

Research Paper

Non-uniform Rectilinear Grid in the Waveguide Modeling
of the Vocal TractTahir Mushtaq QURESHI^{(1)*}, Khalid Saifullah SYED⁽²⁾, Asim ZAFAR⁽¹⁾⁽¹⁾ *Department of Mathematics
COMSATS University Islamabad, Vehari Campus
Vehari (61100), Pakistan*

*Corresponding Author e-mail: tahirmushtaq@ciitvehari.edu.pk

⁽²⁾ *Center for Advanced Studies in Pure and Applied Mathematics (CASPAM)
Bahauddin Zakariya University
Multan (60000), Pakistan**(received August 16, 2019; accepted August 4, 2020)*

For many years, a digital waveguide model is being used for sound propagation in the modeling of the vocal tract with the structured and uniform mesh of scattering junctions connected by same delay lines. There are many varieties in the formation and layouts of the mesh grid called topologies. Current novel work has been dedicated to the mesh of two-dimensional digital waveguide models of sound propagation in the vocal tract with the structured and non-uniform rectilinear grid in orientation. In this work, there are two types of delay lines: one is called a smaller-delay line and other is called a larger-delay line. The larger-delay lines are the double of the smaller delay lines. The scheme of using the combination of both smaller- and larger-delay lines generates the non-uniform rectilinear two-dimensional waveguide mesh. The advantage of this approach is the ability to get a transfer function without fractional delay. This eliminates the need to get interpolation for the approximation of fractional delay and give efficient simulation for sound wave propagation in the two-dimensional waveguide modeling of the vocal tract. The simulation has been performed by considering the vowels /ɔ/, /a/, /i/ and /u/ in this work. By keeping the same sampling frequency, the standard two-dimensional waveguide model with uniform mesh is considered as our benchmark model. The results and efficiency of the proposed model have compared with our benchmark model.

Keywords: non-linear mesh; waveguide; delay lines.

1. Introduction

The sound is the only tool that gives the ability to the human being to communicate with each other in the most natural way. For the communications of the human being, the lungs push the airflow to the glottis and then the vocal tract modulates it, and final airflow is radiated through the lips. Vocal folds and vocal tract are two basic subsystems of the human speech system. The vocal folds are considered as the source of sound in human beings and generate a train of pulses due to lungs pressure. These pulses pass through the vocal tract which modulates them with its resonance. Several research works have been conducted on the modelling of the vocal folds with different variations (BAILLY *et al.*, 2018; FLANAGAN, LANDGRAF, 1968; ISHIZAKA,

FALANAGAN, 1972; 1977; KUMAR, ŠVEC, 2019; MADDOX *et al.*, 2014; QURESHI, SYED, 2011a; 2011b; 2018; RADOLF *et al.*, 2018; SHIMAMURA, TOKUDA, 2016).

The vocal tract has a key role in the human speech system. In this area, the vocal tract is represented by its area function which describes the variations of the cross-sectional area along its center midline. The quality of the sound has a prime concern with the vocal tract. Many works have been dedicated to the modelling of the vocal tract in the literature (BIRKHOFF *et al.*, 2010; KELLY, LOCHBAUM, 1962; MATHUR *et al.*, 2006; MULLEN *et al.*, 2003; QURESHI, ISHAQ, 2019; QURESHI, SYED, 2015; 2019; STORY, 2013; VÄLIMÄKI, KARJALAINEN, 1994; VAMPOLA *et al.*, 2015). To develop the computational model, the approach of the waveguide model is being used for the modelling of

the vocal tract. The digital waveguide models are used for the propagation of sound wave within a domain. Under certain condition, it is resembling the finite difference time domain (KARJALAINEN, ERKUT, 2004; MURPHY, BEESON, 2007). In the simulation, the vocal tract is represented by a regular grid where each node is considered as scattering junctions connected by unit waveguide elements (MORSE, 1981; SMITH, 2002; SPEED *et al.*, 2013a; 2013b). For the physical modelling of the vocal tract, the approach of cylindrical segments (KELLY, LOCHBAUM, 1962; MULLEN *et al.*, 2003; 2006) has been successfully used in the modelling of the waveguide while other approaches such as conical segments (MAKAROV, 2009; STRUBE, 2003; VÄLIMÄKI, KARJALAINEN, 1994) for the better approximation to the vocal tract. The cylindrical segment approach was first implemented by Kelly-Lochbaum for the modelling of the vocal tract (KELLY, LOCHBAUM, 1962) which is known as one-dimensional waveguide model. In this model, propagation of the sound wave is obtained by the modelling of each cylinder as forward- and backward-traveling wave components.

Digital waveguide models are the extensions of the waveguide model and they are very popular due to its realistic and high-quality sound generation in real-time. Now a day, these models are frequently used in recent works such as (COOPER *et al.*, 2006; MULLEN *et al.*, 2006; 2007; MURPHY, BEESON, 2007; SPEED *et al.*, 2013a; 2013b; VAN DUYNE, SMITH, 1993a; 1993b). A digital waveguide has also been used to model the vocal tract of animals in (WILKINSON, REISS, 2016) and it was successfully implemented to reproduce sound effects of a lion's roar and wolf's growl. The Kelly-Lochbaum model of the vocal tract was based on the fixed-length tubes of different cross-sectional areas (KELLY, LOCHBAUM, 1962) and some works have also been dedicated to modifying the Kelly-Lochbaum model by performing fractional elongation of the lengths of the tubes (MATHUR *et al.*, 2006; QURESHI, ISHAQ, 2019; QURESHI, SYED, 2019).

Due to simplicity, the one-dimensional waveguide models are computationally efficient as compare to two- and three-dimensional waveguide models. As far as the accuracy is a concern, the standard two- and three- dimensional waveguides models have better accuracy as compared to one-dimensional waveguide models (BEESON, MURPHY, 2004; CAMPOS, HOWARD, 2000; MURPHY, BEESON, 2007; MURPHY, HOWARD, 2000). There is a number of limitations imposed on the digital waveguide models (CAMPOS, HOWARD, 2005; FONTANA, ROCCHESO, 2001; VAN DUYNE, SMITH, 1996). One of them is the restriction on the uniform structured grid with the same gridlines (SPEED *et al.*, 2013a; 2013b; VAN DUYNE, SMITH, 1993a; 1993b; 1996). Present novel work is focused on the two-dimensional digital waveguide model

of the vocal tract with the structured and non-uniform rectilinear grid in orientation. In the current grid orientation, there are two types of grid lines. One is assumed as smaller grid line and other is considered as larger grid line where the larger grid line is the double of the small grid line. The use of these two different lines in the meshing generates the scheme of the non-uniform rectilinear mesh. This scheme enables us to get a transfer function by propagation wave sound in the non-uniform grid without fractional delay. This eliminates the need to get interpolation for the approximation of fractional delay and give efficient simulation for sound wave propagation in the two-dimensional waveguide modeling of the vocal tract in the non-uniform grid. The current novel work is applied to the vowels /ɔ/, /a/, /i/, and /u/ (STORY *et al.*, 1996). By keeping the same sampling frequency, the standard two-dimensional waveguide model with uniform meshing is considered as our benchmark model. The results and efficiency of the proposed model have compared with the benchmark model.

The present paper is organized as follows. A basic waveguide model of the vocal tract and its mathematical formulation is given in Sec. 2. Section 3 describes the formation of non-uniform rectilinear mesh and way of performing the simulation on this mesh. Results and discussion are mentioned in Sec. 4. Section 5 is reserved for the final conclusions.

2. Physical modeling of the vocal tract

For the numerical solution of the problem, there is a need to discretize the domain of the problem with some physical governing behavior within the system. In the physical modeling of the vocal tract, one-dimensional waveguide model has been implemented to the vocal tract for the real-time solution of the wave propagation (KELLY, LOCHBAUM, 1962). In multi-dimensional wave propagation in the vocal tract, two- and three- dimensional digital waveguide models have been used to obtain better accuracy (SPEED *et al.*, 2013a; 2013b; VAN DUYNE, SMITH, 1993a; 1993b).

2.1. One-dimensional waveguide model

In one-dimensional waveguide model, a vocal tract is approximated by the concatenation of uniform cylindrical segments of different cross-sectional areas. With the help of wave equation, the relationship between the velocity and pressure within the uniform tube has been developed in (MARKEL, GRAY, 1976; RABINER, SCHAFER, 1978). D'Alembert's solution of the one-dimensional wave equation is comprised of the sum of left and right traveling-wave components. The solution of the continuity and momentum equation (KARJALAINEN, ERKUT, 2004) at the junction of two i -th and $(i + 1)$ -th cylinders is given in the form of the

reflection coefficient r_i as in Eq. (1) where A_i is the cross-sectional area of cylinder i :

$$r_i = \frac{A_i - A_{i+1}}{A_i + A_{i+1}}. \quad (1)$$

By using reflection coefficient r_i at each junction of consecutive two cylinders, the iteration of scattering at each junction results in propagation of a traveling wave through the vocal tract.

