

Applying Computational Fluid Dynamics to Speech

With a Focus on the Speech Sounds ‘Pa’ and ‘Sh’

by

Peter J Anderson

B.Sc., The University of Nevada Reno, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

The Faculty of Graduate Studies

(Mechanical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September, 2008

© Peter J Anderson 2008

Abstract

Computational Fluid Dynamics (CFD) are used to investigate two speech phenomena. The first phenomenon is the English bilabial plosive /pa/. Simulations are compared with microphone recordings and high speed video recordings to study the penetration rate and strength of the jet associated with the plosive /pa/. It is found that the dynamics in the first 10ms of the plosive are critical to penetration rate, and the static simulation was not able to capture this effect. However, the simulation is able to replicate the penetration rate after the initial 10ms.

The second speech phenomenon is the English fricative /sh/. Here, the goal is to simulate the sound created during /sh/ to understand the flow mechanisms involved with the creation of this sound and to investigate the simulation design required to predict the sound adequately. A variety of simulation methods are tested, and the results are compared with previously published experimental results. It is found that all Reynolds-Averaged Navier-Stokes (RANS) simulations give bad results, and 2D Large Eddy Simulations (LES) also have poor results. The 3D LES simulations show the most promise, but still do not produce a closely matching spectra. It is found that the acoustic analogy matches the direct measurements fairly well in 3D simulations.

The studies of /pa/ and /sh/ are compared and contrasted with each other. From the findings of the studies, and using theoretical considerations, arguments are made concerning which CFD methods are appropriate for speech research. The two studies are also considered for their direct applications to the field and future research directions which might be followed.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	vi
List of Figures	vii
List of Symbols and Abbreviations	ix
Acknowledgements	xi
Statement of Co-Authorship	xii
Authorship of Chapter 2	xii
Authorship of Chapter 3	xiii
1 Thesis Introduction	1
1.1 Introduction	1
1.2 Background	2
1.2.1 Fluid Mechanics	2
1.2.2 Computational Fluid Dynamics	6
1.2.3 Linguistics	11
1.3 Objectives and Hypothesis	15
1.4 Bibliography	17
2 Characteristics of Air Puffs Produced in English ‘pa’:	
Experiments and Simulations	20
2.1 Introduction	20

Table of Contents

2.1.1	Orifice	21
2.1.2	Flow description	22
2.1.3	Simulation Type	24
2.1.4	Hypotheses	24
2.2	Methods	25
2.2.1	Microphone Experiment	25
2.2.2	High Speed Video Experiments	28
2.2.3	Numerical Simulations	29
2.3	Results	31
2.3.1	Microphone Experiment	31
2.3.2	High Speed Video Experiments	33
2.3.3	Numerical Simulations	33
2.3.4	Comparison of Simulation to Microphone Experiment	36
2.3.5	Comparison of Simulation to High Speed Video Ex- periment	36
2.4	Discussion	40
2.4.1	Future Work	43
2.5	Conclusion	44
2.6	Acknowledgements	45
2.7	Bibliography	46
3	Computational Aeroacoustic Simulations of the English Frica- tive /sh/	49
3.1	Introduction	49
3.2	Methods	52
3.3	Results	56
3.4	Discussion	60
3.5	Conclusion	67
3.6	Acknowledgements	68
3.7	Bibliography	69
4	Thesis Conclusions	71
4.1	Introduction	71

Table of Contents

4.2	Comparison of Papers	71
4.3	Analysis of Chapter 2 and Chapter 3	74
4.3.1	Two-Dimensional Flows	74
4.3.2	RANS Simulations	75
4.3.3	LES Simulations	75
4.3.4	Boundary Layers	76
4.3.5	Acoustic Analogy	76
4.4	Usefulness to Current Research	76
4.5	Further Research	79
4.6	Conclusion	81
4.7	Thesis Contributions	81
4.8	Bibliography	82

List of Tables

2.1	Leading particle edge and lip opening regressions	38
2.2	Time alignment by distance for Figure 2.14	40

List of Figures

2.1	Starting Puff and Starting Jet Comparison	23
2.2	Microphone recordings of ‘pa’ with and withouth microphone pop	27
2.3	Filtering for microphone pop	27
2.4	Transient boundary condition	30
2.5	Experimental pressure fronts	32
2.6	Penetration distances for high-speed video experiments	33
2.7	Simulation validation	34
2.8	Simulation verification	35
2.9	Comparison of simulation particle fronts	35
2.10	Simulation and averaged experimental pressure fronts	36
2.11	Simulation and high speed video particle fronts	37
2.12	Velocity of simulation and high speed video particle fronts . .	38
2.13	Dependance of leading edge on lip opening width	39
2.14	High speed video and simulation comparison	41
2.15	Airflow velocity over time based on the distance from orifice aperture	42
3.1	The 2D domain	54
3.2	Spectrum processing	55
3.3	Determining a statistically steady simulation	56
3.4	A 2D flow snapshot	57
3.5	A 3D flow snapshot	57
3.6	2D simulations	58
3.7	3D simulations	59
3.8	Acoustic analogy results	60

List of Figures

3.9	Comparison of best results	61
3.10	Investigation of far field location	64
3.11	Comparison of LES with no turbulence model	65
3.12	Comparison of 2D simulation with spoken /f/	66

List of Symbols and Abbreviations

Fluid Mechanics Terms

NSE - Navier-Stokes Equations

u - velocity

p - pressure

ρ - density

μ - viscosity (also dynamic or absolute viscosity)

ν - kinematic viscosity, $\nu = \mu/\rho$

Re - Reynolds Number, $Re = \frac{\text{inertial forces}}{\text{viscous forces}} = \frac{\rho V D}{\mu} = \frac{V D}{\nu}$

Computational Fluid Dynamics terms

CFD - Computational Fluid Dynamics

CAA - Computational Aeroacoustics

RANS - Reynolds Averaged Navier Stokes

LES - Large Eddy Simulation

DNS - Direct Numerical Simulation

direct method - measuring sound by directly recording pressure fluctuations

AA - acoustic analogy

FFT - Fast Fourier Transform

Linguistics Terms

formant - a sound resonance

fricative - a consonant speech sound created when airflow is channeled into a turbulent jet

plosive - a consonant speech sound created by stopping then releasing air-

List of Symbols and Abbreviations

flow in the vocal tract

/ʃ/ - the linguistics symbol for ‘sh’ as in ‘shoe’; a fricative

/pa/ - the speech sound ‘pa’ as in ‘paper’; a plosive

Acknowledgements

I want to thank my supervisors prof. Sheldon Green and prof. Sidney Fels for supporting my research, and for providing great insights into the subject and how to approach it, while at the same time patiently letting me develop my own approach. It has also been a pleasure researching with Donald Derrick and prof. Bryan Gick.

I appreciate my friends (and particularly those in our office) who have helped to make this experience so positive.

Finally, thanks to my parents for their constant support and encouragement during this time.

Statement of Co-Authorship

This thesis consists of an introduction, two manuscripts with multiple authors, and a discussion.

Authorship of Chapter 2

The authors of chapter 2 are:

Donald Derrick, Peter Anderson, Bryan Gick, Sheldon Green

My contributions to this project are described below.

Identification and Design of Research Program

Gick was originally interested in studying the airflow of jets that escape from the mouth from a plosive. He also envisioned using CFD to model this. I discussed with Gick and Derrick how to design a simulation which is feasible yet realistic to the speech sound being modeled. Green directed the fluid mechanics of this project, suggesting the underlying theory, characteristics to be focused on, and how to study this experimentally and with simulations.

Performing the Research

My main role in this project was performing the simulations once we established the assumptions to be used and the output to be gathered. In the simulations, I designed and meshed the domain, I set the simulation parameters to be used in Fluent and designed the user-defined function to describe the inlet boundary condition, and ran the simulations until adequate data were gathered.

With direction from Green I studied the theory of jets, including the theory of starting jets and puffs which is described in the article.

Derrick and I performed the high speed video recordings together, including camera setup, experimental setup, and recording the data.

Data Analysis

I processed the simulation data such that it was useful for comparison with the experiments and theory. The high speed video experiments were processed by Derrick. As a group we discussed how to compare the results, the conclusions that can be drawn from the results, and how these findings should be presented in the manuscript.

Manuscript Preparation

The manuscript was largely prepared by Derrick, including the majority of the writing and all the final figures used. I wrote the sections concerning the numerical simulations: simulation theory, simulation methods, simulation results, simulation discussion and errors. I also wrote paragraphs concerning fluid mechanics theory and parts of the discussion. All group members read and revised the manuscript.

Authorship of Chapter 3

The authors of chapter 3 are:

Peter Anderson, Sheldon Green, Sidney Fels

My contributions to this project are described below.

Identification and Design of Research Program

This project is the result of a collaboration between Fels and Green. Fels was interested in acoustics and fluid flow in the vocal tract, and in particular how this may be feasibly applied in the Artisynth project. Green provided the expertise in fluid mechanics guide this goal.

Performing the Research

I performed the simulations under the direction of Fels and Green.

Data Analysis

I analyzed the results, with the suggestions of Fels and Green concerning how to process the data and what results may be of interest.

Manuscript Preparation

I wrote the entire manuscript with revisions and suggestions from Fels and Green.

Chapter 1

Thesis Introduction

1.1 Introduction

For ages turbulence has baffled and frustrated scientists, and today it still remains one of science's great outstanding problems. Many study turbulence out of a theoretical curiosity, but understanding turbulence has many applications because turbulent flows are pervasive, and it helps and hinders many human pursuits. Because of the pervasiveness of turbulence, it isn't surprising that turbulence plays a critical role in speech. Speech, like turbulence, is very complex, and is essential to our everyday lives. Because of the importance that turbulence and speech have in the lives of most people, it isn't surprising that there is a long history of attempts to understand, simplify, simulate, and control both phenomenon. Because of the complexity of both topics, it isn't surprising that both fields have many unanswered questions. As one might expect, there are complex and poorly understood phenomenon where turbulence and speech overlap, and these are the focus of this study.

Such cases are important to understand because they impact the quality of life for many people. Those with speech disorders hope for a remedy, those undergoing surgery in their vocal tract hope to have speech unimpaired afterwards, those modeling speech desire an accurate yet simple model, and those recording speech want to do so clearly and effectively, just to name a few applications. All these wants must be met with a good understanding if solutions are to be found, and thus we find good reason to research the complex fluid mechanics of speech.

The complexity of the topic makes a numerical simulations a good approach for obtaining flow information. While the underlying equations are

too complex to solve analytically, numerical solutions provide a complete data set and a deep understanding of the effect of assumptions, which makes numerical methods a useful research method for building an effective model of speech phenomenon.

1.2 Background

1.2.1 Fluid Mechanics

Fluid flow is described by the Navier-Stokes Equations (NSE), here written in general form with the conservation of mass and conservation of momentum equations:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (1.1)$$

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = -\nabla p + \nabla \cdot \mathcal{T} + \mathbf{f} \quad (1.2)$$

with time t , velocity \mathbf{u} , pressure p , body forces \mathbf{f} , and stress tensor \mathcal{T} . Though not immediately obvious, Eq. 1.2 comes from Newton's law $\frac{\partial m \mathbf{u}}{\partial t} = \mathbf{F}$, with the momentum terms on the left and the force terms on the right. However, in most cases the equations are seen as:

$$\nabla \cdot \mathbf{u} = 0 \quad (1.3)$$

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = -\nabla p + \mu \nabla^2 \mathbf{u} + \mathbf{f} \quad (1.4)$$

which include the assumption that the fluid is Newtonian, that μ is constant, and that ρ is constant, therefore assuming incompressibility. Yet another simplification worth noting is assuming that viscosity is negligible, thus arriving at the Euler equation:

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = -\nabla p + \mathbf{f} \quad (1.5)$$

The Navier Stokes equations are incredibly complex, and cannot in general be solved analytically. Because of the complexity of the NSE and the random

and complex nature of turbulence, it is typically studied statistically. One important feature which distinguishes turbulent from laminar flows is that turbulence has a wide range of scales, and energy cascades from the large scales to the smallest scale. The largest scale is the flow width, and the smallest scale occurs at the point where the dissipation rate (ϵ) equals energy production rate. Thus, a flow with higher energy will have smaller scales. Also, the smaller scales become increasingly isotropic.

One can predict when a flow will be turbulent by considering the Reynolds' number:

$$Re = \frac{\text{inertial forces}}{\text{viscous forces}} = \frac{\rho V D}{\mu} = \frac{V D}{\nu} \quad (1.6)$$

where V = characteristic velocity, D = characteristic length, μ = dynamic viscosity, and ν = kinematic viscosity. The Re which indicates transition from laminar to turbulent flows depends upon the type of flow, but a common case is the smooth pipe which transitions around $Re = 2400$. As Re increases, the smallest scales of turbulence become smaller.

Boundary Layers

When a flow is bounded by a wall one must include the additional complexity of the boundary layer in calculations. For a fluid like air the effects of viscosity are negligible in most parts of the flow, but very close to the wall viscosity plays a crucial role, greatly complicating the theory and numerical simulations. For a detailed description, see [1, 2]. Roughly speaking, the boundary layer is the portion near the wall where viscosity plays an important role. A simple definition of the boundary layer edge would be $u = 0.99U$ with u = velocity and U = main stream flow velocity.

