction of the mesh, including those lying on the symmetric line. In the second stage, we calculate the outgoing pressure to the neighbors of each junction. We iteratively repeat the process of stages 1 and 2 until we reach the number of input samples. In the end, we obtain the transfer function from the output samples.

2. SYMMETRIC WAVEGUIDE MODELING FOR THE VOCAL TRACT

Through the use of the wave equation in a one-dimensional waveguide, uniform cylindrical segments are used to model the vocal tract with varying cross-sectional areas. In a uniform tube, a relationship between speed and pressure has been established [30, 31].

The right and left-moving wave components make up the solution to the one-dimensional wave equation. The scattering junction is defined as the point where two cylindrical tubes with different cross-sectional areas meet. At the intersection of the *i*th and (*i*+1)th cylindrical acoustic tubes, the solution for the continuity equation is created using the reflection coefficient,

$$r_i = \frac{A_i - A_{i+1}}{A_i + A_{i+1}} \tag{1}$$

where A_i and A_{i+1} represent the cross-sectional areas of the cylinders *i* and *i*+1, respectively, measured in units $\left(cm^2\right)$. The reflection coefficient r_i is a dimensionless parameter that illustrates the dispersion of wave spread in the acoustics tubes. If $\left(A_i > A_{i+1}\right)$ then the value of the reflection coefficient will be $\left(r_i > 0\right)$. Otherwise, $\left(r_i < 0\right)$ for $\left(A_i < A_{i+1}\right)$. However, no dispersion of wave spread occurs in the acoustic tubes if $\left(r_i = 0\right)$ or $\left(A_i = A_{i+1}\right)$.

The use of a space-time sample grid can improve the accuracy of wave spread solutions. The accuracy of the one-dimensional waveguide can be improved by extending the waveguide structure, called the waveguide mesh [18, 32]. The current work is focused on the two-dimensional waveguide mesh due to its simple and primary setup. To increase the efficiency of the two-dimensional waveguide model, we impose the symmetric condition at the line of symmetry in the waveguide mesh.

The digital waveguide model uses a variety of meshing topologies [17, 33, 34]. The two-dimensional rectilinear mesh is considered in the proposed work for the sake of simplicity and basic configuration. Each junction in the current rectilinear mesh is linked to its four neighboring junctions at a ninety-degree angle. The current junction is depicted in the diagram below.

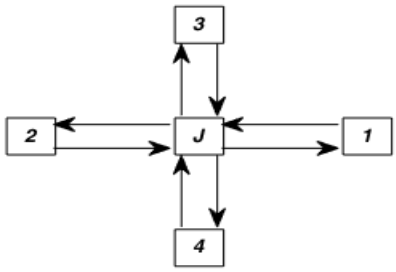

Figure 1. Four-port junction in a two-dimensional waveguide model

The arrows in the figures show bidirectional wave pressure movement from one junction to its neighboring junctions. The wave pressure that comes to junction J from junction 1 is coded as $p^{+}_{J,1}$, arrives from junction 2 is written by $p^{+}_{J,2}$, arrives from junction 3 is represented by $p^{+}_{J,3}$ and comes from junction 4 is denoted by $p^{+}_{J,4}$. However, $p^{-}_{J,1}$, $p^{-}_{J,2}$, $p^{-}_{J,3}$ and $p^{-}_{J,4}$ are denoting the outgoing pressures which originate from junction J and arrive at junctions 1, 2, 3, and 4, respectively. Pressure $p_{J,i}$ on each junction can be formulated as [21].

$$p_J = 2\frac{\sum_{i=1}^{N} Y_i p^{+}_{J,i}}{\sum_{i=1}^{N} Y_i} \tag{2}$$

where Y_i is the dimensionless acoustics impedance in the medium and p_J is the pressure at junction J, measured in pascal (Pa).

The equation given below is used for homogeneous acoustic impedance with four intersecting junctions:

$$p_J = \frac{\sum_{i=1}^{N} p^{+}_{J,i}}{2} \tag{3}$$

Output travelling-wave components at the node are calculated as

$$p^{-}_{J,i} = p_J - p^{+}_{J,i} \tag{4}$$

In the current work, the above-mentioned equations with respect to the digital waveguide model will be applied to the uniform and symmetric mesh of the vocal tract. Some examples of the uniform and non-symmetric mesh of the vocal tract are given below.

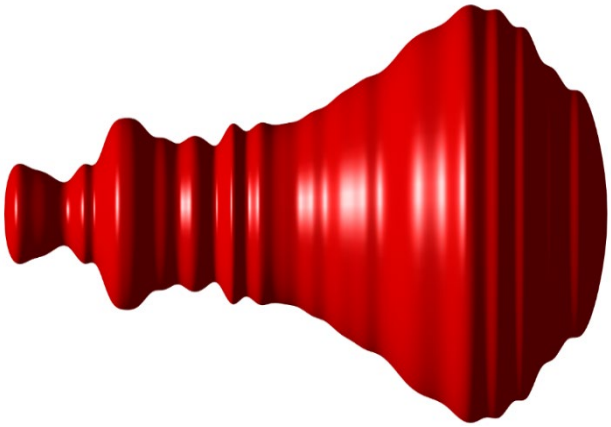

Figure 2. Non-symmetric vocal tract shape for vowel /ʌ/ in a three-dimensional view

