MODELLING SPEECH PRODUCTION USING YEE'S FINITE DIFFERENCE METHOD

Kathleen E. Cummings[1], *James G. Maloney*[2] *and Mark A. Clements*[1]

1School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332
[2]Signature Technology Laboratory, Georgia Tech Research Institute, Atlanta, GA 30332

ABSTRACT

This paper describes a model of speech production based on solving for acoustic wave propagation in the vocal tract using a finite-difference time-domain (FDTD) technique. This FDTD technique was first developed by Yee and utilizes a discretization scheme in which pressure and velocity components are interleaved in both space and time. The specific implementation of this model of speech production, including discretization of the coupled acoustic wave equations, boundary conditions, stability criteria, values of model constants, and method of excitation, are presented in this paper. The accuracy of the model is verified by comparing the FDTD results to the theoretically expected results for a well-known acoustics problem. The FDTD model of speech production has been used in a variety of experiments, and several results, including those that compare the use of several common glottal models as excitation, are presented here.

1. INTRODUCTION

Developing models of speech production has been an important area of research for many years. The many and varied applications of speech processing, including speech synthesis, speech coding, and automatic speech recognition all are dependent on understanding and modelling the mechanisms of speech production. Most current methods of modelling speech production involve either directly analyzing the acoustic speech waveform, which assumes a simplified linear model, or solving the acoustic wave equation in a series of cylindrical tubes representing the vocal tract, which assumes a simplified geometry. Although these models have been highly successful for several years, the pace of new advances has been slow. Many researchers believe that the single most important factor in making further progress in speech processing is to improve the model, and therefore researchers' understanding, of speech production.

This paper will present a model of speech production based on a numerical simulation of acoustic wave propagation in the vocal tract. The model is based on solving for acoustic wave propagation using a finite-difference time-domain (FDTD) technique in a geometry that accurately represents the vocal tract shape. FDTD methods such as the one used in this research inherently provide several important advantages over current speech analysis techniques. First, a FDTD solution provides full knowledge of the acoustic flow at every point in the vocal tract for every time. Second, the finite-difference time-domain method makes no assumption of linearity between the input and the output (note that this is not true for frequency-domain finite-difference methods). Finally, a numerical solution such as the one described in this paper allows many parameters such as the geometry of the vocal tract and the excitation source to be altered easily.

This paper will develop an adaptation of Yee's finite difference scheme [11], which has been used successfully in the fields of electromagnetics and computational fluid dynamics, to the problem of acoustic wave propagation in the vocal tract. A brief description of the theory of speech production upon which the model is based, along with a discussion of the governing equations of acoustic wave propagation in the vocal tract, will be presented. The adaptation of Yee's finite-difference time-domain method to this model of speech production will then be discussed. Issues such as grid placement, boundary conditions, excitation and stability will be presented. In order to validate the accuracy of the method, a rotationally-symmetric three-dimensional unflanged cylinder will be analyzed, and the results will be compared to the theoretical results. Some results of analyzing a two-dimensional vocal tract geometry taken from x-ray data for a vowel using this model of speech production will be presented.

2. THEORY

It is generally accepted that, in speech production, the vocal tract acts as a variable-geometry acoustic resonator that is excited at one end and that radiates from the other end. The glottis acts to regulate the excitation from the lungs, either opening and closing regularly to produce periodic pulses of air, as in voiced speech, or allowing the pressure from the lungs to pass through unimpeded as in unvoiced speech. In unvoiced speech, the excitation of the vocal tract occurs at some point past the glottis. In this case, a partial or complete constriction in the vocal tract causes the build-up of turbulent noise which is released when the constriction is eased. It is possible to have both voiced and unvoiced excitations. The excitation is shaped by the resonant cavities comprising the vocal tract, and the resultant waveform is radiated from the lips and/or nose.

In traditional speech production models, several assumptions are made about the nature of acoustic wave propagation in and the geometry of the vocal tract. Generally, it is assumed that there are no viscous or thermal losses, that only small perturbations about the mean pressure occur, and that only plane wave propagation can occur. Furthermore, the vocal tract is usually assumed to be a series of

[0]This material is based upon work supported under National Science Foundation Grant # CDA-9211129 and a grant from the American Speech-Language-Hearing Association.

