A comparative study of two-dimensional vocal tract acoustic modeling based on Finite-Difference Time-Domain methods

Debasish Ray Mohapatra[1], Victor Zappi[2], Sidney Fels[1]

[1]Electrical and Computer Engineering Department, University of British Columbia, Canada
[2]College of Arts, Media and Design, Northeastern University, USA
debasishray@ece.ubc.ca, v.zappi@northeastern.edu, ssfels@ece.ubc.ca

Abstract

The two-dimensional (2D) numerical approaches for vocal tract modeling can afford a better balance between the low computational cost and accurate rendering of acoustic wave propagation. However, they require a high spatiotemporal resolution in the numerical scheme for a precise estimation of the formants at the expense of the simulation run-time. Recently, a new vocal tract modeling technique, known as 2.5D Finite-Difference Time-Domain (2.5D FDTD), has been introduced by us, which extends the existing 2D FDTD approach by introducing tube depth to its acoustic wave solver. In this work, we exhibit the results from these two methods by comparing their transfer functions and time performance through aero-acoustic simulation of eleven static vowel sounds with different spatial resolutions. We present here the experimental results for vowel sound /ɑ/, which demonstrates the potentials of 2.5D FDTD method to produce precise formants compare to the 2D FDTD at a very low spatial resolution.

Keywords: computational acoustics, vocal tract, FDTD, articulatory speech synthesis

1. Introduction

In recent years, various speech simulation algorithms have been developed to model the sophisticated human upper vocal tract shapes and capture their acoustic characteristics. Though the 3D acoustic analysis (Takemoto, Mokhtari, and Kitamura 2010; Vampola et al. 2015) has more flexibility to represent the intricate geometries, the time complexity is the major issue associated with it. In contrast to 3D, the 1D vocal tract models drastically reduce the computational cost. Nevertheless, their over-simplified representation of the tube structure fails a precise simulation of acoustic wave propagation. As an alternative, the 2D acoustic wave solvers (Arnela and Guasch 2014; Speed, Murphy, and Howard 2009) improve the computational power due to the dimensionality reduction in the numerical grid but without losing much geometrical details of the vocal tract. The most well-known 2D acoustic analysis methods are the finite element method (FEM) and FDTD. Despite the limitation of the FDTD approach in capturing frequency-dependent wall losses, it computes acoustic characteristics much faster than any other method. The next section briefly discusses the numerical implementation of 3D tubes using 2D and 2.5D FDTD techniques.

2. Vocal Tract Modeling

The standard 2D FDTD technique, employed for acoustic simulation, does not facilitate dynamic boundary conditions. Hence, a new scalar field $0 \leq \beta(x, y, t) \leq 1$ was introduced to the solver, which could transit smoothly between $\beta = 1$ (air) and $\beta = 0$ (boundary). At $\beta = 0$, a prescribed velocity boundary condition $v = \boldsymbol{v_b}$ was induced. Specifically, the technique solves the following equations for the acoustic pressure, $p(x, y, t)$ and particle velocities $v(x, t)$ and $v(y, t)$ inside the numerical domain (Zappi et al. 2016):

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

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

The 2.5D FDTD follows the 2D rationale but improves upon it by incorporating a new impedance term in the 2D wave solver, known as tube depth. This approach helps to simulate the acoustic wave propagation in a single-axis symmetric 3D tube. The tube's depth (D), i.e. its continuous extension along the z axis (with x and y axis being the dimension of starting 2D scheme), was derived from the cross-sectional area of the tube and sampled at every grid point for each of the acoustic parameter (p, $\boldsymbol{v_x}$, $\boldsymbol{v_y}$) inside the 2D scheme as shown in Figure 1. Then the resulting depths are mapped to their respective acoustic parameters: $p(x, y, t)$,$v(x, t)$ and $v(y, t)$ as demonstrated by Mohapatra, Zappi, and Fels (2019). Basically, we solve the following equations in the numerical scheme:

$$\frac{\partial p}{\partial t} = -\frac{\rho c^2}{D} \left(\frac{\partial D v_x}{\partial x} + \frac{\partial D v_y}{\partial y} \right) \quad (3)$$

$$\beta \frac{\partial \boldsymbol{v}}{\partial t} + (1-\beta)\,\boldsymbol{v} = -\beta^2 \frac{\boldsymbol{\nabla} p}{\rho} + (1-\beta)\,\boldsymbol{v}_b \quad (4)$$

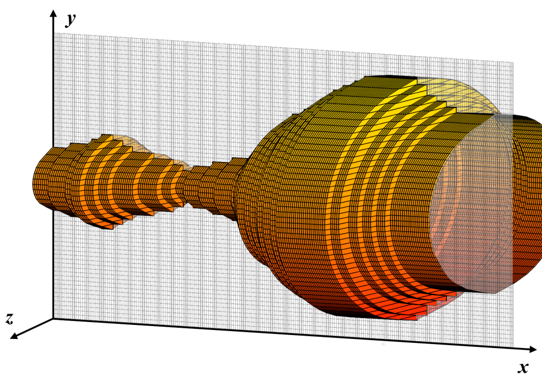

Figure 1: 2.5D - mid-sagittal contour and depth map for vowel /a/

3. Experimental Setup

A comparative study between the 2D and 2.5D schemes is carried out by running the acoustic simulation for the following static vowels: /i/, /ɪ/, /e/, /ɛ/, /æ/, /ɑ/, /o/, /u/, /ʊ/, /ʌ/ and /ɔ/. We use Story's area function dataset (Story 2008) for these vowels to extract vocal tract contours, as it is the ideal one for acoustic analysis and creating 3D tubes with circular cross-section. The 3D vocal tracts with circular cross-sections are defined as flat contours inside a 2D domain.