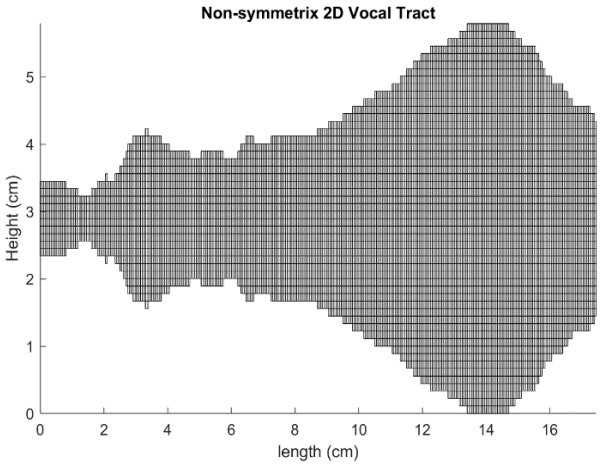

Figure 3. Two-dimensional non-symmetric waveguide mesh for vowel /ʌ/

Figure 2 represents the three-dimensional view of the vocal tract with a non-symmetric structure for better visualization. The orientation of this figure is

taken from the phonetic vowel /ʌ/. A two-dimensional uniform rectilinear grid of the non-symmetric vocal tract for phonetic vowel /ʌ/ is demonstrated in Figure 3. The consideration of a symmetric structure in the three-dimensional vocal tract with two different colors is illustrated in Figure 4. The symmetric line is observed at the intersection of these two colors. This is the line where we consider only the upper or lower part of the vocal tract for computational work. The important point is to define the symmetric condition at this symmetric line. The half mesh of the two-dimensional symmetric vocal tract is also presented in Figure 5. The imposition of symmetric conditions in digital waveguide modelling reduces the computational cost significantly. For the evaluation of the current model, the standard two-dimensional non-symmetric digital waveguide model is taken as a benchmark model.

Let us consider the red color of the upper half part of the vocal tract as illustrated in Figure 4. A four-port junction J is demonstrated in Figure 1 for the rectilinear grid, we now define this four-port junction on the symmetric line. For the symmetry, the neighboring nodes 3 and 4 are very important. The outgoing wave from junction J to port 4 is unimportant, and it is omitted, while the incoming wave from port 4 to junction J is important, and we consider it. To employ symmetric, it must be the same as an outgoing wave from J to port 3. The pressure at the junction J for the symmetric line can be derived as

$$p_J = \frac{\sum_{i=1}^{3} p_{J,i}^{+} + p_{J,3}^{+}}{2} \tag{5}$$

Output travelling-wave components at the node are calculated as

$$p_{J,i}^{-} = p_J - p_{J,i}^{+} \quad i \neq 4 \tag{6}$$

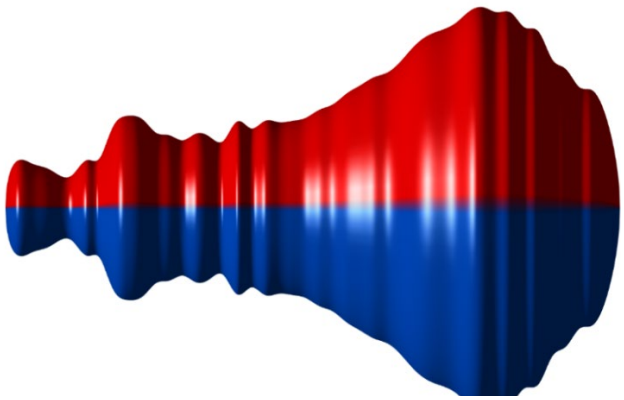

Figure 4. Two-dimensional vocal tract for vowel /ʌ/ with its symmetric parts

On the wall of the vocal tract, the travelling waves are scattered by a wall reflection coefficient r_w. However, at the boundary junction, a proportionate amount of the travelling wave is reflected back into the interior junction. Let p_J be the pressure on the junction of the wall boundary and $p_{J,A}^{+}$ is the pressure wave arriving on the wall boundary from the neighboring junction. Then,

$$p_J = \left(1 + r_w\right) p_{J,A}^{+} \tag{7}$$

We have the following wall boundary condition:

$$p_{J,A}^{-} = r_w p_{J,A}^{+} \tag{8}$$

The inlet of the vocal tract is considered a glottal input boundary condition, which is given as

$$p_{J,i}^{-} = \left(\frac{1 + r_G}{2}\right) U_G + r_G p_{J,i}^{+} \tag{9}$$

where r_G is a glottal reflection coefficient and U_G is a glottal flow sample.

The outlet of the vocal tract is assumed as a lip boundary condition, which is given as

$$p_{J,i}^{-} = r_L p_{J,i}^{+} \tag{10}$$

$$p_{J-1,i}^{-} = \left(1 + r_L\right) p_{J,i}^{+} \tag{11}$$

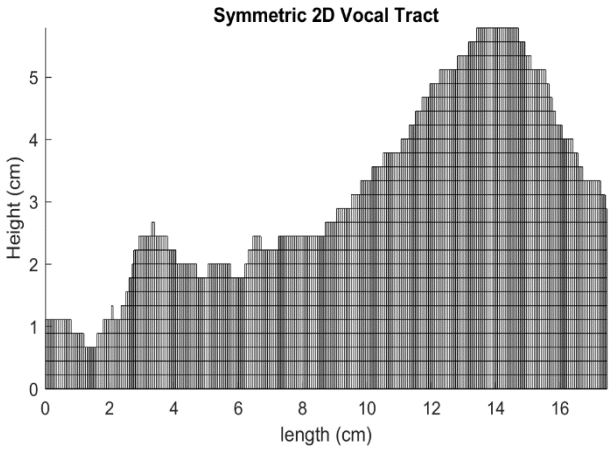

Figure 5. Two-dimensional symmetric waveguide mesh for vowel /ʌ/

3. NUMERICAL SIMULATIONS

The geometrical shapes of the vocal tract for each phonetic vowel of /æ/, /ʌ/, /I/, /ɜ-/, and /o/ are generated for the simulations using a list of cross-sectional areas taken from [29]. For the generation of two-dimensional rectilinear meshing, the cross-

sectional areas have been translated to constant-length acoustic tubes of different radius.

