Interspeech 2022
18-22 September 2022, Incheon, Korea

Three-dimensional finite-difference time-domain acoustic analysis of simplified vocal tract shapes

Debasish Ray Mohapatra[1], *Mario Fleischer*[2], *Victor Zappi*[3], *Peter Birkholz*[4], *Sidney Fels*[1]

[1]Department of Electrical and Computer Engineering, University of British Columbia, Canada
[2]Charité – Universitätsmedizin Berlin, Department of Audiology and Phoniatrics, Berlin, Germany
[3]College of Arts, Media and Design, Northeastern University, Boston, USA
[4]Institute of Acoustics and Speech Communication, TU Dresden, Dresden, Germany

debasishray@ece.ubc.ca

Abstract

The finite-difference time-domain (FDTD) method has been widely used for vocal tract acoustic modelling due to its simplicity and low computational cost. Nevertheless, the method suffers from high discretization error while approximating realistic vocal tract geometries using orthogonal grid elements. Alternatively, simplified vocal tract shapes having regular contours can be used for articulatory models. These geometries can be generated from one-dimensional (1D) area functions, which approximate vocal tracts as concatenated tubes with different cross-sections. To this aim, we modify an existing 3D FDTD model for faster acoustic simulation and synthesize five English vowels with various simplified vocal tract shapes. We implement six geometrical shapes for each vowel, consisting of circular, elliptical and square cross-sections with centric and eccentric tube segment configurations. Vowel transfer functions obtained from these FDTD simulations are compared with a highly accurate finite element (FE) scheme. The acoustic formants of the FDTD model agree well with the corresponding FEM approach for most vowels. The influence of vocal tracts with different geometry approximations remains insignificant for frequencies below 5 kHz. However, vocal tracts with elliptical or eccentric configurations have produced higher-order acoustic modes. This paper characterizes the acoustic properties of simplified vocal tract shapes using the 3D FDTD scheme.

Index Terms: computational acoustics, vocal tract geometry, FDTD, FEM, articulatory speech synthesis

1. Introduction

The complex upper vocal tract geometry plays a significant role in shaping the speech spectrum [1, 2]. Hence, physics-based vocal tract models which can precisely approximate the detailed upper airway geometry [3, 4] are preferred for articulatory speech synthesis. However, acoustic analysis of a realistic vocal tract model is computationally expensive as it requires fine discretization of the vocal tract contour. It also presents various challenges while generating time-evolving vocal tract shapes for diphthong sounds [5]. Alternatively, simplified geometries have also been generated from 1D vocal tract area functions having acoustic properties comparable to the actual vocal tracts [6, 7]. The numerical analysis of simplified vocal tract geometries is computationally cheap, and they can easily be manipulated using various interpolation methods for the dynamic boundary condition [8, 9].

The classic low-dimensional (1D and 2D) acoustic models [7, 10, 11, 12, 13, 14] utilize the geometry simplification approach to represent vocal tracts as straight tubes with circular cross-sections. These oversimplifications can distort the acoustic properties of synthesized speech. In contrast to the 1D and 2D systems, the 3D models [6, 15, 16] provide better geometrical flexibility, and they can generate a range of vocal tract shapes with different levels of detail having the same area functions. Arnela et al. [4, 9] and Vampola et al. [3] previously explored acoustic effects of simplified static and dynamic vocal tract shapes using the 3D FEM method. The FEM scheme can closely approximate the vocal tract boundary using unstructured tetrahedral mesh elements for precise acoustic output. However, the FEM simulations are computationally expensive.

