Proc. of ESCA-EUROSPEECH'95, Madrid, Spain.

Numerical simulations of fluid flow in the vocal tract

G. Richard,† M.Liu,† D. Sinder,† H. Duncan,† Q. Lin,† J. Flanagan,†
S. Levinson,‡ D. Davis,§ S. Slimon,§

†Rutgers Univ. - CAIP Center, Piscataway, NJ 08855, USA.
‡AT&T Bell Laboratories, Murray Hill, NJ 07974, USA.
§General Dynamics, Groton, CT 06340, USA.

Abstract

An alternate approach to speech synthesis based on numerical solution of Navier-Stokes (NS) and Reynolds-Averaged-Navier-Stokes (RANS) equations is described. Unlike the traditional methods based on linear acoustic theory, the NS and RANS formulations are not limited by the assumptions of linearity, negligible viscous effects, and plane wave propagation. In the present formulation, the Navier-Stokes equations are discretized and solved using a finite difference method. Initial applications involve 2-D simulations of flow through ideal channels (straight or curved tubes). In another application, the formulation is applied to the geometry of the three cardinal vowels. Synthetic speech sounds of encouraging quality are obtained for the three vowels.

1 Introduction

Traditional methods for speech synthesis and generation have been based on solutions of a form of the wave equation called the Webster equation. While such approaches (see [4], [1], [2]) have achieved a remarkable level of success, there are still noticeable limitations in the speech produced by these techniques. These limitations must, in part, be attributed to the assumptions of linearity, negligible viscous effects and plane wave propagation made in deriving and applying the Webster equation. It is likely that these assumptions have neglected features of the flow field, and their corresponding influence on the speech signal that are important in synthesizing natural sounding speech.

In this paper, an alternate approach is described. It is based on numerical solution of the Navier-Stokes (NS) equations for laminar flows and the Reynolds-Averaged-Navier-Stokes (RANS) equations for turbulent flows. This approach attemps to couple physically realistic expressions of the phenomena involved (e.g. the Navier-Stokes equations) with realistic descriptions of the vocal tract geometry to provide a "first principles" solution to the speech synthesis problem. This approach is particularly attractive since it does not make the assumptions of linear acoustic theory. Some previous efforts in this direction include [3] and [5] but were limited by the extreme computational requirement for solving the time-dependent Navier-Stokes equations for realistic geometric configurations.

The expected results are high quality synthesis and a new parameterization of speech for applications in automatic recognition and low bit-rate coding resulting from a parsimonious modeling of articulatory shapes and dynamics. To reach this goal, we are aware that critical geometric features such as the curvature of the vocal tract or the local (and one-sided) effect of the different articulators clearly have significant impact on the flow structure in the human vocal tract. However, as a preliminary stage and to gain further insight into the use of Navier-Stokes techniques for such an application, the fluid flow simulations are performed in idealized channels and stylized shapes that represent gross features of the vocal tract for vowels.

The paper is organized as follows: section 2 is devoted to the governing equations and to the flow solver description. Section 3 gives some numerical results on straight tubes, curved tubes and finally on stylized vocal tract shapes. Section 4 proposes some conclusions.

2 Governing equations and flow solver description

The proposed model is based on the fundamental Navier-Stokes (NS) equations which provide a realistic description of fluid motion. The first equation describes the conservation of mass principle. This principle merely states that the mass of a system of fixed identity is constant. The second equation, the momentum conservation equation, is based on Newton's second law which states that the vector sum of all external forces acting on a fluid mass equals the time rate of change of the linear momentum vector of the fluid mass.

The work described in this paper is based on a particular form of the Navier-Stokes (NS) equations

This research is jointly supported by NSF and ARPA under contracts NSF/ARPA IRI-9314946 and ARPA DAST 69-93-C-0064.

called the Reynolds-Averaged-Navier-Stokes (RANS) equations which are given by:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}(\rho u_i) = 0 \qquad (1)$$

$$\frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}(\rho u_i u_j) = -\frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} \qquad (2)$$

with

$$\tau_{ij} = \left[\mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)\right] - \overline{u_i^{'} u_j^{'}}$$

where ρ, u_i, p, μ and $(\overline{u_i^{'} u_j^{'}})$ respectively correspond to density, mean velocity components, mean pressure, viscosity and Reynolds stresses. The Reynolds stresses which account for the effects of the turbulence on the mean flow, result from the time-averaging process which is carried out to render the NS equations computationally tractable.

The solver used to resolve numerically the RANS equations is NFC2D/3D (Natural and Forced Convection in 2 Dimensions and 3 Dimensions). It is based on finite difference discretization and utilizes a generalized multiblock approach for subdividing the flow domain. The multiblock capability accomodates both continuous and discontinuous grid connections at block interfaces to provide a high degree of flexibility in modelling complex domains. This solver can handle steady and time-dependent solutions for both incompressible and slightly compressible flows. The slightly compressible formulation is intended for use in flows where the Mach number[1] is between 0.001 and 0.1 (which correspond to typical values in the human vocal tract). For turbulent flows, a variety of different eddy viscosity-based turbulence models (including algebraic, one-equation and two equation models) are included.

To confirm the results obtained with NFC2D/3D, an ancillary solver, Fluent, has also been used. This solver uses a control-volume based method to resolve the NS equations.