The total samples of the input signals have been set to 10000 to achieve a fine resolution in the graph of the transfer function. For the numerical solution, the mesh is initialized with zeros in the first step. In the second step, the scattering of the wave is computed at the glottal inlet, interior junctions of the mesh, at the wall, and the lips outlet for each node or junction. In the third step, the delays of waves at each junction are passed to its neighboring junctions. In the fourth step, we calculate the output sample at the lips. To measure the complete output samples, we iteratively solve steps 2 to 4 up to 10,000 times.

With the help of Fast Fourier Transformation (FFT), the output samples are converted into the magnitude response of the frequency domain. The simulations have been carried out using MATLAB software. To attain accuracy in the results, double-precision floating-point variables and arrays are used. Reflection coefficient of the glottis r_G, the reflection coefficient of the wall r_W and reflection coefficient lip r_L have been assigned to values 0.97, 1.0, and -0.9, respectively.

4. RESULTS AND DISCUSSIONS

The construction of the symmetric two-dimensional vocal tract model and its implementation process were discussed in the previous sections. We compare our proposed model with our benchmark model with respect to frequency profiles and formant frequencies. In the current work, we consider the first four formant frequencies for each vowel for comparison with our benchmark model. The first column of all tables in the proposed work is populated by six formant frequencies, which are denoted by F_1, F_2, F_3 and F_4. For the validation of the comparison, the sample delay d_s has been taken as that of taking in our benchmark model. The efficiency and accuracy of the current model are compared with the benchmark model in terms of elapsed time and formant frequencies, respectively.

Figure 6 displays the frequency profile of the phonetic vowel /æ/ of the vocal tract for the transfer function up to 6000 Hz. The formant frequencies of the benchmark and the formant frequencies of the symmetric model are identical, as shown in the graph. In other words, the frequency profiles of the two models are the same.

Numerically calculated values are given in Table 1. In this table, column 4 shows the relative error between the benchmark model and the symmetric model. This column presents the zero relative errors in all four format frequencies in the current case of phonetic vowel /æ/. Column seven gives about 93% efficiency of the proposed model, which signifies the high efficiency of the symmetric model.

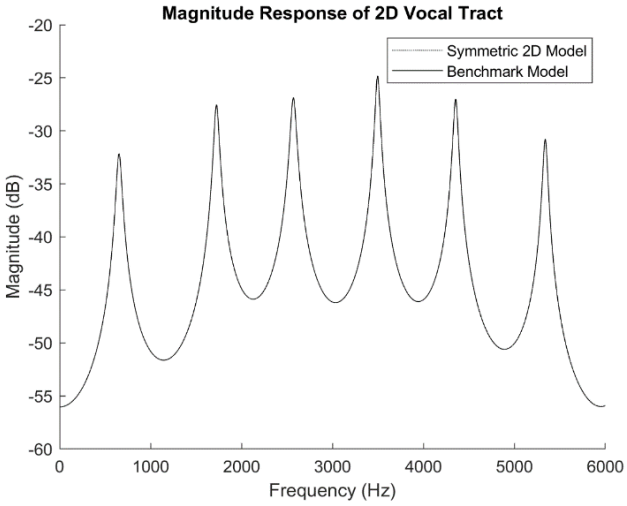

Figure 6. Comparison of the benchmark model with symmetric modeling for vowel /æ/

Table 1. Numerical comparison of benchmark and symmetric models in the case of vowel /æ/

Format frequency	Benchmark Model	Symmetric Model	Error (%)	Elapsed time		Efficiency (%)
				Benchmark model	Symmetric model	
F_1	654	654	0	103.31	53.41	93
F_2	1723	1723	0			
F_3	2549	2549	0			
F_4	3501	3501	0			

The frequency profile of the phonetic vowel /ʌ/ of the vocal tract for the transfer function up to 6000 Hz is shown in Figure 7. This figure shows that there is no difference between the formant frequencies of the benchmark and the formant frequencies of the symmetric model. In other words, the frequency profiles of the two models overlap with each other.

Numerically measured values are mentioned in Table 2. Column 4 illustrates the relative error between the benchmark model and the symmetric model. This column presents the maximum relative error in all four format frequencies as zero percent in the case of the phonetic vowel /ʌ/. Column seven gives about 91% efficiency of the proposed model, which leads to the high efficiency of the symmetric model.

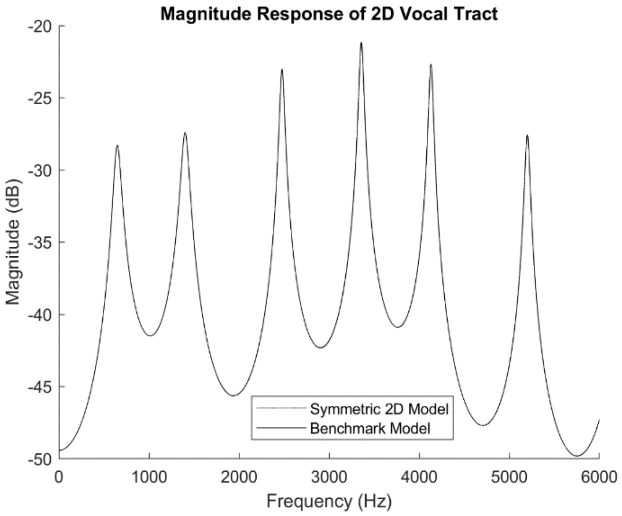

Figure 7. Comparison of the benchmark model with the symmetric model for vowel /ʌ/

Table 2. Numerical comparison of benchmark and symmetric models in the case of vowel /ʌ/

Format frequency	Benchmark Model	Symmetric Model	Error (%)	Elapsed time		Efficiency (%)
				Benchmark model	Symmetric model	
F_1	645	645	0	102.78	53.7	91
F_2	1400	1400	0			
F_3	2477	2477	0			
F_4	3357	3357	0			

The formant frequency profile of the phonetic vowel /I/ of the vocal tract has been presented in Figure 8. We conclude from this figure that there is no difference between the formant frequencies of the benchmark and the symmetric model. In this case, the frequency profiles of the two models are identical to each other. Table 3 shows the numerically measured values of the formant frequencies for the phonetic vowel /I/. Column 4 illustrates the zero relative errors in all four formant frequencies between the benchmark model and the symmetric model. Column seven gives about 97% efficiency of the proposed model.