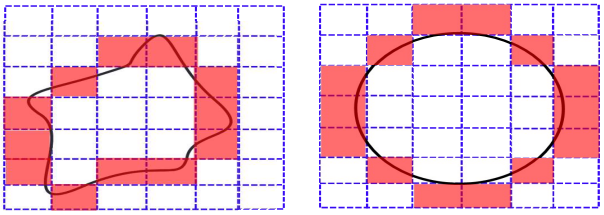

Figure 1: *Frontal view of vocal tract boundary approximation in a "stair-stepped" manner using the orthogonal grid elements of FDTD — irregular shape (left) and regular shape (right).*

Alternatively, the FDTD scheme is known for its simplicity and low computational cost. The conventional FDTD method [17] involves a staggered orthogonal grid for numerical analysis. Using the 3D FDTD scheme, Takemoto et al. [16], and Wang et al. [18] earlier examined the acoustic properties of vocal tracts generated directly from the MRI data of Japanese and Mandarin vowel sounds, respectively. Nevertheless, these realistic vocal tract models require high spatial resolution as the orthogonal lattices in FDTD discretize complex MRI geometries in a "stair-stepped" manner that does not match their non-uniformity. Similar to FEM, the FDTD method can adapt nonorthogonal and unstructured meshes [19, 20, 21] for boundary discretization. Nonetheless, such approaches do not retain the efficiency of the standard FDTD scheme [22, 23], and they may suffer from numerical instabilities [24]. However, less attention has been paid to synthesizing simplified vocal tract shapes using the 3D FDTD scheme. Due to the regular cross-sectional shape of simplified geometries, the orthogonal grid in FDTD produces less discretization error during boundary estimation, as shown in Figure 1. Hence, these geometries can be well-approximated even with a coarse FDTD grid, easing the computational effort while synthesizing precise acoustic output. Thus, this work aims for acoustic characterization of simplified vocal tract shapes using the 3D FDTD method.

In this work, we first present a modified 3D FDTD acoustic wave solver [16] that employs a second-order accurate sten-

10.21437/Interspeech.2022-10649

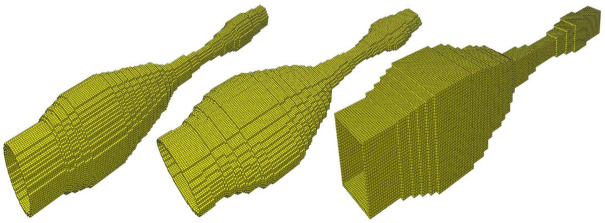

Figure 2: *Simplified vocal tract shapes for vowel sound /a/ having circular-centric (left), elliptical-centric (center) and square-centric (right) cross-sections.*

cil for the spatial and temporal discretization of acoustic components. Compared to the previous higher-order accurate numerical scheme, such an approach reduces the computational memory requirements and the number of numerical operations. Next, we use the proposed FDTD scheme to synthesize static vocal tracts for five English vowel sounds. For each vowel, we implement six simplified vocal tract geometries with different levels of detail and study their acoustic properties by generating vocal tract transfer functions (VTTFs). Results obtained from these FDTD simulations are also compared to a highly precise time-domain FEM acoustic model for evaluation.

2. Method

2.1. Simplified Vocal Tract Shapes

We use Story's 1D area function dataset [25] to generate vocal tract models for five English vowel sounds: /a/, /i/, /u/, /e/ and /o/. This standard dataset describes the vocal tract geometry from the glottal end to the lip end using 44 tube segments of equal length but different cross-sectional areas. It has been widely used for acoustic analysis of static vocal tract models [7, 11, 12]. In this study, we primarily focus on geometry simplifications of the main conduit of the vocal tract model. These simplifications can be categorized based on the vocal tract's cross-sectional shape and its segments' connection type.

We consider vocal tracts as straight tubes with three different cross-sectional shapes – circular, elliptical and square (shown in Figure 2). Vocal tracts with circular cross-sections have been extensively studied using the low-dimensional numerical models [7, 12] and 3D FEM method [4]. Therefore, it makes a good starting point for evaluating the proposed 3D FDTD model. However, the radial symmetry of circular cross-sections oversimplifies the complexity of real vocal tract geometry. Alternatively, elliptical cross-sections can be considered for acoustic analysis of vocal tract models. Since Story's dataset does not provide the geometrical shape of vocal tracts' contour, we empirically set the major to minor axis length ratio as 3:1 to model elliptical cross-sections. This setting avoids the formation of a very narrow constriction along the minor axis of elliptical vocal tracts as it may lead to a high discretization error. Additionally, vocal tracts with square cross-sections are well suited for the 3D FDTD scheme. Its orthogonal grid elements allow a better approximation of the square and rectangular shapes with less discretization error. Therefore, using FDTD, we model vocal tracts with square cross-sections as a special case to analyze their acoustic effects.

