

Research Paper

A Symmetric Approach in the Three-Dimensional Digital Waveguide Modeling of the Vocal Tract

Tahir MUSHTAQ^{(1)*}, Ahmad KAMRAN⁽¹⁾,
Muhammad Zubair Akbar QURESHI⁽²⁾, Zafar IQBAL⁽³⁾

⁽¹⁾ *Department of Mathematics, COMSATS University Islamabad
Vehari Campus, Vehari, Pakistan*

⁽²⁾ *Department of Mathematics, Air University
Islamabad, Pakistan*

⁽³⁾ *Department of Mathematics, Government Graduate College of Science
Multan, Pakistan*

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

(received November 4, 2022; accepted April 4, 2023)

Simulation of wave propagation in the three-dimensional (3D) modeling of the vocal tract has shown significant promise for enhancing the accuracy of speech production. Recent 3D waveguide models of the vocal tract have been designed for better accuracy but require a lot of computational tasks. A high computational cost in these models leads to novel work in reducing the computational cost while retaining accuracy and performance. In the current work, we divide the geometry of the vocal tract into four equal symmetric parts with the introduction of two axial perpendicular planes, and the simulation is performed on only one part. A novel strategy is defined to implement symmetric conditions in the mesh. The complete standard 3D digital waveguide model is assumed as a benchmark model. The proposed model is compared with the benchmark model in terms of formant frequencies and efficiency. For the demonstration, the vowels /ɔ/, /i/, /ε/, /ɑ/, and /u/ have been selected for the simulations. According to the results, the benchmark and current models are nearly identical in terms of frequency profiles and formant frequencies. Still the current model is three times more effective than the benchmark model.

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



Copyright © 2023 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Three-dimensional (3D) acoustic simulations of the vocal tract show tremendous promise for the research of speech acoustics (GULLY, TUCKER, 2019). Many studies have focused on simulating the vocal tract because of its significance in the speech production system (LIM *et al.*, 2019; QURESHI *et al.*, 2020; QURESHI, ISHAQ, 2019; QURESHI, SYED, 2015; 2019; SCHICKHOFER, MIHAESCU, 2020; VAMPOLA *et al.*, 2015). Many researchers have used computational two- and three-dimensional techniques for voice generation based on wave propagation in the vocal tract using the finite element method (FEM) (ARNELA *et al.*, 2016a;

2019; BLANDIN *et al.*, 2021; MOHAPATRA *et al.*, 2019). The most common method for analyzing vocal tract acoustics is FEM. However, the existing technology limits its usage in the real-time application of wave propagation due to the long computation time.

The use of waveguide modeling in computational vocal tract modeling is quite common. Computational models are used in digital waveguide modeling for wave propagation in the vocal tract. It is assumed that each node in the vocal tract grid acts as a scattering junction connected by a unit of waveguide element, which is known as the delay line in the simulation (SPEED *et al.*, 2013a; TREYSSCDE, 2021). High-quality sound generation in real-time is made possible by the

use of digital waveguide models, which are efficient and realistic computational models of the waveguide model (MULLEN *et al.*, 2007; SPEED *et al.*, 2013a). A few attempts have been made to extend the Kelly-Lochbaum model by incorporating fractional delays for the vocal tract's elongation (MATHUR *et al.*, 2006; QURESHI, SYED, 2019). The Kelly-Lochbaum model was based on the same length of each cylindrical segment.

In the literature, the transformation of the vocal tract into concatenated segments with various cross-sectional areas has proven successful for one-dimensional waveguide modeling (MULLEN *et al.*, 2003; 2006), whereas the vocal tract has been converted into better approximated one using conical segmentations (MAKAROV, 2009; STRUBE, 2003). The multi-dimensional digital waveguide model offers more precision than the one-dimensional waveguide model but has a higher computing cost (ARNELA *et al.*, 2016b; BEESON, MURPHY, 2004; BLANDIN *et al.*, 2015; MOHAPATRA *et al.*, 2022; MURPHY, BEESON, 2007; MURPHY, HOWARD, 2000). The restriction of uniform structured mesh is a major limitation in the digital waveguide models (SPEED *et al.*, 2013a; VAN DUYN, SMITH, 1995). The complexity and computational cost of the three-dimensional digital waveguide model for vocal tract modeling limits the interest of the researchers. When accuracy is a concern, the three-dimensional model becomes a suitable candidate for modern computers (SPEED *et al.*, 2013a; 2013b). In recent works, three-dimensional digital waveguide modeling for the vocal tract has been adopted for the uniform linear grid as well as the non-uniform rectilinear grid proposed for voiced stop consonants (GULLY *et al.*, 2017; GULLY, TUCKER, 2019; QURESHI, ISHAQ, 2019; SPEED *et al.*, 2013b).