In boundary layer theory, one typically considers x to be along the wall and y to be perpendicular to the wall directed into the flow. In the y -direction, one may assume that the pressure from the outer flow is constant across a laminar boundary layer. That is $dP/dy = 0$. In the x -direction a pressure gradient may have important consequences. A favorable pressure gradient ($dP/dx < 0$) occurs when the outer flow has an increase in velocity U in the flow direction, and therefore a decrease in pressure. An adverse

pressure gradient ($dP/dx > 0$) means that the outer flow has a decrease in U and thus an increase in P . In an adverse pressure gradient the boundary layer thickens, and if the gradient is strong enough the flow right next to the wall reverses flow direction thus causing the boundary layer to separate. Once the boundary layer is separated the assumptions of boundary layer theory break down [1].

Now if the boundary layer is turbulent the situation is greatly complicated. For one, a turbulent boundary layer will separate later than a laminar boundary layer. More generally, one can no longer define a velocity profile from U . Instead, one must consider the local flow parameters.

$$\text{friction velocity} = u_* = \sqrt{\frac{\tau_0}{\rho}} \quad (1.7)$$

$$y_+ = y \cdot \frac{u_*}{\nu} \quad (1.8)$$

Here, τ_0 is the shear stress at the wall and u_*/ν is the viscous scale used to non-dimensionalize y . The flow regimes in a turbulent boundary layer may be considered as a function of y_+ in the ‘Law of the Wall’. The main regions are:

- Viscous Sublayer: ($0 < y_+ < 5$). Viscous effects dominate;
- Buffer Layer: ($5 < y_+ < 30$). Viscous and Reynolds’ stresses important;
- Logarithmic Layer: ($30 < y_+ < 300$);
- Outer Layer: Largely inviscid;
- Outer Flow.

Two-Dimensional Flows

Turbulence is fundamentally different in 2D flows, which must be taken in to account if one is thinking to model a flow as two-dimensional (this discussion largely follows from [1]). In three dimensional turbulent flows, the energy transfers from the large scales to the small scales via vortex

stretching. Vortex stretching describes how a vortex will spin faster as it is stretched along its axis, thus conserving angular momentum (the same principle applies to a spinning skater who pulls her arms in close to spin faster). The vortex stretching term from the Navier Stokes Eq. 1.2 is:

$$\vec{\omega} \cdot \nabla \vec{u} \quad (1.9)$$

where u is the velocity vector, and $\omega = \nabla \times \vec{u}$ which is the vorticity. However, this term cannot exist in the absence of a third dimension. Thus, the cascade of energy from the large scales to the small scales does not occur in two dimensions.

To consider what does happen in 2D turbulence, it helps to consider the variable $\bar{\omega}^2$ which is termed enstrophy. By assuming an inviscid flow, both energy and enstrophy are conserved, from which one may show that enstrophy cascades to the smaller scales, while energy cascades to the larger scales. In 2D turbulence then, as the eddies grow larger, the viscous effects become less, further promoting growth in energy until the eddies become limited by their constant velocity. Therefore, what happens in 2D is just the opposite of what happens in 3D. There are cases where the flow is approximately 2D (atmospheric flows, for example), but in most cases this approximation should be used cautiously due to the physical consequences.

Acoustics

Sound is a component of an unsteady flow. Sound waves propagate by compression and expansion at the speed of sound. However, an unsteady flow also has pressure fluctuations which respond to the momentum changes in the fluid that are not propagating at the speed of sound. These pressure fluctuations are termed pseudo-sound, because a microphone or ear will interpret it as sound even though it isn't part of the propagating acoustic field [3].

It is typical to consider three different types of sound sources in aeroacoustics: the monopole source, the dipole source, and the quadrupole source [4]. A monopole source is created by a fluctuating mass flow and is the

strongest source type. A dipole is created by fluctuating forces upon a surface, and one may think of this as two equal and opposite monopoles very close to each other, resulting in much cancellation in the radiated sound. The quadrupole comes from fluctuating shear stresses within turbulence, and may be considered as two dipoles very close to each other, thus having the most cancellation making it the weakest sound source.

The acoustic analogy derived by Lighthill [4] arises from the compressible Navier-Stokes equations. Lighthill combined the conservation of mass term (Eq. 1.1) with the conservation of momentum term (Eq. 1.2), and arranged them to appear like the wave equation:

$$\left(\frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x_j^2} \right) \rho = \frac{\partial^2 \mathcal{T}_{ij}}{\partial x_i \partial x_j} \quad (1.10)$$

$$\mathcal{T}_{ij} = \rho u_i u_j + (p - \rho c^2) \delta_{ij} - \tau_{ij} \quad (1.11)$$

The math here involves no approximations, however, Lighthill's analogy of this equation to the wave equation (treating the right hand side as a source term) assumes a static environment just as the wave equation requires [5]. Furthermore, one might approximate $\mathcal{T}_{ij} \approx \rho_0 u_i u_j$ when the mach number is low thus just considering momentum flux [6]. Many similar analogies have been created, each presenting a different definition of the acoustic source term. While these analogies may be successful in some cases, the theory on which they are based involves large approximations, and should be treated with caution [5, 7].

1.2.2 Computational Fluid Dynamics

While the NSE are far from being solved analytically, they can be studied numerically. Computational fluid dynamics (CFD) has evolved with computers, and has grown into a complex topic of its own. To numerically solve a problem one must write the underlying equations in discrete form, which requires approximations to be made. One must also translate the geometry of the problem to a computational domain, and again one must make approximations concerning where the domain should be defined, how

the boundary conditions should be defined, what dimensionality should be used, and how fine the mesh cells should be. A good simulation needs to be carefully designed, and often one has to balance the details to be resolved by the simulation with the computational resources available.

RANS Simulations

When one wants to model turbulence, the simplest approach is to use the Reynolds-Averaged Navier-Stokes (RANS) equations. To do this the variables in the NSE are broken down into average components and fluctuating components:

$$\phi = \bar{\phi} + \phi' \quad (1.12)$$

where ϕ is the flow variable. When the variables are plugged into the NSE, one obtains a term called the Reynolds' stresses which are due to turbulence and are unknown:

$$\text{Reynolds' Stresses} = -\rho \overline{u'_i u'_j} \quad (1.13)$$

To model this term, the Boussinesq approximation [8] is typically made, which allows:

$$-\rho \overline{u'_i u'_j} = -\mu_t \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) - \frac{2}{3} \left(\rho k + \mu_t \frac{\partial u_i}{\partial x_i} \right) \delta_{ij} \quad (1.14)$$

An important assumption that goes into this equation is that the turbulent viscosity μ_t is a scalar (implying isotropy). From here various turbulence models seek to model the terms to close this equation.

One may observe the averaging that is used in the RANS models which may be unacceptable when one wants to observe transient fluctuations (such as sound) in the flow. Also, one may observe the isotropic assumption that goes into the turbulence models such as $k - \omega$ and $k - \epsilon$. The assumption that larger eddies are isotropic is not a good one, and a better alternative is offered in large eddy simulations.

Large Eddy Simulations

The theory behind a large eddy simulation (LES) is that one may resolve the larger eddies with a fine enough mesh, yet model the smaller eddies which cannot be resolved by the mesh. This is a reasonable approach because the large eddies depend upon the flow conditions and geometry while the smaller eddies are more isotropic and thus can be modeled rather than directly resolved. While the grid needed for a LES is finer than those needed for a RANS simulation, it is much coarser than those needed for a direct numerical simulation, thus making it computationally feasible while providing better data than a RANS simulation.

To create a LES, one must filter the Navier-Stokes equations to remove eddies that are below the filter scale. Then one must account for these eddies by modeling them. A generic filter can be written as:

$$\bar{\phi}(x) = \int_D \phi(x') G(x, x') dx' \quad (1.15)$$

In Fluent, this filter is simply taken to be a function of the mesh [8]:

$$\bar{\phi}(x) = \frac{1}{V} \int_V \phi(x') dx', x' \in V \quad (1.16)$$

where V is the volume of a mesh cell. When this is applied to the NSE, the stresses ($\tau_{ij} = \rho \bar{u}_i \bar{u}_j - \rho \bar{u}_i \bar{u}_j$) below the grid scale are unknown and must be modeled. Typically, as in the RANS models, the Boussinesq approximation is applied, but in this case it is a better assumption because it is only applied for the smaller, more isotropic eddies. Thus, one must model μ_t to find the stresses τ from:

$$\tau_{ij} - \frac{1}{3} \tau_{kk} \delta_{ij} = -2\mu_t \bar{S}_{ij} \quad (1.17)$$

where:

$$\bar{S}_{ij} = \frac{1}{2} \left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right) = \text{rate of strain} \quad (1.18)$$

To do this, the Smagorinsky model states [8]:

$$\mu_t = \rho L_s^2 \sqrt{2\bar{S}_{ij}\bar{S}_{ij}} \quad (1.19)$$

where:

$$L_s^2 = \text{minimum}(\kappa d, C_s V^{1/3}) \quad (1.20)$$

Here, κ is the von Karman constant (around 0.4), d is the shortest distance to a wall, C_s is the constant dynamically found, and V is the volume of the cell. This is how the Dynamic Smagorinsky model is applied in Fluent [8].

Direct Numerical Simulation

The most computationally intensive simulation is a direct numerical simulation (DNS), in which the mesh and time step are made small enough to resolve the smallest scales of the flow, thus no turbulence model is needed. However, as discussed above, the smallest scales become smaller as Re increases, which rapidly limits the feasibility of a DNS.

Wall Resolution

With the boundary layer theory given in Section 1.2.1, one may consider how to capture the boundary layer in a simulation. If the flow is laminar one may use U along with boundary layer theory to model what occurs at the wall. Likewise, if the flow is turbulent one may apply the law of the wall to model flow near the wall using a wall function. In Fluent, when one is using a wall function one wants the first few cells next to the wall to be in the log-layer, but the cell centroid should not be in the viscous sublayer or buffer layer. Thus, when using a wall function one can over-resolve the wall. On the other hand, if one is performing a DNS or a LES, then one wishes to resolve the viscous sublayer, which means that the first cell centroid should be around $y_+ = 1$. However, in Fluent if the viscous sublayer is not resolved a wall function is applied [8].

Computational Aeroacoustics

In computational aeroacoustics one typically seeks to resolve the acoustic waves, which creates numerous specialized requirements in addition to the requirements of a standard CFD simulation [9, 6].

Compressibility. Sound propagates through air by compression and expansion. Though the compressibility of air has very little effect on the non-acoustic component of the flow at a small mach number (less than 0.1Ma), sound propagation is impossible without a compressible flow, thus to directly resolve sound one must solve the compressible Navier-Stokes Equations, thus complicating the equations that need to be solved and slowing the simulation. If one is only using the acoustic analogy, the flow does not have to be compressible, but to compare an acoustic analogy with direct measurements it must be compressible.

High-order numerical schemes. High order methods are very helpful for resolving the acoustic waves and seeing that they propagate with minimal dispersion or dissipation. Fluent does not offer such high order numerics, but maintaining a large number of mesh points per wavelength can help to combat this.

Small pressures. Acoustic waves have a very low amplitudes compared to the other pressures that occur in the flow, thus high precision is required to distinguish the acoustic signals above numerical noise.

Non-reflecting boundaries. The standard boundary conditions are insufficient when seeking to resolve acoustics, because these conditions cause reflections. One solution is to find the acoustic component at the boundaries and remove it. A simpler method involves building a buffer layer around the boundary to gradually damp out the waves, but this is computationally expensive.

Range of length scales. The sound often covers a wide range of scales. In order to capture the high frequencies, the low frequencies, and attain an adequate spectrum from an fft, fine mesh and time resolutions are required as well as long simulation runtimes.

Thus, one can understand how resolving the acoustics requires a more sophisticated simulation and higher computational demand.

The acoustic analogy isn't strictly a computational method, but it is heavily used in CAA. The acoustic analogy applied in Fluent is the Ffowcs

Williams - Hawkings (FW-H) model [8]. It is a general derivation, and has the advantage that an arbitrary surface may be used for the source surface, thus allowing completely permeable surfaces which don't affect the flow to be used. The FW-H model has been fairly successful even when the source surface is in the non-linear flow [6], thus defying Lighthill's static medium assumption. The propagation equations that are solved assume propagation into free space.

Verification and Validation

An important part of CFD is verification and validation. In the CFD world, verification means properly solving the equations, while validation checks that the equations solved are in fact true to the physics [10]. Verification is typically done by showing that the solutions converge as the discretizations approach zero. Validation is typically done by comparing the solution with experimental or theoretical results. Both verification and validation are essential for a trustworthy simulation.

1.2.3 Linguistics

The fluid mechanics of a few common speech sounds are considered in this thesis. The simplest sounds to consider are vowels, in which the vibrating vocal chords form a monopole sound source. From the source the sound propagates and resonates in the vocal tract, but the vocal tract is relatively open and the airflow is laminar.