For each cross-sectional shape, we connect vocal tract segments in two ways – centric and eccentric (shown in Figure 3). In the centric configuration, vocal tract segments share the same central axis. This case is of particular interest since it has been commonly used due to its simplicity [4, 12]. For eccentric con-

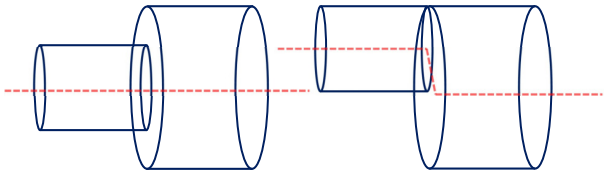

Figure 3: *Illustration of circular-centric (left) and circular-eccentric (right) vocal tract configurations with two tube segments. The dotted red line demonstrates the vocal tract midline (central axis) for the centric and eccentric configurations.*

figuration, the vocal tract segments are placed in such a way to obtain a common upper line that can be viewed as corresponding to the hard palate. These segments do not share the same central axis. Furthermore, each segment of an eccentric vocal tract produces an artificial extension of the contour midline length; consequently, there is a shift in formant frequencies. However, we manually reduce the segment length to adjust the midline length of eccentric vocal tracts and match the actual ones. For simplicity, hereafter, these vocal tract shapes will be termed: a) circular-centric, b) circular-eccentric, c) elliptical-centric, d) elliptical-eccentric, e) square-centric and f) square-eccentric. As the centric and eccentric configurations are expected to have roughly the same formants [2] below 5 kHz, we only model the circular-centric and elliptical-centric shapes with the FEM scheme to evaluate the FDTD simulations.

2.2. Finite-Difference Time-Domain Wave Solver

In order to implement the vocal tract acoustic model, the 3D FDTD scheme discretizes the linear wave equations of continuity and motion (both in space and time) and samples acoustic field components, i.e., pressure (p) at the center and particle velocities (v_x, v_y, v_z) at three sides (facets) of an orthogonal grid cell. During the simulation, the discretized wave equations are updated at each time step using the leapfrog scheme, resulting in an explicit time marching algorithm that can be readily parallelized. In particular, our 3D FDTD wave solver is based on the vocal tract acoustic model proposed by Takemoto et al. [16].

However, Takemoto used the fourth-order spatial differences to update the acoustic components of a grid cell at each time step by accessing the neighbouring 6 cells. Such a higher-order accurate scheme improves the acoustic precision of the analysis model but requires additional computational time due to more numerical operations. Moreover, it may create challenges while parallelizing the discretized wave equations as the scheme requires high computational memory. Therefore, in this work, we implemented the second-order accurate spatial, and temporal discretization of wave equations, which has been previously used for acoustic analysis of tube-like geometries such as 2D vocal tracts and wind instruments [12, 26]. The comparative analysis of the synthesized acoustic outputs using the fourth-order and second-order accurate numerical schemes is essential, although it is beyond the scope of this paper. Like Allen and Raghuvanshi [26] proposed, a scalar parameter β is used to define vocal tract boundary walls. To account for wall impedance, we enforce a semi-reflective boundary condition using the particle velocity ($\boldsymbol{v}_b$) near vocal tract walls [27]. The updated continuous-wave equation can be described as follows,

$$\frac{\partial p}{\partial t} = -\varrho c^2 \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right) \quad (1)$$

$$\beta \frac{\partial \boldsymbol{v}}{\partial t} + (1-\beta)\,\boldsymbol{v} = -\beta^2 \frac{\boldsymbol{\nabla} p}{\varrho} + (1-\beta)\,\boldsymbol{v}_b \quad (2)$$