For the modeling of the vocal tract, the current work is focused on implementing a symmetric 3D digital waveguide model. For this purpose, a 3D geometry of the vocal tract is divided into four equal symmetric parts with the introduction of two axial perpendicular planes. The modeling of the single symmetric part is referred to as a 3D symmetric digital waveguide model of the vocal tract. A novel strategy has been devised to implement symmetric conditions on the two perpendicular axial sides of the meshing.

The proposed approach will reduce the high computation cost in the simulation while keeping the model's accuracy. For the validation of our approach, we choose a complete standard three-dimensional waveguide as a benchmark model in the proposed work. For the demonstrations of our model, we consider cross-sectional areas of the different vowels /ɔ/, /i/, /ε/, /α/, and /u/ from (STORY *et al.*, 1996). In the proposed work, the developed model is compared with the benchmarked model in terms of formant frequencies and efficiency. As a result of the simulations, the symmetric model is found to match the benchmark model

in terms of accuracy and is three times more efficient than the benchmark model.

2. Three-dimensional symmetric waveguide modeling of the vocal tract

For the realistic modeling of the vocal tract, the accuracy of the wave propagation in the vocal tract using 3D waveguide modeling makes it popular in the field of speech production. However, a significant restriction in the application of the three-dimensional waveguide model is the computational cost. Not many works have been reported in the literature to reduce the computational cost using the three-dimensional waveguide model of the vocal tract. Furthermore, the current work is focused on reducing computational cost in the vocal tract modeling. The computational cost is greatly reduced with the help of implementing symmetric three-dimensional waveguide modeling of the vocal tract. Using two axial perpendicular planes, a 3D geometry of the vocal tract is divided into four equal symmetric parts and the modeling of the single symmetric part is referred to as a symmetric three-dimensional digital waveguide model of the vocal tract. The symmetric conditions are implemented on the symmetric sides of the meshing. The visualization of the current approach in the proposed work is demonstrated in Figs. 1 and 2.

Figure 1 shows the three-dimensional geometry of the vocal tract orientation in the case of the vowel /ɔ/, where the length of the vocal tract is along the x -axis. Figure 2 presents the symmetric part with the green color for the vocal tract, as shown in Fig. 1. The re-

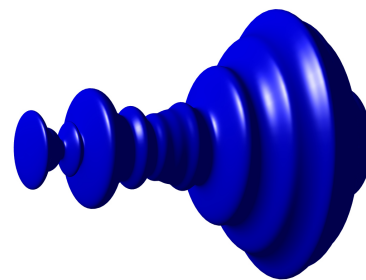


Fig. 1. Three-dimensional geometrical view of the vocal tract.

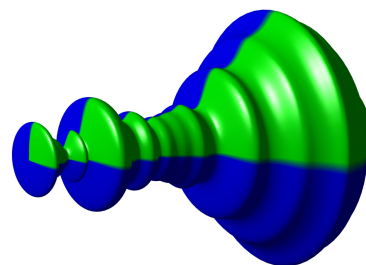


Fig. 2. Three-dimensional symmetric part of the vocal tract.

maining three symmetric parts are shown with the same blue color in the figure, which is not considered in the simulation of wave propagation in the vocal tract.

A vocal tract is a cylindrical tube of varying cross-sectional area in a one-dimensional waveguide model. The total vocal tract length is divided into multiple cylindrical tube segments with different cross-sectional areas. The wave equation is used to formulate the relationship between velocity and pressure in this tube (MARKEL, GRAY, 1976; RABINER, SCHAFER, 1978). D'Alembert's process helps to solve one-dimensional wave equations as a sum of left and right wave components. In the continuity and momentum equation at the intersection of two cylinders k -th and $(k+1)$ -th (KARJALAINEN, ERKUT, 2004), R_k is the reflection coefficient, A_k is the cross-sectional area of the k -th cylinder, and A_{k+1} is the cross-sectional area of the $(k+1)$ -th cylinder:

