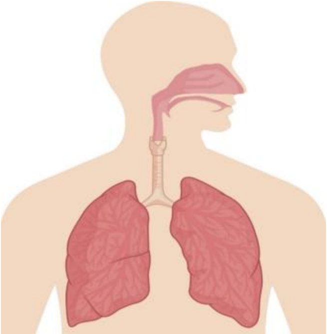
An Extended Two-Dimensional Vocal Tract Model For Fast Acoustic Simulation Of Single-Axis Symmetric Three-Dimensional Tubes

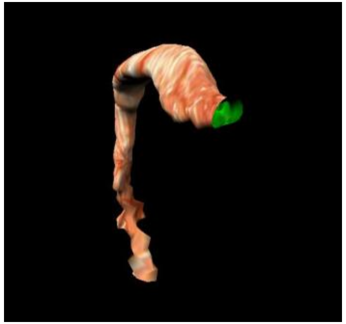
Debasish Ray Mohapatra[1], Victor Zappi[2], Sidney Fels[1]

[1]Human Communication Technologies Lab, The University of British Columbia
[2]College of Arts, Media and Design, Northeastern University

FUNDAMENTAL COMPONENTS IN ARTICULATORY VOCAL SYNTHESIS

- Source Model - Vocal Fold
- Filter Model - Upper Vocal Tract
- Coupling of Vocal fold & Vocal tract





RESEARCH OBJECTIVES

Presenting a new vocal tract model (2.5D) that targets:

Precision: Precise simulation of formants of realistic vocal tract

Time Performance: Computationally light-weight model which could achieve real-time/quasi real-time performance

3D VOCAL TRACT MODEL

Advantage:

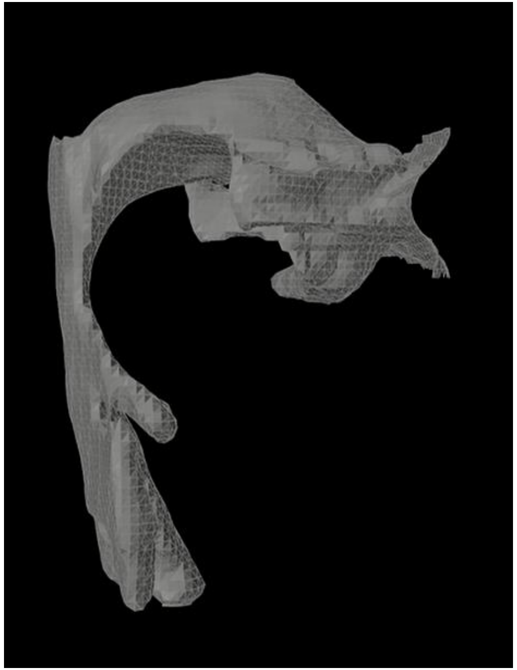
- Simulate airwaves propagation inside the realistic vocal tract geometry

- Produce very precise results

Drawbacks:

- Simulation is extremely slow - Requires several hours to simulate few milliseconds of audio

- Not convenient for many applications like clinical research or designing singing synthesis



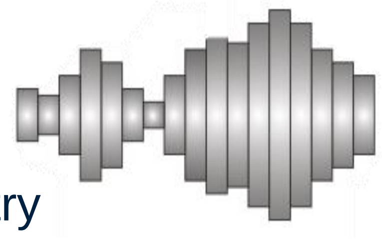
Marc Arnela, "Numerical production of vowels and diphthongs using finite element methods", 2016

1D VOCAL TRACT MODEL

Advantage:

- Simulate acoustic wave propagation in 1D
- 1D models are computationally much faster and can be simulated in real-time.

Drawbacks:



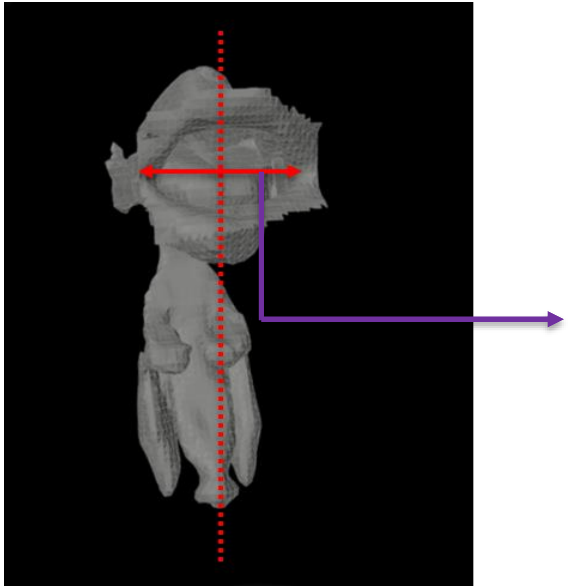
- Oversimplified representation of the actual vocal tract geometry
- Model can correctly simulate the formants up to 5kHz