0-7803-2431-5/95 $4.00 © 1995 IEEE

contiguous rotationally-symmetric tubes of variable cross-sectional area.

The equations that describe acoustic wave propagation are derived from the equations of momentum and continuity:

$$\delta_0 \frac{\partial \vec{u}}{\partial t} = -\nabla p. \tag{1}$$

$$\frac{\partial p}{\partial t} = -\delta_0 c^2 \nabla \cdot \vec{u} \tag{2}$$

where $\vec{u}$ is the gas particle velocity, comprised of velocity components in each coordinate direction, p is the deviation from ambient pressure, c is the speed of sound in the medium, and δ_0 is the density of the gas at rest. These equations assume small perturbations from rest and negligible viscosity.

In rectangular coordinates, these two equations can be expanded to

$$\delta_0 \frac{\partial \vec{u}}{\partial t} = -\frac{\partial p}{\partial x}\hat{x} - \frac{\partial p}{\partial y}\hat{y} - \frac{\partial p}{\partial z}\hat{z} \tag{3}$$

and

$$\frac{\partial p}{\partial t} = -\delta_0 c^2 \left[\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z}\right]. \tag{4}$$

Similarly, in cylindrical coordinates, the two equations can be expanded to:

$$\delta_0 \frac{\partial \vec{u}}{\partial t} = -\left[\frac{\partial p}{\partial r}\hat{r} + \frac{1}{r}\frac{\partial p}{\partial \phi}\hat{\phi} + \frac{\partial p}{\partial z}\hat{z}\right] \tag{5}$$

and

$$\frac{\partial p}{\partial t} = -\delta_0 c^2 \left[\frac{1}{r}\frac{\partial}{\partial r}(r u_r) + \frac{1}{r}\frac{\partial u_\phi}{\partial \phi} + \frac{\partial u_z}{\partial z}\right]. \tag{6}$$

3. METHOD

3.1. Discretization of Differential Equations

The uniqueness of the model described in this paper is the manner in which these equations are discretized and solved for numerically, in the time-domain and in a geometry that accurately represents the vocal tract. In a finite-difference method, differential equations are usually discretized such that a first derivative is approximated as a first difference. In Yee's finite-difference time-domain method, the coupled equations in p and $\vec{u}$ are discretized and solved numerically using a space and time grid in which the samples are interleaved. Yee's method is sometimes called the "leap-frog" method because of the way in which p and $\vec{u}$ samples are interleaved in both time and space by one-half of a grid cell.

A single three-dimensional rectangular grid cell is shown in Figure 1. As can be seen, p is located in the center of each cell, while the $\vec{u}$ components are each located in the center of the appropriate cell face. The differential equations are discretized as follows. In this notation, the superscripts refer to the location in space.

In rectangular coordinates,

$$\left(\frac{u_x^{i-.5,j,k}(n+.5) - u_x^{i-.5,j,k}(n-.5)}{\Delta t}\right) = \frac{-1}{\delta_0}\left(\frac{p^{i,j,k}(n) - p^{i-1,j,k}(n)}{\Delta x}\right) \tag{7}$$

$$\left(\frac{u_y^{i,j-.5,k}(n+.5) - u_y^{i,j-.5,k}(n-.5)}{\Delta t}\right) = \frac{-1}{\delta_0}\left(\frac{p^{i,j,k}(n) - p^{i,j-1,k}(n)}{\Delta y}\right) \tag{8}$$

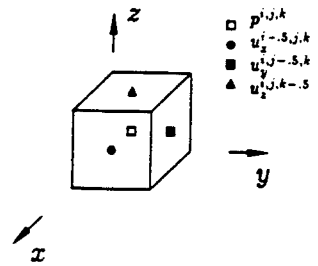

Figure 1. Three-dimensional rectangular grid cell.