$$R_k = \frac{A_k - A_{k+1}}{A_k + A_{k+1}}. \quad (1)$$

The accuracy is increased by getting the three-dimensional structure in the waveguide model, which improves spatio-temporal connectivity in the sample grid. In this case, the mesh of the waveguide model increases with the expansion of the spatio-temporal connection in the sample grid of the three-dimensional (BEESON, MURPHY, 2004; MULLEN *et al.*, 2006; VAN DUYNÉ, SMITH, 1993). In the standard three-dimensional waveguide model, every node is called a scattering junction, where each wave is modified and moves in many directions. Every scattering junction is situated at the same delay lines in the grid and each scattering junction has six more neighboring scattering junctions at a point of 90 degrees, as illustrated in Fig. 3. except at the walls of the vocal tract. The interconnection of multiple junctions is shown in Fig. 4 with a three-dimensional view.

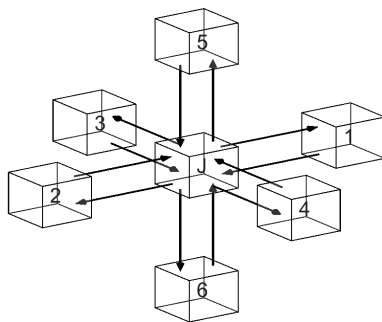


Fig. 3. Single junction with six-neighboring junctions.

Multiple grid topologies are utilized in the literature for the three-dimensional digital waveguide modeling of the vocal tract. However, we selected a three-dimensional uniform rectilinear grid topology in the work because of its simplicity.

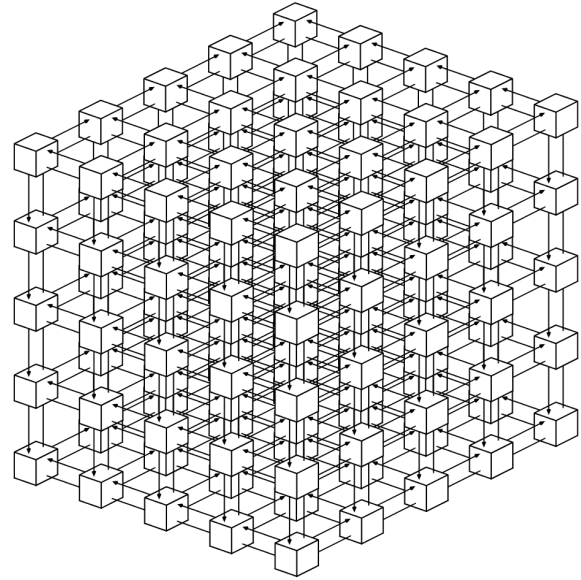


Fig. 4. Interconnection of multiple junctions.

At each junction J , the incoming wave pressures are symbolized as $p_{J,1}^+$, $p_{J,2}^+$, $p_{J,3}^+$, $p_{J,4}^+$, $p_{J,5}^+$, and $p_{J,6}^+$, respectively, and the outgoing pressures are denoted by $p_{J,1}^-$, $p_{J,2}^-$, $p_{J,3}^-$, $p_{J,4}^-$, $p_{J,5}^-$, and $p_{J,6}^-$, respectively. Finally, the pressure $p_{J,i}$ on each scattering junction J can be formulated as (MULLEN, 2006; VAN DUYNÉ, SMITH, 1993):

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

For N -port neighboring junctions of the junction J , the total pressure p at the junction J can be written as (MULLEN, 2006; VAN DUYNÉ, SMITH, 1993):

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

For a three-dimensional case with uniform impedance, the aforementioned equation can be rewritten as:

$$p_J = \frac{\sum_{i=1}^6 p_{J,i}^+}{3}. \quad (4)$$

The outgoing pressure components can be considered as:

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

The current work is focused on the symmetric three-dimensional waveguide mesh. To increase the efficiency of the three-dimensional waveguide model, we impose the symmetric conditions on the symmetric sides of the waveguide mesh. In the configuration of the rectilinear mesh in Fig. 3, we consider that each junction has six neighboring junctions at an angle of 90 degrees from one another. Junctions 1 and 2 are

along the x -axis, junctions 3 and 4 are along the y -axis, while junctions 5 and 6 are along the z -axis. Consider that the length of the three-dimensional vocal tract is along the x -direction while the cross-sectional area is along the yz -plane. In our work, we consider the part of the vocal tract that rests in the first quadrant in the coordinate axes. To make a symmetric model, we impose symmetric boundary conditions on the xy - and xz -planes.

