Flow Visualization in Stylized Vocal Tracts

D. Sinder, G. Richard, H. Duncan, J. Flanagan — *CAIP Center, Rutgers University*
Piscataway, NJ 08855, USA
M. Krane, S. Levinson — *Lucent Technologies Bell Laboratories*
Murray Hill, NJ 07974,USA
S. Slimon, D. Davis — *Electric Boat Corporation*
Groton, CT 06340, USA

Abstract

An approach to speech synthesis based on computational fluid dynamics is being used to simulate and visualize air flow patterns in stylized vocal tract configurations. While traditional methods for speech synthesis have had remarkable success using linear acoustic models, such models do not generate flow induced effects such as noise generation (as in the /s/ in *sea*). The fluid-based approach inherently encorporates non-linear, non-plane wave, and viscous effects and may provide insight into three-dimensional effects as well as flow-induced phenomena. A two-pronged approach is taken comprising both numerical and physical simulation of flow. The numerical simulation involves solving a slightly compressible form of the Reynolds-Averaged Navier-Stokes (RANS) equations which describe fluid motion. Flow quantities, such as pressure and velocity, are computed on a body-fitted grid as they evolve with time in the vocal tract geometry. Concurrent measurements from acoustic simulation on a physical model of the same geometry allows direct measurement of real flow quantities. While numerical results generate flow visualization and sound synthesis, physical measurements provide validation of the numerical results through both quantitative comparisons and qualitative listening tests.

1 Introduction

The objective of the research described in this paper is to obtain a fundamental understanding of speech generation by using a computational model based on first principles of fluid dynamics. This new approach to speech synthesis is multi-disciplinary, requiring expertise in computational fluid dynamics, aeroacoustics, and signal processing. It uses numerical simulation to visualize flow quantities such as pressure and velocity as they evolve with time in a vocal tract model. An improved understanding of the speech production process is expected to lead to a more compact parameterization of the speech signal that will simultaneously advance automatic speech recognition, low bit rate coding, and speech synthesis. These advances will aid development of natural language interfaces for computers which will become increasingly important as computers and their capabilities grow in sophistication.

Traditional methods for speech synthesis have had remarkable success using a linearized form of the wave equation called the Webster equation ([2, 4, 6]). This equation models wave propagation in the vocal tract as a one-dimensional plane wave propagation in an inviscid fluid. While this model is quite accurate for voiced vowel sounds, other classes of speech sounds involve three-dimensional viscous effects where the convective flow and its interaction with the acoustic field generates the dominant sounds. Traditional synthesizers generate these sounds in a manner which is non-physical. Namely, a noise source is modulated by the Reynolds number[1] computed in each section of the vocal tract ([3]).

A new approach which simultaneously computes both the convective and acoustic fields involves numerical solution of the Reynolds-Averaged Navier-Stokes (RANS) equations – a form of the equations describing fluid motion. The RANS equations are derived from first principles of physics: conservation of mass and conservation of momentum (Newton's second law). Thus, the approach is not limited by assumptions of plane-wave propagation and negligible viscous effects. Numerical results are supplemented and validated by physical aeroacoustic simulation in the same geometries since the RANS method contains far fewer approximations than other computational models of speech production. Some previous efforts using a fluid flow approach to speech synthesis include [5] and [10] but were limited by the extreme computational requirement for solving time-dependent Navier-Stokes equations. Recent advances in high-performance computing now makes solution of this problem within reach.

A number of early validation experiments were performed to verify viability of the fluid dynamic approach and the RANS solution methods. These results can be found in [8]. This paper presents a summary of more recent results. Section 2 describes the numerical approach used for the simulations and methods for validating results. Section 3 outlines experiments on stylized vowel geometries. Section 4 discusses initial results on stylized fricative geometries. Finally, Section 5 discusses future objectives of this research.

2 Numerical Approach

The software package used to numerically solve the RANS equations is NFC (for Natural and Forced Convection). This package, developed at Electric Boat Corporation, is a multi-block, finite difference based solver which is second order accurate in both space and time.

