First ETRW on
Speech Production Modeling
Autrans, France
May 20-24, 1996

A Fluid Flow Approach to Speech Generation

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

[†]Rutgers University- CAIP Center, Piscataway, NJ 08855-1390
[‡]Bell Laboratories, Murray Hill, NJ 07974
[§]Electric Boat Corporation, Groton, CT 06340

Abstract

A fluid dynamic formulation of speech generation may lead to an improved understanding of the physics of speech production. Unlike more traditional linear acoustic methods of speech synthesis, this alternate approach aims to capture more of the relevant physics by numerically solving a form of the Reynolds-Averaged Navier-Stokes equations describing fluid motion. Though computationally intensive, the method is not limited by assumptions of linearity and plane wave propagation inherent in linear acoustic analysis. Numerical simulations of flows in stylized vocal tract shapes, as well as measurements on physical flows are described. Special attention is given to fricative generation, since the physiological understanding, and subsequent synthesis, of these sounds stands to gain the most from this approach.

1 Introduction

As computers become pervasive in our society, the need for more natural human-computer interfaces is magnified. This need demands improvements in speech synthesis, recognition, and coding. To that end, a better understanding of the physics of speech production will be essential. The research described herein seeks to acquire such insight.

The approach aims to simulate the process of speech generation using computational fluid dynamics (CFD). The equations of fluid motion are put in a form of the Reynolds-Averaged Navier-Stokes (RANS) equations. Numerical solution of the RANS equations allows for a first principles analysis of speech production.

Unlike traditional methods of synthesis which have been successful using a linearized form of the wave equation called the Webster equation ([1],[2],[3]), the RANS solution is not limited by assumptions of linearity and plane wave propagation. In order to verify the viability of the fluid dynamic approach and the methods used in numerically solving the RANS equations, a number of early validation experiments were performed ([4]). These included investigations of the suitability of a slightly compressible formulation at low Mach numbers[1], comparison with linear acoustics for simple geometries (*e.g.*, a straight tube), and grid and time-step independence trials.

The applicability of this approach has been demonstrated by synthesizing vowels of encouraging quality. For vowel generation using stylized geometries, the CFD approach compares favorably with linear acoustic synthesis ([4]) which is capable of accurately producing these sounds. However, linear acoustic synthesis generates fricative sounds in a manner which is non-physical. Namely, a noise source is modulated by the Reynolds number[2] computed in each section of the vocal tract ([5]). One goal of this research is to develop improved models for fricative generation.

This paper presents ongoing research using this alternate approach, focusing on fricative generation for which understanding is still rather limited. Section 2 describes the numerical approach, section 3 discusses the issue of boundary conditions and section 4 provides initial results on fricative generation using a combined experimental and computational approach. Finally, section 5 proposes some conclusions and discusses future objectives.

2 Numerical Approach

The software package used to numerically solve the RANS equations is NFC (for Natural and Forced Convection) which was developed at Electric Boat Corporation. NFC is a multi-block finite difference based solver which is second order accurate in both space and time. It is based on 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, u_i are the mean velocity components, p is the static pressure, ν is the kinematic

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

[1] Mach number is defined as the ratio of fluid flow velocity to the sound speed.

[2] 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).

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. However, research to date has used only laminar representation as an initial step towards the eventual inclusion of turbulence.

3 Boundary Conditions

3.1 Description

In order to solve the RANS equations, conditions must be specified on the boundaries of the geometry being studied. Specifying these conditions involves imposing certain values or relationships on flow variables. Inappropriate boundary conditions may result in nonphysical flows. Therefore, boundary condition specification plays an integral part in obtaining accurate results. Figure 1 shows the boundaries on which boundary conditions must be specified for a typical vocal tract flow computation.

