# MODAL ANALYSIS OF ACOUSTIC WAVE PROPAGATION IN THE VOCAL TRACT USING A FINITE-DIFFERENCE SIMULATION

Gordon Ramsay and Li Deng
Dept. of Electrical & Computer Engineering, University of Waterloo, Canada.

## ABSTRACT

A time-domain simulation of wave propagation in the vocal tract is outlined, using finite-differences to map a system of partial differential equations onto a state-space recursion. It is shown that the eigenvalues and eigenvectors of the time-varying system matrix can be used to determine the resonant modes of the vocal tract at any desired time instant. The effect of glottal vibration on formant structure is examined using this technique.

## INTRODUCTION

The correspondance between the formants of the speech signal and the resonant modes of the vocal tract is well known, as is the role of the spatial pressure and volume-velocity distributions associated with each mode in determining the response of the vocal tract to an acoustic source located at any point within the tract. Previous attempts to determine the spatial and temporal modes using acoustic models [1] [2] have relied largely on frequency-domain methods which are limited to static tract shapes, and which do not account for coupling between the glottis and the supra-glottal cavities. In this paper, a time-domain simulation method is described which allows the resonant modes of the modelled vocal tract to be determined at every sample point. The method relies on the use of a finite-difference approximation to convert a system of acoustic partial differential equations into state-space form. The eigenstructure of the resulting matrix recursion determines the spatial and temporal modes of the vocal tract, and can be calculated at each time instant. Since the model is valid for time-varying glottal and supra-glottal area functions, it can be used to examine the effect of fast variations in glottal shape on the formants. This is difficult to investigate using spectral analysis techniques due to the short time scales involved, as reported in [3][4].

## THE ACOUSTIC MODEL

The acoustic model adopted here is based on a previous model due to Maeda [5]. Define a bounded region $\Omega \subset \mathcal{R}^2$ to represent all points $(x,t)$ in the vocal tract at distance $x$ from the trachea entrance at time $t$. $\Omega$ is divided into three sub-domains $\Omega_1$, $\Omega_2$, $\Omega_3$ representing the trachea, glottis, and oral cavity respectively, which are separated and enclosed by boundaries $\partial\Omega_{t0}$, $\partial\Omega_{tT}$, $\partial\Omega_{x0}$, $\partial\Omega_{xL}$, $\partial\Omega_{12}$, $\partial\Omega_{23}$.

$$\begin{aligned}
&\Omega = \{(x,t) : 0 \le x \le L(t), 0 \le t \le T\},\\
&\Omega_1 = \{0 < x < X_T\},\\
&\Omega_2 = \{X_T < x < X_G\},\\
&\Omega_3 = \{X_G < x < L(t)\},\\
&\partial\Omega_{t0} = \{t = 0\},\ \partial\Omega_{tT} = \{t = T\},\\
&\partial\Omega_{x0} = \{x = 0\},\ \partial\Omega_{xL} = \{x = L(t)\},\\
&\partial\Omega_{12} = \{x = X_T\},\ \partial\Omega_{23} = \{x = X_G\},
\end{aligned}$$

$L(t)$ is the time-varying tract length; $X_T$ and $X_G$ are the distances from the trachea entrance to the entrance and exit of the glottis, assumed constant.

Suppose that the vocal tract may be approximated by a non-uniform time-varying elastic tube of equivalent circular cross-section $A(x,t)$ and circumference $S(x,t)$. Under the usual assumptions that the processes governing fluid flow are laminar and isentropic, the state of the air in the vocal tract can adequately be described by the pressure $p(x,t)$ and volume velocity $u(x,t)$ at all points along the mid-line, and the system of linearized acoustic equations governing one-dimensional planar wave propagation can then be derived from conservation of mass and momentum. Additional losses must be included to account for viscous friction and wall vibration. Assuming that the wall surfaces are locally-reacting, the displacement $y(x,t)$ from an equilibrium radius can be conveniently modelled as the response of a second-order linear mechanical system to local pressure variations. The equation for Poiseuille flow in a cylindrical duct can be used to provide the viscous loss term. The system of partial differential equations for $p$,$u$,$y$ in $\Omega_1$ and $\Omega_3$ is then given by:

$$\begin{aligned}
\frac{\partial p}{\partial x} + \frac{\partial}{\partial t}\left(\frac{\rho u}{A}\right) + \frac{8\pi\mu}{A^2}u &= 0,\\
\frac{\partial u}{\partial x} + \frac{\partial}{\partial t}\left(\frac{Ap}{\rho c^2}\right) + \frac{\partial}{\partial t}(A + Sy) &= 0,\\
m\frac{\partial^2 y}{\partial t^2} + b\frac{\partial y}{\partial t} + ky - Sp &= 0.
\end{aligned}$$

Flow within the glottis is assumed incompressible, and wall vibration may be neglected, leading to the following modified system of equations for $\Omega_2$,

$$\begin{aligned}
\frac{\partial p}{\partial x} + \frac{\partial}{\partial t}\left(\frac{\rho u}{A}\right) + \frac{12\mu l_g^2}{A^3}u &= 0,\\
\frac{\partial u}{\partial x} &= 0,
\end{aligned}$$

where $l_g$ is the glottal length and $\rho$, $\mu$ the density and viscosity of air.