By contrast, the fricative is formed in a very different manner. The fricative is formed when the vocal tract converges to a narrow constriction, forcing the air into a turbulent jet which strikes an obstruction (such as the teeth or roof of the mouth), which generates the sound. The vocal chords may be creating sound as well (in a voiced fricative), but fricatives don't depend on the vocal chords generating sound (as seen in voiceless fricatives). The driving principle behind a fricative is dipole sound generation by turbulence striking an obstacle. The turbulent jet will have a quadrupole sound source as well, which comes from the turbulent fluctuations within the jet

itself rather than forces between the jet and the wall, but in most cases the quadrupole is expected to be weaker than the dipole.

Yet another very different sound generation mechanism in speech is the plosive. This occurs when the vocal tract closes off entirely allowing pressure to be built up behind the closure. When the closure is broken, a sudden sound is created as well as a turbulent burst of air. An example of this is the bilabial plosive, in which both lips form the closure, and when it is broken the speaker creates a sound such as /ba/ or /pa/.

Not only do speech sounds differ greatly between one another, but also within themselves. A culture or language may tend to make a sound in a different manner than another, and likewise there will be variations between people, not just because they have unique articulations, but also because they have unique physical characteristics thus creating variation in vocal tract shape and the airflow produced therein. More than that, an individual may pronounce the same speech sound in a different way depending on the context of its use. For example, the /sh/ in ‘shoe’ is said differently than the /sh/ in ‘ash’ because the transitions in and out of the speech sound vary, and the same word will be pronounced differently depending on the context its used in. This great diversity of speech sounds causes a great challenge for linguists to understand what it is about the speech sound that allows the listener to perceive it correctly when one realization may differ so much from the next [11].

Once a sound is created, it will propagate through the vocal tract and be modified along the way: some frequencies will be diminished while others are enhanced. The mechanism for this is resonance. In a rigid cavity or tube the air velocity at the wall must be zero, which defines a node in the acoustic wave. Due to this limitation, the waves with wavelengths such that nodes fall at the walls will fit naturally into the cavity and constructively interfere thus creating a resonant frequency. On the other hand, wavelengths that don’t fit naturally will interfere destructively, thus being filtered out [12]. With this principle in mind, Steven’s [13] considers a couple of simplified cases. First, the natural frequencies of a tube open at one end and closed

at the other are given by:

$$f_n = \frac{c(2n - 1)}{4l} \quad (1.21)$$

where c is the speed of sound and l is the length of the tube and $n = 1, 2, \dots, \infty$ which represents each vocal tract resonance or formant. A narrow tube which has a cavity at the end (or a Helmholtz resonator) has the natural frequency:

$$f_1 = \frac{c}{2\pi} \sqrt{\frac{A_2}{A_1 l_1 l_2}} \quad (1.22)$$

where A and l correspond to the area and length of the big and narrow cavities. When such a cavity is linked by a narrow tube, it will have a shifted resonant frequency:

$$f'_0 = f_0 \left(1 \pm \frac{2}{\pi} \sqrt{\frac{A_2}{A_1}} \right) \quad (1.23)$$

where f_0 is the frequency of the cavity were it uncoupled.

Of course, the resonance in the vocal tract isn't nearly this simple, but one can build a concept of what is happening in the vocal tract from this theory. For example, from Eq. 1.21 one might estimate $f_1 = 472\text{Hz}$ and $f_2 = 1417\text{Hz}$ for a vocal tract of $l = 18\text{cm}$ and $c = 340\text{m/s}$. This open tube estimate is confirmed in Titze's map of vowel formants, where this is a common value for the first formant [12]. Likewise, if one wishes to make an estimate for the resonance created by the cavity below the tongue in some fricatives, a rough estimate can be made with Eq. 1.22. Estimating $A_1 = 12.7\text{cm}^2$, $l_1 = 2\text{cm}$, $A_2 = 1\text{cm}^2$, and $l_2 = 2\text{cm}$, one finds $f_1 = 900\text{Hz}$. In this way, one may estimate what frequencies to expect during speech, the effect that a cavity may have on the spectrum, and how cavities may affect each other. One must be aware that many idealizing assumptions go into these equations, and in reality the energy losses at the walls, non-ideal shapes of the cavities, and relative scales of the cavities will cause these simplifications to break down, but they do provide a means for first approximations.

1.2. Background

Taking the sound sources to depend just upon the airflow and not on the surrounding sections of the vocal tract is commonly done. In other words, one assumes that the source is independent from that which modifies the sound. This assumption allows a source-filter approach and is commonly used (see [14] for example). With this, one may find the pressure spectrum:

$$p_r(f) = S(f)T(f)R(f) \quad (1.24)$$

Where S is the source function, T is the transfer function, and R is the radiation function [13]. Note that these are functions of frequency f . The transfer function describes the spectral filtering that the vocal tract applies to the source, and thus depends on the shape of the vocal tract. The radiation function describes the sound radiation beyond the mouth, and can be a complex issue of itself.

Outside of the mouth, the simplest approach is to treat the mouth as a point source. An acoustic wave will be omnidirectional when its wavelength is much larger than the dimensions of its source. Viewing the mouth as a point source assumes that the frequencies have wavelengths much larger than the mouth. This assumption is good until about 1000Hz ($\lambda \approx 34cm$) and workable until 4000Hz ($\lambda \approx 8.5cm$) [13]. This also assumes that the receiver is in the ‘far field’ which would be at least a few centimeters away, or more ideally one wavelength away at the longest wavelength being observed. These assumptions clearly have limitations, but they allow one to make quick approximations using:

$$p_r(t) = \frac{\rho}{4\pi r} \frac{\partial \dot{V}(t - \frac{r}{c})}{\partial t} \quad (1.25)$$

Here, the pressure at a distance r from source is given as a function of time t . The source has a volume velocity \dot{V} , and c is the speed of sound.

One classic speech simulation model considers the vocal tract as a series of cylinders of variable radius with which one can derive the transfer function [15]. A more sophisticated approach uses the smoothly changing area of the vocal tract in conjunction with 1D equations [16]. A source-filter approach

like this is reasonable for vowels and allows for fast computation, but it cannot include turbulent effects that only exist in three dimensions and thus to be fully modeled requires a 3D fluid mechanics simulation. In fact, CFD studies in linguistics are not common, but some good examples include a study of the acoustics at the vocal chords [17], a study of airflow in the vocal tract with considerations for obstructive sleep apnea [18], and a study of flow as it escapes the lips during a plosive [19].

Fricatives have been heavily studied, and Shadle [20] discusses numerous experimental findings. Of particular interest is the experiment studying the fricative sound ‘sh’. Shadle started with geometry from mid-sagittal x-ray data provided by Fant [21], and used that shape to form a domain made from plexiglass. The depth of the domain (perpendicular to the mid-sagittal plane) was kept as a constant 2.54cm, thus ignoring the true 3D form of the vocal tract. However, the constriction formed between the tongue and the roof of the mouth was narrowed to a realistic constriction depth using clay. In an anechoic chamber, air was blown through this model and the sound was measured in the acoustic far-field. To record sound at various location inside the mouth, small microphones were inserted in the plexiglass flush to the wall.

Experiments such as these are useful to a CFD user because they are well-known to fricative researchers, and provide methods and results which can be easily compared to simulations. Shadle’s geometry was simple and well-defined, such that it is relatively easy to replicate as a simulation geometry, with the only exception being the shape of the constriction which isn’t known exactly. Shadle also gives the input flow rate and the acoustic results, thus detailing an experiment that is both repeatable for simulation and close to speech.

1.3 Objectives and Hypothesis

In light of these facts, this study seeks to apply CFD to linguistics with three distinct goals in mind. First, it is important to design simulations which will properly replicate the fluid flow observed in speech. This involves

designing computational domains and using proper numerical methods to capture the aspect(s) of speech which are desired. If CFD cannot do this, then it has little to offer to linguistics. Second, it is useful to find the simplest simulation methods which still yield good results. A very complex and complete simulation will likely find a good solution, but such simulations can have a very large computational cost, thus the goal is to find the fastest method which still works. Third, with working simulations one can apply the data gathered to learn about linguistics. In the end, the goal of using CFD in linguistics is to apply the data towards some end, be it voice modeling, surgery planning, theory development, or entertainment. This research will seek to draw conclusions from the simulations which will advance the understanding in linguistics and beyond.

CFD has often been used to simulate both turbulence and acoustics, thus it is expected that simulations can be created to capture the phenomenon studied. For example, Dejoan et al [22] demonstrate the ability of a LES of a turbulent wall jet, and Lai et al [23] demonstrate the application of LES for simulating acoustics. However, such examples also show cases where simulation results are wrong, and not always for an obvious reason, thus correct results cannot be taken for granted and must be interpreted properly. Such accurate simulations will probably have to be three dimensional. The geometry of the vocal tract is three-dimensional, and turbulent flows are also three dimensional, thus it is unlikely that two dimensional simulations can capture the flows being studied. However, it is hoped that 2D flows will provide some insights into the flow, which can be useful in some applications, if not all.

1.4 Bibliography

- [1] Pijush K. Kundu and Ira M. Cohen. *Fluid Mechanics*, volume 2. Academic Press, San Diego, 2002.
- [2] Stephen B. Pope. *Turbulent Flows*. Cambridge University Press, 2000.
- [3] J. E. F. Williams. Hydrodynamic Noise. *Annual Review of Fluid Mechanics*, 1:197–&–, 1969.
- [4] M. J. Lighthill. On Sound Generated Aerodynamically. I. General Theory. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 211(1107):564–587–, 1952.
- [5] C. K. W. Tam. Computational aeroacoustics examples showing the failure of the acoustic analogy theory to identify the correct noise sources. *Journal of Computational Acoustics*, 10(4):387–405–, 2002.
- [6] T. Colonius and S. K. Lele. Computational aeroacoustics: progress on nonlinear problems of sound generation. *Progress in Aerospace Sciences*, 40(6):345–416–, 2004.
- [7] A. T. Fedorchenko. On some fundamental flaws in present aeroacoustic theory. *Journal of Sound and Vibration*, 232(4):719–782–, 2000.
- [8] Fluent. *Fluent 6.2 Documentation*, 2005.
- [9] C. K. W. Tam. Computational aeroacoustics: An overview of computational challenges and applications. *International Journal of Computational Fluid Dynamics*, 18(6):547–567–, 2004.
- [10] P. J. Roache. Quantification of uncertainty in computational fluid dynamics. *Annu. Rev. Fluid Mech.*, 29:123–160, 1997.
- [11] A. Jongman, R. Wayland, and S. Wong. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108(3):1252–1263, 2000.

1.4. Bibliography

- [12] Ingo R. Titze. *Principles of Voice Production*. Prentice Hall, Inc., Englewood Cliffs, 1994.
- [13] Kenneth Stevens. *Acoustic Phonetics*. The MIT Press, Cambridge, 2000.
- [14] S. Narayanan and A. Alwan. Noise source models for fricative consonants. *Ieee Transactions on Speech and Audio Processing*, 8(3):328–344–, 2000.
- [15] Siddharth Mathur. *Variable-Length Vocal Tract Modeling for Speech Synthesis*. PhD thesis, University of Arizona, Tucson, 2003.
- [16] K van den Doel and Ascher U. M. Real-Time Numerical Solution To Webster’s Equation On A Non-Uniform Grid. Preprint, 2007.
- [17] W. Zhao, C. Zhang, S. H. Frankel, and L. Mongeau. Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms. *Journal of the Acoustical Society of America*, 112(5):2134–2146, 2002.
- [18] P. Nithiarasu, O. Hassan, K. Morgan, N. P. Weatherill, C. Fielder, H. Whittet, P. Ebdon, and K. R. Lewis. Steady flow through a realistic human upper airway geometry. *International Journal for Numerical Methods in Fluids*, 57(5):631–651, 2008.
- [19] X. Pelorson, G. C. J. Hofmans, M. Ranucci, and R. C. M. Bosch. On the fluid mechanics of bilabial plosives. *Speech Communication*, 22(2-3):155–172–, 1997.
- [20] Christine H. Shadle. Articulatory-Acoustic Relationships In Fricative Consonants. In *Speech Production and Speech Modelling*, pages 187–209–. Kluwer Academic, Netherlands, 1990.
- [21] Gunnar Fant. *Acoustic Theory of Speech Production*, volume 2. Mouton, The Hague, 1970.

1.4. Bibliography

- [22] A. Dejoan and M. A. Leschziner. Large eddy simulation of a plane turbulent wall jet. *Physics of Fluids*, 17(2):–, 2005.
- [23] H. Lai and K. H. Luo. A three-dimensional hybrid LES-Acoustic analogy method for predicting open-cavity noise. *Flow Turbulence and Combustion*, 79(1):55–82, 2007.