The formant frequencies of the benchmark and the formant frequencies of the symmetric model do not differ, as shown in the graph of Figure 9 for the phonetic vowel /ɝ/. The frequency profiles of the two models are identical. Table 4 lists the numerically measured values. Column 4 of the table presents the maximum relative error in all four format frequencies is zero percent in the current case of phonetic vowel /ɝ/. Column seven gives about 93% efficiency of the proposed model.

The formant frequency profile of the phonetic vowel /o/ of the vocal tract has been presented in Figure 10. It is noted from this figure that there is no difference between the formant frequencies of the benchmark and the formant frequencies of the symmetric model. The numerically measured values of the formant frequencies for the phonetic vowel /o/ are listed in Table 5. The zero relative errors in all four formant frequencies between the benchmark model and the symmetric model are observed in the current case. Column seven gives about 92% efficiency of the proposed model.

To study the smooth transition from one vowel to another vowel in the symmetric modelling of the waveguide model, we present the cross-section areas of three vowels /I/, /æ/, and /ʌ/ in a three-dimensional view as shown in Figure 11. In this figure, the cross-sectional areas of the first vowel /I/ are inserted at position 1 on the vowel axis, while the cross-sectional areas of the vowels /æ/ and /ʌ/ are inserted at positions 2 and 3, respectively, on the vowel axis. In this figure, there is a linear transition from one vowel to another.

In Figure 12, the two-dimensional cubic interpolation is used to generate the smooth variation of cross-sectional areas from vowel /I/ to vowel /æ/ and from vowel /æ/ to vowel /ʌ/, which exhibit the smooth graph on the vowel axis. In the current figures, blue and yellow colors are both selected as the basic colors in the current graph. The dark blue color represents for the minimum value of the graph, while the maximum value is denoted by the yellow color. The gradient colors from dark blue to yellow describe the intermediate values between the minimum and maximum values.

Figure 13 represents the three-dimensional visualization of the frequency domain up to 6000 Hz corresponding to the variation in cross-sectional areas of the three vowels from vowel /I/ to vowel /æ/ and from vowel /æ/ to /ʌ/. We note in the current figure that there is a smooth transition of the frequency curve from one vowel to another vowel with respect to the corresponding variation in the cross-sectional areas of the said vowels' transition. To get frequency profiles for the interpolated cross-sectional areas in the transition of one vowel to another vowel, we accomplish this by taking the intersection of the different planes with the present graph with their fixed position on the vowel-axis. The intersection points of planes with the current graph are denoted by the yellow color. The curve traced out by the intersection points is known as the transfer function in the form of the frequency profile corresponding to the interpolated cross-sectional areas. In the current work, three plans are drawn, which are fixed at 1.5,

2.2, and 2.8 on the vowel axis. For better visualization, we also draw the above-mentioned three planes separately in Figure 14 along with the intersection point, which traces out the frequency profiles corresponding to cross-sectional areas fixed at 1.5, 2.2, and 2.8 on the vowel axis.

The yellow-colored curves outlined by the peaks in the graph of Figure 13 form the six formant frequency profiles as shown in Figure 15. The first yellow-colored curve along the vowel axis gives the variation in the first formant frequency during the transition from vowel /I/ to vowel /æ/ and from vowel /æ/ to /ʌ/. Similarly, the other five yellow-colored curves along the vowel axis produce the variation in five formant frequencies during the transition from vowel /I/ to vowel /æ/ and from vowel /æ/ to /ʌ/.

Figure 16 demonstrates the cross-section areas of three vowels /ʌ/, /o/, and /ɝ/ positioned at 1, 2, and 3, respectively, on the vowel axis. By using two-dimensional interpolation, the variation of cross-sectional areas from vowel /I/ to vowel /æ/ and then from vowel /æ/ to vowel /ʌ/ is produced, which illustrates the smooth transition between the vowels.

Figure 18 denotes the three-dimensional visualization of the frequency domain up to 6000 Hz corresponding to the variation in cross-sectional areas of the three vowels from vowel /ʌ/ to vowel /o/ and then from vowel /o/ to /ɝ/. There is a smooth transition of the frequency curve from one vowel to another vowel with respect to the corresponding variation in the cross-sectional areas of the above-mentioned vowels' transition. To obtain frequency profiles for the interpolated cross-sectional areas in the transition from one vowel to another, we take the intersection of the various planes with the present graph at their fixed position on the vowel axis. The yellow color represents the intersection points of planes with the current graph. The intersection points trace out the curves, which are named as the transfer function in the form of the frequency profile corresponding to the interpolated cross-sectional areas. The three planes are used and kept fixed at 1.3, 2.3, and 2.9 on the vowel axis.

We also illustrate the above-mentioned three planes separately in Figure 19 along with the intersection point, which traces out the frequency profiles corresponding to cross-sectional areas fixed at 1.3, 2.3, and 2.9 on the vowel axis.

In Figure 20, the six yellow-colored curves along the vowel axis produce the variation in six formant frequencies during the transition from vowel /ʌ/ to vowel /o/ and then from vowel /o/ to vowel /ɝ/.