$$\left(\frac{u_z^{i,j,k-.5}(n+.5) - u_z^{i,j,k-.5}(n-.5)}{\Delta t}\right) = \frac{-1}{\delta_0}\left(\frac{p^{i,j,k}(n) - p^{i,j,k-1}(n)}{\Delta z}\right) \tag{9}$$

$$\begin{aligned}\left(\frac{p^{i,j,k}(n+1) - p^{i,j,k}(n)}{\Delta t}\right) = \\ -\delta_0 c^2 \left(\frac{u_x^{i+.5,j,k}(n+.5) - u_x^{i-.5,j,k}(n+.5)}{\Delta x}\right) \\ -\delta_0 c^2 \left(\frac{u_y^{i,j+.5,k}(n+.5) - u_y^{i,j-.5,k}(n+.5)}{\Delta y}\right) \\ -\delta_0 c^2 \left(\frac{u_z^{i,j,k+.5}(n+.5) - u_z^{i,j,k-.5}(n+.5)}{\Delta z}\right)\end{aligned} \tag{10}$$

Each $\vec{u}$ component is dependent on the p value in both the given space-cell and in the previous space-cell for the previous time and on the respective $\vec{u}$ component in the given space-cell for the previous time-step. Each p value is dependent on the $\vec{u}$ values in the given space-cell and the adjacent space-cell for the current time and on the p value in the given space-cell for the previous time-step. For each time-step, all space values are calculated.

A similar, though more complex, discretization can be made for the acoustic wave equations in cylindrical coordinates. Because of the manner in which the pressure and velocity components are interleaved in space, it is important to note that the discretized r in these equations has a different value (i.e., the respective distance from the r axis) depending on the component value being calculated.

3.2. Model Specifications

In addition to discretizing the differential equations, there are several issues that must be addressed in implementing a finite-difference solution. These issues include boundary conditions within the geometry of the problem, boundary conditions that substitute for the infinite space outside of a given geometry, stability criteria and the values of constants, and excitation of the geometry. This section will discuss briefly each of these issues in terms of the implementation of this model of speech production.

3.2.1. *Boundary Conditions*

There are two types of boundary conditions that must be specified in a numerical model. The first is the definition of the boundaries of the geometry that is being modelled. In a model of speech production, this means defining the boundary conditions at the vocal tract walls and at the glottis. In all of the examples considered in this paper, the vocal tract walls will be assumed to be rigid (currently, a formulation of the model is being developed that allows for yielding walls). That is, the normal velocity at each wall boundary is assumed to be zero. The boundary condition at the glottis involves the injection of the excitation source, which will be discussed later in this section.

The second type of boundary condition that must be defined are the boundary conditions outside of the model geometry. In order to simulate the infinite space outside the lips with a finite numerical grid, an *Absorbing Boundary Condition* (ABC) is used to truncate the grid at the outer edges of the computational space. There are a variety of ABCs that have been proposed. In the model of speech production developed here, it is possible to extend the computational space well beyond the end of the vocal tract model. Therefore, a relatively simple ABC has been implemented to truncate the numerical grid space. The boundary cell is simply assigned the value of the cell one previous space cell and two previous time-steps [8].

3.2.2. Stability Criterion

In order for the solution to be stable, the space and time grid increments must be chosen to satisfy the well-known "domain of dependence condition." Also called the "Courant-Friedrichs-Lewy" condition, in a two-dimensional model, the grid increments must be selected to satisfy

$$c\,\Delta t \leq \frac{1}{\sqrt{\frac{1}{\Delta x^2}+\frac{1}{\Delta y^2}}}. \qquad (11)$$

With Δx equal to Δy, selecting the quantity $\frac{c\Delta t}{\Delta x}$ equal to 0.5 satisfies this condition.

In this model of speech production, the two-dimensional grid size was selected as $\Delta x = 0.125$ cm and $\Delta y = 0.125$ cm. The constant values used in this model are:

$$\begin{aligned} \Delta x &= 0.125 \text{ cm} \\ \Delta y &= 0.125 \text{ cm} \\ c &= 353.1442 \text{ m/sec} \\ \delta_0 &= 0.15 \text{ kg}/m^3 \end{aligned}$$

Note that because the average temperature inside the mouth is greater than $30°C$, c, the speed of sound, is calculated to be 353.1445 m/sec (assuming a temperature inside the mouth of $37°C$). The sampling rate, $1/\Delta t$, is calculated to be 565.031 kHz. Clearly, this model requires a high level of computational intensity.