Formants	/a/	/i/	/u/	/e/	/o/
f_1	4 Hz (0.6%) 34 Hz (5.0%)	10 Hz (3.8%) 6 Hz (2.3%)	9 Hz (3.6%) 16 Hz (6.4%)	3 Hz (0.7%) 14 Hz (3.1%)	3 Hz (0.6%) 43 Hz (8.9%)
f_2	14 Hz (1.3%) 45 Hz (4.2%)	51 Hz (2.4%) 70 Hz (3.3%)	38 Hz (5.3%) 33 Hz (4.6%)	4 Hz (0.2%) 64 Hz (3.2%)	26 Hz (3.3%) 52 Hz (6.6%)
f_3	101 Hz (3.3%) 14 Hz (0.5%)	68 Hz (2.2%) 72 Hz (2.4%)	43 Hz (1.9%) 207 Hz (9.0%)	56 Hz (2.3%) 15 Hz (0.6%)	62 Hz (2.6%) 188 Hz (7.8%)
f_4	82 Hz (2.0%) 114 Hz (2.8%)	34 Hz (0.8%) 160 Hz (3.9%)	83 Hz (2.3%) 271 Hz (7.5%)	17 Hz (0.5%) 126 Hz (3.4%)	183 Hz (5.2%) 142 Hz (4.1%)
f_5	44 Hz (0.9%) 209 Hz (4.2%)	191 Hz (3.9%) 6 Hz (0.1%)	156 Hz (3.7%) 174 Hz (4.1%)	59 Hz (1.2%) 108 Hz (2.2%)	103 Hz (2.2%) 391 Hz (8.5%)
f_6	100 Hz (1.8%) 75 Hz (1.3%)	307 Hz (5.4%) 105 Hz (1.9%)	98 Hz (2.0%) 355 Hz (7.1%)	272 Hz (4.8%) 84 Hz (1.5%)	109 Hz (2.0%) 448 Hz (8.2%)
f_7	108 Hz (1.6%) –	179 Hz (2.7%) –	162 Hz (2.6%) –	72 Hz (1.1%) –	170 Hz (2.8%) –
f_8	121 Hz (1.6%) –	147 Hz (1.9%) –	225 Hz (3.4%) –	153 Hz (2.0%) –	275 Hz (3.8%) –
Average	(1.6%) (3.0%)	(2.9%) (2.3%)	(3.1%) (6.4%)	(1.6%) (2.3%)	(2.8%) (7.4%)

Table 1: *Absolute positional differences ($|\Delta F_i| = |F_{i,FDTD} - F_{i,FEM}|$) and percentage of deviation ($|\Delta F_i|/F_{i,FDTD}$) of the first eight formants computed for circular-centric and circular-eccentric vocal tract shapes with respect to the 3D FEM model.*

Formants	/a/	/i/	/u/	/e/	/o/
f_1	22 Hz (3.2%) 37 Hz (5.5%)	5 Hz (1.9%) 13 Hz (5.0%)	8 Hz (3.2%) 5 Hz (2.0%)	6 Hz (1.3%) 17 Hz (3.8%)	3 Hz (0.6%) 19 Hz (4.0%)
f_2	23 Hz (2.2%) 3 Hz (0.3%)	2 Hz (0.1%) 58 Hz (2.7%)	36 Hz (5.1%) 1 Hz (0.1%)	31 Hz (1.5%) 3 Hz (0.2%)	52 Hz (6.6%) 17 Hz (2.2%)
f_3	24 Hz (0.8%) 34 Hz (1.1%)	53 Hz (1.8%) 14 Hz (0.5%)	55 Hz (2.4%) 68 Hz (3.0%)	62 Hz (2.5%) 37 Hz (1.5%)	42 Hz (1.8%) 50 Hz (2.1%)
f_4	144 Hz (3.6%) 68 Hz (1.7%)	75 Hz (1.8%) 23 Hz (0.6%)	79 Hz (2.2%) 98 Hz (2.7%)	76 Hz (2.1%) 10 Hz (0.3%)	59 Hz (1.7%) 64 Hz (1.8%)
f_5	33 Hz (0.7%) 112 Hz (2.3%)	115 Hz (2.3%) 0 Hz (0.0%)	129 Hz (3.1%) 24 Hz (0.6%)	68 Hz (1.4%) 35 Hz (0.7%)	69 Hz (1.5%) 133 Hz (2.9%)
f_6	135 Hz (2.4%) 22 Hz (0.4%)	114 Hz (2.0%) 265 Hz (0.5%)	98 Hz (2.0%) 102 Hz (2.1%)	126 Hz (2.2%) 10 Hz (0.2%)	157 Hz (2.9%) 24 Hz (0.4%)
f_7	137 Hz (2.0%) –	37 Hz (0.6%) –	214 Hz (3.5%) –	52 Hz (0.8%) –	136 Hz (2.2%) –
f_8	199 Hz (2.8%) –	210 Hz (2.7%) –	252 Hz (3.8%) –	154 Hz (2.0%) –	202 Hz (2.9%) –
Average	(2.2%) (1.9%)	(1.7%) (1.6%)	(3.2%) (1.8%)	(1.7%) (1.1%)	(2.5%) (2.2%)