Chapter 2

Characteristics of Air Puffs Produced in English ‘pa’: Experiments and Simulations

2.1 Introduction

The release burst and aspiration or ‘pop’ associated with voiceless aspirated plosive consonants (e.g. /p^h/, /t^h/ and /k^h/) in many languages is a potentially important cue in the perception of these sounds [1] and a well-known challenge for audio engineers and microphone manufacturers [2, 3]. Plosive release burst and aspiration contain both sound and what has been termed pseudo-sound [4, 5]. While sound waves propagate through air at the speed of sound ($c = \sqrt{\gamma \cdot P/\rho}$ for an ideal gas), pseudo-sounds are pressure fluctuations within the flow that are detectable by an ear or microphone. Audio engineers typically want to record only the sound and not the pseudo-sound, while speech perception may make use of both. The present paper seeks to characterize the properties of flow (as opposed to sound) associated with English aspirated ‘p’.

While a good deal is known about the properties of air flow from an orifice in industrial application, there exist a number of problems peculiar to modeling oral aspiration in speech that have not been previously addressed,

A version of this chapter has been submitted for publication.

Derrick, D. and Anderson, P. and Gick, B. and Green, S. Characteristics of Air Puffs Produced in English ‘pa’: Experiments and Simulations

including properties of the orifice, flow description and simulation type.

2.1.1 Orifice

During the production of the labial plosive ‘p’ release, the lips constitute a highly complex orifice, being elastic and continuously changing in geometry and rigidity. During English and Japanese bilabial stop releases, Westbury and Hashi used Westbury’s X-ray micro-beam data to demonstrate that the lips accelerate away from each other after the release, reaching a maximum velocity of about 200 mm/s at 25 ms, and then decelerate until they reach an average opening of ~ 20 mm after ~ 200 ms [6]. While this mouth opening time is quick, it is not negligible compared to the time scales under study. Pelorson et al. argued for the importance of modeling changes in the lip opening over time, [7], but did not model or measure it due to the limitations of computational speed at the time. The rate of lip opening is likely to have a large effect on initial flow rate as air flows faster through constrictions in a tube.

In an engineering setting there have been few studies of the effects of variable orifice geometry on the fluid mechanics. One such study, by Dabiri and Gharib [8], considered the starting jet formed by a circular orifice of time-varying diameter. They studied the effects of changing nozzle diameter on the flow, and found that a temporally increasing nozzle diameter causes the leading vortex ring to have the strongest vorticity at a larger radius from the centerline than for a constant nozzle diameter, but they did not measure jet penetration distance, which is a primary quantity of interest here.

During the production of ‘pa’s, lip aperture geometry is close to an ellipse [7], and so there is the need to consider whether modeling the general elliptical shape of the mouth opening is important in simulating airflow after a bilabial release. Non-circular jets have been previously studied as a means of providing passive flow control. Research results, both numerical and experimental, show significant differences between circular and elliptical jets [9, 10], thus simulating the general shape of the lip opening is likely to be important for accurate simulations.

Deformation of lip shape due to the elasticity of lips, and the fluid-structure interaction between the lips and air, are much smaller in amplitude than the general trend of the lips to open to 20 mm throughout the progress of the release of a labial plosive. Such disturbances at the source are expected to have little effect on the general flow [11, 12].

2.1.2 Flow description

After a bilabial stop release into a vowel, the pressure in the mouth drops asymptotically to approximately 1/10th of its initial value in the first 60ms of the flow (similar to Figure 2.4) [13, 14, 7]. The pressure in the mouth is sufficiently great that the flow out of the mouth during an utterance such as ‘pa’ is turbulent. In a turbulent flow a large range of scales is present, as opposed to the smaller range present in a smooth, laminar flow. One can confirm that a ‘pa’ is turbulent by considering the Reynolds number, which is a dimensionless parameter important for characterizing flows:

$$Re = \frac{\rho \cdot V \cdot D}{\mu}$$

In this expression, ρ is the fluid density, V is the mean fluid velocity at the orifice, D is the orifice diameter, and μ is the dynamic viscosity. For a Reynolds number greater than 1000, round jets become turbulent a short distance from the nozzle [15]. D is approximated as 6.1 mm by finding the hydraulic diameter for the mouth (see Stevens for similar hydraulic diameters [14]), $V=20\text{m/s}$ as a conservative estimate, and using typical values of air of $\rho=1.2\text{kg/m}^3$ and $\mu = 1.8 \cdot 10^{-5} \text{ N} \cdot \text{s/m}^2$, the Re is ~ 8100 , so this flow is turbulent.

Turbulent starting jets (the initiation of a continuous flow from an orifice) and puffs (in which flow at the orifice is cut off soon after initiation) have been heavily studied for other applications such as fuel injection, and are typically studied with round nozzles. Sangras et al. [11] (note correction [16]), provides a nice summary of starting jet and puff research. The leading

2.1. Introduction

edge of the burst follows the equation (dropping the virtual origin):

$$X = c \cdot T^n$$

where $X = x/D$ = non-dimensional distance, c is an experimentally determined constant, $T = t \cdot V/D$ = non-dimensional time, and $n = 1/2$ for starting jets and $n = 1/4$ for puffs. Figure 2.1 shows the difference between a starting jet and a puff using the range of ‘ c ’ reported in Sangras et al. [11]. The puff and starting jet penetration distances diverge significantly for $T > 100$. Using the characteristic diameter and velocity of a ‘pa’ estimated above, puffs and jets would penetrate noticeably different distances after about 30ms. Since there is a need to understand ‘pa’ behaviour to 100ms or more, it is clearly necessary to model the actual transient pressure driving the flow.

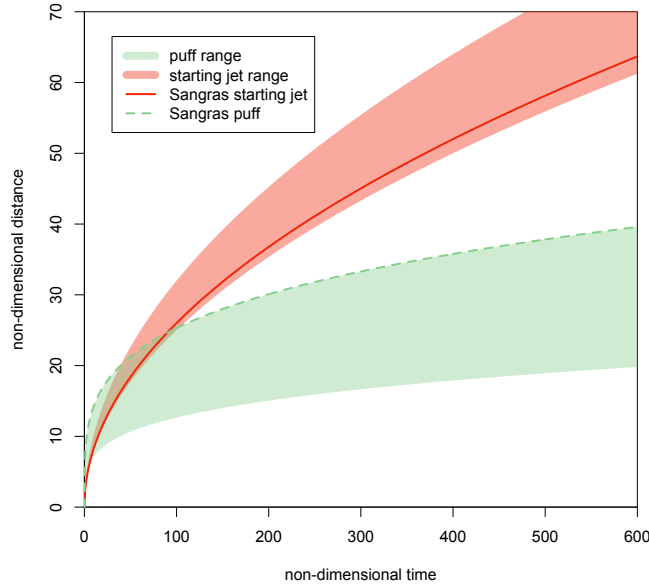


Figure 2.1: (color online) Starting Puff and Starting Jet Comparison.

Assuming the room temperature to be 22C and the air jet to 37C (body temperature), then from the ideal gas law one finds the ratio $\rho_0/\rho_{jet} = 1.05$. Diez et al. [17] found the effects of buoyancy to be small for the temporal and spatial range considered in this study, and, while Diez et al. considered

buoyant forces acting along the streamwise direction, in speech the buoyant force will be roughly perpendicular to the jet, presumably resulting in an even smaller effect on streamwise penetration. Temperature will also cause the jet viscosity to be $\sim 5\%$ higher than the surrounding air, but this difference should also produce negligible effects on a flow at this Re.

2.1.3 Simulation Type

One must consider whether the problem can be modeled in 2D, or if a more complex 3D model is required. If the domain is 2D, then the mouth would have to be treated as a plane jet, as was done by Pelorson et al.[7]. Although turbulence is a 3D flow, it is possible to consider a 2D RANS (Reynolds-Averaged Navier-Stokes) turbulence model. However, Reichert [18] and Stanley [19], among others, report significant inaccuracies in 2D simulations of plane jets. Finally, RANS models average the flow, but the turbulent fluctuations themselves are of interest to us, therefore a more sophisticated technique such as Large Eddy Simulation (LES) is needed. Thus, both the geometry and the flow compel us to model plosive aspiration in 3D using LES.

2.1.4 Hypotheses

Based on the above discussion, it is proposed that to adequately simulate air flow from the mouth after the release of a bilabial stop into a vowel, one needs to take into account the known decrease in air pressure following the release. It is also hypothesized that the mouth can be adequately modeled as a 2D narrow ellipse. Computational limitations require a static geometry. The validity of this assumption must be measured against lip aperture over time from the high-speed video experiments. Due to the fact that the airflow throughout most of the release is turbulent, it is necessary to resolve the turbulent properties, and because the lip aperture exists in 3D space, one needs a 3D LES simulation to accurately model air flow after a bilabial release.

2.2 Methods

These hypotheses were tested by comparing the results of two sets of experiments with simulations. The first experiment used a microphone located at varying distances from a participant repeating the syllable 'pa' to record pressure fronts corresponding to microphone 'pops'. The second experiment used high-speed video to record smoke particles. The microphone pops were compared to the simulation pressure front. The leading edge of the smoke particles recorded in the high speed video was compared to the leading edge of the simulation particle front.

2.2.1 Microphone Experiment

Data Recording

For the microphone 'pop' experiment, a single male participant was seated in a sound-proof room. Two microphones were placed in the room, one dummy microphone at 50 cm away from the mouth of the participant, and one SHURE SM58 set 5 cm away from the mouth of the participant. The cover of the microphone was removed to increase the effect of the pop on the recording, and the microphone was plugged into a Sound Devices USBPre microphone pre-amplifier plugged into a 1.42GB dual processor PowerMac G4 with 512 MB of ram running Mac OSX 10.4.10 and recording with Audacity 3.3 at sampling rate of 44,100 kHz. Both microphones were lined up and placed at exactly the mouth height of the participant.

The participant wore a set of Direct Sound Extreme Isolation Headphones plugged into the USBPre and set to monitor microphone input in real-time. The self-monitoring allowed the participant to adjust his speaking angle to make sure that microphone pops were being picked up by the Shure SM58 microphone, a particularly difficult task at distances past 20 cm.

The participant was handed a thin rigid tube to place in the corner of his mouth. The tube was attached to a SCICON Macquiner 516 airflow meter set to record the mouth pressure of the participant during the experiment. The airflow meter was attached to the same powerMac and using Macquiner

8.9.5.

The participant was asked to say the word ‘pa’ fifteen times while focusing on the dummy microphone set 50 cm away. The experiment was repeated with the microphone moved back at 5 cm increments from 5 to 40 cm away from the participant.

Data Analysis

For each token the maximum air pressure just prior to the release burst of ‘pa’ was recorded along with the difference in time from the onset of the sound of each ‘pa’ and the beginning of a microphone pop. Airflow perturbations, or microphone pops, affect microphone output through the production of a very low frequency wave caused by the air-flow, and high frequency aperiodic sound. To measure how long the airflow took to reach the microphone, the time between the onset of the release burst and the onset of the first significant low frequency perturbation that looks and sounds like microphone ‘pop’ was used, as illustrated in token 75 in Figure 2.2(a).

However, these perturbations are difficult to isolate, particularly from a sound signal for distances from 20 to 40 cm due to overlap with the high amplitude vocalic portion of the sound wave. Fortunately, microphone pops are also associated with turbulence at higher frequencies. The high frequency aperiodic sound is hard to isolate in the waveform, but easy to detect by listening to the sound. Therefore each token was also examined by listening for the onset of pops using a set of high-quality Sennheiser HD650 headphones and a Total Bithead preamp. This turbulent sound helped isolate the onset of the microphone pop. A good example of a straightforward measurement can be found in token 72 in Figure 2.2. For cases where neither listening nor examining the original wave worked, the original sound file was band-pass filtered using a band pass elliptic filter set from 30 to 100 Hz in MATLAB with 30 Hz skirts. These frequencies are produced in the sounds of speech, but microphone pops produce these frequencies at higher amplitude making the leading edge of the microphone pop easier to detect.

The time between the onset of the original sound wave and the onset

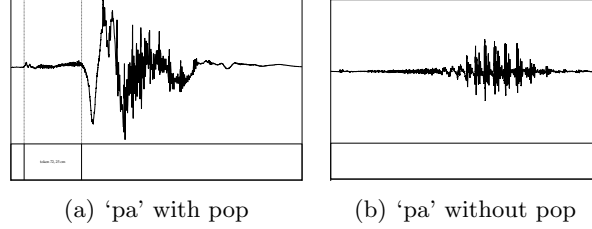


Figure 2.2: Microphone recordings of 'pa' with and without microphone pop. (183 ms clip)

of the first visibly larger peak was selected, but only when there was an obvious increase in the amplitude of these low frequency waves clustered together. This filtering method can reduce the accuracy of measurements because it excludes relevant frequencies that cannot be used because they overlap the fundamental frequency and first harmonic. However, in some cases the method was very helpful, as in token 7 shown in Figure 2.3 where it is hard to see the onset of the pop in the unfiltered waveform, but easy to see in the low-pass filtered waveform.