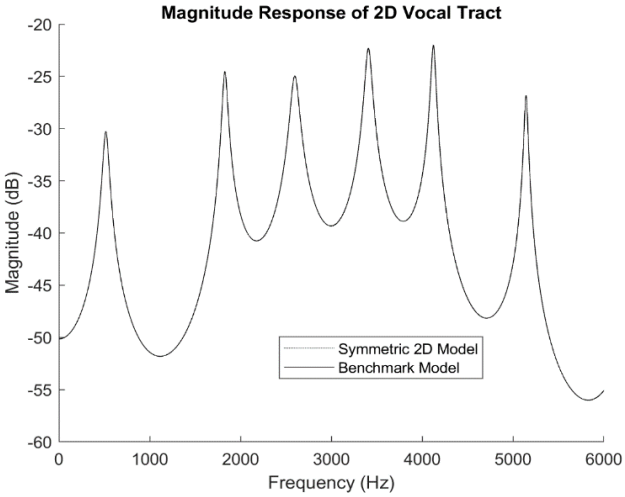

Figure 8. Comparison of the benchmark model with symmetric modeling for vowel /I/

Table 3. Numerical comparison of benchmark and symmetric models in the case of vowel /I/

Format frequency	Benchmark Model	Symmetric Model	Error (%)	Elapsed time		Efficiency (%)
				Benchmark model	Symmetric model	
F_1	516	516	0	113.8	57.68	97
F_2	1828	1828	0			
F_3	2598	2598	0			
F_4	3408	3408	0			

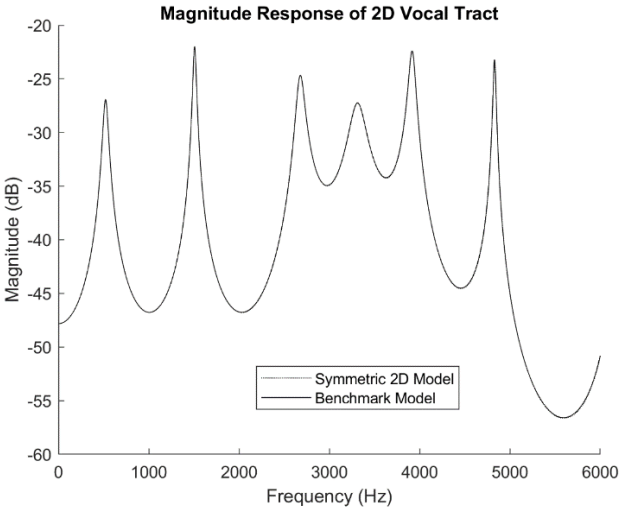

Figure 9. Comparison of the benchmark model with symmetric modeling for vowel /ɝ/

Table 4. Numerical comparison of benchmark and symmetric models in the case of vowel /ɜ/

Format frequency	Benchmark Model	Symmetric Model	Error (%)	Elapsed time		Efficiency (%)
				Benchmark model	Symmetric model	
F_1	520	520	0	105.9	54.86	93
F_2	1518	1518	0			
F_3	2679	2679	0			
F_4	3312	3312	0			

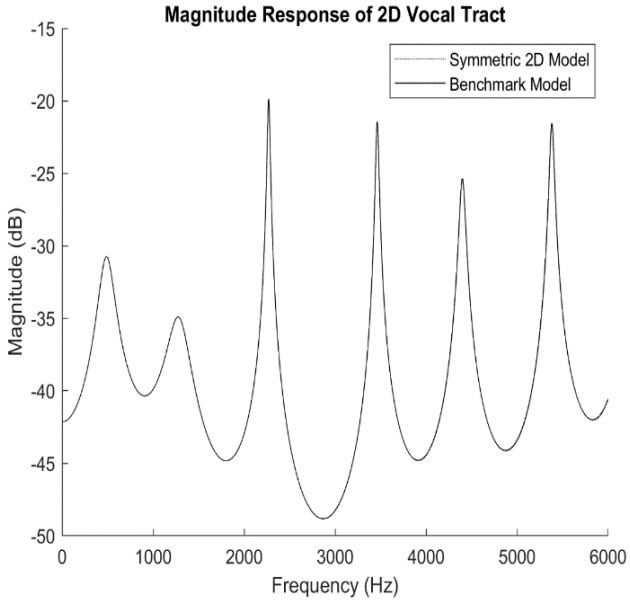

Figure 10. Comparison of the benchmark model with symmetric modeling for vowel /o/

Table 5. Numerical comparison of benchmark and symmetric models in the case of vowel /o/

Format frequency	Benchmark Model	Symmetric Model	Error (%)	Elapsed time		Efficiency (%)
				Benchmark model	Symmetric model	
F_1	487	487	0	108.2	56.13	92
F_2	1274	1274	0			
F_3	2270	2270	0			
F_4	3463	3463	0			

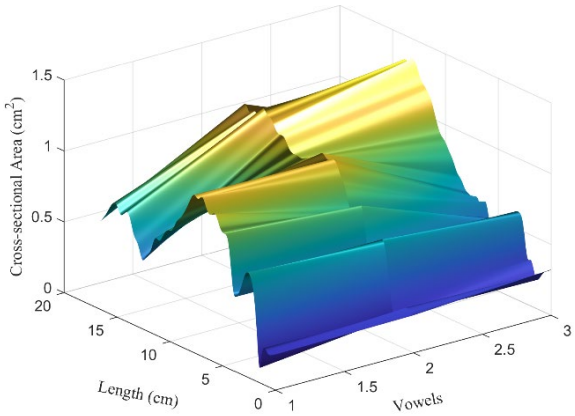

Figure 11. Cross-sectional areas of three vowels /I/, /æ/, /ʌ/ in three-dimensional view