3.2.3. Excitation

The vocal tract model is excited by assuming that the normal velocity at the glottis is equal to the volume velocity of a known glottal excitation volume velocity pulse. The glottal opening is assumed to be approximately 3/8 cm long [6]. Therefore, u_y in three adjacent grids cells at the location of the glottis is set equal to the value of a known glottal volume velocity pulse at successive time samples. A two-dimensional grid for the Russian vowel /e/ is shown in Figure 4. The three grid cells in which the excitation is injected are annotated. It is therefore possible to use any one of several glottal models to generate the glottal volume velocity with which the vocal tract model is excited.

4. RESULTS

Having defined the current model specifications, the remainder of this paper is devoted to presenting several results obtained using this model. The first part of this section will present a comparison of the model results to the theoretical results for a well-known acoustics problem. The second part of this section will present the results of having modelled vowel production with the two-dimensional finite difference grid model.

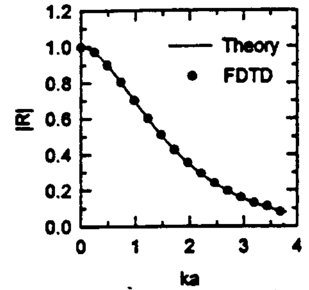

Figure 2. Comparison of $|R|$, finite-difference time-domain solution and theoretical solution.

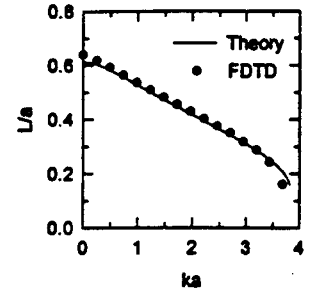

Figure 3. Comparison of l/a, finite-difference time-domain solution and theoretical solution.

4.1. Theoretical Verification

The traditional method of verifying the accuracy of a numerical model is to select a problem with a well-known theoretical solution and to compare the results of using the numerical model to the theoretically expected results. For this paper, a three-dimensional unflanged, open-ended circular cylinder was modelled and excited with a Gaussian pulse. The reflection coefficient was calculated from the numerical solution of the propagation of the Gaussian pulse in the cylinder over time and compared to the theoretically derived reflection coefficent [7]. The comparison of the experimentally calculated and the theoretically expected values $|R|$ and l/a versus ka (where k is the wave number, $2\pi/\lambda$, and a is the radius of the cylinder) are shown in Figures 2 and 3. As can be seen, the agreement between the finite difference and the theoretical results is excellent. The accuracy of this finite difference time domain model is therefore verified.

4.2. Vocal Tract Models

Several experiments have been performed using the finite difference model of speech production. The production of several vowels has been modelled using two-dimensional grids based on published x-ray data [10]. One such grid,

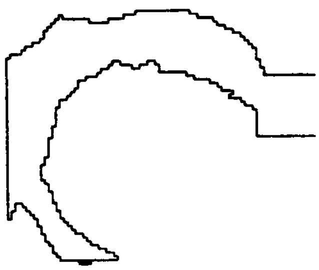

Figure 4. Two-dimensional rectangular grid, /e/.

	mea-sured	Inv filt	Beta fcn	Hed	Rose trig	Rose poly
F1 (Hz)	512	500	500	500	500	500
F2 (Hz)	1600	1535	1535	1535	1535	1535
F3 (Hz)	2320	2345	2345	2345	2345	2345
F4 (Hz)	3379	3603	3621	3586	3586	3586

Table 1. Measured and calculated formant locations.

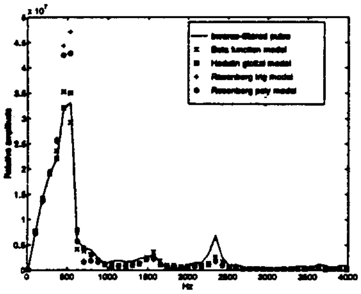

Figure 5. Comparison of the magnitude of the resulting speech spectra given five different excitations.

for the Russian vowel /e/ is shown in Figure 4. One of the many experiments that has been performed using the model was an experiment in which the model was excited with five different, common glottal excitations.

The FDTD model was excited with five different volume velocity pulses: an inverse-filtered glottal waveform, a *normal* beta function glottal model, a Rosenberg trigonometric glottal model pulse [9], a Rosenberg polynomial glottal model pulse [9], and a Hedelin glottal model pulse [5]. These are shown in Figure 5. The model parameters for these last four were determined by performing a least-squares best-fit to the inverse-filtered glottal waveform. The *normal* beta function glottal model is one that was developed in previous research to model various styles of glottal excitation [2] [3] [4].