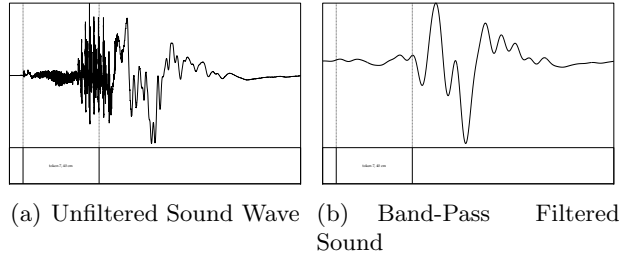


Figure 2.3: Measurements from sound token 7, distance = 40 cm, 437 ms clip

If none of these three techniques produced a discernible result, the token was not used because the microphone did not record a loud enough pop to isolate.

The microphone pop timing corresponds to the leading pressure front recorded in the air puff simulation.

2.2.2 High Speed Video Experiments

Two sessions of high speed video of the participant from the ‘pop’ experiment saying the word ‘pa’ while expelling white smoke were made. The smoke had a similar density as air, and was close to body temperature or approximately 37 °C at the time of expiration.

For the first round, digital videos of three productions of ‘pa’ were recorded with black foam board in the background and a standard tape measure pasted to the board. The camera was placed approximately 460 cm away and focused on the tape measure such that the shot was 52.8 cm wide at the focal point. Bright sunlight was used to provide lighting. The participant then stood to the edge of the black bristol board such that their mouth opened just above the tape measure. The participant inhaled white smoke prior to the production of the ‘pa’ so that the expelled air from the production of the ‘pa’ would be visible during filming. Video was captured using a Bassler 504kc high speed color digital video camera with a Micro-Nikkor 70-180mm Telephoto Zoom Lens. The camera was plugged in to an EPIX PIXCI CL3 SD frame grabber card with 1GB of PC133 MHz memory in a P4 computer with 1GB of ram running Windows XP. Digital video was captured into the frame buffer using X-Cap Lite set to capture at 1024x768 resolution at 500 fps at maximum light gain and exported frame by frame into 1280x1024 32bit TIFF files.

For the second round, digital video of twelve productions of ‘pa’ were captured using black foam board background and meter sticks for scale. The camera was placed approximately 330 cm away and focused on the tape measure such that the shot was 53.0 cm wide at the focal point. A film light was placed facing the speaker to clearly illuminate the smoke particles. Video was captured using a Phantom v12 high speed monochrome digital video camera with a Navitar 6.5x lens. Digital video was transferred from the camera’s built-in memory to 1280x800 resolution jpegs at 2000 fps.

Data analysis

For both rounds, the point of the opening of the mouth was captured using ImageJ's point capture utility, and the leading edge of the white smoke was recorded frame by frame for 150 ms of recorded time. The points were converted to distance in cm and analyzed statistically (see Figure 2.14).

For both rounds, exact measurements of initial mouth pressure could not be made because the air flow apparatus would have interfered with the visual recording of air puff travel. However the pressure can be inferred from Kenneth Stevens data on initial intra-oral air pressure during the production of aspirated stops at normal volume and the previous recordings of louder 'pa's during the microphone study which used the same subject (See Figure 2.4[14]).

For the second round, the rate of lip opening was also captured using ImageJ's point capture utility. The position of the top of the mucous membrane of the upper lip and the crease that intersects the mental protuberance and the skin below the lower lip were recorded frame by frame for 40 ms for each of the 12 recordings. These points provide stable landmarks for measuring the rate of lip opening. The points were converted to distance in mm and analyzed statistically.

2.2.3 Numerical Simulations

For the base numerical study, a domain of physical dimensions 350x100x100 mm³ which is meshed with 721,800, non-uniform, hexahedral control volumes was used. The mouth is shaped like a narrow ellipse in the x=0 plane, with $r_y=2\text{mm}$ and $r_z=15\text{mm}$. A rough integration of upper and lower lip pellet velocity from Westbury [6] shows the lips have a y radius of 2 mm ~ 17 ms after they begin to separate.

Stevens [14] shows the intraoral pressure quickly dropping after the release burst for 'pa', thus the mouth was modeled as a transient pressure inlet which quickly drops to 1/10th of its initial value as shown in Figure 2.4. In the simulation, the mouth lies in a plane that is modeled as a wall, while the rest of the boundaries are pressure outlets set to atmospheric pressure.

2.2. Methods

The air is incompressible and initially still. Nitrogen particles were injected and tracked as a dye, thus defining the leading edge of the jet. An implicit,

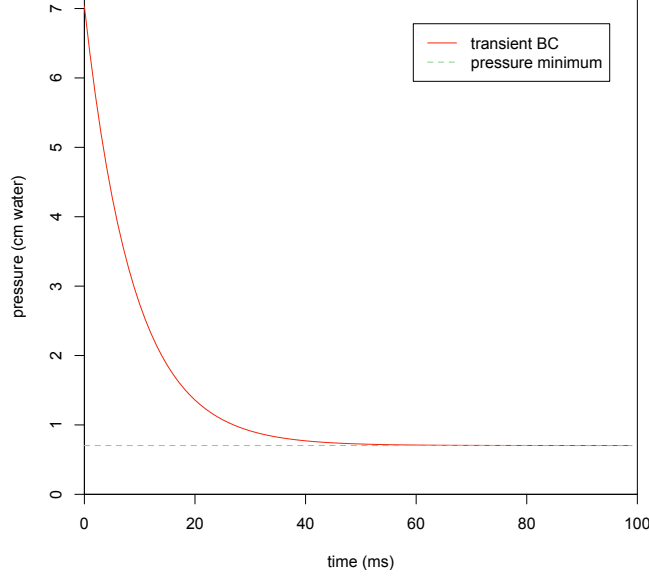


Figure 2.4: (color online) Transient boundary condition

bounded central differencing spatial discretization and an implicit second-order time discretization, with a large eddy simulation (LES) to model this turbulent flow was used. LES resolves the large eddies within the flow, but eddies smaller than the mesh scale are resolved by a turbulence model (in this case dynamic Smagorinsky). The model was performed over 4000 time steps of size $\Delta t = 0.025$ ms ($t_{\text{final}} = 100$ ms). Using Fluent as the solver, and running on three parallel processors, this process took ~ 6 days.

To explore the quality of the simulation methods and initial assumptions, numerous variations to this baseline simulation were run:

Variation 1: A grid refinement study was performed with the standard hexahedral mesh using a simulation with a medium mesh of 88,380 control volumes and a course mesh of 11,925 control volumes.

Variation 2: A similar simulation replacing the mouth-shaped and time-varying inlet with a circular and constant velocity inlet, thus modeling

a starting jet from a circular nozzle, was run in order to validate the numerical methods. See Roache (1997) [20] for general discussion of verification and validation.

Variation 3: A simulation with the starting inlet pressure three times higher than normal (24 cm/H₂O) yet falling to the same final value (0.703 cm/H₂O) was run to simulate a loud utterance.

Variation 4: A simulation with a constant pressure inlet of 7.03 cm/H₂O (690Pa), which is the same initial pressure of the baseline simulation was also run to test the importance of the transient pressure inlet.

Variation 5: A simulation where the initial pressure was raised by 1Pa was run. This slight change has little effect on the physics, but it does cause the numerics to change slightly, thus providing a second realization of the turbulent flow.

Variation 6: A simulation was run where the inlet pressure condition was unchanged, but the initial domain was perturbed with small velocities, thus providing realistic disturbances in the air which are greater than machine zero.

Some preliminary simulations in 2D, using LES and RANS, were also conducted, but these soon proved to be inadequate.

2.3 Results

The results of the microphone and high speed video experiments, along with the numerical simulations, are described below:

2.3.1 Microphone Experiment

Of 120 tokens recorded, 90 had discernible pops according to the standards described in Section 2.2.1. Individual measurements were highly variable, as seen in Figure 2.5; the fit line is based on a loglinear quadratic fit with an assumed zero intercept. The fit line is highly significant, with an $F(2,88)$

2.3. Results

= 4386, $p < 0.001$ for each coefficient, and adjusted $R^2 = 98.9\%$. Linear, quadratic, cubic and loglinear statistical models produced less significant results. As a result of trying to produce microphone pops in a microphone

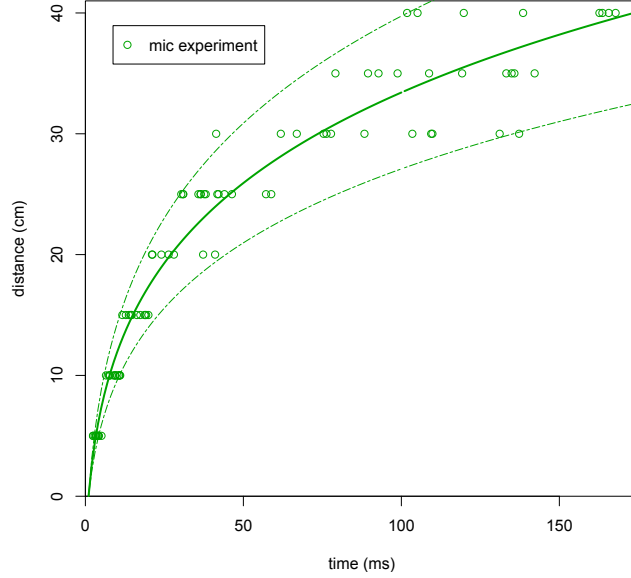


Figure 2.5: (color online) Experimental pressure fronts

50 cm away, the average intra-oral pressure was ~ 25 centimeters of water, or three times higher than normal, with high variability. This variability is largely a question of repeatability. It is almost impossible for a person to produce a repeatable mouth shape, initial air pressure, rate of decrease of air pressure, rate and degree of mouth opening, and orientation of the mouth to the microphone.

Many of these variables could not be measured, and even initial mouth pressure could not be isolated from the other variables as no significant relationship was found between rate of air travel and intra-oral pressure prior to the release burst.

Nevertheless the effect of many of these variables is known. Lower initial air pressure, faster rate of decrease of air pressure from the flow source, larger mouth opening, puff orientation away from the microphone, and per-

turbations in the air all decrease the rate of flow penetration. These effects combined can be quite significant.

2.3.2 High Speed Video Experiments

For the pilot round, three high-speed tokens were recorded, but only one was produced at a normal volume and voicing quality for an English ‘pa’ syllable. This token was selected for comparison with the numerical simulations. For the second round of recordings, all 12 recordings were produced at a normal volume and voicing quality for an English ‘pa’ syllable.

Results of measuring the leading edge of the smoke particles for each recording are shown in Figure 2.6.

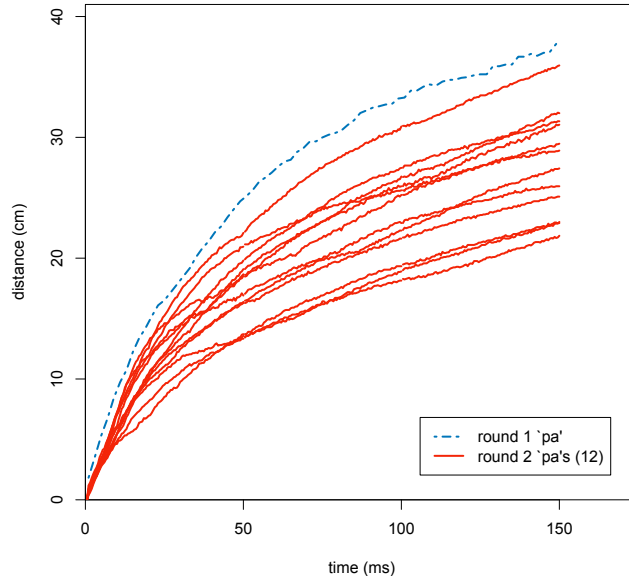


Figure 2.6: (color online) Penetration distances for high-speed video experiments

2.3.3 Numerical Simulations

The validation study (Variation 2) gives fine agreement with previous jet experiments described in the introduction, as shown in Figure 2.7. Figure 2.8

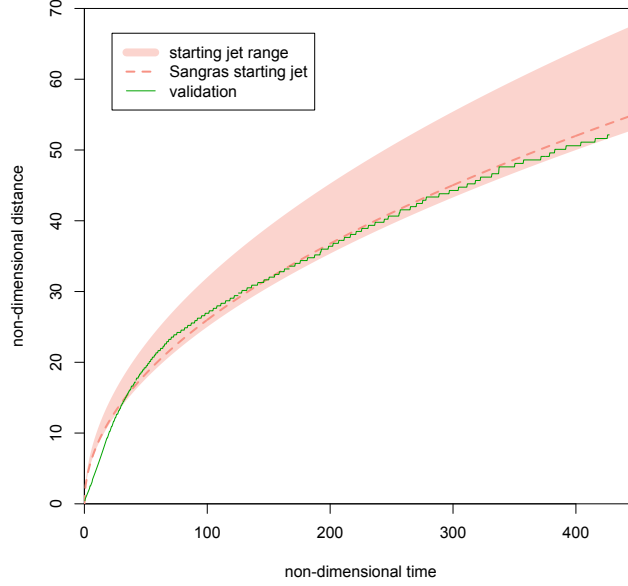


Figure 2.7: (color online) Simulation validation: starting jet range is defined by the constants reported in Sangras’ summary [11]

shows the grid refinement study (Variation 1), along with the perturbed inlet simulation (Variation 5) and the perturbed domain simulation (Variation 6). The convergence is oscillatory, but outside of the asymptotic range. See Celik (1997) [21] and Celik (2005a) [22] for discussion of oscillatory convergence and complications of LES verification respectively. A comparison of the baseline numerical simulation, the simulation of the loud utterance (Variation 3), and the constant inlet pressure simulation (Variation 4) are presented in Figure 2.9. As suggested in the introduction, 2D simulations did not yield realistic results; generally they resulted in a jet penetration rate that was too fast. The loss of the 3D geometry caused the flow to be that of a plane jet rather than a jet from a nozzle. The loss of the 3D flow meant that turbulence could not be truly modeled by LES, and the time-averaging of the RANS simulations removed flow details that are of interest. Use of 2D simulations was quickly dropped, therefore those results are not presented in detail here.