2.2. Two-dimensional digital waveguide model

The accuracy of the one-dimensional waveguide can be increased by an extension to generate a lattice structure of the waveguides in form of increasing connectivity of the spatio-temporal sample grid called a digital waveguide mesh (MULLEN *et al.*, 2003; 2006; VAN DUYNÉ, SMITH, 1993a; 1993b). The scattering of one-dimensional waveguide model is performed at each node of the waveguide mesh which leads to implicit solution the higher dimensional wave equation. Several topologies of the mesh may be implemented in the digital waveguide model. In the current work, the two-dimensional rectilinear mesh has been considered due to its simple and basic configuration. In the current configuration of the rectilinear mesh, the junctions are positioned at same intervals on a grid in such a way that each junction has four neighboring junctions at an angle of 90 degrees from one another. Figure 1 illustrates the orientation of the simplest rectilinear mesh and scattering junctions while Fig. 2 represents

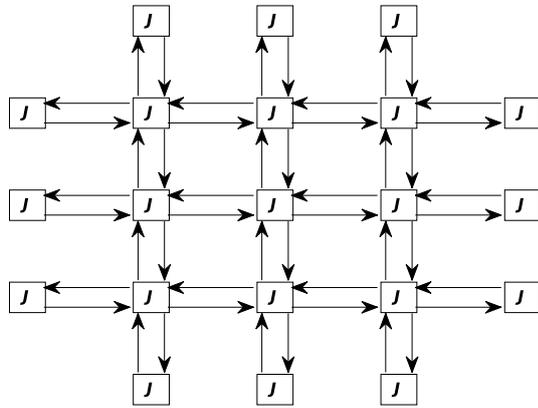


Fig. 1. A view of the simplest rectilinear grid with boundary junctions.

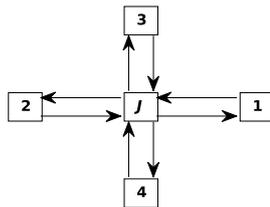


Fig. 2. A view of single-junction with its neighboring junctions.

the single junction with its neighboring junctions. We note from the figures that there are two types of arrows constituted in horizontal and vertical directions, respectively.

The directions of the arrows show the flow of the wave pressure from one junction to its neighboring junctions. In the literature, the wave pressure arrived at the junction is represented by the sign of positive pressure while it is denoted by the sign of negative pressure when it leaves the junction. From Fig. 2, we note that the single junction denoted by J is interconnected with junction 1 from the east, junction 2 from the west, junction 3 from the north and junction 4 from the south. The wave pressures come at junction J from junction 1 is denoted by $p_{J,1}^+$, from junction 2 is denoted by $p_{J,2}^+$, from junction 3 is denoted by $p_{J,3}^+$ and from junction 4 is denoted by $p_{J,4}^+$. The outgoing pressures from junction J to junctions 1, 2, 3, and 4 are denoted by $p_{J,1}^-$, $p_{J,2}^-$, $p_{J,3}^-$, and $p_{J,4}^-$, respectively. So, the pressure $p_{J,i}$ on each waveguide can be calculated as (MULLEN, 2006; VAN DUYNÉ, SMITH, 1993b)

$$p_{J,i} = p_{J,i}^+ + p_{J,i}^-. \quad (2)$$

The total pressure p at each junction J for N -port junctions can be written in Eq. (3) as (MULLEN, 2006; VAN DUYNÉ, SMITH, 1993b)

$$p_J = 2 \frac{\sum_{i=1}^N Y_i p_{J,i}^+}{\sum_{i=1}^N Y_i}, \quad (3)$$

where Y_i is the acoustics impedance in the medium.

In the proposed work, each junction J at rectilinear grid has four intersecting junctions with the assumption of homogeneous acoustic impedance, then the Eq. (3) can be modified as

$$p_J = \frac{\sum_{i=1}^4 p_{J,i}^+}{2}. \quad (4)$$

Output traveling-wave components are determined by taking the difference between the pressure at each node and the incoming traveling wave at each connection as shown in the Eq. (5):

$$p_{J,i}^- = p_J - p_{J,i}^+. \quad (5)$$

The traveling waves arrived at the mesh nodes on the boundary of the vocal tract are scattered with help of wall reflection coefficients r_w . In other words, a proportional amount of traveling wave that arrived at the boundary node is reflected back into the interior mesh. For the case of single connection at node A , we have

$$p_{J,A}^- = r_w p_{J,A}^+. \quad (6)$$

In the N -dimensional waveguide model, the sampling frequency f_s is found by using the distance between two consecutive nodes d and speed of sound c (MULLEN *et al.*, 2007)

$$f_s = \frac{c\sqrt{N}}{d}. \quad (7)$$

In the current work, the value of N is 2 and the speed of sound is assumed as $c = 345$ m/s. In rectangular meshing arrangement, the path between any two junctions of the mesh can either be even or odd that limit the sampling frequency valid up to $f_s/4$ (VAN DUYNÉ, SMITH, 1993b).

2.3. Non-linear rectilinear grid meshing

There are numerous grid topologies which can be used for the waveguide modeling of the vocal tract. A few of them are proposed as square, triangular, hexagonal, rectilinear and tetrahedral for two- or three-dimensional waveguide models (FONTANA, ROCCHESO, 1995; 2001; SAVIOJA *et al.*, 1994; VAN DUYNÉ, SMITH, 1993b; 1996). The drawback of these topologies in the waveguide model is the limitation of fixed unit-length digital waveguide branches by means of lossless scattering junctions. The easiest way to handle the topologies is the rectilinear topologies. In these topologies, the wave propagation speed from a junction to its neighbor through fixed distance is relative to $1/\sqrt{N}$ where N is the number of dimensions (MULLEN *et al.*, 2007). In digital waveguide models, the propagation wave travels from junction to another junction through a path that always has either only even or odd numbers of unit delays (MULLEN *et al.*, 2007; VAN DUYNÉ, SMITH, 1993b) which leads to sampling frequency as $f_s/4$.

We choose rectilinear grid topology due to its simplicity in the present work. We introduce a novel approach in the generation and implementation of the non-uniform rectilinear grid in the digital waveguide models of the vocal tract. Using this approach, the comparable formant frequencies are obtained with less computational cost. In the current grid orientation, there are two types of grid lines. One is the smaller-grid line and other is the larger-grid line which is the double of the smaller-grid line. The scheme of using the combination of both smaller- and larger-delay lines generates the non-uniform rectilinear two-dimensional waveguide mesh. This scheme enables us to get a transfer function by propagation wave sound in the non-uniform grid without fractional delay. This eliminates the need to get interpolation for the approximation of fractional delay and give efficient simulation for sound wave propagation in the two-dimensional waveguide modeling of the vocal tract in the non-uniform grid. An example of simple uniform rectilinear mesh is shown in Fig. 3. The current figure has the simple topology

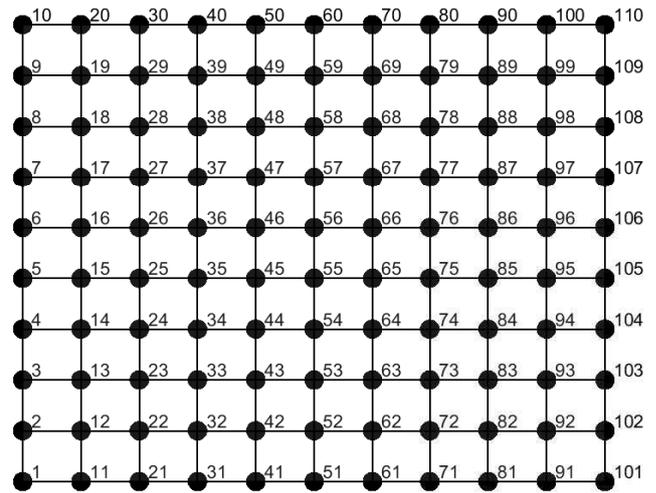


Fig. 3. A simple example of the uniform rectilinear mesh.

of 10×11 ordered mesh with the total number of 110 nodes. Each node in the grid is located by its integer number and these nodes are used for the scattering of the wave propagation. The interior nodes of the mesh are orientated as four-port junctions. The boundary nodes in the mesh are three-port junctions while boundary nodes at corners two-port junctions. We note that there are 199 edges of the same length in total that act as the fixed-sample delay for the wave propagation. Figure 4 is representing the non-uniform rectilinear mesh grid which is obtained from the uniform mesh grid shown in Fig. 3 by deleting some nodes. In the present figure, we delete the following nodes numbered as 12, 32, 52, 72, 92, 14, 34, 54, 74, 94, 16, 36, 56, 76, and 96 from the mesh shown in Fig. 3. We observed that there are two types of squares in which some squares are smaller than the other one. The smaller squares are one-fourth of the big one. In the current orientation of the mesh, all nodes are not used for the purpose of the scattering of the wave propaga-

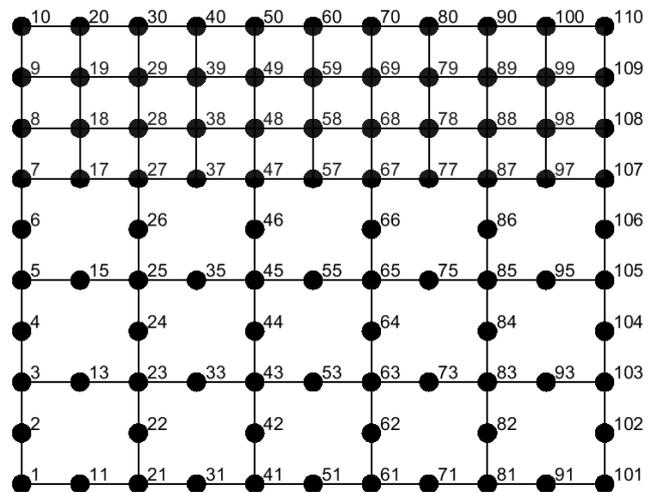


Fig. 4. A simple example of the non-uniform rectilinear mesh.