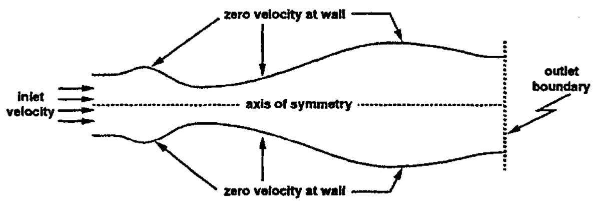

Figure 1: *Boundary conditions.*

The inlet (glottis end) is handled by specifying the velocity of the flow. This may be a pulse train (for voiced sounds), white noise (for whispered sounds), or a step function (to compute a transfer function or generate fricatives). At the outlet (mouth end), the boundary condition is selected to model some variation on a radiation condition. Thus, a pressure relief condition may be imposed for no radiation, a radiation impedance may be specified for partial radiation, or an anechoic termination may be used for a completely non-reflective boundary.

Simulations may be run axisymmetrically, in which case a symmetry boundary is specified. Alternatively, simulations may be run fully three dimensional. After specifying the inlet and outlet (and possibly symmetry) boundaries, the remaining boundary is defined as a rigid wall. The velocity along the wall is set to zero to satisfy the no-slip condition.

3.2 Outlet Boundary Condition Improvements

Previous work on synthesis of vowel sounds used a pressure relief condition for which pressure was fixed at the outlet boundary ([4]). However, for shapes with large mouth openings, radiation impedance is small and the pressure relief is inappropriate. Thus, a radiation condition has been implemented which models the mouth opening as a piston in an infinite plane baffle ([3],[5]). The result is an impedance with both resistive and reactive components.

Figure 2 compares the spectra resulting from simulation of a straight tube with the impedance boundary condition to a simulation with the pressure relief condition. As is desired, the impedance condition has the effect of decreasing formant frequencies and increasing bandwidths, especially at high frequencies.

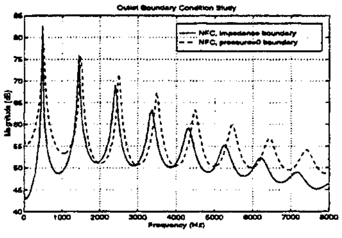

Figure 2: *Comparison of spectra using impedance boundary condition and pressure relief at outlet.*

In the case of fricative generation, it is possible that the pressure relief and impedance conditions may cause non-physical reflections at the outlet. Therefore, future efforts will include the addition of a completely non-reflective outlet condition to aid in studying these effects.

4 Towards Fricative Generation

4.1 Unvoiced Vowels

Although one of our objectives is to study flow induced noise in the vocal tract (such as unvoiced fricatives), it was decided to perform a preliminary numerical experiment to verify that known random quantities will be accurately propagated (i.e with little numerical dissipation). Using an approach similar to that used for noise generation in linear acoustics, whispered vowels have been produced using the RANS solver by specifying the inlet velocity as white noise. Figure 3 shows the spectrum of whispered speech generated using this method for the vowel /a/ (as in *hot*). The peaks of this spectrum correspond to the natural resonances of the simulated geometry. In addition, the effects of the random inlet velocity are seen over the entire outlet pressure spectrum. This is significant since there were initial concerns that the numerical scheme used in NFC would dissipate these relatively small signals over the length of the vocal tract.

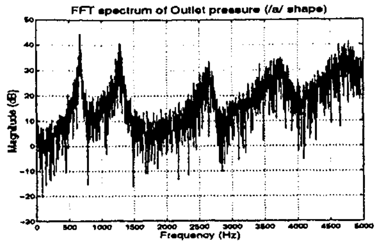

Figure 3: *Spectrum of whispered speech from stylized /a/ geometry generated using velocity noise at inlet.*

4.2 Stylized Fricative Topologies

For fricative sounds, comparison with linear acoustics provides little insight since it does not generate flow induced noises. Therefore, a set of experiments has been designed to allow for computational to experimental comparisons. This set is meant to evaluate the ability of a RANS formulation to compute the sound produced by the flow.

4.2.1 Experimental Apparatus

The experimental investigation is being performed in the Acoustics and Audio Communications Department of Bell Laboratories. The experimental apparatus shown on Figure 4 consists of a nozzle followed by a 129cm long circular duct. Three obstruction geometries are being evaluated - no obstruction, a mild obstruction and a sharp obstruction. The obstructions are located 10cm from the inlet to the circular duct and have an axisymmetric Gaussian thickness distribution. The ratio of the area of maximum constriction to that of the circular duct is 0.42 for the mild obstruction and 0.09 for the sharp obstruction.