For the symmetry on the xy -plane, the neighboring nodes 5 and 6 are between the line of symmetry along the y -axis. The outgoing wave from junction J to port 6 is unimportant and is omitted, while the incoming wave from port 6 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 5. In Eq. (4), the pressure at the junction J for the symmetric line can be derived as:

$$p_J = \frac{\sum_{\substack{i=1 \\ i \neq 6}}^6 p_{J,i}^+ + p_{J,5}^+}{3}. \quad (6)$$

Output traveling-wave components at the node are calculated as:

$$p_{J,i}^- = p_J - p_{J,i}^+, \quad i \neq 6. \quad (7)$$

For the symmetry on the xz -plane, the neighboring nodes 3 and 4 are between the line of symmetry along the z -axis. The outgoing wave from junction J to port 4 is unimportant and 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. In Eq. (4), the pressure at the junction J for the symmetric line can be derived as:

$$p_J = \frac{\sum_{\substack{i=1 \\ i \neq 4}}^6 p_{J,i}^+ + p_{J,3}^+}{3}. \quad (8)$$

Output traveling-wave components at the node are calculated as:

$$p_{J,i}^- = p_J - p_{J,i}^+, \quad i \neq 4. \quad (9)$$

3. Numerical simulation

A series of cross-sectional areas are acquired from the particular geometric configurations of the vocal tract. In the proposed work, we consider the different configurations of the vocal tract for the vowels /ɔ/, /i/, /ɛ/, /a/, and /u/. The cross-sectional areas of these vowels are taken from (STORY *et al.*, 1996) with varying vocal tract lengths of 17.46, 15.88, 16.67, 17.46, and 18.25 cm, respectively. For the production of the three-dimensional mesh, the cross-sectional areas are interpolated into the smooth shape of the three-dimensional vocal tract by using a piecewise third-order spline.

For the best resolution of the graphs, the length of the impulse response is chosen as 25 000 input samples for every vowel considered in this work. All input samples have zero values except the first input sample, which has a value of 1. The numerical solution of the current vocal model comprises two basic steps for each iteration. The first step is used to find the scattering of the waves at each junction of the mesh, while the wave delay of each junction is passed to the neighboring junctions in the second step. In the first step, we apply the glottal boundary condition (SPEED *et al.*, 2013a) on each junction that lies on the inlet of the vocal tract. For the scattering of the wave, Eq. (4) is used to calculate the total pressure on each interior junction of the mesh and we also measure the outgoing pressure in all six directions with the help of Eq. (5). The scattering of the waves at outlet junctions (SPEED *et al.*, 2013a) is also computed. For the junctions that lie on the symmetric xy -plane, Eqs. (6) and (7) are employed to calculate total and outgoing pressures, respectively. Similarly, for the case of a symmetric xz -plane, Eqs. (8) and (9) are used. In the second step, all components of the wave delay of each junction are passed to its neighboring junctions. We iterate this process according to the total number of input samples to get output samples. The frequency response is then obtained from the transfer function by applying fast Fourier transformation (FFT) and natural logarithms on the output samples. The simulations of the current work are compiled in the high-level computer programming language MATLAB 2020. MATLAB computer language also offers the use of C++-coded files to increase the efficiency of the simulations. We also developed a C++-coded file used in MATLAB to increase the efficiency for the high-cost computational work in the 3D symmetric meshing. The reflection coefficient of the glottis r_G , the reflection coefficient of the wall r_W and the reflection coefficient lip r_L are assigned with values of 0.97, 1.0, and -0.9 , respectively.

4. Results and discussion

In this section, we demonstrate our current approach in the execution of the symmetric three-dimensional digital waveguide model of the vocal tract. Our benchmark model is used to compare and validate our symmetric model. Based on accuracy and efficiency, the comparison between the symmetric and the benchmark models is established. The first three lowest formant frequencies are known to be sufficient for every vowel to be recognized. In the current work, we take six formant frequencies of the benchmark and symmetric models and compare them with each other so that we can get a better idea about the relative error, and check the efficiency of our symmetric model. The six formant frequencies are denoted by f_1 , f_2 , f_3 ,

f_4, f_5, f_6 , which are mentioned in the first column of all tables.