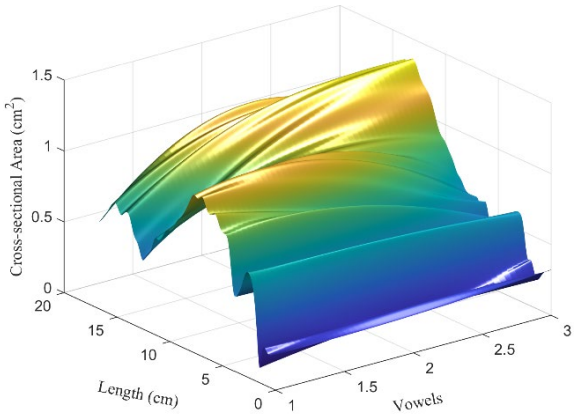

Figure 12. Smooth transition from vowel /I/ to vowel /æ/ and then to vowel /ʌ/

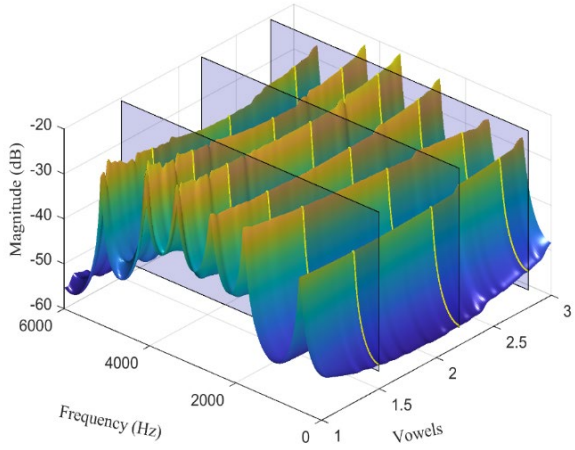

Figure 13. Three-dimensional frequency domain in the transition of three vowels /I/, /æ/ and /ʌ/

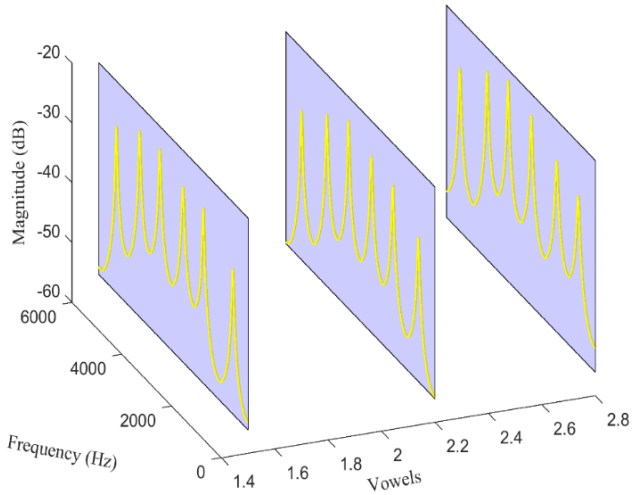

Figure 14. Three planes along with their frequency profiles in the transition of vowels /I/, /æ/, and /ʌ/

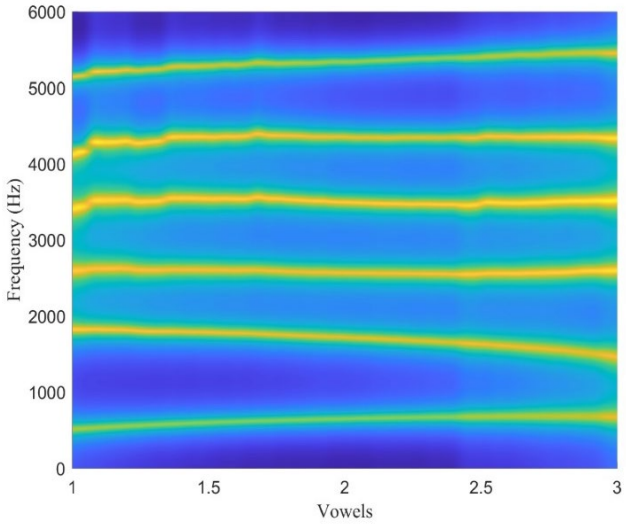

Figure 15. Six formant frequency profiles in the transition of vowels /I/, /æ/, and /ʌ/

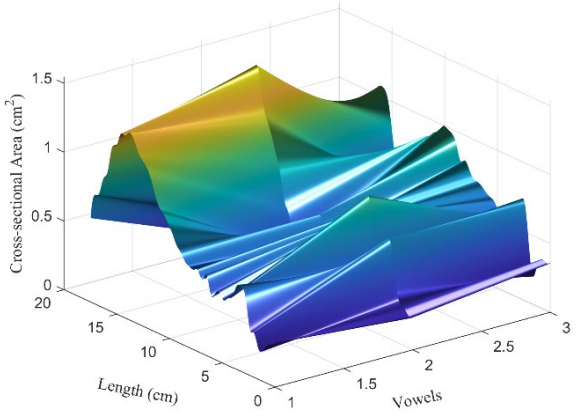

Figure 16. Cross-sectional areas of three vowels /ʌ/, /o/, and /ɝ/ in three-dimensional view