2.1 Governing Equations

NFC solves a particular form of the RANS equations which is given by (in tensor notation):

$$M^2 \frac{\partial p}{\partial t} + \frac{\partial u_i}{\partial x_i} = 0$$

$$\frac{\partial u_i}{\partial t} + \frac{\partial u_i u_j}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j}\left[\frac{(\nu + \nu_t)}{Re}\frac{\partial u_i}{\partial x_j}\right]$$

where M is the Mach number[2], u_i are the mean velocity components, p is the static pressure, ν is the kinematic viscosity, and ν_t is the turbulent eddy viscosity coefficient.

In these equations, the flow is assumed to be slightly compressible and an isentropic assumption has been used to relate pressure and density. These equations are applicable to time-dependent, turbulent, low Mach number flows, such as those which exist in the human vocal tract. NFC also contains a wide variety of user selected eddy viscosity based turbulence models. These models range from simple algebraic models to more complex two-equation models. All of these turbulence models are low Reynolds number models allowing for integration of the governing equations down to solid walls.

2.2 Boundary Conditions

For numerical solution of the RANS equations, conditions must be specified at the boundaries of the computational geometry. Physiologically, the vocal tract is a curved, non-symmetric tube with moving, compliant walls. Data from MRI is available for obtaining three-dimensional representations of the vocal tract (see, for example, [1]). However, such complex geometries require even more complex computational grids– possibly resulting in unstable computational

---

[1] In the linear acoustic model, the Reynolds number, Re, is used as a measure of turbulence intensity ($Re = \frac{UL}{\nu}$, where U and L respectively represent the characteristic flow velocity and length and ν represents the kinematic viscosity).

[2] Mach number is defined as the ratio of fluid flow velocityto the sound speed.

solutions. Thus, smooth, stylized, axisymmetric geometries are used until it is better understood which flow phenomena must be adquately resolved.

Although the effects of the moving, compliant walls can be significant, at the present stage of research, rigid walls are assumed, and a no-slip (zero velocity) condition is specified at the wall. The outlet (corresponding to the lips) is set to a fixed pressure (pressure relief). This is only accurate for small mouth openings, however, acoustic radiation models are difficult to apply in the RANS formulation because they equally effect the convective flow.

At the inlet, which corresponds to the vocal cord opening, a step of low velocity can used to obtain the acoustic response of the geometry. For voiced vowel sounds, a velocity pulse-train is specified at the inlet, corresponding to pulses of air from the lungs emitted as the vocal cords vibrate). For fricatives in speech, a quasi-steady air flow from the lungs is forced through a tight constriction in the vocal tract. Therefore for fricative geometries, a step of high velocity is specified at the inlet.

2.3 Validating Numerical Results

Since a primary objective of this research is to obtain a fundamental understanding of speech production, verifying that numerical results represent physical flow behavior is a major component of the work. To that end, several forms of comparison are drawn. For vowel geometries, linear acoustics is capable of high quality simulation since the plane-wave approximation in these cases is accurate. Therefore, results from RANS simulations can be compared with linear acoustics. However, for fricative geometries, the RANS solution involves sound generated by fluid motion which is three-dimensional, viscous, and rotational. In these cases, linear acoustics does not provide a useful standard for comparison. For this reason, numerical simulations are coordinated with physical acoustic experiments to measure the flow in the same geometry. Finally, for all geometries, qualitative listening tests are regularly performed. This is an important aspect for developing parameterizations of the speech signal which can be used to produce speech synthesis with a high degree of human naturalness.

3 Vowel Geometries

Initial investigations using the RANS approach for speech synthesis concentrated on vowel geometries since results could be compared with linear acoustics which provides an accurate solution for these shapes. These experiments produced vowel sounds of encouraging quality which compared favorable with linear acoustics ([8]).