For the comparison process, the sample delay d_s is taken the same for both models. In the current work, vowels /ɔ/, /i/, /ε/, /α/, and /u/ are taken with different vocal tract lengths. The efficiency and accuracy of the current model are compared with the benchmark model in the collapsed time, frequency profiles and formant frequencies, respectively. At least three formant frequencies are required for each vowel. In our work, we consider the first six formant frequencies produced by the benchmark model and our proposed model (ARNELA *et al.*, 2016b).

Figure 5 shows the frequency profiles generated by a transfer function of a vowel /ɔ/ of the vocal tract up to 6000 Hz. We observe that the frequency profiles of benchmark and symmetric models overlap. There is no difference between the profiles of the benchmark and the symmetric models. In other words, the frequency profiles of both models are identical to each other. Numerically calculated values are given in Table 1. In this table, column 4 shows the relative error between the benchmark and symmetric models. This column presents the zero relative errors in all six formant frequencies in the current case of vowel /ɔ/. Column 7 shows that the proposed model is about 3 times more efficient than the benchmark model due to the high efficiency of the symmetric model.

The frequency profiles of the vowel /i/ of the vocal tract obtained by its transfer function are shown in Fig. 6. It is also noted that the frequency profile of the symmetric model matches the frequency profile of the benchmark model. Measured values of formant frequencies are presented in Table 2. Column 4 of this table shows the zero relative errors in all six formants frequencies of vowel /i/. Column 7 shows that

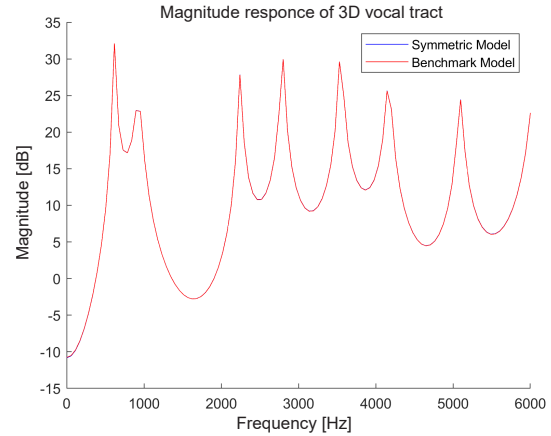


Fig. 5. Comparison between the frequency profiles of the benchmark and symmetric models in the vowel /ɔ/ case.

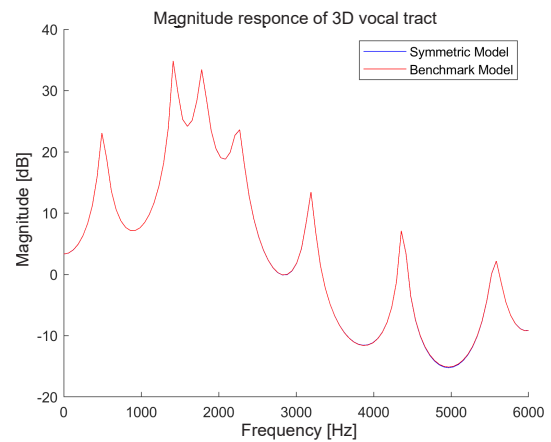


Fig. 6. Comparison between the frequency profiles of the benchmark and symmetric models in the vowel /i/ case.

the symmetric model, in this case, is about 3 times more efficient than the benchmark model.

Table 1. Comparison between the formant frequencies of the benchmark and symmetric models in the vowel /ɔ/ case.

Format frequency	Benchmark model [Hz]	Symmetric model [Hz]	Error [%]	Elapsed time (average) [sec]		Efficiency
				Benchmark model	Symmetric model	
f_1	597	597	0	627.323	209.391	2.995
f_2	894	894	0			
f_3	2234	2234	0			
f_4	2755	2755	0			
f_5	3573	3573	0			
f_6	4243	4243	0			

Table 2. Comparison between the formant frequencies of the benchmark and symmetric models in the vowel /i/ case.