tion. Some nodes are assumed as the local-nodes of the long edges in the horizontal and the vertical directions. In the present figure, the nodes which are numbered as 11, 13, 15, 31, 33, 35, 51, 53, 55, 71, 73, 75, 91, 93, 95 are the local-nodes of the long edges in horizontal direction while nodes 2, 4, 6, 22, 24, 26, 42, 44, 46, 62, 64, 66, 82, 84, 86, 102, 104, and 106 are the local-nodes of the long edges in vertical direction. These nodes are not used for the scattering of the wave propagation but these are used for keeping to track the movement of waves locally within the long edges of the grid. In the rest of the present paper, we refer the above-mentioned nodes as local-nodes while other nodes are referred to as junctions. Each local-node in the horizontal long edges has two-way movement which may be named as east and west respectively. Similarly, each local-node in the vertical long edges has two-way movement which may also be named as north and south respectively. There are three steps involved in the simulation of wave propagation for the vocal tract. In the first step, we find the scattering of the waves at each junction of the mesh by using Eqs (4) and (5). The movement of the waves at each local-node of the mesh is performed in the second step which is described as,

$$\left\{ \begin{array}{l} \left(\begin{array}{l} p_{i,E}^+ = p_{i,W}^- \\ p_{i,W}^+ = p_{i,E}^- \end{array} \right) \text{ for each local-node} \\ \qquad \qquad \qquad \text{in the horizontal direction,} \\ \left(\begin{array}{l} p_{i,S}^+ = p_{i,N}^- \\ p_{i,N}^+ = p_{i,S}^- \end{array} \right) \text{ for each local-node} \\ \qquad \qquad \qquad \text{in the vertical direction,} \end{array} \right. \quad (8)$$

where i denotes the local-node and E , W , N , and S are representing the east, west, north, and south, respectively. In the last step, we pass the delay at each junction and the propagation waves are also arrived at each local-node from its neighboring junctions.

Let d_L be the length of single-long edge formed by the connection of three nodes, for example, nodes 23, 33, and 43 in the mesh form a single-long edge and d_s denotes the length of the small edge formed by the connection of two nodes such as 28 and 38 in Fig. 4, then we have

$$d_L = 2d_s. \quad (9)$$

We also note that the length of each edge in the mesh is either the length of d_L or the length of d_s . In this work, the sampling frequency is obtained by using the smallest length d_s in Eq. 7, then

$$f_s = \frac{c\sqrt{2}}{d_s}, \quad (10)$$

where c is described earlier.

3. Numerical work for simulations

The basic approach for the two-dimensional non-uniform mesh of the vocal tract has been presented in

the previous section. The numerical approach for the simulation of the current work will be demonstrated in this section. The shape of the vocal tract is converted into a series of equivalent cross-sectional areas along the length of the vocal tract. The series of the cross-sectional areas for the vocal tract shape of the vowels /ɔ/, /a/, /i/, and /u/ have been taken from (STORY *et al.*, 1996). In order to translate the shape of the vocal tract into a two-dimensional mesh, the cross-sectional areas are converted into length-wise acoustics tubes of different radii. To express the shape of the vocal tract by a smooth curve, piece-wise third-order spline interpolation is taken in this work. A two-dimensional non-uniform mesh is then constructed with the help of above-mentioned smooth shape of the vocal tract as illustrated in Fig. 5. The non-linear meshing with finer-mesh along the boundaries of the vocal tract has been shown in Fig. 5, The smallest sample delay size in all demonstrations of the current work for the digital waveguide model of the two-dimensional mesh has been opted as $d_s = 0.0585$ cm which results in the sampling frequency of $f_s = 834$ kHz.

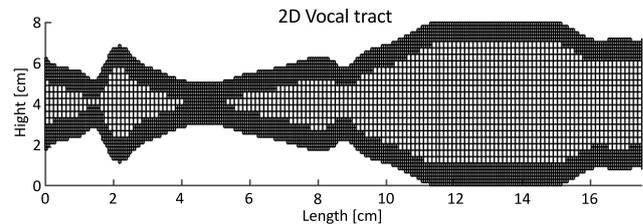


Fig. 5. The finer-boundary meshing of the non-uniform mesh of the vocal tract.

Boundary reflection coefficients on the glottis, walls and lip radiation are assigned as $r_G = 0.97$, $r_w = 1.0$, and $r_L = -0.9$, respectively. For transfer function of the vocal tract, the length of input signals is taken as 500000 samples long. The large size of the input samples leads to a good resolution at the frequency profile. The output samples are transformed into magnitude response with the help of Fast Fourier Transformation (FFT). The simulation has been performed in the Matlab 2017 as a programming language. Double precision floating point variables and arrays are used to maintain the accuracy of the results. A C++ file is also developed for highly cost computational algorithms to increase the efficiency of the computational work in the current computer language. Compilation of the C++ file is performed within the Matlab environment by using Visual C++ compiler. This process generates another file which can easily be linked with Matlab for the purpose of efficient simulations.

4. Results and discussion

In the previous sections, we have defined the construction of the non-uniform rectilinear mesh and its

implementation in the waveguide model of the vocal tract. We also compare our current model with the benchmark model. The standard two-dimensional waveguide model with uniform mesh is assumed as our benchmark model. The comparison between the two models has been made on the basis of formant frequencies. Each vowel needs at least three formant frequencies for its recognition (MULLEN, 2006). However, we consider the first six formant frequencies generated by our proposed model and benchmark model in all our demonstrations. These formant frequencies are represented by F1, F2, F3, F4, F5, and F6 in the tables of the current work.

For the sake of comparison, the smallest sample delay d_s in all demonstrations of the present model has been kept the same as that of our benchmark model. The series of the cross-sectional areas for different vocal tract shapes of the vowels are mentioned in (STORY *et al.*, 1996) and we choose the vowels /ɔ/, /a/, /i/, and /u/ for the presentations of the current work. The following vowels /ɔ/, /a/, /i/, and /u/ have the vocal tract shapes of length 17.46 cm, 17.46 cm, 16.67 cm and 18.25 cm respectively. The computational efficiency in term of elapsed time and accuracy of the current model in term of formant frequencies has been compared with the benchmark model. The proposed model allows finer mesh along with the selected layers of the wall boundaries of the vocal tract and there is a coarser mesh along with the interior layers of the mesh. In this style, the number of layers for the finer along with the boundaries is specified by the parameter B_L . This case is shown in Fig. 5 with four finer layers near to the boundaries of the vocal tract, i.e., $B_L = 4$.

Figures 6–8 represent the frequency profiles of the vocal tract transfer function for the vowel /ɔ/ up to 6000 Hz. These three figures are generated by the dif-

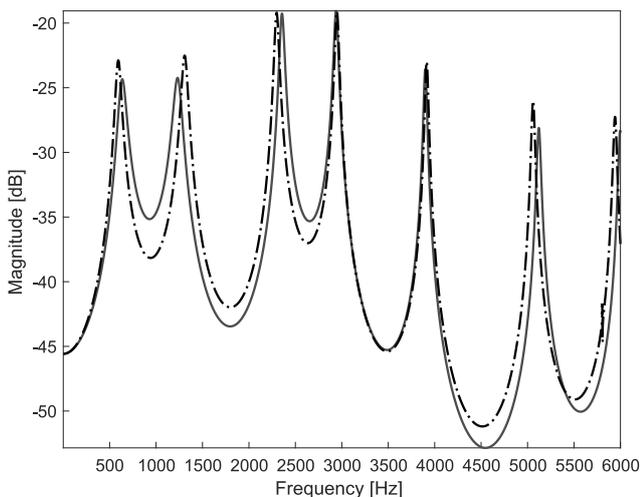


Fig. 6. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /ɔ/, for $B_L = 4$.

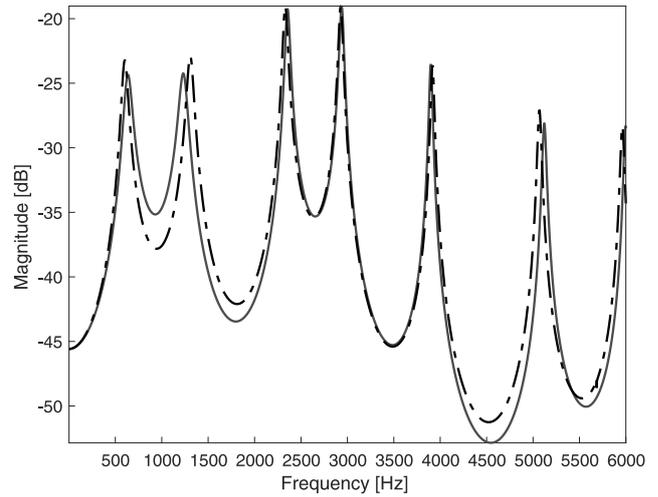


Fig. 7. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /ɔ/, for $B_L = 8$.