The boundary condition at the trachea entrance $\partial\Omega_{x0}$ is provided by the lung pressure $P_{lung}(t)$.

The boundary condition for $\partial\Omega_{xL}$ at the lips is supplied by the radiation impedance, which can be modelled using Flanagan's approximation,

$$\frac{\partial u}{\partial t} - \frac{9\pi^2}{128\rho c}\frac{\partial}{\partial t}(Ap) - \frac{3\pi\sqrt{\pi A}}{8\rho}p = 0.$$

Continuity is assumed across the internal boundaries $\partial\Omega_{12}$ and $\partial\Omega_{23}$, and all quantities assume their equilibrium values initially along $\partial\Omega_{t0}$.

The mixed initial/boundary-value problem must now be solved on $\Omega$ using numerical methods. Applying the finite-difference technique, the continuous domain $\Omega$ is sampled on a grid of points $\{(x_j,t_k) : j = 1\ldots M, k = 1\ldots N\}$, which need not be uniform. The partial derivatives in the original continuous-domain equations are replaced by difference operators, yielding simultaneous linear algebraic equations linking the quantities $p$, $u$, $y$ sampled at neighbouring grid points over several time steps. In this way, the continuous-domain equations are translated into a system of linear difference equations defining a recursion on a finite-dimensional state-space, which can be solved to yield an approximation to the true solution. The solution of the discretized system will converge to a solution of the original continuous-domain system if it can be shown that the discretization is *consistent* and the recursion is *stable*. A two-level implicit difference scheme has been carefully constructed from the above equations to guarantee convergence for a slowly time-varying grid defined on $\Omega$. The details are omitted here due to lack of space.

Denoting by $Z_k$ the vector of values for $p(x_j,t_k)$, $u(x_j,t_k)$, $y(x_j,t_k)$, $y'(x_j,t_k)$ on all grid points at time $t_k$, and taking $Z_0 = 0$, the resulting implicit recursion may be written as

$$P_k Z_{k+1} = Q_k Z_k + F_k$$

where $P_k$, $Q_k$ are sparse banded matrices whose coefficients are functions of $A(x,t)$ determined by the difference scheme, and $F_k$ is a vector driving function derived from the boundary condition on $\partial\Omega_{x0}$.

This is a standard generalized eigenvalue problem, and the properties of the recursion are clearly entirely determined by the eigenstructure of the matrices $P_k^{-1}Q_k$. In particular, at any time $t_k$, the solution $Z_k$ of the recursion can be expressed as a modal sum involving the eigenvalues $\lambda_k^i$ and eigenvectors $\phi_k^i$ of $P_k^{-1}Q_k$.

Although the original equations do not possess a well-defined system of eigenfunctions due to the time-varying tract length, the finite-difference scheme can be shown to approximate the original PDEs in the limit as the grid dimensions tend to zero, and the changing eigenstructure of the corresponding matrix recursion can be taken to represent a "local" approximation to the evolution of the vocal tract resonant modes. The eigenvalues of the system matrix represent the time-varying poles of the vocal tract, and can be used to calculate the formants and their bandwidths, while the eigenvectors represent the changing spatial distributions of pressure and volume velocity associated with each formant.

By examining the modal structure of matrices $P_k^{-1}Q_k$, calculated for any particular time-varying area function $A(x,t)$ on $\Omega$, it is therefore possible to arrive at a complete characterization of the behaviour of the modelled vocal tract for every discrete time sample.

## EXPERIMENTAL RESULTS

As a useful application of this technique, consider the dynamic changes in glottal shape which occur during phonation. It is well known that formant motion occurs during the glottal cycle, but previous investigations [3][4] have found difficulties in determining the exact nature of this effect using conventional spectral analysis techniques.

The modal analysis method circumvents some of these problems. Figure 1 shows the variation in frequency and bandwidth for the first three formants, calculated every 0.2ms during two glottal cycles over a period of 20ms for the vowel /a/, together with the associated normalized modal pressure and volume velocity distributions. The area function was generated from an articulatory model; 6 grid points were used for the trachea, 7 in the glottis, and 79 for the oral tract, and the glottal section areas were assumed to execute co-phasic sinusoidal oscillations about a slightly-abducted rest position.