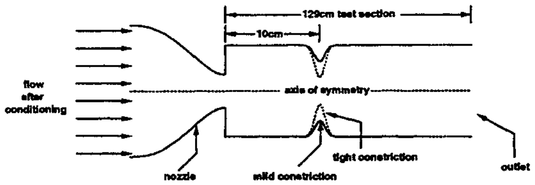

Figure 4: *Schematic diagram of the experimental apparatus.*

The three obstruction geometries were selected since the flow structure and resulting acoustic spectra were expected to be significantly different for each case. In addition, the mild and sharp obstruction cases are stylized representations of a dental fricative topology (i.e., an obstruction representative of the teeth is located downstream of a narrow circular inlet representative of the tongue-palate constriction). The flow passage is scaled up from the dimensions of the vocal tract in order to have sufficient spatial resolution in both wall pressure and velocity measurements. A more detailed description of the experimental set-up and instrumentation may be found in [6].

4.2.2 Initial Results: Mild Constriction

Initial results have been obtained for the mild constriction case from both experimental and numerical methods (results for the sharp constriction case may be found in [6]). The numerical simulations performed to date have been carried out in an axisymmetric geometry without the use of a model for subgrid motions (turbulence model). These assumptions are somewhat inconsistent with the expected physical flow (e.g. the flow is expected to be turbulent and asymmetric). However, these assumptions provide for a baseline analysis that allows comparison with follow-on simulations that will account for flow asymmetry, and turbulence.

The response to an acoustic input was measured by exciting the tube with a chirp signal using a small loudspeaker mounted at the pipe inlet. This no-flow condition is compared with the solution computed from the Webster equation[3] on Figure 5. The spectral resonances match well for this case, as expected. With flow excitation, however, additional spectral characteristics appear. Figure 6 shows the spectrum resulting from a 20m/s inlet flow excitation along with the no-flow spectrum. An additional peak is seen between the second and third harmonics of the fundamental, as well as between the third and fourth harmonics.

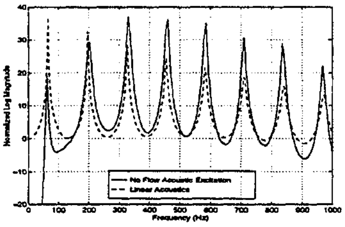

Figure 5: *Spectra from acoustically-excited (no-flow) physical apparatus and numerical linear acoustic solution.*

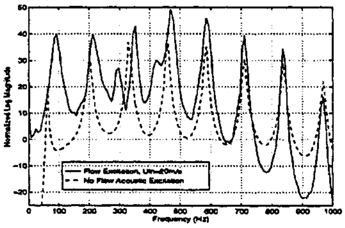

Figure 6: *Spectrum from 20m/s flow excitation compared with no-flow spectrum.*

Results from the RANS solver show the flow to be unsteady with vortices shed from the inlet impinging on the constriction as is seen in the time history of velocity magnitude shown in Figure 7. The computed outlet pressure has been Digital-to-Analog converted to an audio signal. This signal has a perceptual broadband and

[3] Linear acoustic results are obtained using TractTalk ([3]) and incorporate viscous loss, heat loss, and radiation impedance at the mouth.

tonal quality consistent with the experiment. The outlet pressure spectrum is shown in Figure 8 in comparison to the measured pressure spectra from the acoustic and 20m/s inlet flow excitations. This figure reveals that the RANS solution does not capture the low frequency formant shifts due to the flow excitation. This may be due to the differences in outlet conditions and inlet velocity profiles.