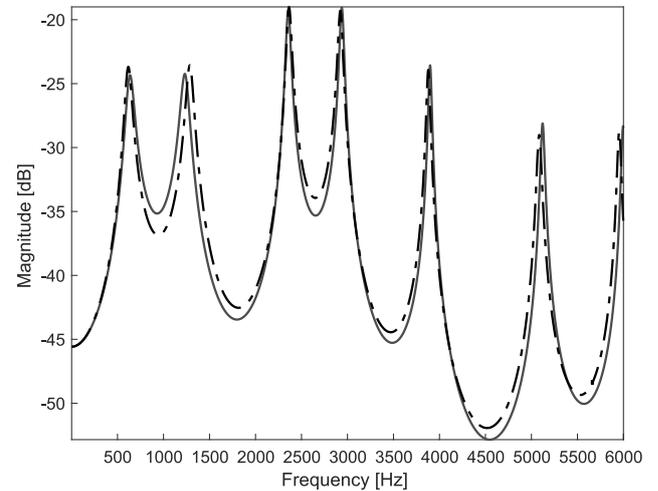


Fig. 8. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /ɔ/, for $B_L = 12$.

ferent values of the parameter B_L which are assumed as 4, 8, and 12, respectively. The solid curves in the figures are the frequency profiles generated by the transfer function of the benchmark model and the frequency profiles generated by the transfer function of the proposed model are shown with the dash-dotted curve. The formant frequencies in the profiles are identified by the peaks in the frequency profiles. In the later discussion, we will use these formants frequencies as a reference in the figures. Corresponding to these three figures, Tables 1–3 represent the comparison of the measured formant frequencies of the vowel /ɔ/ with the different values of the parameter B_L . First, second, third, fourth, fifth, and sixth formant frequencies are represented by F1, F2, F3, F4, F5, and F6, respectively, in the first column of the tables. The columns 2 and 3 represent formant frequencies of the benchmark

Table 1. Comparison of the current model with the benchmark model for vowel /ɔ/, for $B_L = 4$.

Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	636	603	5.2	149.08	122.55	22
F2	1232	1302	5.7			
F3	2356	2301	2.4			
F4	2938	2945	0.2			
F5	3898	3911	0.3			
F6	5121	5060	1.1			

Table 2. Comparison of the current model with the benchmark model for vowel /ɔ/, for $B_L = 8$.

Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	636	610	4.1	149.08	129.12	15
F2	1232	1294	5.0			
F3	2356	2331	1.1			
F4	2938	2928	0.3			
F5	3898	3903	0.13			
F6	5121	5073	0.94			

Table 3. Comparison of the current model with the benchmark model for vowel /ɔ/, for $B_L = 12$.

Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	636	622	2.2	149.08	134.79	10
F2	1232	1272	3.2			
F3	2356	2363	0.3			
F4	2938	2930	0.3			
F5	3898	3879	0.27			
F6	5121	5096	0.49			

model and the proposed model, respectively and column 4 shows the relative error of the current model with the benchmark model. Elapsed times and the efficiency of the benchmark and the proposed models are given in the columns 5-7, respectively, and the unit of the elapsed time is second.

Figure 6 shows the frequency profile of the vocal tract transfer function for the vowel /ɔ/ up to 6000 Hz. In this case the values of B_L is chosen as 4. We see that some formant frequencies of the current model are very closed to that of the benchmark model and some formant frequencies have a little deviation from that of the benchmark model. This leads to the fact that the frequencies profiles of the current model and benchmark model are close to each other. With respect to Fig. 6, numerically calculated formant frequencies of the current model and benchmark model are given

in Table 1. The relative errors of the present model with respect to the benchmark model are mentioned in the fourth column of the table. The relative errors of the first six formant frequencies are shown as 5.2%, 5.7%, 2.4%, 0.2%, 0.3%, and 1.1%, respectively. A maximum relative error of the present model is found at the second formant frequency with the value of 5.7% and it has a minimum relative error at fourth formant frequency with the value of 0.2%. The relative errors show that the current model and benchmark model have negligible errors and these profiles are close to each other. The time taken by the current and the benchmark models are mentioned in the table as 149.08 and 122.55, respectively, which shows that the current model is about 22% more efficient than that of the benchmark model. It is concluded from Table 1 that the current model has very comparable formant fre-

quencies with that of the benchmark model by using the value of $B_L = 4$ and there is also a significant reduction in the computational cost.

Figure 7 illustrates the similarities of the frequency profile of the proposed model with that of the benchmark model for the case of $B_L = 8$. The matching of the frequency profiles of the current model to that of the benchmark model has better than the previous case. Some formant frequencies of the current model are well aligned to that of the benchmark model while other formant frequencies are close to it. There is also a good matching of the current model to the benchmark model with respect to their formant frequencies. The measured formant frequencies of the present model and the benchmark model are shown in Table 2. The relative errors of the first six formant frequencies are given as 4.1%, 5.0%, 1.1%, 0.3%, 0.13%, and 0.94%, respectively. A maximum relative error of the proposed model is 5.0% located at the second formant frequency while the minimum relative error is found as 0.13% at fourth formant frequency. The negligible relative errors show that the current model and the benchmark model are also close to each other and the present model has better accuracy as compared to the previous case. The elapsed times obtained from the current model and the benchmark model are given in the table as 149.08 and 129.12, respectively. The last column of the current table exhibits that the current model is approximately 15% more efficient than that of the benchmark model. It is also noted that the current model has better accuracy as compared to the previous one with little increase in the computational cost.

Similarly, in the case of $B_L = 12$, the comparison of the frequency profiles of the two model are demonstrated in Fig. 8. It is observed that the accuracy of the current model with that of the benchmark model is much better than the previous two cases concerning the formant frequencies. The frequency profile of the proposed model is much closer to that of the benchmark model. Table 3 is illustrating the measured formant frequencies of the present case. The relative errors are given as 2.2%, 3.2%, 0.3%, 0.3%, 0.27%, and 0.49%, respectively. The maximum and minimum relative errors are found as 3.2% and 0.3%, respectively. The measured elapse times by the current model and the benchmark model are given in the table as 149.08 and 134.79, respectively. The current model is approximately 10% more efficient than that of the benchmark model corresponding to the last column of the current table. It is concluded from the above three tables that the accuracy of the proposed model increases and efficiency of the current model decreases with the increase in the value of the parameter B_L .

The two- and three-dimensional views for the changes over the formant frequencies and the frequencies profiles along with the parameter B_L are shown in

Figs 9 and 10, respectively. In Fig. 9, the bright color lines are representing the changes in the six formant frequencies along with the parameter B_L . It is noted from the figure that the formant frequencies have insignificant changes after the value of $B_L = 15$. The irregular and smooth less transition is observed for the value of B_L ranging from 1 to 3 while the significant and smooth changes are noted from values ranging 4 to 14. Similarly, the three-dimensional view of the current case is displayed in Fig. 10.

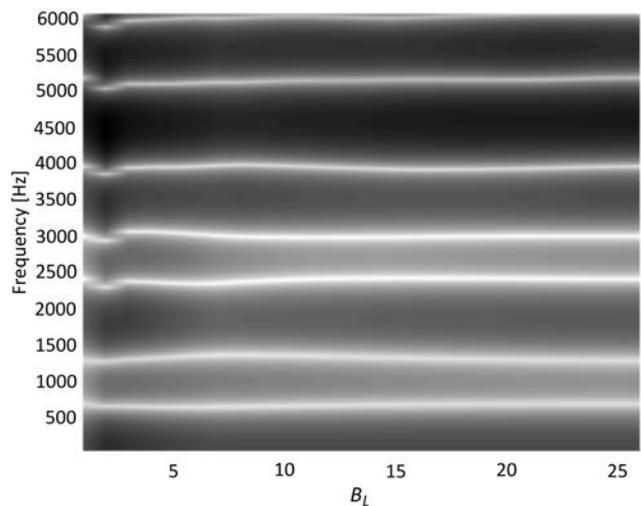


Fig. 9. Two-dimensional representation for the variation of formant frequencies with the variation in the values of the parameter B_L for vowel /ɔ/.

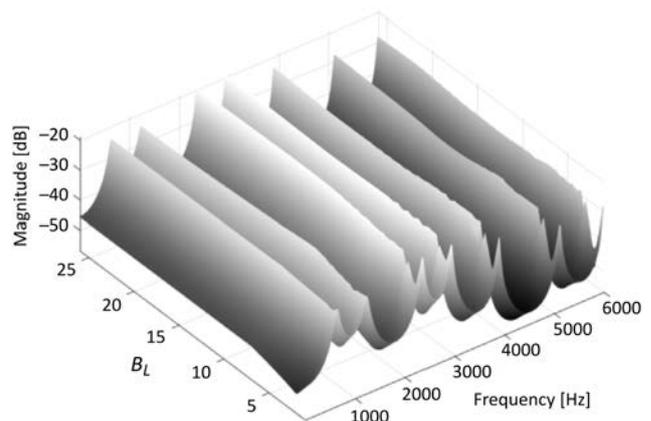


Fig. 10. Three-dimensional representation for the variation of formant frequencies with the variation in the values of the parameter B_L for vowel /ɔ/.