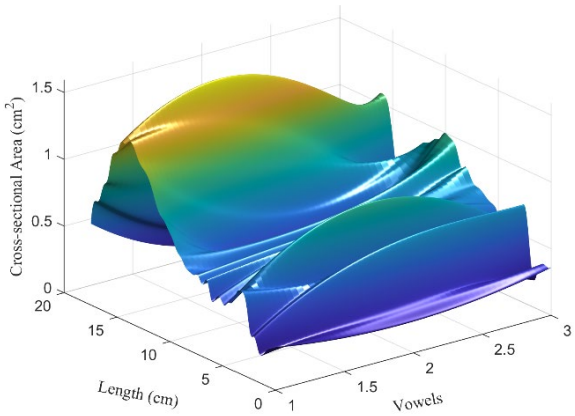

Figure 17. Smooth transition from vowel /ʌ/ to vowel /o/ and then to vowel /ɝ/

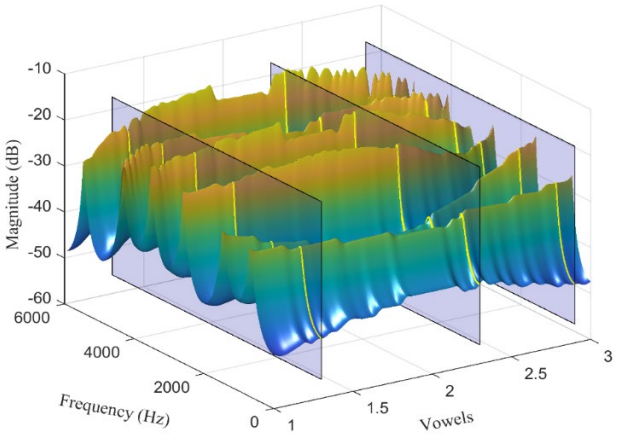

Figure 18. Three-dimensional frequency domain in the transition of three vowels /ʌ/, /o/, and /ɝ/

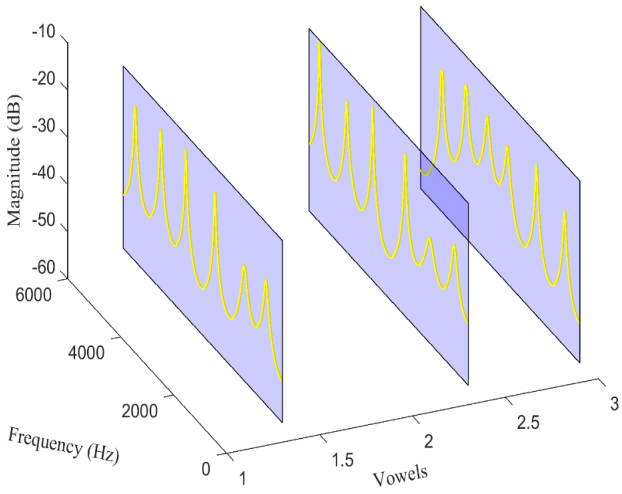

Figure 19. Three planes along with their frequency profiles in the transition of /ʌ/, /o/, and /ɝ/

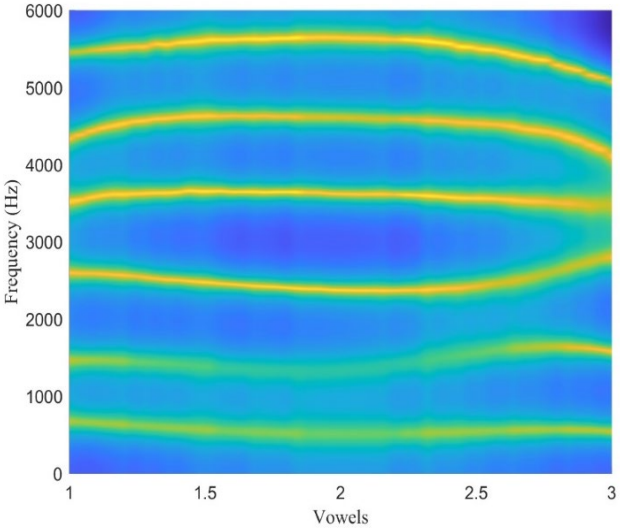

Figure 20. Six formant frequency profiles in the transition of vowels /ʌ/, /o/, and /ɝ/

5. CONCLUSIONS

In the proposed work, the symmetric mesh is used for modeling of two-dimensional waveguide model of the vocal tract. The upper half part of the vocal tract is taken with uniform meshing. The symmetric conditions are applied on the line of symmetry in the meshing of the vocal tract. Some selected phonetic vowels /æ/, /ʌ/, /I/, /ɝ/, and /o/ have been taken in the simulation. By tables and figures, we draw the following conclusions:

- Successful implementation of symmetric conditions in the modeling of the two-dimensional digital waveguide for the vocal tract.
- The symmetric model presents about the same formant frequencies that are obtained from the benchmark model.
- The frequency profiles of the present model are approximately overlapping with those of the benchmark model.
- The symmetric model is more efficient than the benchmark model. In all cases, the proposed model is more than 90% more efficient than that of the benchmark model.
- Successful generation of formant curves with the variation of the interpolated cross-sectional areas in the transition from one vowel to another vowel.

In general, the implementation of the symmetric waveguide model can lead to more efficient and accurate simulations in various applications such as speech synthesis, speech processing, medical imaging, acoustic modeling, music synthesis, etc.

REFERENCES

[1] Flanagan J., Landgraf L., Self-oscillating source for vocal-tract synthesizers, *Audio and Electroacoustics, IEEE Transactions on,* Vol. 16, No. 1, 1968, pp. 57-64.

[2] Ishizaka K., Falanagan J. L., Synthesis of voiced sounds from a two-mass model of the vocal cords, *Bell System Technical Journal,* Vol. 51, No. 6, 1972, pp. 1233-1268.

[3] Qureshi T. M., Syed K. S., Fulcrum-Point Based Self-Oscillatory Glottal Model with Numerical Flow Simulation, *International Journal of Acoustics & Vibration,* Vol. 23, No. 4, 2018,

[4] Kumar S. P., Švec J. G., Kinematic model for simulating mucosal wave phenomena on vocal folds, *Biomedical Signal Processing and Control,* Vol. 49, 2019, pp. 328-337.

[5] Qureshi T. M., Syed K. S., A New Approach to Parametric Modeling of Glottal Flow, *Archives of Acoustics,* Vol. 36, No. 4, 2011, pp. 695-712.

[6] Kelly J. L., Lochbaum C. C., Speech synthesis. Proceedings of the Stockholm Speech Communications Seminar, RIT, Stockholm, Sweden, 1962, pp. 1-4.