Format frequency	Benchmark model [Hz]	Symmetric model [Hz]	Error [%]	Elapsed time (average) [sec]		Efficiency
				Benchmark model	Symmetric model	
f_1	492	492	0	550.188	178.868	3.076
f_2	1557	1557	0			
f_3	1885	1885	0			
f_4	2294	2294	0			
f_5	3195	3195	0			
f_6	4423	4423	0			

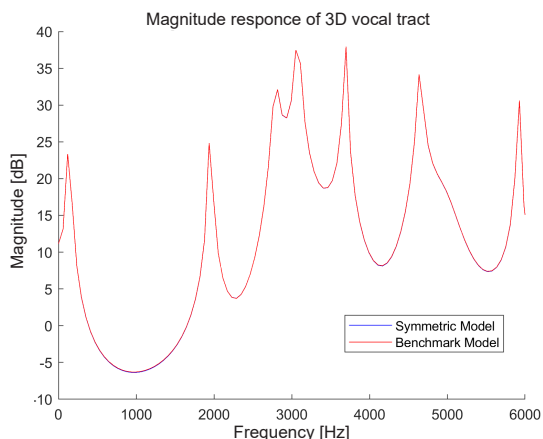


Fig. 7. Comparison between the frequency profiles of the benchmark and symmetric models in the vowel /ε/ case.

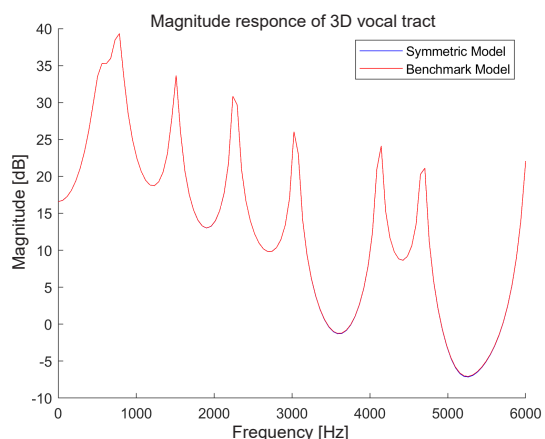


Fig. 8. Comparison between the frequency profiles of the benchmark and symmetric models in the vowel /α/ case.

Table 3. Comparison between the formant frequencies of the benchmark and symmetric models in the vowel /ε/ case.

Format frequency	Benchmark model [Hz]	Symmetric model [Hz]	Error [%]	Elapsed time (average) [sec]		Efficiency
				Benchmark model	Symmetric model	
f_1	157	157	0	513.045	155.675	3.296
f_2	1955	1955	0			
f_3	2815	2815	0			
f_4	3127	3127	0			
f_5	3752	3752	0			
f_6	4690	4690	0			

Table 4. Comparison between the formant frequencies of the non-symmetric and symmetric models in the vowel /α/ case.

Format frequency	Benchmark model [Hz]	Symmetric model [Hz]	Error [%]	Elapsed time (average) [sec]		Efficiency
				Benchmark model	Symmetric model	
f_1	819	819	0	674.704	210.66	3.202
f_2	1638	1638	0			
f_3	2307	2307	0			
f_4	3052	3052	0			
f_5	4168	4168	0			
f_6	4764	4764	0			

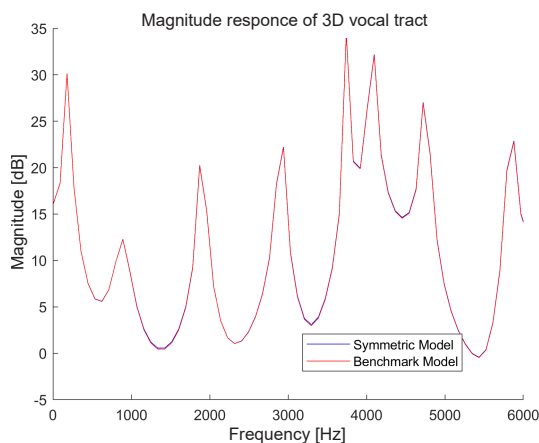


Fig. 7. Comparison between the frequency profiles of the benchmark and symmetric models in the vowel /u/ case.

Table 5. Comparison between the formant frequencies of the benchmark and symmetric models in the vowel /u/ case.