More recent results using a high velocity pulse train at the inlet (peak velocity = 20 m/s) can be used to visualize flow patterns for voiced vowels. Figure 1 shows the velocity magnitude from this case. The pulses of high velocity can be seen convecting down the length of the vocal tract. Qualitative listening tests are performed by collecting pressure data from the outlet of the geometry and converting it from digital to analog. The spectrum of this signal is shown in Figure 2.

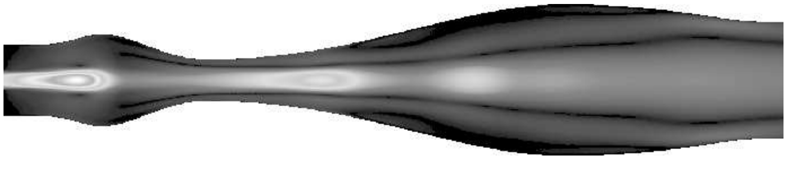


Figure 1: *Snapshot of velocity magnitude distribution in the shape /a/ (as in* hot*) when excited by a high velocity pulse train. Flow is from vocal cord opening at the left to lip opening at the right.*

Additional experiments have been performed whereby unvoiced vowels (*i.e.,* whispered speech) are synthesized by specifying a random velocity at the inlet. These experiments are discussed in [9].

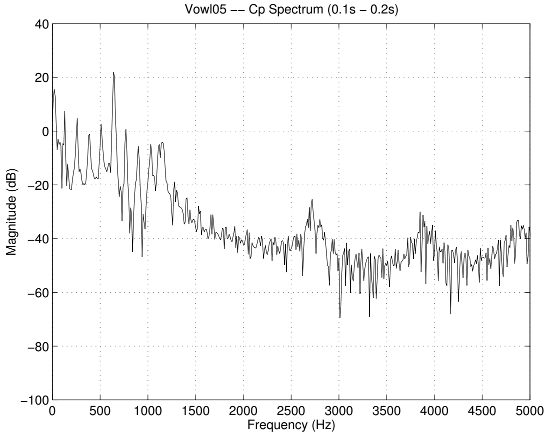


Figure 2: *Spectrum of voiced vowel sound corresponding to a stylized shape for the sound /a/ as in* hot*.*

4 Stylized Fricative Geometries

In the case of fricative geometries, the geometry is typically more complex than for vowel cases. A tight constriction is formed in the vocal tract (*e.g.,* between the tongue and the palate), and air from the lungs is forced through the constriction forming a jet. The jet may impinge on an obstacle further down the vocal tract (*e.g., the teeth*). In order to study these sounds from a fluid-dynamic perspective a triplet of stylized geometries were designed. Each is based on a straight tube where a jet is formed at the inlet. In two of the cases , a gaussian-shaped obstacle (one “mild” constriction and one “tight” constriction) is placed 10cm downstream of the inlet. For the third case, no obstacle is present. A schematic diagram illustrating the three configurations is shown in Figure 3.

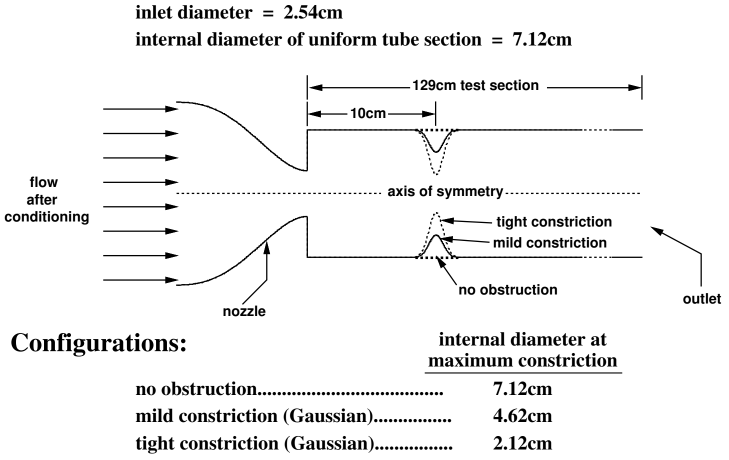


Figure 3: *Schematic diagram of the experimental apparatus used for physical acoustic simulation.*