[7] Välimäki V., Karjalainen M., Improving the kelly-lochbaum vocal tract model using conical tube sections and fractional delay filtering techniques. ICSLP, 1994, vol. 2, pp. 615-618.

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

[9] Vampola T., Laukkanen A.-M., Horáček J., Švec J. G., Vocal tract changes caused by phonation into a tube: a case study using computer tomography and finite-element modeling, *J The Journal of the Acoustical Society of America,* Vol. 129, No. 1, 2011, pp. 310-315.

[10] Qureshi T. M., Syed K. S., Improved vocal tract model for the elongation of segment lengths in a real time, *Computer Speech & Language,* Vol. 57, 2019, pp. 41-58.

[11] Qureshi T. M., Ishaq M., Real-Time Vocal Tract Model for Elongation of Segment Lengths in a Waveguide Model, *Archives of Acoustics,* Vol. 44, No. 2, 2019, pp. 287–300.

[12] Mathur S., Story B. H., Rodríguez J. J., Vocal-tract modeling: Fractional elongation of segment lengths in a waveguide model with half-sample delays, *IEEE Transactions on audio, speech, and language processing,* Vol. 14, No. 5, 2006, pp. 1754-1762.

[13] Smith J. O., Principles of digital waveguide models of musical instruments, *Applications of digital signal processing to audio and acoustics,* 2002, pp. 417-466.

[14] Speed M., Murphy D., Howard D., Three-Dimensional Digital Waveguide Mesh Simulation of Cylindrical Vocal Tract Analogs, *IEEE Transaction on audio, speech, and language processing,* Vol. 21, No. 2, 2013, pp. 449-454.

[15] Mullen J., Howard D., Murphy D., Acoustical simulations of the human vocal tract using the 1D and 2D digital waveguide software model. Proceedings of the 7th International Conference on Digital Audio Effects, Naples, Italy, 2004, vol. 4, pp. 5-8.

[16] Mullen J., Howard D. M., Murphy D. T., Real-time dynamic articulations in the 2-D waveguide mesh vocal tract model, *Audio, Speech, and Language Processing, IEEE Transactions on,* Vol. 15, No. 2, 2007, pp. 577-585.

[17] Murphy D. T., Howard D. M., 2-D digital waveguide mesh topologies in room acoustics modelling. Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx), 2000, pp. 211-216.

[18] Mullen J., Howard D. M., Murphy D. T., Waveguide physical modeling of vocal tract acoustics: flexible formant bandwidth control from increased model dimensionality, *Audio, Speech, and Language Processing, IEEE Transactions on,* Vol. 14, No. 3, 2006, pp. 964-971.
[19] Strube H. W., Are conical segments useful for vocal-tract simulation? (L), *The Journal of the Acoustical Society of America,* Vol. 114, No. 6, 2003, pp. 3028-3031.
[20] Makarov I., Approximating the vocal tract by conical horns, *Acoustical Physics,* Vol. 55, No. 2, 2009, pp. 261-269.
[21] Mullen J., "Physical modelling of the vocal tract with the 2D digital waveguide mesh," University of York, 2006.
[22] Qureshi T. M., Syed K. S., Two dimensional featured one dimensional digital waveguide model for the vocal tract, *Computer Speech & Language,* Vol. 33, No. 1, 2015, pp. 47-66.
[23] Qureshi T. M., Syed K. S., Zafar A. J. A. o. A., Non-uniform Rectilinear Grid in the Waveguide Modeling of the Vocal Tract, Vol. 45, No. 4, 2020, pp. 585–600.
[24] Murphy D. T., Beeson M., The KW-boundary hybrid digital waveguide mesh for room acoustics applications, *Audio, Speech, and Language Processing, IEEE Transactions on,* Vol. 15, No. 2, 2007, pp. 552-564.
[25] Beeson M. J., Murphy D. T., RoomWeaver: A digital waveguide mesh based room acoustics research tool. Proceedings of the Seventh International Conference on Digital Audio Effects, Naples, Italy, 2004, pp. 268-273.
[26] Campos G. R., Howard D. M., On the computational efficiency of different waveguide mesh topologies for room acoustic simulation, *Speech and Audio Processing, IEEE Transactions on,* Vol. 13, No. 5, 2005, pp. 1063-1072.
[27] Fontana F., Rocchesso D., Signal-theoretic characterization of waveguide mesh geometries for models of two-dimensional wave propagation in elastic media, *Speech and Audio Processing, IEEE Transactions on,* Vol. 9, No. 2, 2001, pp. 152-161.
[28] Campos G., Howard D., A parallel 3D digital waveguide mesh model with tetrahedral topology for room acoustic simulation. Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx),(Verona, Italy), 2000, pp. 73-78.
[29] Story B. H., Titze I. R., Hoffman E. A., Vocal tract area functions from magnetic resonance imaging, *The Journal of the Acoustical Society of America,* Vol. 100, No. 1, 1996, pp. 537-554.
[30] Markel J. E., Gray A. H., *Linear prediction of speech*. Springer-Verlag New York, Inc., 1976.
[31] Rabiner L. R., Schafer R. W., Digital processing of speech signals., *Prantice-Hall,/nc,* 1978,
[32] Van Duyne S. A., Smith J. O., The 2-D digital waveguide mesh. Applications of Signal Processing to Audio and Acoustics, 1993. Final Program and Paper Summaries, 1993 IEEE Workshop on, New Paltz, NY, 1993, pp. 177-180: IEEE.
[33] Murphy D. T., Digital waveguide mesh topologies in room acoustics modelling, 2005,
[34] Murphy D., Kelloniemi A., Mullen J., Shelley S. J. I. S. P. M., Acoustic modeling using the digital waveguide mesh, Vol. 24, No. 2, 2007, pp. 55-66.