http://www.vocaltractlab.de/index.php?page=background-articulatory-synthesis

2.5D – AN EXTENSION OF 2D

2D Representation

Inclusion of Depth for 2.5D

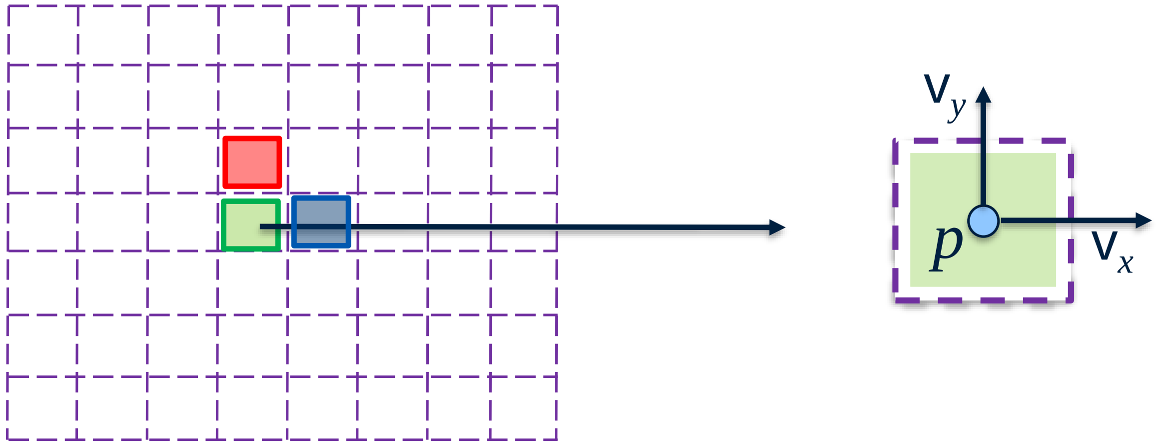


Depth across mid-sagittal plane

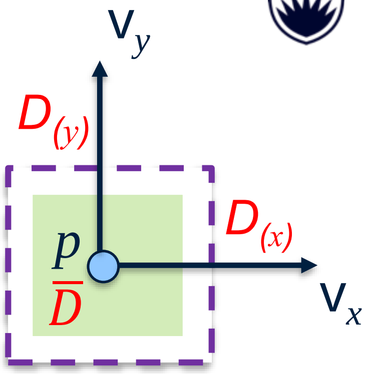
Zappi et al., "Towards real-time two-dimensional wave propagation for articulatory speech synthesis", 2016

2D FDTD – ACOUSTIC PARAMETERS REPRESENTATION

Yee Grid Representation



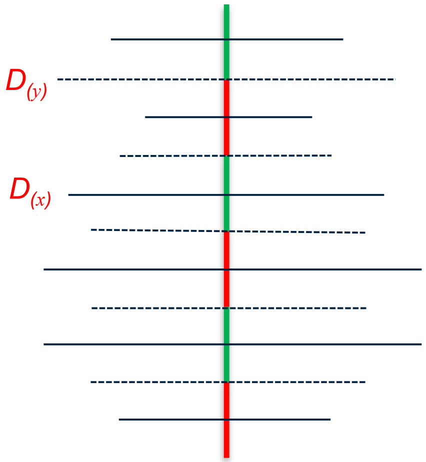
2.5D – DEPTH REPRESENTATION



- Three depth components ($\bar{D}$,$D_{(x)}$ and $D_{(y)}$) are computed

- $D_{(x)}$ and $D_{(y)}$ are aligned with velocity components

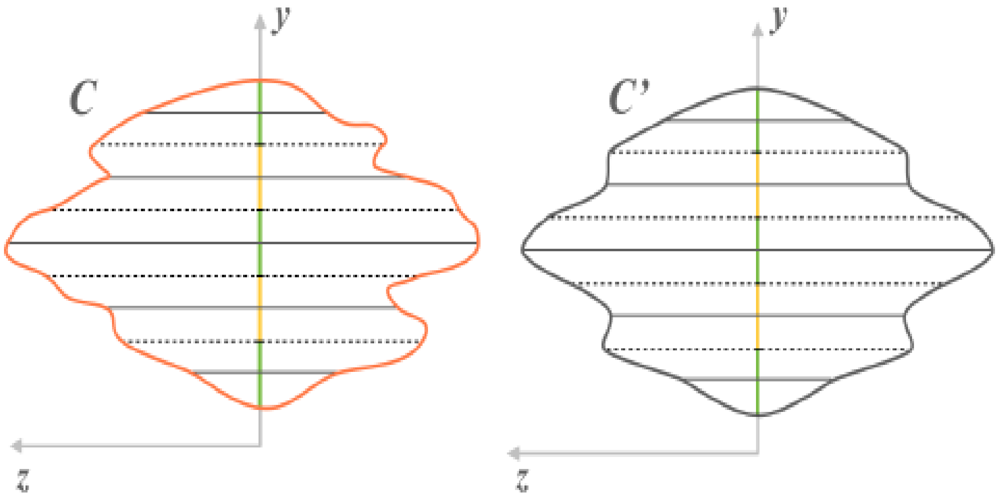
- $\bar{D}$ is aligned with the pressure field