2.3. Results

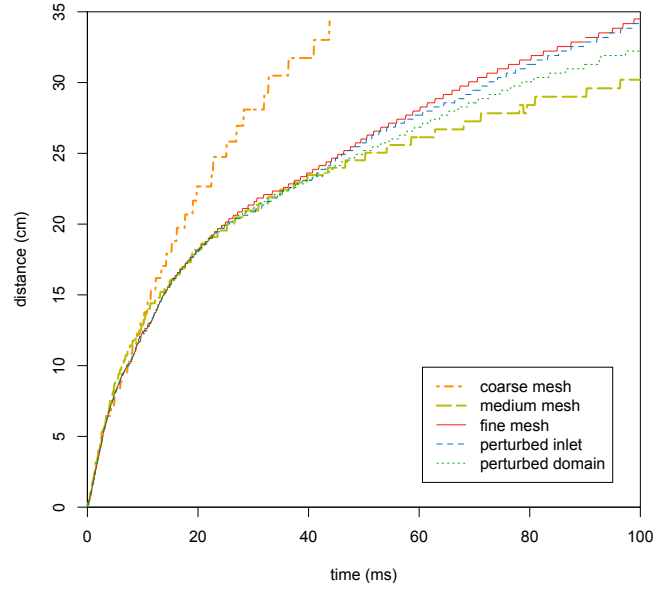


Figure 2.8: (color online) Simulation verification

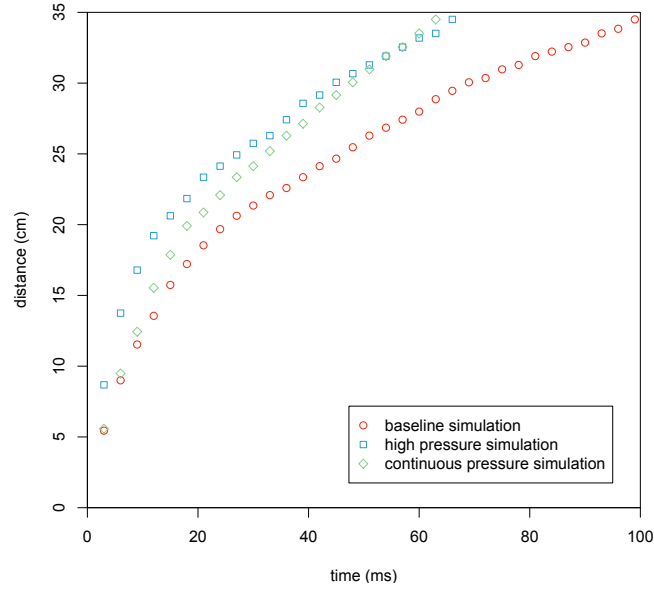


Figure 2.9: (color online) Comparison of leading particle front for baseline, high pressure onset and continuous pressure simulation

2.3.4 Comparison of Simulation to Microphone Experiment

The simulation pressure front is defined as the distance at which the absolute value of the pressure reaches $1/10^{th}$ of the maximum pressure for each time step. The simulation pressure front was compared to the results from the microphone experiment (Figure 2.10). The simulation pressure front falls within the 95% confidence interval of the experiment.

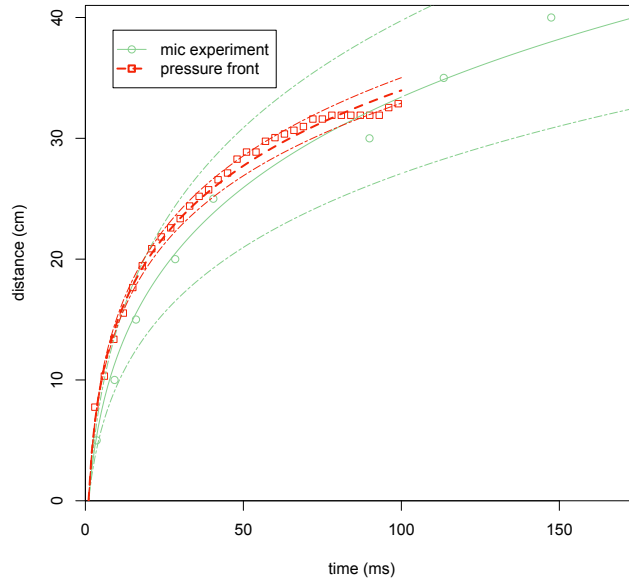


Figure 2.10: (color online) Simulation and averaged experimental pressure fronts

2.3.5 Comparison of Simulation to High Speed Video Experiment

The particle front from the high speed video recordings and the numerical simulation are compared in Figure 2.11. The graph shows the loglinear quadratic fit lines for the pilot puff ($F(2,89) = 1.125e+05$, $p < 0.001$, adjusted $R^2 = 99.9\%$), second round puff average ($F(3,3598) = 7.479e+04$, $p < 0.001$, adjusted $R^2 = 97.7\%$), and numerical simulation ($F(2,1208) = 1.816e+06$, $p < 0.001$, adjusted $R^2 = 99.9\%$). A comparison graph be-

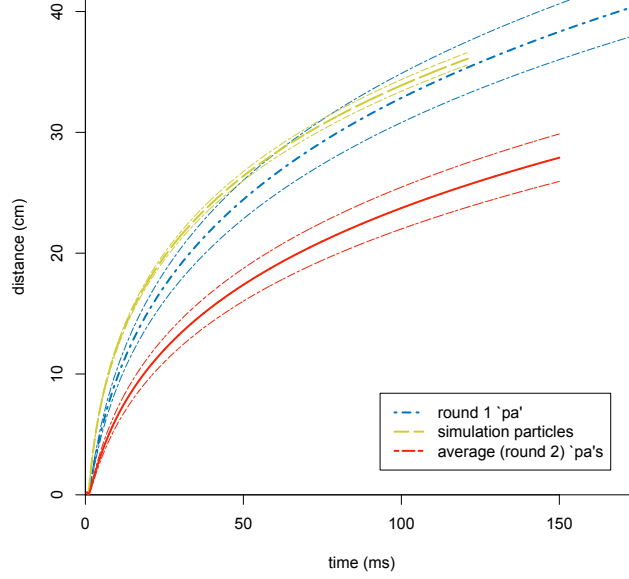


Figure 2.11: (color online) Simulation and high speed video particle fronts

tween the velocities over time of the particle front from the high speed video recordings and the numerical simulation appears in Figure 2.12. Note that the differences decrease dramatically after 40 ms, as shown in the inset graph of Figure 2.12.

There is a strong negative relationship between the rate of lip opening and leading particle edge distance travelled for the first 20 ms, decreasing after 30 ms and losing significance by 40 ms, as shown in Figure 2.13. Both the significance and the t-value of the partial regression coefficient decrease over time as the leading edge of the puff moves away from the mouth opening. The results can be seen in Table 2.1

Images were aligned such that the times at which the high speed film's particle flow penetrate 5, 10, 15, 20, 25, 30 and 35 cm are matched with the same times in the simulation. Because frames are spaced 2 ms apart, the first frame with visible particle flow is assumed to occur ~ 1 ms after lip opening. This time averaging, combined with the observation that the simulation flow rate matches closely, but not exactly, with the high-speed

2.3. Results

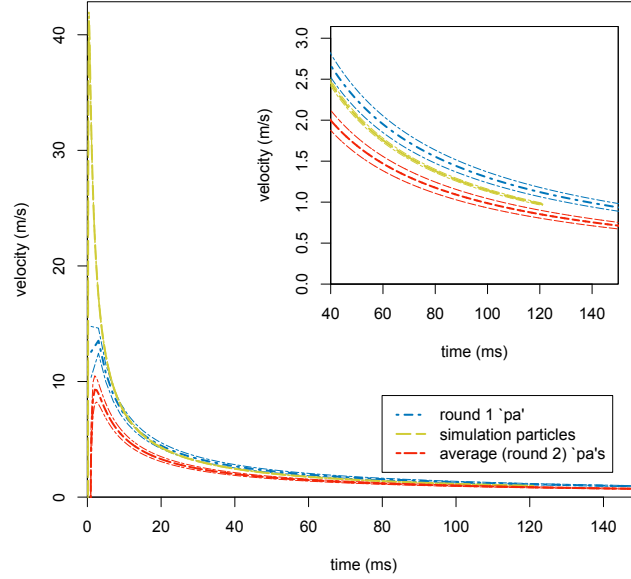


Figure 2.12: (color online) Velocity of simulation and high speed video particle fronts

coefficient	time span	Estimate	Std. Err.	t	p
puff travel	10 ms	-1.69	0.34	-4.98	* < 0.001
distance by	20 ms	-0.81	0.18	-4.46	* < 0.001
lip opening	30 ms	-0.36	0.11	-3.18	* = 0.002
width	40 ms	-0.09	0.08	-1.03	0.304

Table 2.1: Partial regressions of the interaction between the leading particle edge and lip opening averaged over 10, 20, 30 and 40 ms

2.3. Results

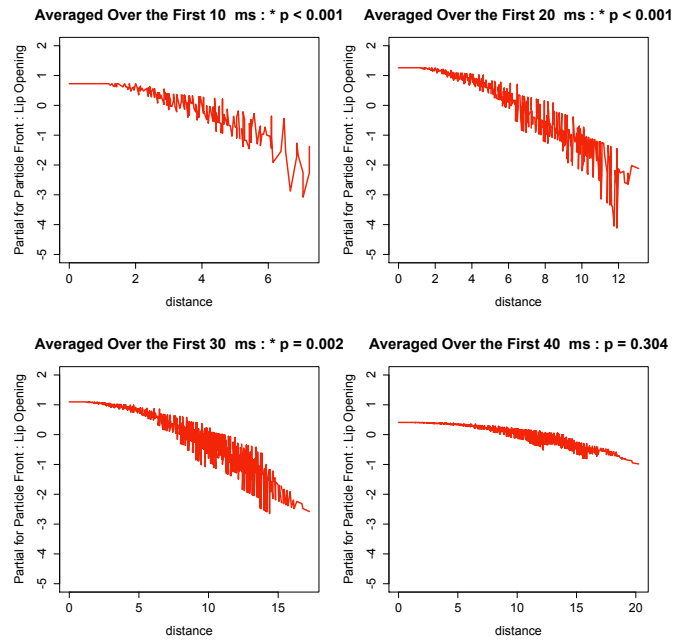


Figure 2.13: (color online) Negative partial regression between the width of lip opening and leading edge distance

2.4. Discussion

time (ms)	particle distance (cm) by data source	
	high speed video	simulation
5	5.3	8.1
11	9.7	12.8
21	15.2	18.5
35	20.1	22.3
51	25.1	26.3
75	30.0	31.0
121	35.0	>34.8

Table 2.2: Time alignment by distance for Figure 2.14

video, creates distance alignment differences. As a result, the images do not align by particle front distance, and the differences can be seen in Table 2.2. The velocity field, instead of the particle field, is shown because Fluent does not export the particle data in a usable format and because the particle field can be inferred from the velocity field. In the high speed video, most of the smoke is expelled in the first 30 ms, so the air expelled after that time is not as visible in the video frames.

A graph of the simulated airflow velocity as a function of time in which each curve shows the velocity at a particular distance from the front of the orifice is shown in Figure 2.15. The data are spatially averaged over a 1cm radius in the xz plane and 2.1ms in time. These lines reveal velocity oscillations around 100 Hz that were not smoothed out by the averaging. The oscillations are caused by large eddies in the flow that are resolved by the LES simulations, but which would not have been resolved with a RANS simulation.

2.4 Discussion

These results show some significant discrepancies between the microphone experiment, the high speed video experiment, and the simulations, but upon examination, these errors make sense in light of the assumptions and experimental methods used.