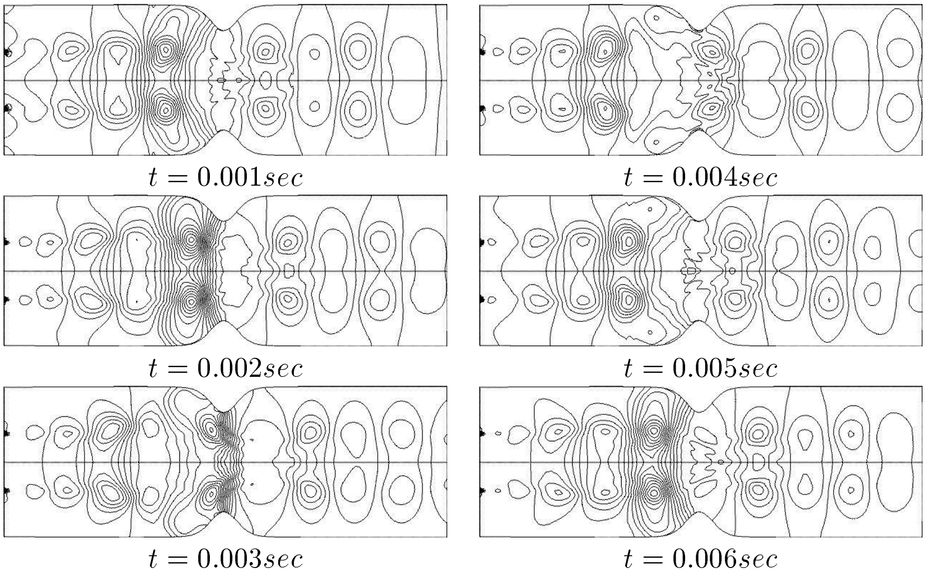
4.1 Physical Aeroacoustic Simulation

Since the convective field is critical for producing the noise-like sound, it is crucial that the RANS solution is verified as representing a physical flow and that components of the flow field which generate sound are resolved. Therefore, numerical simulations on fricative geometries are coordinated with physical aeroacoustic simulation. These physical simulations are performed in the Acoustics and Audio Communications Department of Bell Laboratories.

In the experimental set-up flow from a room air supply (100 psi) enters a muffler which is followed by a flow conditioning section consisting of a honeycomb and three fine-mesh screens. The flow then passes through a 16:1 area ratio nozzle into the main test section in one of the three configurations described above and shown schematically in Figure 3. The apparatus is composed of several sections which allow for the obstruction sections to be interchanged. Additionally, the sections are machined with access holes for hot-wire probes and pressure transducers for recording velocity (three components) and mean and unsteady pressures. Microphones near the outlet, both interior and exterior to the tube, record outlet sound pressure. Initial results comparing numerical and physical experiments have been obtained and are discussed in [9].

4.2 Numerical Simulation

Numerical simulation on each of the three geometries have been performed with the assumption of laminar flow. In each case, results show the flow to be unsteady even though the inlet velocity is constant. Figure 4 shows a time-history of pressure contours for the mild constriction case. The circular regions are vortex rings shed from the inlet which convect down the length of the tube with the unsteady jet formed in the center. The velocity magnitude distribution shown in Figure 5 help illustrate the unsteady jet.


$t = 0.001 sec$ $t = 0.004 sec$

$t = 0.002 sec$ $t = 0.005 sec$

$t = 0.003 sec$ $t = 0.006 sec$

Figure 4: *Time history of pressure contours for the mild constriction case (laminar flow). Flow is from left to right.*

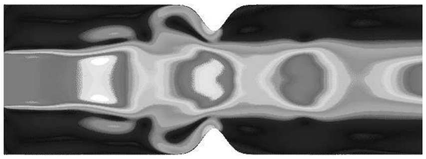


Figure 5: *Snapshot of the velocity magnitude distribution in the mild constriction for laminar flow.*

When a turbulence model (Briley-McDonald-Fish, see [7]) is encorporated into the computations, the large vortical structures are not evident. Figure 6a shows a single frame of pressure contours for the mild constriction case computed with a turbulence model. In this case the quasi-circular region just upstream of the constriction is a circulation zone which does not convect with the jet. The velocity magnitude distribution shown in Figure 6b helps to illustrate that the jet is nearly steady.