2.5D – DEPTH REPRESENTATION

<u>Single-Axis Symmetrical Representation</u>

- Depth map introduces a single-axis symmetry
- Curvature is preserved and most of the cross sectional details



2.5D – WAVE SOLVER EQUATION

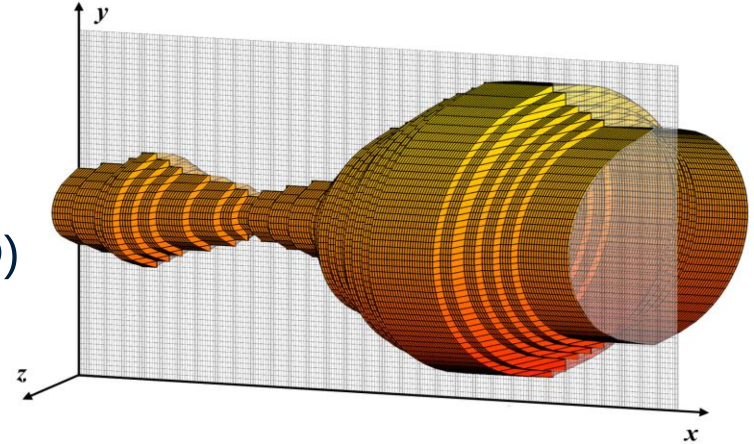
$$p^{n+1} = \frac{\bar{D}p^{(n)} - \rho c^2 \Delta t \; \tilde{\boldsymbol{\nabla}} \cdot \boldsymbol{V}^{(n)}}{\bar{D}(1 + \rho' \Delta t)}$$

$$\boldsymbol{v}^{(n+1)} \;\; = \;\; \frac{\beta \boldsymbol{v}^{(n)} - \beta^2 \Delta t \; \tilde{\nabla} p^{(n+1)}/\rho \; + \; \Delta t(1-\beta)\boldsymbol{v_b}}{\beta + \Delta t(1-\beta)}$$

$$where, \boldsymbol{V} \;\; = \;\; (D_{(x)} v_x, \;\; D_{(y)} v_y \;)$$

2.5D – MODEL VALIDATION

- Used Story’s area-function to construct the vocal tract geometry
- 2.5D FDTD acoustic wave solver was implemented in MATLAB environment
- The grid resolution was set to the best possible minimum value.
- 2D FDTD (same simulation parameters as 2.5D) for time performance comparison.



Mid-sagittal contour and depth map for vowel /a/

B.H. Story, “Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002”, 2008

SIMULATION RESULT – PRECISION

Comparison with slow but precise 3D FEM for static vowel sounds /a/, /i/, /u/

Percentage error in terms of formant positions			
Formants	/a/	/i/	/u/
f1	0.57	-1.14	0.38
f2	-2.62	2.32	-4.88
f3	-0.36	1.66	1.59
f4	1.37	1.98	0.47
f5	-1.17	0.01	2.56
f6	-0.40	-0.22	-2.01
f7	-0.07	-1.44	-0.75
f8	0.32	0.45	0.39

Marc Arnela, "Numerical production of vowels and diphthongs using finite element methods", 2016

SIMULATION RESULT – TIME PERFORMANCE

Comparison with fast but unprecise 2D FDTD model

- MATLAB implementation of 2.5D FDTD proves to be extremely lightweight

- Both 2D and 2.5D models showcased a comparable time complexity (around 15 mins to produce 50 milli-seconds of audio)

- A highly optimized parallel implementation for the 2D FDTD model in GPU has already been proven to be quasi real-time in terms of time complexity.

DISCUSSION & MODEL LIMITATION

- 2.5D VT model is as precise as 3D model in a simple scenario. And it has the potential to run close to real-time.

- The model needs to be tested in more complex scenario

- Losses are only modeled across the 2D contour but not along z-axis

FUTURE WORK

- Test the precision of the model while simulating bent tube with circular cross-sections

- Add boundary condition to test irregular cross-sections.

- Implementing the model using GPU to target real-time simulation

THANK YOU

THE UNIVERSITY OF BRITISH COLUMBIA

Northeastern University
College of Arts, Media and Design

QUESTIONS PLEASE