2.4. Discussion

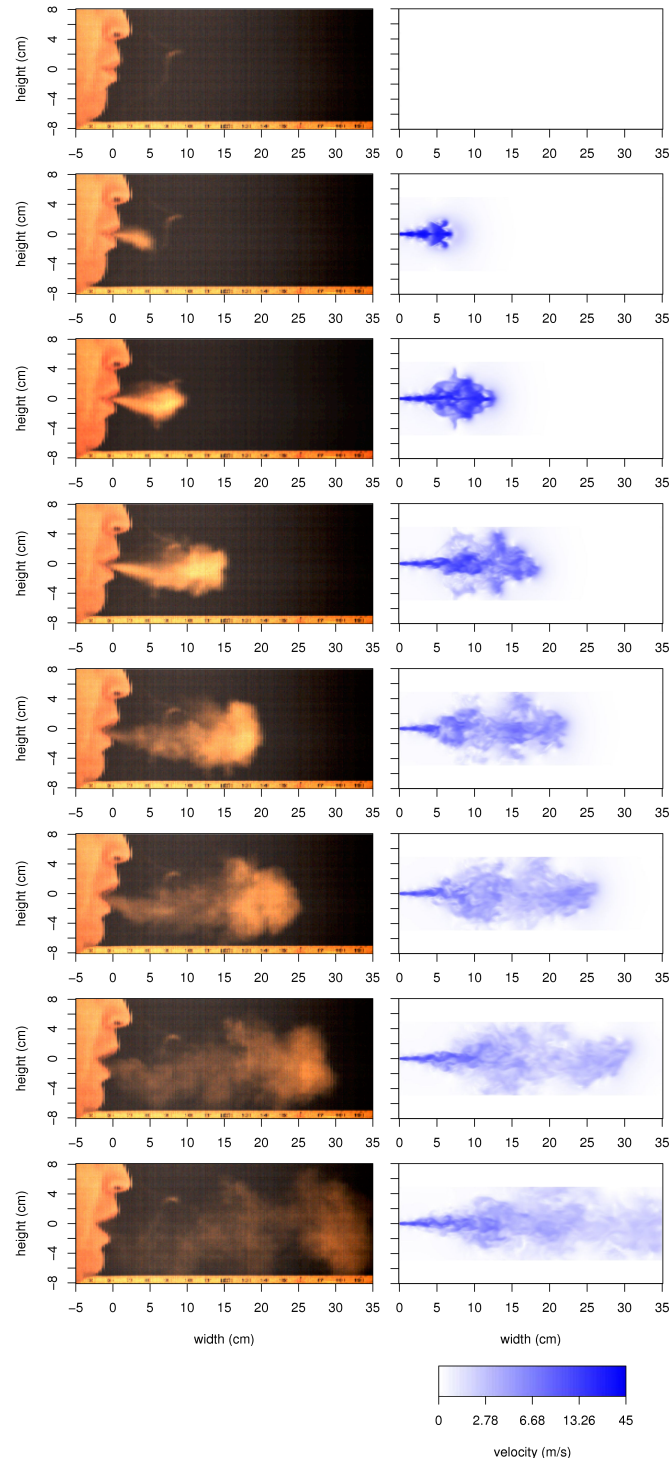


Figure 2.14: (color online) ‘Pa’ on high speed video (left) compared with numerical simulation velocity field (right). From top to bottom, the time (in ms) of the image is: 0, 5, 11, 21, 35, 51, 75 and 121.

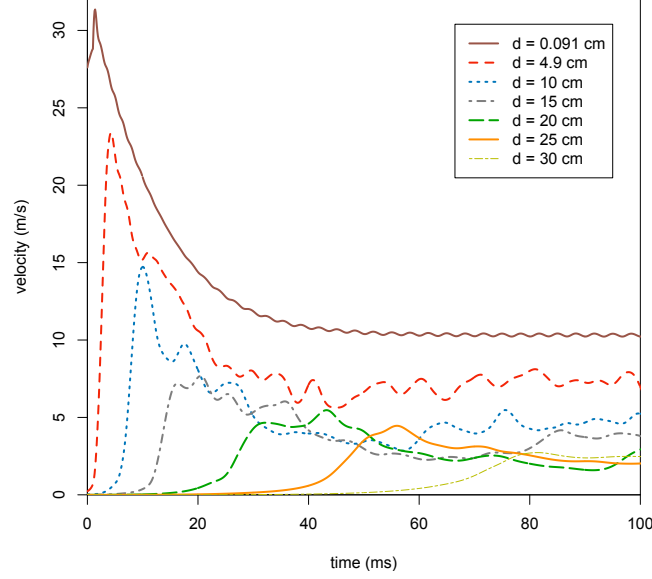


Figure 2.15: (color online) Airflow velocity over time based on the distance from orifice aperture

The microphone experiments were expected to show faster penetration than normal because the average intra-oral pressure was three times higher than normal, which was needed to attain good recordings. This impact, however, was not expected to be too large because velocity scales with the square root of pressure, as derived from Bernoulli's principle. Also, the air-pressure measurements showed that it quickly fell to the normal level predicted in Stevens book [14]. Therefore, the effect of the intra-oral air pressure would be less than one might expect, and the differences would be most significant at distances closest to the mouth.

The microphone experiment had a high variance due in part to the difficulty in capturing the microphone pops, especially at increasing microphone distances. Because of this high variance, the simulations fell within the range of results from the microphone experiment.

The high speed video experiments, on the other hand, were captured at pressures reasonable for speech, and were deemed trustworthy.

The measurements taken from the high-speed video were much more

accurate than those taken from the microphone pop experiment as there were no visual artifacts interfering with the visibility of the leading edge of the smoke comparable to the interference of the acoustic waves on the capture of microphone pops. Most of the variability in recorded results was seen in the rate of particle penetration during the first 40 ms, and could be largely attributed to the rate of lip opening during the first 20 ms. The faster the lips opened, the slower the initial penetration. Differences within a few tenths of a millimeter over a few milliseconds were significant.

The strong negative relationship between the rate of lip opening and velocity of the jet's leading edge makes sense if one considers a constant flow rate through the geometry behind the lips. The viscous effects of the boundary layer will slow the penetration rate, yet may also cause a higher velocity in the core of the jet (out of the boundary layer).

As discussed in the introduction, there were a number of simplifying assumptions made for the simulation, particularly concerning the mouth. Not including the lips in the model meant that the boundary layer effects were not modeled; these effects slow the jet. Pelorson et al.[7] discuss the dominant role of viscous effects at the lips in the first milliseconds of a plosive, and Fujimura [23] also emphasizes the rapidity of the change in the first 10ms. Figure 2.12 shows that most of the simulation error occurs in the first 10-20ms of the burst where the simulation velocity is much higher than experiment, and the data in Figure 2.13 and Table 2.1 confirm that the simulations suffer this error. This error largely accounts for the differences between the high speed video experiment and the simulation.

2.4.1 Future Work

The velocity data in this paper can be used to identify the maximum distance a perceiver can be separated from a speaker and still detect puffs of air from labial plosives during their speech (though the minimum velocity at which skin receptors can detect air flow is as yet unknown), or as a basis for identifying the minimum distance a microphone needs to be from a speaker based on the microphone's sensitivity to air-flow velocity.

While the simulations and the experiments match closely after 40 ms, the simulations predict faster airflow at the onset of the puff than found in the experiments. This difference was partially related to the fact that the mouth shape expands during the production of the ‘pa’ syllable, but not in the simulation. Simulation of the change in oral aperture size would require changing the mesh throughout the simulation. This would be a challenging problem for further research. In addition, mesh and time step refinement may improve the quality of the simulations.

2.5 Conclusion

The results show that the hypotheses regarding the need for 3D LES simulation with a mouth-shaped orifice and decreasing air pressure at the orifice are all reasonably valid for the accurate simulation of air-flow after the release of an aspirated labial plosive. While the static elliptical orifice provided an adequate basis for simulation, the static and anatomically incorrect mouth shape contributed to the observed discrepancies in the results. Simulations involving a change in the orifice shape throughout the simulated time period, corresponding to known mouth shape changes in the production of labial plosives, may solve this discrepancy.

By validating air-flow simulations with experimental data, it is possible to plot mean velocity in time as a function of downstream distance. This information can be used with experimental data to identify the distance away from the orifice or the time from the beginning of a speech release burst at which a person can perceive the air-flow or a given microphone can pick up a ‘pop’.

These results provide the groundwork upon which future research in microphone manufacturing, sound engineering, speech perception research and aerodynamic modeling of speech may be conducted.

2.6 Acknowledgements

The authors appreciate the financial support of NSERC and the advice and guidance provided by Professors Sid Fels and Kees van den Doel. We also wish to thank Laurie McCleod and Walker Peterson for their help in conducting this research.

2.7 Bibliography

- [1] Leigh Lisker and Arthur S. Abramson. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20:384–422, 1964.
- [2] G. W. Elko, J. Meyer, S. Backer, and J. Peissing. Electronic pop protection for microphones. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 46–49, October 21-24 2007.
- [3] M. Schneider. Transients in microphones: Pop and impulse. In *Proceedings of the Audio Engineering Society UK Conference: Microphones & Loudspeakers*, 1998.
- [4] M. J. Lighthill. The bakerian lecture, 1961. sound generated aerodynamically. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 267(1329):147–182, 1962. Article type: Full Length Article, Full publication date: May 8, 1962 (19620508). Copyright 1962 The Royal Society.
- [5] J. E. Williams. Hydrodynamic noise. *Annual Review of Fluid Mechanics*, 1:197–222, 1969.
- [6] John R. Westbury and Michiko Hashi. Lip-pellet positions during vowels and labial consonants. *Journal of Phonetics*, 25(4):405–419, October 1997.
- [7] X. Pelorson, G. C. J. Hofmans, M. Ranucci, and R. C. M. Bosch. On the fluid mechanics of bilabial plosives. *Speech Communication*, 22(2-3):155–172, —1997—.
- [8] J. O. Dabiri and M. Gharib. Starting flow through nozzles with temporally variable exit diameter. *Journal of Fluid Mechanics*, 538:111–136, 2005.
- [9] E. J. Gutmark and F. F. Grinstein. Flow control with noncircular jets. *Annual Review of Fluid Mechanics*, 31:239–272, 1999.

2.7. Bibliography

- [10] R. S. Miller, C. K. Madnia, and P. Givi. Numerical-simulation of non-circular jets. *Computers & Fluids*, 24(1):1–25, 1995. 34 PERGAMON-ELSEVIER SCIENCE LTD.
- [11] F. J. Diez, R. Sangras, O. C. Kwon, and G. M. Faeth. Self-preserving properties of unsteady round nonbuoyant turbulent starting jets and puffs in still fluids (vol 124, pg 460, 2002). *Journal of Heat Transfer-Transactions of the ASME*, 125(1):204–205, 2003. 1.
- [12] S. C. Crow and Champagne F. H. Orderly structure in jet turbulence. *Journal of Fluid Mechanics*, 48(AUG16):547–, 1971. 37 CAMBRIDGE UNIV PRESS.
- [13] Kenneth Stevens. Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America*, 50(4B):1180–1192, 1971.
- [14] Kenneth Stevens. *Acoustic Phonetics*. MIT Press, Cambridge, Massachusetts, 2000.
- [15] S. J. Kwon and I. W. Seo. Reynolds number effects on the behavior of a non-buoyant round jet. *Experiments in Fluids*, 38(6):801–812, 2005. 18 SPRINGER.
- [16] R. Sangras, O. C. Kwon, and G. M. Faeth. Self-preserving properties of unsteady round nonbuoyant turbulent starting jets and puffs in still fluids. *Journal of Heat Transfer-Transactions of the ASME*, 124(3):460–469, 2002. 32.
- [17] F. J. Diez, R. Sangras, G. M. Faeth, and O. C. Kwon. Self-preserving properties of unsteady round buoyant turbulent plumes and thermals in still fluids. *Journal of Heat Transfer-Transactions of the Asme*, 125(5):821–830, —2003—.
- [18] R. S. Reichert and S. Biringen. Numerical simulation of compressible plane jets. *Mechanics Research Communications*, 34(3):249–259, 2007.

2.7. Bibliography

- Reichert, R. S. Biringen, S. 32 PERGAMON-ELSEVIER SCIENCE LTD.
- [19] S. Stanley and S. Sarkar. Simulations of spatially developing two-dimensional shear layers and jets. *Theoretical and Computational Fluid Dynamics*, 9(2):121–147, 1997. 65 SPRINGER VERLAG.
- [20] P. J. Roache. Quantification of uncertainty in computational fluid dynamics. *Annual Review of Fluid Mechanics*, 29:123–160, 1997. 101 ANNUAL REVIEWS INC.
- [21] I. Celik and O. Karatekin. Numerical experiments on application of richardson extrapolation with nonuniform grids. *Journal of Fluids Engineering-Transactions of the ASME*, 119(3):584–590, 1997. 17 ASME-AMER SOC MECHANICAL ENG.
- [22] I. B. Celik, Z. N. Cehreli, and I Yavuz. Index of resolution quality for large eddy simulations. *Journal of Fluids Engineering-Transactions of ASME*, 127:949–958, 2005.
- [23] Osamu Fujimura. Bilabial stop and nasal consonants: a motion picture study and its acoustical implications. *Journal of Speech and Hearing Research*, 4(3):233–247, —1961—.

Chapter 3

Computational Aeroacoustic Simulations of the English Fricative /sh/

3.1 Introduction

Fricatives are produced when air is channeled through a constriction, thus forming a jet, which strikes an obstacle and produces sound. The fricative may combine sound produced at the vocal chords with the sound produced at the constriction (a voiced fricative), or it may be only the sound produced at the constriction (a voiceless fricative). While a fricative can be produced with a relatively static vocal tract, in the context of normal speech the fricative occurs between other speech sounds, and thus includes vocal tract dynamics.