The frequency profiles of the transfer function for vowel /a/ has been demonstrated in Figs 11–13 for three different values of the parameter B_L taken as 4, 8, and 12, respectively. The numerically calculated formant frequencies of the vowel /a/ corresponding to three different values of the parameter B_L have been mentioned in Tables 4–6. These three figures show that

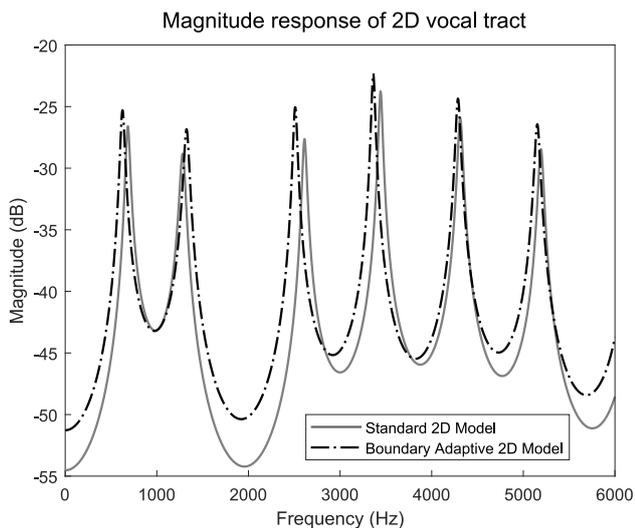


Fig. 11. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /a/, for $B_L = 4$.

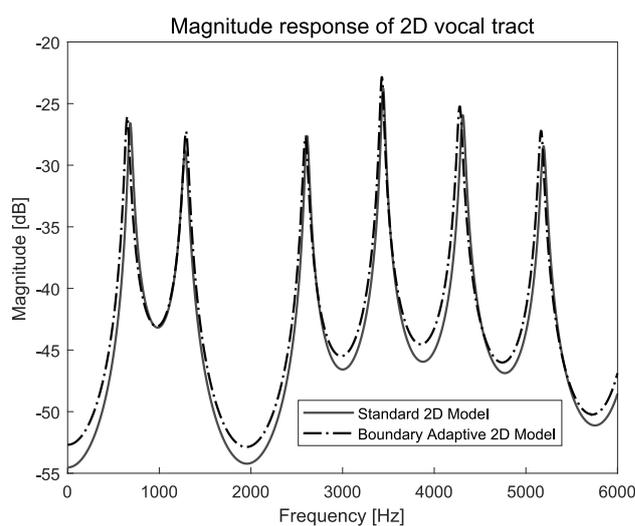


Fig. 13. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /a/, for $B_L = 12$.

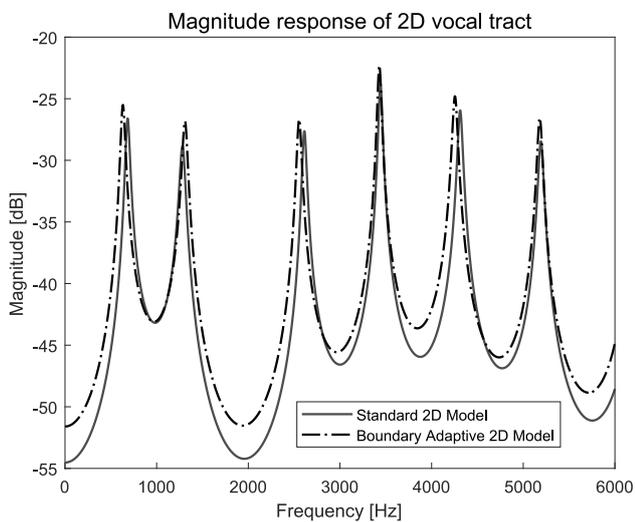


Fig. 12. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /a/, for $B_L = 8$.

the convergence of the current model to the benchmark model increases with the decrement in the values of the parameter B_L . In the case of Fig. 11, mea-

sured formant frequencies of the current model and benchmark model are given in Table 4 with the efficiency. The relative errors are found as 8.0%, 3.3%, 3.8%, 2.3%, 0.5%, and 0.8%, respectively, for first six formant frequencies. A maximum relative error of the current model is shown as 8.0% while a minimum relative error is given as 0.5% for the current value of the parameter B_L . The little higher errors are measured in this case as compared to the same case for vowel /ɔ/. The elapsed time of the current and the benchmark models are given in the table as 184.78 and 144.05, respectively, which leads to the efficiency of 28% of the current model. In the view of Table 4, we observe that the formant frequencies of the current model are comparable with that of the benchmark model with high-efficiency value.

Table 5 gives the comparison and efficiency of the current model with that of the benchmark model corresponding to Fig. 12. The relative errors for first six formant frequencies are calculated as 6.9%, 2.0%, 2.1%, 0.3%, 1.3%, and 0.2%, respectively. The maximum and minimum relative errors are respectively 6.9% and 0.2% for the value of $B_L = 8$. An improve-

Table 4. Comparison of the current model with the benchmark model for vowel /a/, for $B_L = 4$.

Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	683	628	8.05	184.78	144.08	28
F2	1282	1324	3.27			
F3	2611	2511	3.83			
F4	3442	3362	2.32			
F5	4311	4288	0.53			
F6	5193	5153	0.77			

Table 5. Comparison of the current model with the benchmark model for vowel /a/, for $B_L = 8$.

Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	683	636	6.88	184.78	153.13	21
F2	1282	1307	1.95			
F3	2611	2555	2.14			
F4	3442	3430	0.35			
F5	4311	4255	1.30			
F6	5193	5181	0.23			

Table 6. Comparison of the current model with the benchmark model for vowel /a/, for $B_L = 12$.

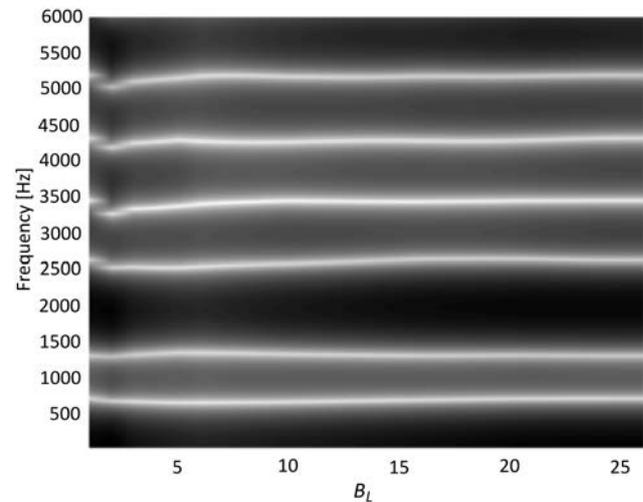
Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	683	656	3.95	184.78	160.78	15
F2	1282	1292	0.78			
F3	2611	2595	0.61			
F4	3442	3429	0.38			
F5	4311	4278	0.77			
F6	5193	5166	0.52			

ment in relative errors is observed with the increasing value of B_L . The computational times measured for the current model and the benchmark model are, respectively, 184.78 and 153.13. The current table shows that the formant frequencies of the proposed model are closed to that of the benchmark model with the efficiency of 21%.

Corresponding to Fig. 13, the relative errors have been measured as 4.0%, 0.78%, 0.61%, 0.38%, 0.77%, and 0.5%, respectively, for the value of $B_L = 12$ which are specified in Table 6. It is observed that the accuracy of the present model with that of the benchmark model is much better than the previous two cases concerning the formant frequencies. The formant frequencies of the proposed model are much closer to that of the benchmark model with the efficiency of 15%.

The visual tracking of the formant frequencies and the frequencies profiles for vowel /a/ along with the parameter B_L are demonstrated with two- and three-dimensional graphs in Figs 14 and 15, respectively. We note that there is no significant change in the six formant frequencies profiles after $B_L = 15$.

Figures 16–18 demonstrates the frequency profiles of the transfer function of vowel /i/ for the same different values of the parameter B_L as mentioned before. The convergence of the present model to benchmark model may be observed in these figures. Tables 7–9 describe the calculated formant frequencies of the vowel /i/ corresponding to the different values of the parameter B_L from the profiles of Figs 16–18. Table 7 defines

Fig. 14. Two-dimensional representation for the variation of formant frequencies with the variation in the values of the parameter B_L for vowel /a/.

the following relative errors as 9.7%, 7.0%, 3.5%, 2.9%, 1.2%, and 0.5% by using the value of $B_L = 4$ while improved relative errors are obtained as 4.1%, 3.9%, 0.1%, 0.8%, 0%, and 0.1% by increasing the value of B_L as 8 that are found in Table 8. However, the formant frequencies of the current model are very closed to that of the benchmark model by taking the value of $B_L = 12$ as given in Table 9 in which a maximum relative error is 1.2% and a minimum relative error

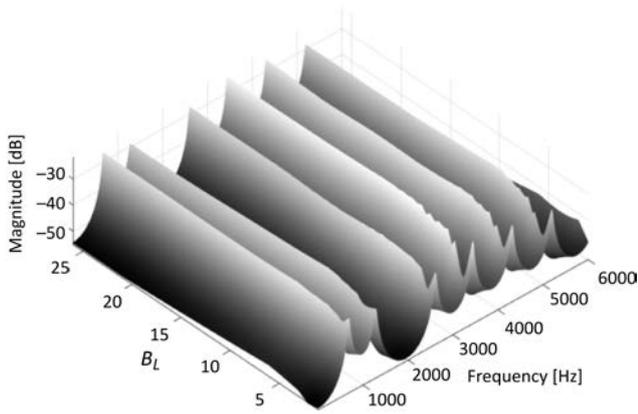


Fig. 15. Three-dimensional representation for the variation of formant frequencies with the variation in the values of the parameter B_L for vowel /a/.