During each glottal period, the vocal tract poles were found to execute "teardrop-shaped" movements in the complex plane, with the path shape depending on the area function. The formant frequencies appear to increase monotonically with glottal area, as found in [3][4], but the effect on the bandwidths is more complicated, with positive and negative excursions. The modal distributions when the glottis is closed are similar to those calculated in [1][2], and the distortion that occurs due to glottal opening is most pronounced in the lower formants.

## CONCLUSIONS

A time-domain technique for simulating the time-varying resonant modes of the vocal tract has been described, and used to examine the effect of rapid changes in glottal shape on modelled formant positions. Results derived by previous authors [3][4] using spectral analysis have been largely confirmed, but the calculation method applied here is somewhat more reliable, and clarifies the precise movement of the tract modes during simulation. It remains to be verified, however, whether the linear acoustic model is indeed a valid approximation to real speech production.

## REFERENCES


[1] Fant G., Pauli S., (1975) "Spatial characteristics of vocal tract resonance modes," *Proc. Speech Communication Seminar, Stockholm 1974*, pp. 121–132.

[2] Mrayati M., Carré R., (1976) "Relations entre la forme du conduit vocal et les caractéristiques acoustiques des voyelles françaises," *Phonetica 33* pp. 285–306.

[3] Cranen B., Boves L., (1987) "Spectral consequences of a time-varying glottal impedance," *Proc. International Congress of Phonetic Sciences, 1987*, pp.361–366.

[4] Meyer P., Strube H.W., (1984) "Calculations on the time-varying vocal tract," *Speech Communication 3* pp. 109–122.

[5] Maeda S., (1982) "A digital simulation method of the vocal-tract system," *Speech Communication 1* pp. 199–229.




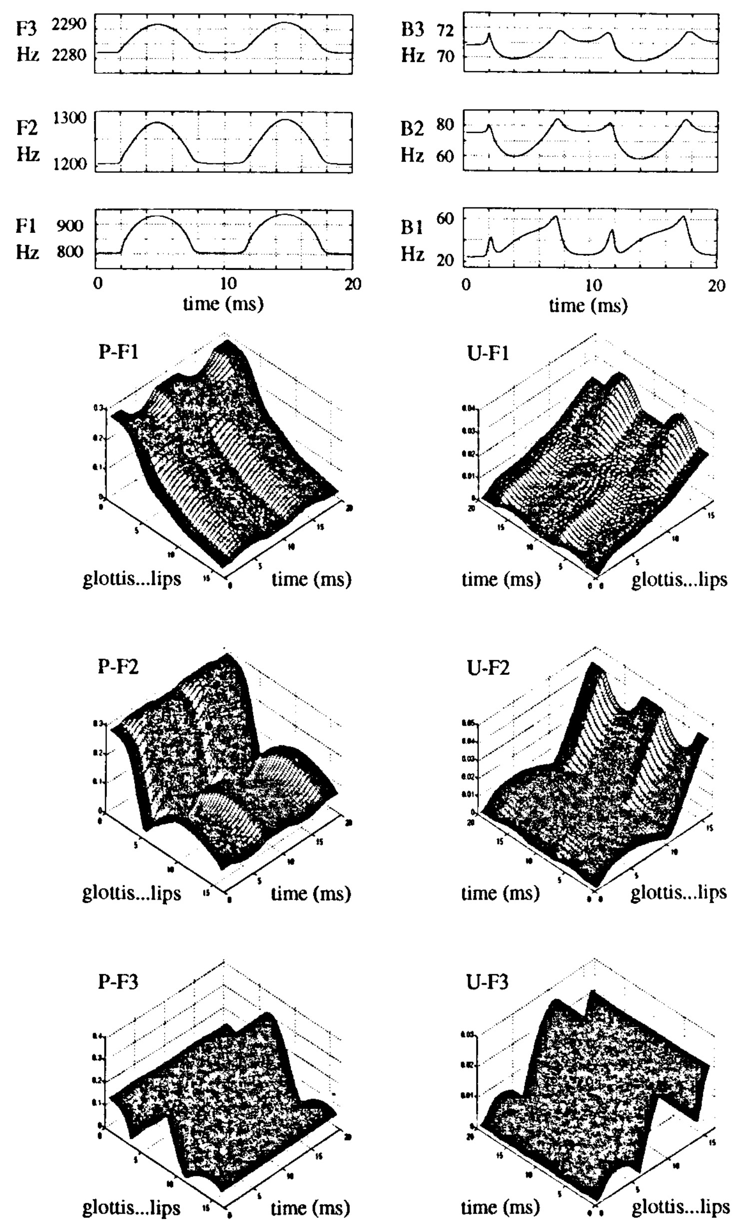


Figure 1 : Formant frequencies, bandwidths, and pressure/volume-velocity distributions for /a/.