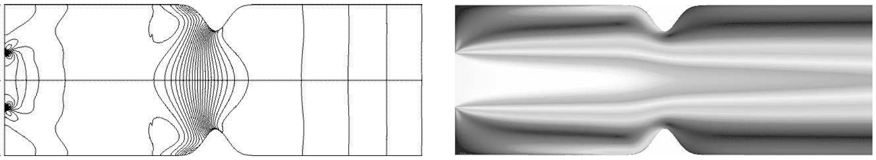

(a) (b)

Figure 6: *Snapshot of pressure contours (a) and velocity magnitude distribution (b) in the mild constriction computed with a turbulence model.*

The pupose of the turbulence model is approximate the effect of sub-grid-scale fluid motions which are not resolved by the grid. When the model is used, the energy in the grid-scale vortical structures is quickly distributed more evenly through out the flow. One question which remains to be answered is whether turbulence models will be critical to capturing the flow induced sound that is of primary interest in these studies. Future plans include acoustic comparisons of laminar and turbulent simulations.

5 Conclusions

Numerical simulation of fluid flow in the vocal tract may prove to be a powerful tool towards developing an improved understanding of human speech production. In particular, sounds in which flow-induced noise contributes significantly to the acoustics are being studied using this method. Future investigations will include investigations of various turbulence models and, eventually, full three-dimensional simulations.

Acknowledgment

This research is jointly supported by the National Science Foundation (NSF) and the Advanced Research Project Agency (ARPA) under contracts IRI-931496 and ARPA DAST 69-93-C-0064.

References

[1] T. Baer, J.C. Gore, L.C. Gracco, and P.W. Nye. Analysis of vocal tract shape and dimension using magnetic resonance imaging: Vowels. *J. Acoust. Soc. Am.*, 90(2):799–828, 1991.

[2] C.H. Coker, N. Umeda, and C.P. Browman. Automatic synthesis from ordinary english text. *IEEE Trans. Audia Electroacoust.*, AU-21:293–298, 1973.

[3] James L. Flanagan. *Speech Analysis Synthesis and Perception.* Springer-Verlag, New York, second edition, 1972.

[4] J.L. Flanagan, Ishizaka K., and Shipley K.L. Signal models for low bit-rate coding of speech. *J. Acoust. Soc. Am.*, 68(3):780–791, 1980.

[5] G.C. Hegerl and H. Hoge. Numerical simulation of the glottal flow by a model based on compressible navier-stokes equations. In *Proc. ICASSP*, pages 477–480, 1991.

[6] Q. Lin. *Speech Production Theory and Articulatory Speech Synthesis.* PhD thesis, Royal Inst. of Technology, Stockholm, Sweden, 1990.

[7] H. McDonald and R. W. Fish. Practical calculations of transitional boundary layers. *Journal of Heat and Mass Transfer*, 16:1729–1744, 1973.

[8] G. Richard, M. Liu, D. Sinder, H. Duncan, Q. Lin, J. Flanagan, S. Levinson, D. Davis, and S. Slimon. Numerical simulations of fluid flow in the vocal tract. In *Proceedings of EUROSPEECH95*, pages 1297–1300, Madrid, Spain, 1995.

[9] D. Sinder, G. Richard, H. Duncan, Q. Lin, J. Flanagan, S. Levinson, D. Davis, and S. Slimon. A fluid flow approach to speech generation. In *Proceedings of the First ESCA Tutorial and Research Workshop on Speech Production Modeling: From control strategies to Acoustics.* European Speech Communication Association, May 1996.

[10] T.J. Thomas. A finite element model of fluid flow in the vocal tract. *Computer Speech and Language*, 1:131–151, 1986.