A comparison of the magnitude of the spectrum of the resulting speech waveform is shown in Figure 5. The speech waveform was taken as the pressure at the point (290,41), approximately 28 cm in front of the lips.

From the resulting speech waveforms for each of the five excitations, the formant locations were determined and compared to each other and to the measured formant locations for this vocal tract shape. The formant locations are shown in Table 1, along with the measured formants for this vocal tract shape [10]. The resonance locations produced using this new model of speech production are extremely accurate. In fact, for the first three formants, the percent error between the measured and the model formant locations was less than 5%. For the fourth formant, the percent error was around 6.5%. As is observed in the plot of the speech spectra, all of the excitations resulted in largely the same spectral content.

Generally, the results for this model of speech production are viewed as movies. Pressure and velocity are observed for the entire grid versus time. Several animated versions of results from several different modelling experiments will be shown at the conference. Additionally, the model has recently been expanded to three dimensions. MRI data that has been collected as part of this project will also be available at the conference.

5. CONCLUSION

A model of speech production based on a finite-difference time-domain solution of wave propagation in the vocal tract has been presented in this paper. The model is based on a "leap-frog" FDTD method, first developed by Yee [11], in which the pressure and velocity components are interleaved in both time and space. The discretization of the coupled acoustic wave equations for this FDTD implementation were presented. Other specifics of the model, including boundary conditions, stability criteria, constant values, and model excitation were discussed. The accuracy of this FDTD model was confirmed by modelling a well-known acoustics problem, reflection in an open-ended, circular cylinder. The results calculated using the FDTD model were extremely accurate when compared to the theoretical solution. Finally, the geometry that was used to model a Russian vowel, /e/, was presented. The results of exciting the model with five common glottal volume velocity pulses, along with the pressure versus time for a given vowel model and a given excitation, were shown. The primary method of interpreting the results of this model is with animated versions of the pressure and velocity. Examples of these will be shown at the conference.

REFERENCES

[1] Cummings, K. E. and Clements, M. A. "Analysis of Glottal Waveforms Across Stress Styles," *Proceedings, IEEE ICASSP*, 1990, pp 369-372.

[2] Cummings, K. E. and Clements, M. A. "Improvements to and Applications of Analysis of Stressed Speech Using Glottal Waveforms," *Proceedings, IEEE ICASSP*, 1992, vol 2, pp 25-28.

[3] Cummings, K. E. and Clements, M. A., "Application of the Analysis of the Glottal Excitation of Stressed Speech to Speaking Style Modification," *Proceedings, IEEE ICASSP*, 1993, vol 2, pp 207-210.

[4] Cummings, K. E., *Analysis, Synthesis, and Recognition of Stressed Speech* , Ph. D. Dissertation, Georgia Institute of Technology, September, 1992.

[5] Hedelin, P., "High Quality Glottal LPC-Vocoding," *Proceedings, IEEE ICASSP*, 1986, vol 1, pp 465-468.

[6] Hegerl, G., and Hode, H., "Numerical Simulation of the Glottal Flow by a Model Based on the Compressible Navier-Stokes Equations," *Proceedings, IEEE ICASSP*, 1991, vol 1, pp 477-480.

[7] Levine, H. and Schwinger, J., "On the Radiation of Sound from an Unflanged Circular Pipe," *Physical Review*, Vol 3, No 4, February 15, 1948, pp 383-406.

[8] Maloney, J. G., *Analysis and Synthesis of Transient Antennas Using the Finite-Difference Time-Domain (FDTD) Method*, Ph. D. dissertation, Georgia Institute of Technology, November, 1992.

[9] Rosenberg, A. E., "Effect of Glottal Pulse Shape on the Quality of Natural Vowels," *Journal of the Acoustical Society of America*, vol 49, no 2, pp 583-590, 1971.

[10] Sorokin, V. N., "Determination of vocal tract shapes for vowels," *Speech Communication*, Volume 11, pp 71-85, March, 1992.

[11] Yee, K. S., "Numerical Solution of Initial Boundary Value Problems Involving Maxwells' Equations in Isotropic Media," *IEEE Trans. Antennas Propagat.*, vol AP-14, pp 302-307, May, 1966.