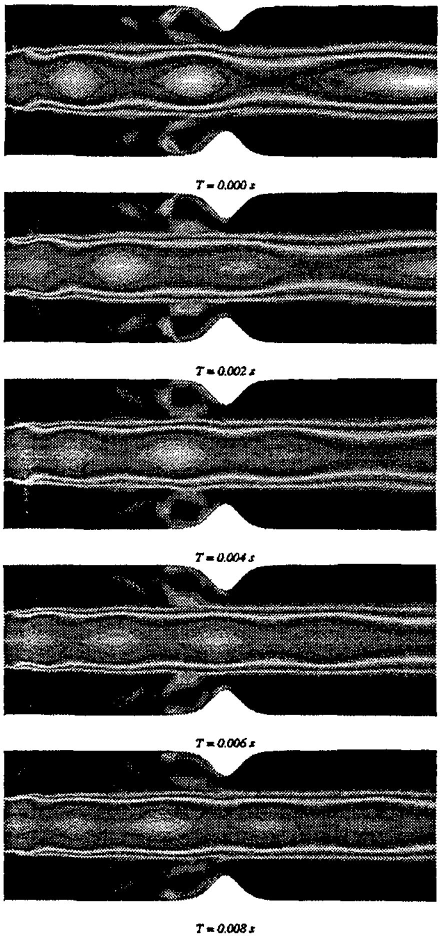

Figure 7: *Time history of velocity magnitude computed from RANS solver.*

5 Conclusions

A fluid dynamic formulation for speech synthesis provides a means to capture fluid flow behavior in vocal tract shapes. This information is of primary importance for synthesizing sounds for which flow induced acoustics play a large role– namely, unvoiced fricatives. Preliminary results have been obtained for the acoustic response of a stylized dental fricative due to flow using an experimental approach and a computational fluid dynamics approach.

Future work will include appropriate improvements to the numerical approach in order to better compare with measured data. Inlet and outlet conditions will be modified to create a better match with experimentally measured conditions. In addition, it is planned to use a non-reflective boundary condition in order to characterize the flow induced noise sources since in this case there will be no resonances in the facility. Comparisons will be made with the experiment being performed with an anechoic termination. Three dimensional simulations will also be performed to account for asymmetries in the flow. In addition, various turbulence models will be used.

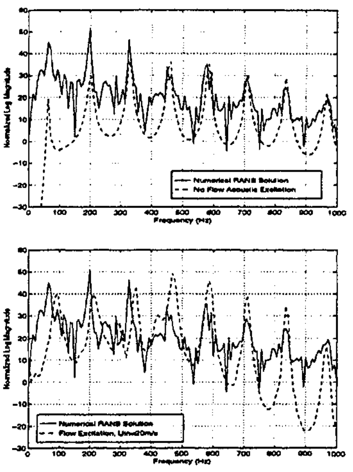

Figure 8: *Spectrum from numerical solution compared with (top) measured pressure spectrum of no-flow case (bottom) measured pressure spectrum in the 20 m/s inlet flow excitation case.*

References

[1] Coker, C.H., Umeda, N. & Browman, C.P. (1973). "Automatic Synthesis from Ordinary English Text," *IEEE Trans. Audio Electroacoust,* AU-21, pp. 293-298.

[2] Flanagan, J.L., Ishizaka, K., and Shipley, K.L., (1980). "Signal Models for Low Bit-rate Coding of Speech," *J. Acoust. Soc. Am.* 68(3), pp. 780-791

[3] Lin, Q. (1990). *Speech Production Theory and Articulatory Speech Synthesis,* Ph.D Dissert., Royal Inst. of Techn., Stockholm, Sweden.

[4] Richard, G., Liu, M., Sinder, D., Duncan, H., Lin, Q., Flanagan, J., Levinson, S., Davis, D., Slimon, S. (1995). "Numerical Simulations of Fluid Flow in the Vocal Tract," *Proc. of EUROSPEECH95,* Madrid, Spain, pp 1297-1300.

[5] Flanagan, J.L. (1972), *Speech Analysis, Synthesis, and Perception, 2nd ed.,* New York, Springer Verlag.

[6] Slimon, S., Davis, D., Levinson, S., Krane, M., Richard, G., Sinder, D., Duncan, H., Lin, Q., Flanagan, J. (1996). "Low Mach Number Flow Through a Constricted, Stylized Vocal Tract," *Proceedings of the 17th AIAA96 Aeroacoustics Conference,* State College, PA, May 6-8 1996.