At the glottal end, a band-pass velocity pulse having frequency range 2Hz-20kHz is applied. We place a virtual microphone 3mm inside the mouth opening to record the pressure variation. We have not included the mouth radiation effect for FDTD methods. Hence, the open-end termination at the mouth end is implemented by imposing Dirichlet boundary conditions as done by Arnela and Guasch (2014). Courant-Friedrichs-Lewy (CFL) stability condition in two dimensions: $\Delta t \leq \Delta s/\sqrt{2}c$ is imposed, where c is the speed of sound. The vocal tract wall reflection is implemented, as illustrated by Takemoto, Mokhtari, and Kitamura (2010) to incorporate the time-dependent boundary losses.

The simulation generates 50ms of synthesized audio for every vowel sounds. During the simulation, we set the following physical parameters fixed: air density $\rho = 1.14$ kg/m^{2}s, boundary admittance $\mu = 0.005$ and sound speed $c = 350$ m/s. We implement both models in MATLAB environment and discretize the acoustic components using the 2D Yee scheme (Yee 1966). At each time-step, the solver iterates across the complete grid to sample acoustic components at each grid point. The application runs on a workstation equipped with an Intel Core i7-8700K processor.

4. Model Validation

We simulate each vowel sound under three different spatial grid resolution for both 2D and 2.5D FDTD methods: low ($\Delta s =$ 0.74mm), mid ($\Delta s = 0.37$mm) and high ($\Delta s = 0.28$mm). The high spatial resolution yields better fidelity vocal tract geometry, which in turn provides precise acoustic features but at the expense of simulation run-time. First, we compare the formants' positions for vowel sounds /ɑ/,/i/ and /u/ of 2D and 2.5D FDTD models simulated at a low spatial resolution with a 3D FEM model results (Arnela, Dabbaghchian, et al. 2016). This shows the capability of these two models to simulate acoustic wave propagation like a high quality 3D model. Then we increase the resolution as a variable to measure the changes in formants' position and record simulation run-time, which provides how high do we need to set it to get better acoustic features in both of these models. At the end, we measure the formants of rest of the vowels and compare them to formant frequencies measured from the recorded speech as done by Story (2008). We expect a large variation in the result for both of these models as they do not include mouth radiation.

5. Result

The transfer function for all the vowels are computed by applying Fast Fourier Transformation (FFT) to the pressure waves, recorded at the virtual microphone and formants' positions are extracted. We record the first three formants as they are mainly responsible for determining and distinguishing the vowel and vowel categories (Vampola et al. 2015). The formants for 3D FEM model for vowel /ɑ/ can be found in Arnela's PhD dissertation (Arnela Coll 2015).

Model Type	F1	F2	F3
3D FEM	696	1068	3031
2D FDTD (% Error)	680(2.29)	1040(2.62)	3000(1.02)
2.5D FDTD (% Error)	700(0.57)	1040(2.62)	3020(0.36)

Table 1: Comparing the first three formants (Frequencies in Hz) of FDTD methods to 3D FEM model for vowel /ɑ/. Simulation duration ≈ 276 seconds

6. Discussion & Conclusion

The Table 1 shows that even at a very low spatial resolution ($\Delta s = 0.74mm$ and sample rate=661,500Hz) there are hardly any differences between 3D FEM and 2.5D FDTD methods.

Generating synthesized speech in real-time is quite challenging. Moreover, the quality of such speech sounds depends upon various factors. We continue to work and improve our new 2.5D FDTD method for vocal tract modelling. In future, we plan to include the mouth radiation and couple it with a vocal fold model to produce synthesized speech.

7. References

Arnela Coll, Marc (2015). "Numerical production of vowels and diphthongs using finite element methods". PhD thesis. Universitat Ramon Llull.

Arnela, Marc, Saeed Dabbaghchian, Rémi Blandin, Oriol Guasch, Olov Engwall, Annemie Van Hirtum, and Xavier Pelorson (2016). "Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds". In: *The Journal of the Acoustical Society of America* 140.3, pp. 1707–1718.

Arnela, Marc and Oriol Guasch (2014). "Two-dimensional vocal tracts with three-dimensional behavior in the numerical generation of vowels". In: *The Journal of the Acoustical Society of America* 135.1, pp. 369–379.

Mohapatra, Debasish Ray, Victor Zappi, and Sidney Fels (2019). "An extended two-dimensional vocal tract model for fast acoustic simulation of single-axis symmetric three-dimensional tubes". In: *arXiv preprint arXiv:1909.09585*.

Speed, Matt, Damian Murphy, and David M Howard (2009). "Characteristics of two-dimensional finite difference techniques for vocal tract analysis and voice synthesis". In: *Tenth Annual Conference of the International Speech Communication Association*.

Story, Brad H (2008). "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002". In: *The Journal of the Acoustical Society of America* 123.1, pp. 327–335.

Takemoto, Hironori, Parham Mokhtari, and Tatsuya Kitamura (2010). "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method". In: *The Journal of the Acoustical Society of America* 128.6, pp. 3724–3738.

Vampola, Tomáš, Jaromır Horáček, Anne-Maria Laukkanen, and Jan G Švec (2015). "Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on CT measurement". In: *Logopedics Phoniatrics Vocology* 40.1, pp. 14–23.

Yee, Kane (1966). "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media". In: *IEEE Transactions on antennas and propagation* 14.3, pp. 302–307.

Zappi, Victor, Arvind Vasuvedan, Andrew Allen, Nikunj Raghuvanshi, and Sidney Fels (2016). "Towards real-time two-dimensional wave propagation for articulatory speech synthesis". In: *Proceedings of Meetings on Acoustics 171ASA*. Vol. 26. 1. ASA, p. 045005.