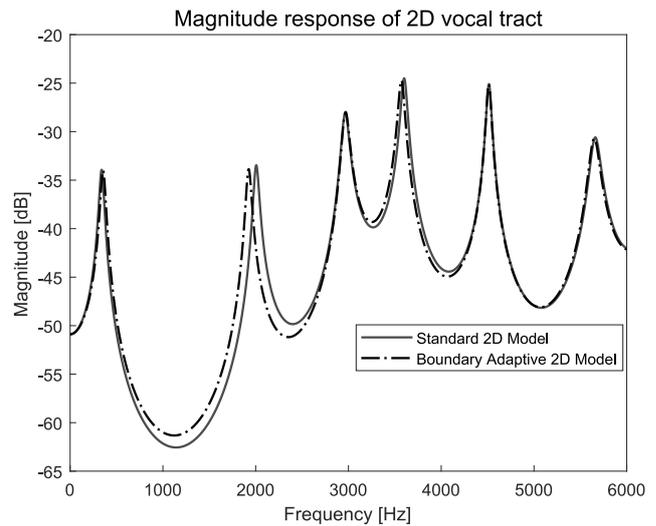


Fig. 17. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /i/, for $B_L = 8$.

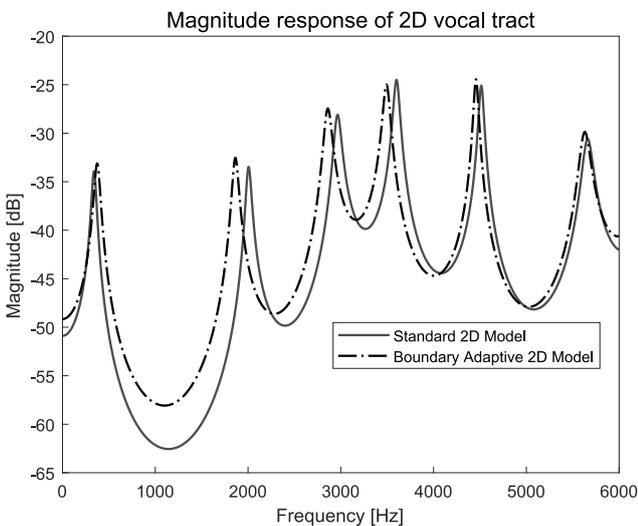


Fig. 16. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /i/, for $B_L = 4$.

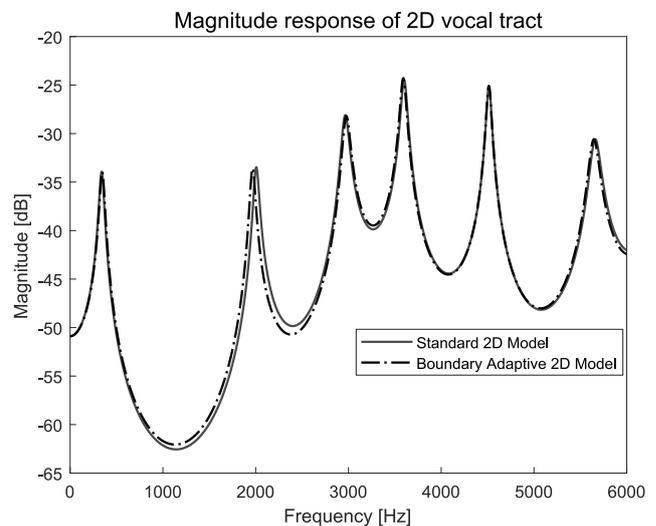


Fig. 18. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /i/, for $B_L = 12$.

is 0%. The efficiencies have been calculated as about 19%, 15% and 9%, respectively, with the values of 4, 8, and 12 for the parameter B_L in the present case. Figures 19–20 illustrate the two- and three-dimensional

graphs for the convergence of the proposed model with the increasing values of the parameter B_L .

Table 7. Comparison of the current model with the benchmark model for vowel /i/, for $B_L = 4$.

Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	340	373	9.70	149.85	125.90	19
F2	2005	1864	7.03			
F3	2965	2861	3.50			
F4	3600	3495	2.92			
F5	4513	4459	1.19			
F6	5660	5632	0.49			

Table 8. Comparison of the current model with the benchmark model for vowel /i/, for $B_L = 8$.

Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	340	354	4.11	149.85	130.31	15
F2	2005	1926	3.94			
F3	2965	2967	0.07			
F4	3600	3573	0.75			
F5	4513	4513	0			
F6	5660	5652	0.14			

Table 9. Comparison of the current model with the benchmark model for vowel /i/, for $B_L = 12$.

Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	340	344	1.18	149.85	137.55	9
F2	2005	1969	1.79			
F3	2965	2975	0.34			
F4	3600	3593	0.19			
F5	4513	4513	0			
F6	5660	5648	0.21			

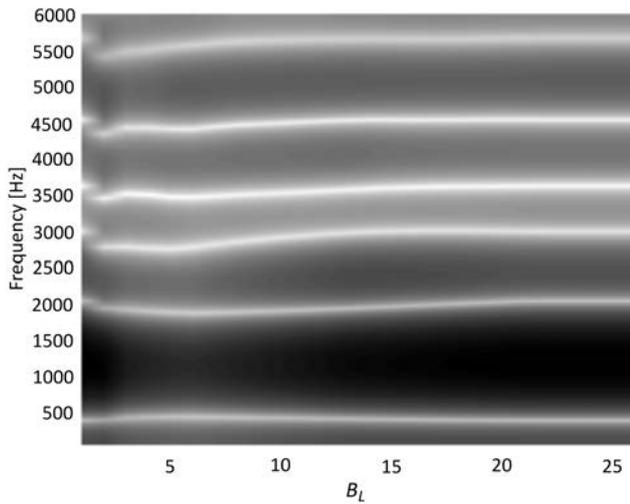


Fig. 19. Two-dimensional representation for the variation of formant frequencies with the variation in the values of the parameter B_L for vowel /i/.

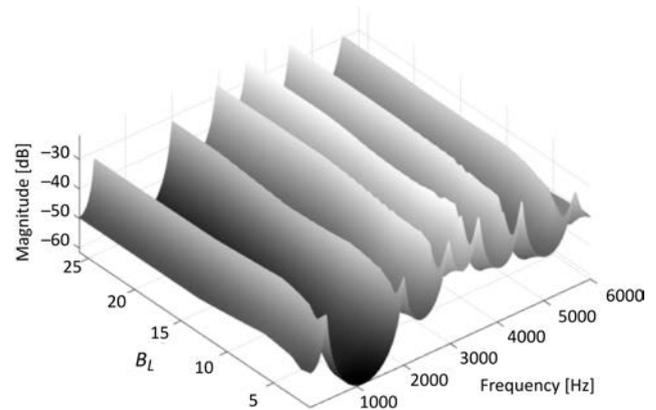


Fig. 20. Three-dimensional representation for the variation of formant frequencies with the variation in the values of the parameter B_L for vowel /i/.

For the case of vowel /u/, the frequency profiles have been shown in Figs 21–23 respectively for three different values of the parameter B_L . It is noted that the increment in the values of B_L bring the improvement in the convergence of the present model to the benchmark model. The calculated formant frequencies from Figs 21–23 are mentioned in Tables 10–12, respectively, in the current case of the vowel. Table 10 has been observed with the following relative errors

8.7%, 2.1%, 2.1%, 3.8%, 3.0%, and 1.7% by fixing the value of $B_L = 4$. These relative errors are further improved by 5.9%, 1.8%, 3.0%, 1.6%, 0.9%, and 1.6% for assuming the value of the parameter $B_L = 8$ as shown in Table 11. By fixing the value of $B_L = 12$, the more enhancement in the formant frequencies is achieved as 0.84%, 0.33%, 2.64%, 0.63%, 0.07%, and 0.95% by observing in Table 12 which exhibits the good approximation of current model to the benchmark model. From Tables 10–12, the efficiencies of the proposed model are measured as about 23%, 17%, and

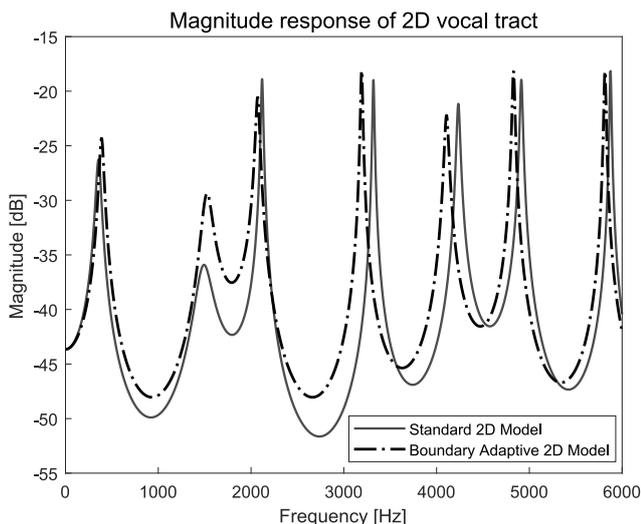


Fig. 21. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /u/, for $B_L = 4$.