Format frequency	Benchmark model [Hz]	Symmetric model [Hz]	Error [%]	Elapsed time (average) [sec]		Efficiency
				Benchmark model	Symmetric model	
f_1	144	144	0	559.679	185.035	3.024
f_2	855	855	0			
f_3	1924	1924	0			
f_4	2921	2921	0			
f_5	3704	3704	0			
f_6	3989	3989	0			

Figures 7, 8, and 9 show the frequency profiles of vowels /ε/, /α/, and /u/ of the vocal tract, respectively. In all these cases, the frequency profiles of the benchmark and symmetric models overlap. Numerically measured values in Tables 3, 4, and 5 of column 4 present zero relative errors in all 6 formant frequencies in all cases. The efficiency of our model over the benchmark model, as shown in Figs. 7, 8, and 9, is about 3.3, 3.2, and 3 times better, respectively.

5. Conclusions

In the proposed work, the symmetric model was used for modeling three-dimensional waveguide model of the vocal tract. The computational cost was greatly reduced by employing a symmetric approach in this work. The simulation was performed on the vowels /ɔ/, /i/, /ε/, /α/, and /u/. By studying tables and figures, we draw the following conclusions:

- successful implementation of the symmetric approach in the three-dimensional waveguide modeling of the vocal tract was conducted;
- the formant frequencies of the symmetric model are the same as the benchmark model;
- the frequency profiles of the current model overlap with that of the benchmark model;
- the symmetric model is more efficient than the benchmark model. In all the cases, the proposed model is about 3 times more efficient than the benchmark model.

We conclude that the symmetric model presents itself as a highly efficient and accurate three-dimensional waveguide model of the vocal tract. The proposed model provides an opportunity to efficiently investigate the vocal tract's frequency response for speech production. We believe the symmetric model may serve as a useful vocal tract model in speech synthesizers and provide a new dimension for further investigations.

References

1. ARNELA M. *et al.* (2016a), Influence of lips on the production of vowels based on finite element simulations and experiments, *The Journal of the Acoustical Society of America*, **139**(5): 2852–2859, doi: [10.1121/1.4950698](https://doi.org/10.1121/1.4950698).
2. ARNELA M. *et al.* (2016b), Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds, *The Journal of the Acoustical Society of America*, **140**(3): 1707–1718, doi: [10.1121/1.4962488](https://doi.org/10.1121/1.4962488).
3. ARNELA M., DABBAGHCHIAN S., GUASCH O., ENGWALL O. (2019), MRI-based vocal tract representations for the three-dimensional finite element synthesis of diphthongs, *IEEE/ACM Transactions on Audio, Speech, Language Processing*, **27**(12): 2173–2182, doi: [10.1109/TASLP.2019.2942439](https://doi.org/10.1109/TASLP.2019.2942439).
4. BEESON M.J., MURPHY D.T. (2004), RoomWeaver: A digital waveguide mesh based room acoustics research tool, [in:] *Proceedings of the Seventh International Conference on Digital Audio Effects*, pp. 268–273.
5. BLANDIN R. *et al.* (2015), Effects of higher order propagation modes in vocal tract like geometries, *The Journal of the Acoustical Society of America*, **137**(2): 832–843, doi: [10.1121/1.4906166](https://doi.org/10.1121/1.4906166).
6. BLANDIN R., FÉLIX S., DOC J.-B., BIRKHOLZ P. (2021), Combining multimodal method and 2D finite elements for the efficient simulation of vocal tract acoustics, [in:] *Proceedings of the 27th International Congress on Sound and Vibration*.
7. GULLY A.J., DAFFERN H., MURPHY D.T. (2017), Diphthong synthesis using the dynamic 3D digital waveguide mesh, *IEEE/ACM Transactions on Audio, Speech, Language Processing*, **26**(2): 243–255, doi: [10.1109/TASLP.2017.2774921](https://doi.org/10.1109/TASLP.2017.2774921).
8. GULLY A.J., TUCKER B. (2019), Modeling voiced stop consonants using the 3D dynamic digital waveguide mesh vocal tract model, [in:] *Proceedings of the International Congress of Phonetic Sciences 2019*, Australasian Speech Science and Technology Association Inc.
9. KARJALAINEN M., ERKUT C. (2004), Digital waveguides versus finite difference structures: Equivalence and mixed modeling, *EURASIP Journal on Applied Signal Processing*, **2004**(7): 978–989, doi: [10.1155/S1110865704401176](https://doi.org/10.1155/S1110865704401176).
10. LIM Y., ZHU Y., LINGALA S.G., BYRD D., NARAYANAN S., NAYAK K.S. (2019), 3D dynamic MRI of the vocal tract during natural speech, *Magnetic Resonance in Medicine*, **81**(3): 1511–1520, doi: [10.1002/mrm.27570](https://doi.org/10.1002/mrm.27570).
11. MAKAROV I.S. (2009), Approximating the vocal tract by conical horns, *Acoustical Physics*, **55**(2): 261–269, doi: [10.1134/S106377100902016X](https://doi.org/10.1134/S106377100902016X).