Table 2: *Absolute positional differences ($|\Delta F_i| = |F_{i,FDTD} - F_{i,FEM}|$) and percentage of deviation ($|\Delta F_i|/F_{i,FDTD}$) of the first eight formants computed for elliptical-centric and elliptical-eccentric vocal tract shapes with respect to the 3D FEM model.*

where c is the speed of sound and ϱ is the air density. For time-domain analysis, Equations 1 and 2 are discretized as follows,

$$p^{(n+1)} = \frac{p^{(n)} - \varrho c^2 \Delta t \, \widetilde{\boldsymbol{\nabla}} \cdot \boldsymbol{v}^{(n)}}{1 + \Delta t(1-\beta)} \tag{3}$$

$$\boldsymbol{v}^{(n+1)} = \frac{\beta \boldsymbol{v}^{(n)} - \beta^2 \Delta t \, \widetilde{\nabla} p^{(n+1)}/\varrho \; + \; \Delta t(1-\beta)\boldsymbol{v_b}}{\beta + \Delta t(1-\beta)} \tag{4}$$

where $\widetilde{\boldsymbol{\nabla}}$ denotes the second-order discrete spatial derivatives. The code for the 3D FDTD vocal tract model is provided here.[1]

2.3. Experimental Setup

For comparative analysis, a time-domain 3D FEM acoustic model was employed [28]. The FEM approach approximates the wave propagation inside the vocal tract geometry by directly solving the irreducible wave equation for the acoustic pressure. For the 3D FEM simulation, the tetrahedral mesh elements were used to model the vocal tract airway. The spatial and temporal resolutions of the FEM numerical scheme were set to $\Delta s = 3$ mm and $\Delta t = 0.005$ ms, respectively.

For both the 3D FEM and FDTD simulations, we set the air density to $\varrho = 1.14$ kg/m^3 and sound speed to $c = 350$ m/s. A constant boundary admittance coefficient of $\mu = 0.005$ was used for both numerical models to simulate vocal tract wall absorption [6]. The semi-reflective boundary condition was applied by deriving the normal acoustic impedance and the corresponding particle velocity as suggested by Yokota et al. [27]. We implemented an open-end termination condition (i.e., $p = 0$) at the surface of the mouth end by imposing the Dirichlet boundary condition. This approach reduces computational time and memory but does not consider radiation losses at lips. The study of radiation impedance using the 3D FDTD vocal tract model is still in progress. We applied a Gaussian velocity pulse at the start of the vocal tract tube (i.e., glottal end) as the source