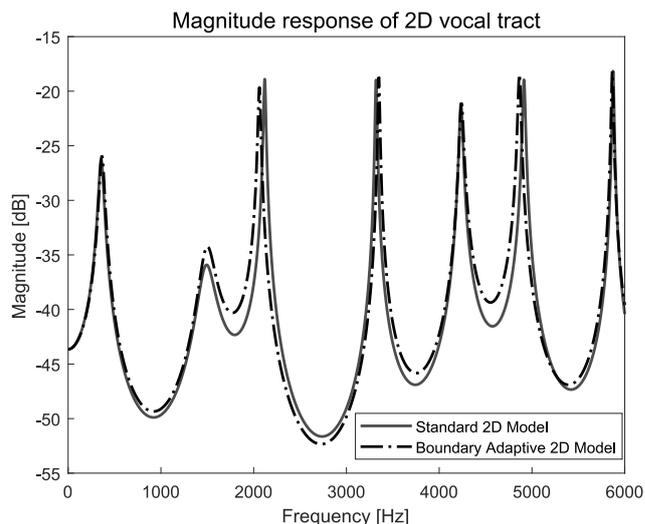


Fig. 23. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /u/, for $B_L = 12$.

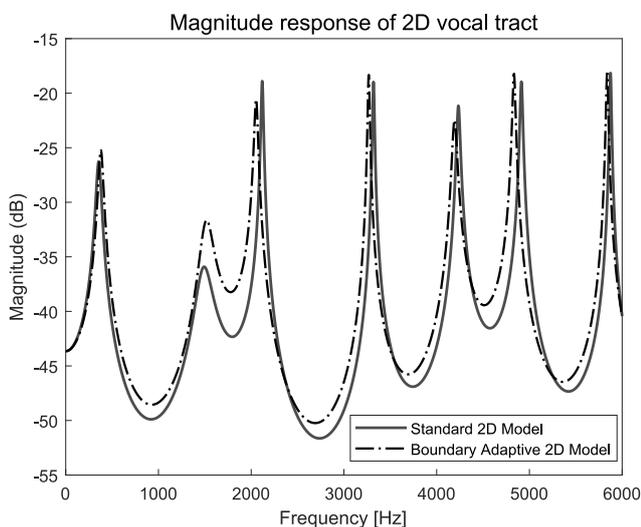


Fig. 22. Comparison of formant frequencies of the proposed model with the benchmark model for the vowel /u/, for $B_L = 8$.

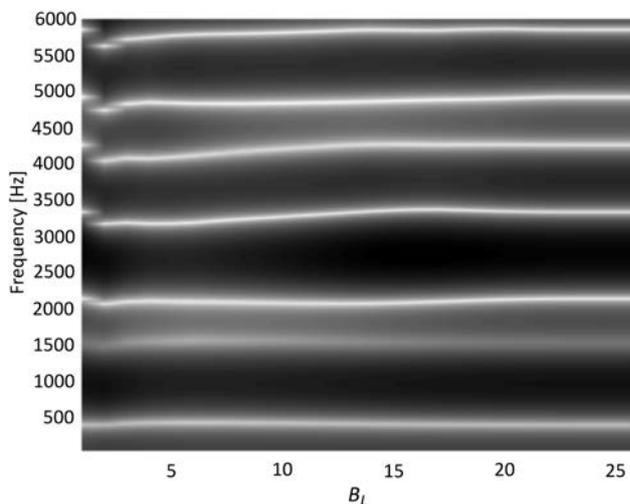


Fig. 24. Two-dimensional representation for the variation of formant frequencies with the variation in the values of the parameter B_L for vowel /u/.

13%, respectively, with the values of 4, 8, and 12 for the parameter B_L . Figures 24 and 25 show the two-

and three-dimensional graphs for the convergence of the current model with the increasing values of the parameter B_L .

Table 10. Comparison of the current model with the benchmark model for vowel /u/, for $B_L = 4$.

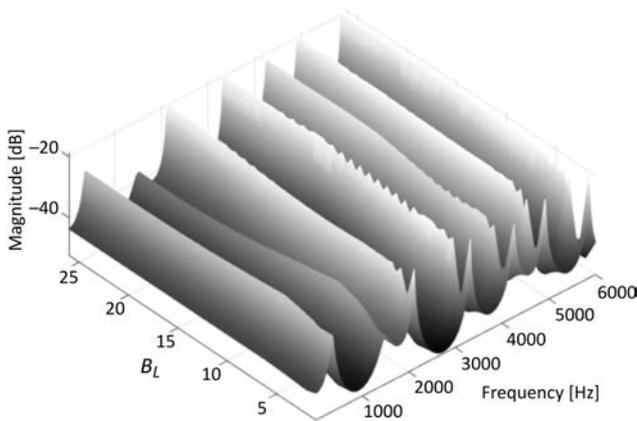
Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	357	388	8.68	190.07	153.98	23
F2	1494	1526	2.14			
F3	2118	2072	2.17			
F4	3319	3193	3.79			
F5	4234	4108	2.98			
F6	4913	4831	1.67			

Table 11. Comparison of the current model with the benchmark model for vowel /u/, for $B_L = 8$.

Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	357	378	5.88	190.07	162.16	17
F2	1494	1521	1.80			
F3	2118	2054	3.02			
F4	3319	3267	1.57			
F5	4234	4194	0.94			
F6	4913	4834	1.61			

Table 12. Comparison of the current model with the benchmark model for vowel /u/, for $B_L = 12$.

Formant frequency	Benchmark model	Proposed model	Relative error [%]	Elapsed time		Efficiency [%]
				Benchmark model	Proposed model	
F1	357	360	0.84	190.07	167.65	13
F2	1494	1499	0.33			
F3	2118	2062	2.64			
F4	3319	3240	0.63			
F5	4234	4237	0.07			
F6	4913	4866	0.96			

Fig. 25. Three-dimensional representation for the variation of formant frequencies with the variation in the values of the parameter B_L for vowel /u/.

5. Conclusions

In this work, a novel approach of non-uniform meshing has been used for the modeling of the vocal tract. In this approach, the delay lines among the nodes of the grid are not of the same delay. These delay lines are integer multiple of each other. In this approach, smaller-delay lines and larger-delay lines are two types of delay lines where the larger-delay lines are the double of the smaller-delay lines. For the demonstration, the selected number of mesh layers along with the walls of the vocal tract has been taken as the finer mesh by the use of smaller-delay lines while the rest of mesh

layers have been constructed coarser mesh by the use of larger-delay lines. These selected number of layers for finer mesh has been denoted by the parameter B_L . This approach allows us to find the transfer function of the vocal tract in a non-uniform rectilinear mesh without fractional delay. This eliminates the need to get interpolation for the approximation of fractional delay and give efficient simulation for sound wave propagation in the two-dimensional waveguide modeling of the vocal tract in the non-uniform mesh. The simulation of the model has been performed on the vowels /ɔ/, /a/, /i/, and /u/ with the variation in the values of the parameter B_L . The current vowels /ɔ/, /a/, /i/, and /u/ have the vocal tract shapes of length 17.46 cm, 17.46 cm, 16.67 cm, and 18.25 cm, respectively. By keeping the same sampling frequency, the standard two-dimensional waveguide model with uniform meshing is assumed as the benchmark model. The results and efficiency have been compared with the benchmark model. In all cases, we draw the following conclusions:

- Formant frequencies of the proposed model of the vocal tract are comparable to that of the benchmark model.
- Relative errors of formant frequencies of the proposed model with the benchmark model are reduced with the increase in the value of the parameter B_L .
- The efficiency of the proposed model is reduced with the increase in the value of the parameter B_L .

- There are the smooth profiles of formant frequencies with the change in the values of the parameter B_L .
- For the large values of the parameter B_L , the profiles of the formant frequencies become constant.

Hopefully, the proposed model in the present work may serve as a useful vocal tract model in speech synthesizers.