12. MARKEL J.E., GRAY A.H. (1976), *Linear Prediction of Speech*, Springer.
13. 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](https://doi.org/10.1109/TSA.2005.858550).
14. MOHAPATRA D.R., FLEISCHER M., ZAPPI V., BIRKHOLZ P., FELS S. (2022), Three-dimensional finite-difference time-domain acoustic analysis of simplified vocal tract shapes, [in:] *Proceedings of Interspeech*, pp. 764–768, doi: [10.21437/Interspeech.2022-10649](https://doi.org/10.21437/Interspeech.2022-10649).
15. MOHAPATRA D.R., ZAPPI V., FELS S. (2019), An extended two-dimensional vocal tract model for fast acoustic simulation of single-axis symmetric three-dimensional tubes, [in:] *Proceedings of Interspeech 2019*, pp. 3760–3764, doi: [10.21437/Interspeech.2019-1764](https://doi.org/10.21437/Interspeech.2019-1764).
16. MULLEN J. (2006), *Physical modelling of the vocal tract with the 2D digital waveguide mesh*, Ph.D. Thesis, The University of York.
17. MULLEN J., HOWARD D.M., MURPHY D.T. (2003), Digital waveguide mesh modeling of the vocal tract acoustics, [in:] *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 119–122, doi: [10.1109/ASPAA.2003.1285834](https://doi.org/10.1109/ASPAA.2003.1285834).
18. 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](https://doi.org/10.1109/TSA.2005.858052).
19. 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](https://doi.org/10.1109/TASL.2006.876751).
20. 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](https://doi.org/10.1109/TASL.2006.881681).
21. MURPHY D.T., HOWARD D.M. (2000), 2-D digital waveguide mesh topologies in room acoustics modelling, [in:] *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx)*, pp. 211–216.
22. 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](https://doi.org/10.24425/aoa.2019.128492).
23. 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](https://doi.org/10.1016/j.csl.2014.12.004).
24. 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**(4): 41–58, doi: [10.1016/j.csl.2019.02.001](https://doi.org/10.1016/j.csl.2019.02.001).
25. QURESHI T.M., SYED K.S., ZAFAR A. (2020), Non-uniform rectilinear grid in the waveguide modeling of the vocal tract, *Archives of Acoustics*, **45**(4): 585–600, doi: [10.24425/aoa.2020.135247](https://doi.org/10.24425/aoa.2020.135247).
26. RABINER L.R., SCHAFFER R.W. (1978), *Digital Processing of Speech Signals*, Prentice-Hall.
27. SCHICKHOFFER L., MIHAESCU M. (2020), Analysis of the aerodynamic sound of speech through static vocal tract models of various glottal shapes, *Journal of Biomechanics*, **99**: 109484, doi: [10.1016/j.jbiomech.2019.109484](https://doi.org/10.1016/j.jbiomech.2019.109484).
28. 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](https://doi.org/10.1109/TASLP.2013.2294579).
29. 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): 449–455, doi: [10.1109/TASL.2012.2224342](https://doi.org/10.1109/TASL.2012.2224342).
30. 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](https://doi.org/10.1121/1.415960).
31. 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](https://doi.org/10.1121/1.1623789).
32. TREYSSCÖDE F. (2021), A model reduction method for fast finite element analysis of continuously symmetric waveguides, *Journal of Sound and Vibration*, **508**: 116204, doi: [10.1016/j.jsv.2021.116204](https://doi.org/10.1016/j.jsv.2021.116204).
33. 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](https://doi.org/10.3109/14015439.2013.775333).
34. VAN DUYN S.A., SMITH J.O. (1993), Physical modeling with the 2-D digital waveguide mesh, [in:] *Proceedings of the International Computer Music Conference*.
35. VAN DUYN S.A., SMITH J.O. (1995), The tetrahedral digital waveguide mesh, [in:] *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 234–237, doi: [10.1109/ASPAA.1995.482998](https://doi.org/10.1109/ASPAA.1995.482998).