[1] https://github.com/Debasishray19/fdtd-vocaltract-model

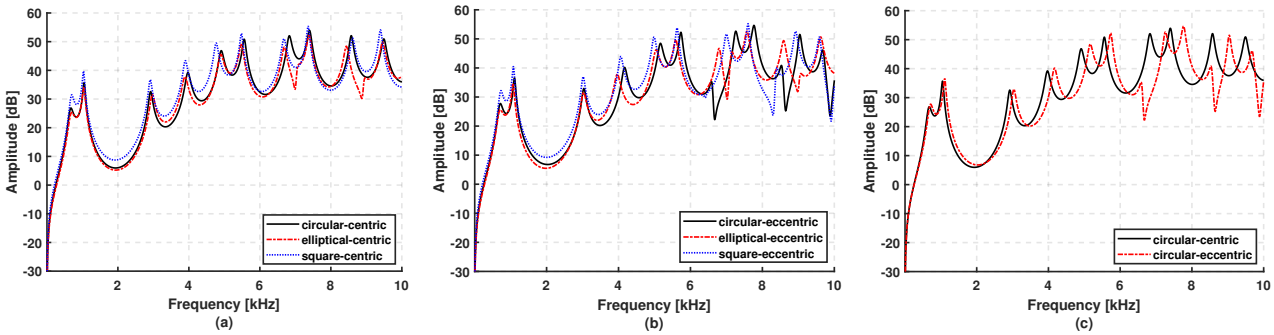

Figure 4: *Transfer functions of simplified vocal tract geometries for vowel /a/ computed using the 3D FDTD numerical scheme.*

excitation function. The outgoing pressure waves are recorded inside the vocal tract at each time step to generate audio samples of 1000 ms.

For FDTD simulations, the grid resolution was set to $\Delta s \approx 0.92$ mm; this is the largest value that preserves the geometrical details of vocal tract tubes while generating precise VTTFs. This value was determined empirically. The temporal resolution Δt of the 3D FDTD simulation was restricted by the Courant-Friedrichs-Lewy condition, $\Delta t \leq \Delta s/\sqrt{3}c$. Hence, we set $\Delta t = 1.51 \times 10^{-6}$ s, equal to a simulation sampling rate of 661.5 kHz. The sampling rate of our 3D FDTD wave solver was considerably lower than Takemoto's model (i.e., 1000 kHz), therefore reducing the simulation run-time.

3. Result and Discussion

For acoustic characterization of FDTD simulations, we obtained VTTFs, i.e., the Fast Fourier Transform (FFT) ratio of the excitation pulse at the glottal end to the recorded pressure samples at the mouth end. To validate the proposed 3D FDTD method and the transfer function measurement, we applied them to a one-end open cylindrical tube and compared the obtained resonance frequencies with the corresponding analytical solutions. The analytical tube was designed with a length of $L = 17.09$ cm to roughly represent the vocal tract length of vowel /a/. The tube's cross-sectional area was made sufficiently small, i.e., $A = 0.97$ cm^2 to fulfill the planar wave propagation condition. The first eight resonance frequencies obtained from the FDTD simulation and the analytical calculation match very well with a maximum difference of 0.6%.

We first validate the circular-centric geometry by comparing the acoustic formants generated from the FDTD and FEM simulations of five vowels. Table 1 shows their absolute positional differences in percentage (i.e., percentage of deviation). Overall, the formants of FDTD simulations agree well with the corresponding FEM models with average percentage differences of below 3% for most of the vowels. These positional differences could be considered perceptually identical [29]. The existing mismatch in formant frequencies could be due to the orthogonal grid elements of FDTD scheme. The influence of the orthogonal grid on acoustic outputs can be studied by directly comparing the FDTD and FEM simulations of a vocal tract with square or rectangular cross-sections. Understanding this phenomenon is of great interest, although we do not provide such comparative analysis in this work as we primarily focus on studying the effect of vocal tract geometry simplification.

The radial symmetry of circular-centric configuration does not allow the appearance of higher-order modes (i.e., antiresonances) in transfer functions. In contrast, if we break this symmetry and generate vocal tracts with elliptical cross-sections from 1D area functions, higher-order acoustic modes become apparent for frequencies above 5 kHz as shown in Figure 4a. Like circular-centric, the acoustic formants of elliptical-centric vocal tract configuration also match well with the corresponding FEM approach with average percentage differences of below 3% for most of the vowels (shown in Table 2). Contrary to elliptical cross-sections, the square-centric configuration has a geometrical symmetry similar to the circular-centric vocal tract shape. Therefore, we did not observe any antiresonances in VTTFs (shown in Figure 4a) of square-centric vocal tract models. Moreover, the FDTD simulation of circular-centric and square-centric configurations produced approximately similar acoustic formants. Below 5 kHz, we did not find any significant influence of different vocal tract shapes on VTTFs as the plane wave assumption holds up.