References

1. BAILLY L. *et al.* (2018), 3D multiscale imaging of human vocal folds using synchrotron X-ray microtomography in phase retrieval mode, *Scientific Reports*, **8**(1): 14003, doi: 10.1038/s41598-018-31849-w.
2. BEESON M.J., MURPHY D.T. (2004), RoomWeaver: A digital waveguide mesh based room acoustics research tool, *Proceedings of the Seventh International Conference on Digital Audio Effects*, pp. 268–273, Naples, Italy, <http://www.mattmontag.com/auralization/media/RoomWeaver.pdf>.
3. BIRKHOLZ P., KRÖGER B.J., NEUSCHAEFER-RUBE C. (2010), Articulatory synthesis and perception of plosive-vowel syllables with virtual consonant targets, *Proceedings of the 11th Annual Conference of the International Speech Communication Association INTER-SPEECH 2010*, pp. 1017–1020, Chiba, Japan.
4. CAMPOS G., HOWARD D. (2000), 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)*, pp. 73–78, Verona, Italy.
5. CAMPOS G.R., HOWARD D.M. (2005), On the computational efficiency of different waveguide mesh topologies for room acoustic simulation, *IEEE Transactions on Speech and Audio Processing*, **13**(5): 1063–1072, doi: 10.1109/TSA.2005.852015.
6. COOPER C., MURPHY D., HOWARD D., TYRRELL A. (2006), Singing synthesis with an evolved physical model, *IEEE Transactions on Audio, Speech, and Language Processing*, **14**(4): 1454–1461, doi: 10.1109/TSA.2005.860844.
7. FLANAGAN J., LANDGRAF L. (1968), Self-oscillating source for vocal-tract synthesizers, *IEEE Transactions on Audio and Electroacoustics*, **16**(1): 57–64, doi: 10.1109/TAU.1968.1161949.
8. FONTANA F., ROCCHESO D. (1995), A new formulation of the 2D-waveguide mesh for percussion instruments, *Proceedings of the XI Colloquium on Musical Informatics*, pp. 27–30, Bologna, Italy.
9. FONTANA F., ROCCHESO D. (2001), Signal-theoretic characterization of waveguide mesh geometries for models of two-dimensional wave propagation in elastic media, *IEEE Transactions on Speech and Audio Processing*, **9**(2), 152–161, doi: 10.1109/89.902281.
10. ISHIZAKA K., FLANAGAN J.L. (1972), Synthesis of voiced sounds from a two-mass model of the vocal cords, *The Bell System Technical Journal*, **51**(6): 1233–1268, doi: 10.1002/j.1538-7305.1972.tb02651.x.
11. ISHIZAKA K., FLANAGAN J. (1977), Acoustic properties of longitudinal displacement in vocal cord vibration, *The Bell System Technical Journal*, **56**(6): 889–918, doi: 10.1002/j.1538-7305.1977.tb00546.x.
12. KARJALAINEN M., ERKUT C. (2004), Digital waveguides versus finite difference structures: Equivalence and mixed modeling, *EURASIP Journal on Applied Signal Processing*, **2004**(7): 561060, doi: 10.1155/S110865704401176.
13. KELLY J.L., LOCHBAUM C.C. (1962), Speech synthesis, *Proceedings of the Stockholm Speech Communications Seminar*, RIT, Stockholm, Sweden.
14. KUMAR S.P., ŠVEC J.G. (2019), Kinematic model for simulating mucosal wave phenomena on vocal folds, *Biomedical Signal Processing and Control*, **49**: 328–337, doi: 10.1016/j.bspc.2018.12.002.
15. MADDOX A., OREN L., KHOSLA S., GUTMARK E. (2014), Prediction of pressure distribution between the vocal folds using Bernoulli's equation, *The Journal of the Acoustical Society of America*, **136**(4): 2126–2126, doi: 10.1121/1.4899655.
16. MAKAROV I. (2009), Approximating the vocal tract by conical horns, *Acoustical Physics*, **55**(2): 261–269, doi: 10.1134/S106377100902016X.
17. MARKEL J.E., GRAY A.H. (1976), *Linear prediction of speech*, New York: Springer-Verlag, Inc.
18. MATHUR S., STORY B.H., RODRÍGUEZ J.J. (2006), Vocal-tract modeling: fractional elongation of segment lengths in a waveguide model with half-sample delays, *IEEE Transactions on Audio, Speech, and Language Processing*, **14**(5): 1754–1762, doi: 10.1109/TSA.2005.858550.
19. MORSE P. (1981), Vibration and Sound, *The Journal of the Acoustical Society of America*, **71**(6): 1623, doi: 10.1121/1.387830.
20. MULLEN J. (2006), *Physical modelling of the vocal tract with the 2D digital waveguide mesh*, PhD Thesis, Department of Electronics, University of York.
21. MULLEN J., HOWARD D.M., MURPHY D.T. (2003), Digital waveguide mesh modeling of the vocal tract acoustics, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, New Paltz, NY, USA, 2003, pp. 119–122, doi: 10.1109/ASPAA.2003.1285834.
22. MULLEN J., HOWARD D.M., MURPHY D.T. (2006), Waveguide physical modeling of vocal tract acoustics: flexible formant bandwidth control from increased model dimensionality, *IEEE Transactions on Audio, Speech, and Language Processing*, **14**(3): 964–971, doi: 10.1109/TSA.2005.858052.
23. MULLEN J., HOWARD D.M., MURPHY D.T. (2007), Real-time dynamic articulations in the 2-D waveguide mesh vocal tract model, *IEEE Transactions on Audio, Speech, and Language Processing*, **15**(2): 577–585, doi: 10.1109/TASL.2006.876751.
24. MURPHY D.T., BEESON M. (2007), The KW-boundary hybrid digital waveguide mesh for room acoustics applications, *IEEE Transactions on Audio,*

- Speech, and Language Processing*, **15**(2): 552–564, doi: 10.1109/TASL.2006.881681.
25. MURPHY D.T., HOWARD D.M. (2000), 2-D digital waveguide mesh topologies in room acoustics modelling, *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx)*, pp. 211–216.
 26. QURESHI T., SYED K. (2011a), A one-mass physical model of the vocal folds with seesaw-like oscillations, *Archives of Acoustics*, **36**(1): 15–27, doi: 10.2478/v10168-011-0002-3.
 27. QURESHI T.M., ISHAQ M. (2019), Real-time vocal tract model for elongation of segment lengths in a waveguide model, *Archives of Acoustics*, **44**(2): 287–300, doi: 10.24425/aoa.2019.128492.
 28. QURESHI T.M., SYED K.S. (2011b), A new approach to parametric modeling of glottal flow, *Archives of Acoustics*, **36**(4): 695–712, doi: 10.2478/v10168-011-0047-3.
 29. QURESHI T.M., SYED K.S. (2015), Two dimensional featured one dimensional digital waveguide model for the vocal tract, *Computer Speech & Language*, **33**(1): 47–66, doi: 10.1016/j.csl.2014.12.004.
 30. QURESHI T.M., SYED K.S. (2018), Fulcrum-point based self-oscillatory glottal model with numerical flow simulation, *International Journal of Acoustics & Vibration*, **23**(4): 516–528, doi: 10.20855/ijav.2018.23.41235.
 31. QURESHI T.M., SYED K.S. (2019), Improved vocal tract model for the elongation of segment lengths in a real time, *Computer Speech & Language*, **57**: 41–58, doi: 10.1016/j.csl.2019.02.001.
 32. RABINER L.R., SCHAFER R.W. (1978), *Digital processing of speech signals*, Prantice-Hall, Inc.
 33. RADOLF V., HORÁČEK J., BULA V., KOŠINA J., ŠVEC J. (2018), Experimental simulation of unilateral paralysis of human vocal folds, *34th Conference on Computational Mechanics*, pp. 87–88.
 34. SAVIOJA L., RINNE T.J., TAKALA T. (1994), Simulation of room acoustics with a 3D finite difference mesh, *Proceedings of International Computer Music Conference*, Aarhus, Denmark, pp. 463–466.
 35. SHIMAMURA R., TOKUDA I.T. (2016), Effect of level difference between left and right vocal folds on phonation: physical experiment and theoretical study, *The Journal of the Acoustical Society of America*, **140**(4): 3393–3394, doi: 10.1121/1.4970869.
 36. SMITH J.O. (2002), Principles of digital waveguide models of musical instruments, [in:] Kahrs M., Brandenburg K. [Eds], *Applications of Digital Signal Processing to Audio and Acoustics. The International Series in Engineering and Computer Science*, Vol. 437, pp. 417–466, Springer, Boston, MA, doi: 10.1007/0-306-47042-X_10.
 37. SPEED M., MURPHY D., HOWARD D. (2013a), Modeling the vocal tract transfer function using a 3D digital waveguide mesh, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **22**(2): 453–464, doi: 10.1109/TASLP.2013.2294579.
 38. SPEED M., MURPHY D., HOWARD D. (2013b), Three-dimensional digital waveguide mesh simulation of cylindrical vocal tract analogs, *IEEE Transaction on Audio, Speech, and Language Processing*, **21**(2): 449–454, doi: 10.1109/TASL.2012.2224342.
 39. STORY B.H. (2013), Phrase-level speech simulation with an airway modulation model of speech production, *Computer Speech & Language*, **27**(4): 989–1010, doi: 10.1016/j.csl.2012.10.005.
 40. STORY B.H., TITZE I.R., HOFFMAN E.A. (1996), Vocal tract area functions from magnetic resonance imaging, *The Journal of the Acoustical Society of America*, **100**(1): 537–554, doi: 10.1121/1.415960.
 41. STRUBE H.W. (2003), Are conical segments useful for vocal-tract simulation? (L), *The Journal of the Acoustical Society of America*, **114**(6): 3028–3031, doi: 10.1121/1.1623789.
 42. VÄLIMÄKI V., KARJALAINEN M. (1994), Improving the Kelly-Lochbaum vocal tract model using conical tube sections and fractional delay filtering techniques, *Proceedings of the International Conference on Spoken Language Processing*, pp. 615–618, https://www.isca-speech.org/archive/archive_papers/icslp_1994/i94_0615.pdf.
 43. VAMPOLA T., HORÁČEK J., LAUKKANEN A.-M., ŠVEC J.G. (2015), Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on CT measurement, *Logopedics Phoniatrics Vocology*, **40**(1): 14–23, doi: 10.3109/14015439.2013.775333.
 44. VAN DUYN S.A., SMITH J.O. (1993a), The 2-D digital waveguide mesh, *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, 1993, pp. 177–180, doi: 10.1109/ASPAA.1993.379968.
 45. VAN DUYN S.A., SMITH J.O. (1993b), Physical modeling with the 2-D digital waveguide mesh, *Proceedings of the International Computer Music Conference*, pp. 40–47, Tokyo, Japan.
 46. VAN DUYN S.A., SMITH J.O. (1996), The 3D tetrahedral digital waveguide mesh with musical applications, *Proceedings of the 1996 International Computer Music Conference*, pp. 9–16, Hong Kong.
 47. WILKINSON W., REISS J.D. (2016), A synthesis model for mammalian vocalization sound effects, *61st International Conference of Audio Engineering Society: Audio for Games*, London, UK, <https://www.eecs.qmul.ac.uk/~josh/documents/2016/wilkinson%20reiss%20-%202016.pdf>.