3 Numerical simulations

3.1 Fundamental characterizations

Although our objective is to synthesize speech using an NS approach, it is necessary to investigate some key computational issues using extremely simple geometries. These issues include:

- the applicability of the slightly compressible RANS formulation to very low Mach number flows (typical values in the vocal tract are in the range [0.1,0.001]);
- the variability of computational grid spacings and time steps to adequately resolve relevant features of both the flow field and the acoustic field;
- the comparison between linear acoustic results and fluid flow simulation results to evaluate the numerical accuracy of the computational algorithms being used.

[1] $M = \frac{u_i}{c}$ where c is the speed of sound

Initial studies involve simulations of 2D flows in a straight circular tube with rigid walls and dimensions representative of those in the vocal tract of an adult male (*17 cm* long with a radius of *1.25 cm*). The inflow is represented by a step function in axial velocity at the inlet and a constant zero pressure at the outlet (i.e free space condition). Furthermore, a zero velocity at walls is assumed (see Figure 1).

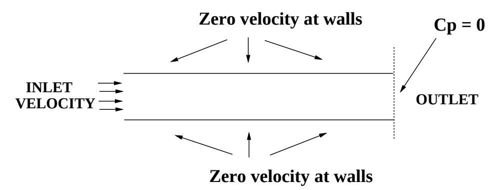

Figure 1: *Boundary conditions.*

3.1.1 Slightly compressible RANS formulation

To demonstrate the applicability of the slightly compressible RANS formulation to speech synthesis, computations using fully compressible form of NS equations are performed at Mach numbers ranging from 0.001 to 0.5. For very low Mach numbers, the compressibility effects are negligible. For Mach numbers greater than approximately 0.2 (well beyond the range of Mach number typically seen in speech), compressibility effects become more important as seen in the Mach=0.5 result (see Figure 2). The solutions obtained were numerically stable (i.e no spurious oscillations). This is a significant result since there were initial concerns relative to stability of the slightly compressible RANS formulation at very low Mach numbers.

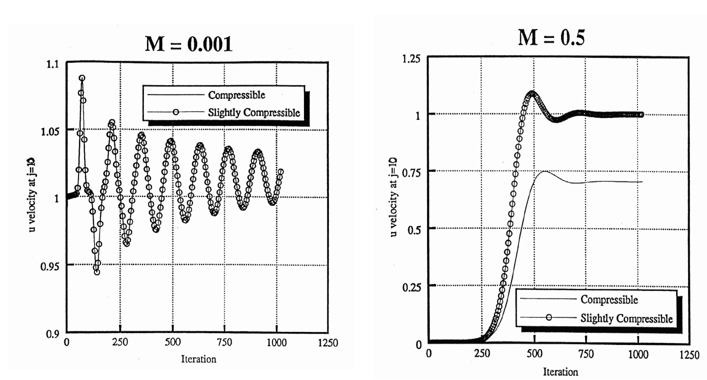

Figure 2: *Effect of medium compressibility as a function of the Mach number.*

3.1.2 Grid and time step

In all of the initial calculations, the time step is set using an inviscid CFL[2] condition. A computational study is performed to investigate the sensitivity of the solutions to the time step. Figure 3 shows the results for a parametric set of simulations in which the time step and axial grid spacing (maintaining a CFL number of 1) are sequentially halved. As seen in this figure, a time step independent solution is never obtained. In these simulations, a first order discretization in time is used. Since these results indicate that extremely small time steps are required to achieve time step independent solutions, a supplemental parametric study is conducted using a secon order accurate time integration scheme.

For solutions of second order scheme, a measure of first order damping is added to control dispersion error present in true second order scheme. The second order time integration scheme with a damping factor of 0.4 (values of 0.5 and 0 respectively correspond to a true second order scheme and a true first order scheme) shows a better accuracy than the first order scheme for the same number of grid nodes.

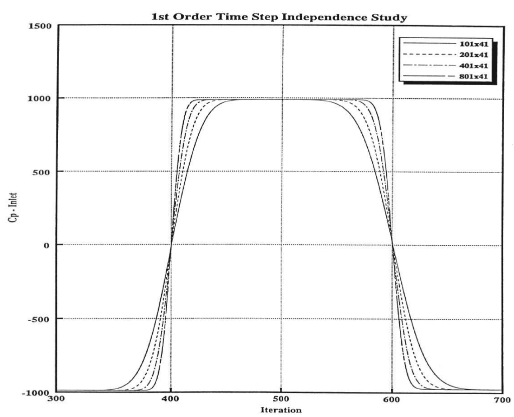

Figure 3: *First order time step independence study of the effect of grid size in the axial direction on the inlet pressure wave form.*

3.1.3 Accuracy of the solutions

To obtain preliminary feedback on the general level of accuracy being provided by the flow-based solutions, a long run time simulation using the second order scheme (with low first order damping) is performed on a straight channel. For such a channel (*17 cm* long with a radius of *1.25 cm*), the linear acoustic theory (in loss-less conditions) predicts the first resonances to be located at *500,1500,2500,3500,4500 Hz*. Figure 4 displays the outlet pressure spectrum obtained using an NS approach and shows the good agreement between the flow-based solutions and the predicted resonances.

[2] CFL $= \frac{c\Delta t}{\Delta x}$, where Δt is the time step, Δx is the axial grid spacing, and c is the speed of sound