Alternatively, eccentric configuration with different cross-sectional shapes can produce higher-order acoustic modes (Figure 4b). However, they do not present an equivalent antiresonance (i.e., dips in VTTFs at different positions). Below 5 kHz, there are no significant differences between centric and eccentric cases, as shown in Figure 4c. Tables 1 and 2 also present a comparative analysis of FDTD simulations with respect to the FEM method for the circular-eccentric and elliptical-eccentric vocal tracts. We avoid comparing higher formants (i.e., f_7 and f_8) as the appearance of antiresonances in the eccentric case can produce misleading results. Compared to vowels /u/ and /o/, the average error percentage of formants is relatively low for other vowels. These differences could be due to the narrow constrictions of vowels /u/ and /o/ near the mouth end, which produce a high discretization error, and it could be avoided by increasing the sampling rate of the FDTD simulation.

4. Conclusion

The 3D FDTD simulation of simplified vocal tract shapes generates transfer functions similar to those obtained with the FEM method at a low spatiotemporal resolution. Moreover, it synthesizes the higher-order modes for vocal tracts with elliptical cross-sections or eccentric configurations. Thus, FDTD seems a promising method for studying vocal tract acoustic with simplified geometries. We plan to incorporate bent vocal tract shapes with side branches for future work to emulate the acoustic properties of realistic vocal tracts. The proposed model also needs to consider radiation impedance near the mouth end to simulate the effect of mouth radiation. Finally, numerical simulation of these simplified vocal tract shapes needs to be directly compared with realistic vocal tract models to understand the influence of geometry simplifications using the FDTD scheme.

5. Acknowledgement

This work is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canadian Institutes for Health Research (CIHR).

6. References

[1] J. Dang and K. Honda, "An improved vocal tract model of vowel production implementing piriform resonance and transvelar nasal coupling," in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, vol. 2. IEEE, 1996, pp. 965–968.

[2] R. Blandin, M. Arnela, R. Laboissière, X. Pelorson, O. Guasch, A. V. Hirtum, and X. Laval, "Effects of higher order propagation modes in vocal tract like geometries," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 832–843, 2015.

[3] T. Vampola, J. Horáček, A.-M. Laukkanen, and J. G. Švec, "Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on ct measurement," *Logopedics Phoniatrics Vocology*, vol. 40, no. 1, pp. 14–23, 2015.

[4] M. Arnela, S. Dabbaghchian, R. Blandin, O. Guasch, O. Engwall, A. Van Hirtum, and X. Pelorson, "Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds," *The Journal of the Acoustical Society of America*, vol. 140, no. 3, pp. 1707–1718, 2016.

[5] M. Arnela, S. Dabbaghchian, O. Guasch, and O. Engwall, "Mri-based vocal tract representations for the three-dimensional finite element synthesis of diphthongs," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2173–2182, 2019.

[6] T. Vampola, J. Horáček, and J. G. Švec, "Fe modeling of human vocal tract acoustics. part i: Production of czech vowels," *Acta acustica united with Acustica*, vol. 94, no. 3, pp. 433–447, 2008.

[7] M. Arnela and O. Guasch, "Two-dimensional vocal tracts with three-dimensional behavior in the numerical generation of vowels," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, pp. 369–379, 2014.

[8] O. Guasch, M. Arnela, R. Codina, and H. Espinoza, "A stabilized finite element method for the mixed wave equation in an ale framework with application to diphthong production," *Acta Acustica united with Acustica*, vol. 102, no. 1, pp. 94–106, 2016.

[9] M. Arnela and O. Guasch, "Finite element synthesis of diphthongs using tuned two-dimensional vocal tracts," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 10, pp. 2013–2023, 2017.

[10] P. Birkholz, D. Jackèl, and B. J. Kroger, "Simulation of losses due to turbulence in the time-varying vocal system," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1218–1226, 2007.

[11] V. Zappi, A. Vasuvedan, A. Allen, N. Raghuvanshi, and S. Fels, "Towards real-time two-dimensional wave propagation for articulatory speech synthesis," in *Proceedings of Meetings on Acoustics 171ASA*, vol. 26, no. 1. Acoustical Society of America, 2016, p. 045005.

[12] D. R. Mohapatra, V. Zappi, and S. Fels, "An extended two-dimensional vocal tract model for fast acoustic simulation of single-axis symmetric three-dimensional tubes," *Proc. Interspeech 2019*, pp. 3760–3764, 2019.

[13] D. R. Mohapatra, "Talking tube: a novel approach for vocal tract acoustic modelling using the finite-difference time-domain method," Ph.D. dissertation, University of British Columbia, 2021.

[14] D. R. Mohapatra, V. Zappi, and S. Fels, "A comparative study of two-dimensional vocal tract acoustic modeling based on finite-difference time-domain methods," *arXiv preprint arXiv:2102.04588*, 2021.

[15] M. Arnela and O. Guasch, "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method," *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. 4197–4209, 2013.

[16] H. Takemoto, P. Mokhtari, and T. Kitamura, "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method," *the Journal of the acoustical society of America*, vol. 128, no. 6, pp. 3724–3738, 2010.

[17] K. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media," *IEEE Transactions on antennas and propagation*, vol. 14, no. 3, pp. 302–307, 1966.

[18] Y. Wang, H. Wang, J. Wei, and J. Dang, "Mandarin vowel synthesis based on 2d and 3d vocal tract model by finite-difference time-domain method," in *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE, 2012, pp. 1–4.

[19] S. Gedney, F. S. Lansing, and D. L. Rascoe, "Full wave analysis of microwave monolithic circuit devices using a generalized yee-algorithm based on an unstructured grid," *IEEE Transactions on Microwave Theory and techniques*, vol. 44, no. 8, pp. 1393–1400, 1996.

[20] A. Bossavit and L. Kettunen, "Yee-like schemes on a tetrahedral mesh, with diagonal lumping," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 12, no. 1-2, pp. 129–142, 1999.

[21] R. Palandech, R. Mittra *et al.*, "Modeling three-dimensional discontinuities in waveguides using nonorthogonal fdtd algorithm," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 2, pp. 346–352, 1992.

[22] A. Gansen, M. El Hachemi, S. Belouettar, O. Hassan, and K. Morgan, "A 3d unstructured mesh fdtd scheme for em modelling," *Archives of Computational Methods in Engineering*, vol. 28, no. 1, pp. 181–213, 2021.

[23] N. K. Madsen, "Divergence preserving discrete surface integral methods for maxwell's curl equations using non-orthogonal unstructured grids," *Journal of Computational Physics*, vol. 119, no. 1, pp. 34–45, 1995.

[24] S. D. Gedney and J. A. Roden, "Numerical stability of nonorthogonal fdtd methods," *IEEE Transactions on Antennas and Propagation*, vol. 48, no. 2, pp. 231–239, 2000.

[25] B. H. Story, "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002," *The Journal of the Acoustical Society of America*, vol. 123, no. 1, pp. 327–335, 2008.

[26] A. Allen and N. Raghuvanshi, "Aerophones in flatland: Interactive wave simulation of wind instruments," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, pp. 1–11, 2015.

[27] T. Yokota, S. Sakamoto, and H. Tachibana, "Visualization of sound propagation and scattering in rooms," *Acoustical science and technology*, vol. 23, no. 1, pp. 40–46, 2002.

[28] P. Birkholz, S. Kürbis, S. Stone, P. Häsner, R. Blandin, and M. Fleischer, "Printable 3d vocal tract shapes from mri data and their acoustic and aerodynamic properties," *Scientific data*, vol. 7, no. 1, pp. 1–16, 2020.

[29] J. L. Flanagan, *Speech analysis synthesis and perception*. Springer Science & Business Media, 2013, vol. 3.