Fricatives cannot be understood without understanding the fluid dynamics in the vocal tract, which are themselves not completely understood because they involve turbulent flow through a complex geometry. In addition, there is no complete theory of sound generation by turbulence. However, some general principles are known which makes this problem approachable.

When the flow passes through the constriction it greatly increases in velocity and becomes turbulent either around the constriction or when it

A version of this chapter will be submitted for publication.

Anderson, P. and Green, S. and Fels, S. Computational Aeroacoustic Simulations of the English Fricative /sh/

strikes the obstacle. The unsteady turbulent flow generates sound in a number of ways. The dominant sound source typically comes from the forcing between the fluid and the obstacle it strikes (a dipole sound source). However, sound may also be generated by the fluid having an unsteady flow rate in the constriction (a monopole sound source) or by shear within the turbulence itself (a quadrupole source) [1, 2]. Once the sound is created it will propagate through the remainder of the vocal tract, being modified along the way by resonators such as the sublingual cavity, and then escape past the lips to free space where it may be detected by an ear or microphone. The location of the sound creation and the nature of the modification are still a matter of recent studies [3].

Part of the problem in understanding fricatives comes from our lack of understanding of sound. Sound is a component of fluid flow, and as such is described by the Navier-Stokes equations [4]. However, an unsteady flow also has pressure variations that respond to the changing momentum in the flow [5]. Such pressure fluctuations (often called pseudo-sound) make it difficult to separate the propagating sound field from the rest of the flow, and in fact there is no exact method known to isolate the sound component from the rest of the flow. Therefore, aeroacoustic calculations (such as the acoustic analogy [6]) require various approximations and assumptions, often with a dubious basis [7, 8].

Despite this fundamental problem, we can still learn much about acoustics through computational aeroacoustics (CAA). In CAA, as in CFD, one simulates the Navier Stokes equations (or derived equations using simplifying assumptions), but in CAA one takes extra measures to insure that the sound field is adequately resolved and propagated. The particular challenges that CAA faces have been discussed in great detail [9, 10, 11], so here we'll consider the most relevant issues.

Resolving sound waves requires a wide scale of resolution. In speech, the frequencies 100Hz to 12000Hz are important; these have respective periods of 0.01s to 0.000083s, and respective wavelengths of 3.4m to 0.02833m in air at room temperature. The numerical method needs to adequately resolve these temporal and spatial scales. A high order finite difference scheme may be

able to resolve a wavelength with 7 mesh points, but a second-order scheme common to CFD requires approximately 20 mesh points per wavelength [9], thus providing an upper limit on the highest frequency that can be resolved. Likewise, the time scale must be adequately resolved. However, a finer resolution requires more RAM, more hard disk storage, and a longer simulation runtime.

When the flow velocities are well below the speed of sound, the acoustic fluctuations in the flow are less than the non-acoustic pressures by orders of magnitude, thus making it a challenge to resolve the acoustic amplitudes. Upon recovering a signal that seems to be sound, one must be aware that it may contain numerical artifacts such as those caused by improper boundary conditions, or it may be pseudo-sound.

The boundary conditions of a standard CFD simulation are not adequate for CAA as they cause waves to reflect back into the domain. Therefore non-reflecting boundary conditions must be developed.

When designing a CAA simulation, one must find a balance between simulation quality and computational limitations. The quality of the simulation depends upon the factors mentioned above, but it also depends greatly on whether the simulation is 2D or 3D; whether one uses an acoustic analogy, direct method, or other methods; and the numerical integration scheme used. Despite these numerous difficulties, a CAA simulation may provide excellent flow and sound data, thus shedding light on the underlying phenomenon, and is particularly useful in cases where experiments are difficult to perform, such as the human vocal tract.

Therefore, we seek to get a better understanding of fricatives using computational fluid dynamics. In particular, we use a standard CFD software package (Fluent) to see if we can adequately simulate the English fricative ‘sh’, or /ʃ/. From there we intend to draw conclusions about the theory of /ʃ/ and other fricatives, and also to draw conclusions regarding the type and quality of simulations needed for fricative simulations.

We don’t expect Fluent to be as efficient or accurate as specialized CAA code, but that it will provide a reasonable flow simulation from which we can learn about fricatives and proper simulation methods. We expect that

2D simulations won't be adequate to capture the essence of the flow and the sound, and that 3D large eddy simulations will be needed for accurate results.

3.2 Methods

To investigate the capabilities of CAA, we choose to compare with the 'level 3' experimental case of /f/ that Shadle describes [12]. The advantage of this case is that Shadle provides experimental results for a fairly simple geometry, thus we can create a comparable simulation. The disadvantage, however, is that this is a simplified geometry, and Shadle concludes that the deviation of her results from recordings of spoken /f/ are probably due to geometrical simplifications. Thus our test case, while modeling something close to the human /f/, is expected to sound wrong as in Shadle's experiment.

In the hopes of finding a minimal yet adequate simulation method we will use a variety of simulation methods. We will perform 2D and 3D simulations, in both cases using a large eddy simulation and a simpler RANS turbulence model, the $k-\omega$ SST; chosen to combine the strengths of the $k-\epsilon$ and $k-\omega$ models. The resulting 'sound' from these simulations will be recorded using the direct method (measuring pressure change at 20cm from the mouth) and using an acoustic analogy on a variety of source surfaces. Thus we will have a range of simulation complexity.

We will perform the simulations using Fluent, a common commercial CFD software. The advantage of using Fluent is that it is readily accessible CFD, but the disadvantages are that it is designed to be a robust and general CFD package; consequently it is slow, and accuracy is compromised for the sake of stability, nor does it contain high-order schemes and boundary conditions needed for high-performance CAA.

An image of the 2D domain can be seen in Figure 3.1, which is derived from Shadle's geometry. The 3D domain is extruded 25.4mm in the third dimension and narrowed at the constriction as Shadle did. The 2D domain has 71,216 cells with boundary layer cells as fine as $\Delta = 0.01\text{mm}$ and the coarsest cells being no larger than $\Delta = 2\text{mm}$. The 3D domain has 1,262,021

3.2. Methods

cells, but computational limitations force the cells to be significantly larger, thus the finest cells were $\Delta = 0.2\text{mm}$ at the most critical flow regions and the largest cell being no larger than $\Delta = 4\text{mm}$. The inlet is defined as a mass flow inlet (alternates were later tested in 2D—see below) with a flow rate of $0.000804[\text{kg/s}]$ in 3D and $0.0316535[\text{kg/s}]$ in 2D (that is, the 3D rate with 0.0254m divided out). To attain non-reflecting boundary conditions, a buffer zone was created inside the pressure outlet which gradually damps the waves. Such a condition can itself cause reflections if not done gradually [10], and by trial and error we found adequate performance by damping pressure according to:

$$P = P - F \cdot (P - P_0) \quad (3.1)$$

where

$$F = \frac{R - r}{R}, r < R \quad (3.2)$$

where background pressure $P_0 = 0$, r is the distance from wall, and the damper width $R = 10\text{cm}$. This buffer region was implemented in Fluent with user-defined functions. It is worth noting that Fluent does provide non-reflecting boundary conditions, but they did not work with the settings required for this simulation.

The simulations were run with a constant time step of 0.00001s until the spectrum became fairly steady. The spatial and temporal integration schemes are all 2^{nd} order accurate. The flow is compressible, as is required to directly measure the sound waves. For all simulations the acoustic analogy data is recorded concurrently with the direct method, thus the two cases are comparable. The pressure probes are 20cm from the lips.

The sound reported by an acoustic analogy is calculated from the flow coincident with the source surface. One may consider different source surfaces, and thus investigate the contribution of each source surface upon the final sound. However, the acoustic analogy as implemented in Fluent, is only relevant for sound propagating to free space, thus an acoustic analogy result of sound created at the constriction will not consider the modifications that the sound will undergo between the constriction and when it escapes beyond the teeth and lips. Thus we can find the sound as generated by the source

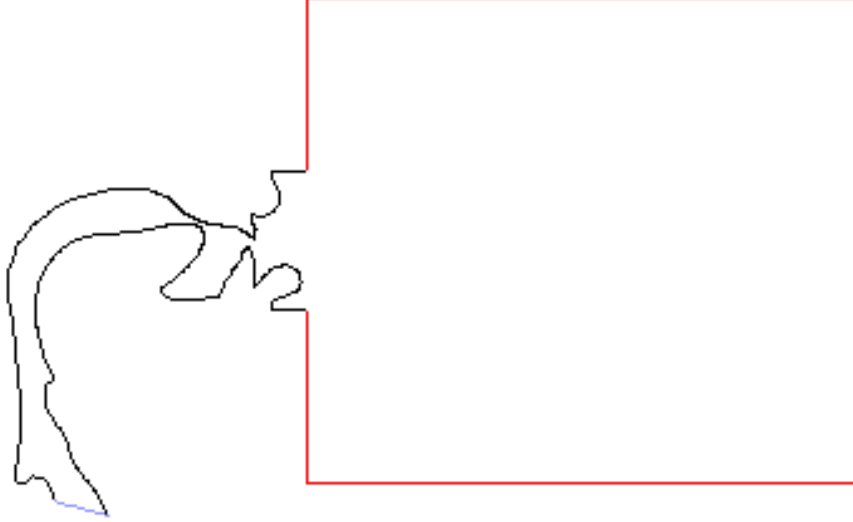


Figure 3.1: The 2D domain. The inlet is the paler line in the bottom left. The outlet is the rectangle enclosing the free space beyond the lips

surface and compare it with the sound recorded by the direct method to understand the contribution of that source surface to the final sound (Shadle did a similar study of coherence to find the important source surface). The source surfaces used are: the constriction, the cavity, the lower tooth, the upper tooth, the lower lip, and the upper lip.

The spectral analysis uses a 0.02048s hanning window (2048 data points). Ideally, these simulations would obtain about 5s of data (thus 500,000 time steps) which allows averaging many spectra to obtain a smooth spectrum, but such a long run time isn't feasible, so we obtain an averaged-spectrum with the signal, but also apply a smoothing algorithm to supplement the averaging. Figure 3.2 shows an unsmoothed and unaveraged fft compared with a smoothed and averaged fft to demonstrate the ability of this method to capture the essence of the spectrum.

Our temporal and spatial resolutions are lower than those recommended in the theory section. The time step we used only allows for 10 samples per period of a 10,000Hz wave, and the largest cell size in 3D only allows for 8.5 samples per wavelength of a 10,000Hz wave. While it is true that

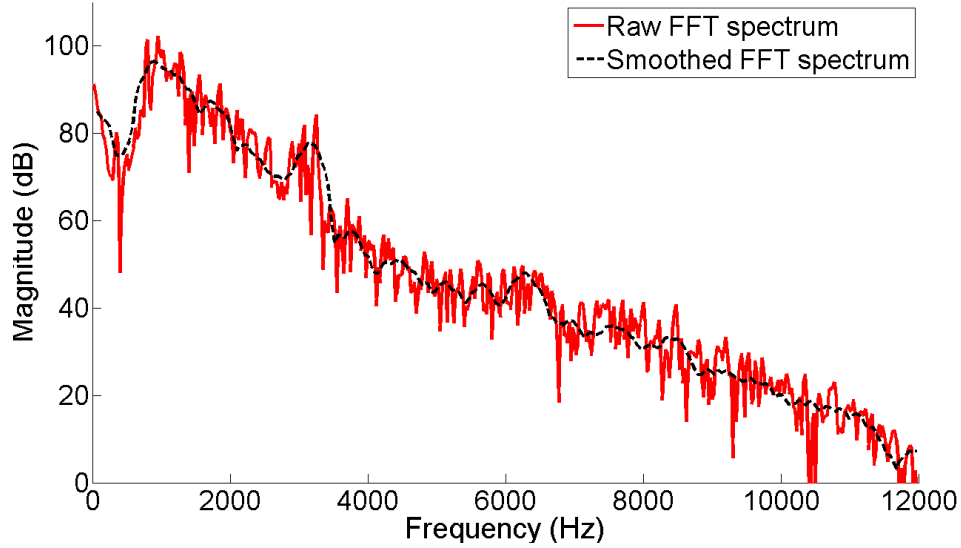


Figure 3.2: Spectrum processing.

these resolutions are well below the desired level, in some simplified tests we found them adequate for wave propagation over short distances. Thus, this simulation should adequately resolve the desired scales, but with the warning that the higher frequencies are not as well resolved as would be hoped.

When a simulation starts, just like a physical flow, it takes some time to reach its steady state. When the flow is unsteady as these simulated flows are, the flow will never reach a steady state, but it will reach a statistically steady state, which occurs when the long-term average flow is steady though it contains unsteady fluctuations. In this study, a statistically steady state is judged from the spectra. In Figure 3.3, one may see that the first spectrum (2048 samples) varies significantly from the last spectrum of the signal. One may also see the spectra that result from the averaging and smoothing (as described above), and the first spectrum that is considered to be statistically steady. Ideally the statistically steady spectrum would be taken after a longer time, but the runtime of the simulations limits this greatly.

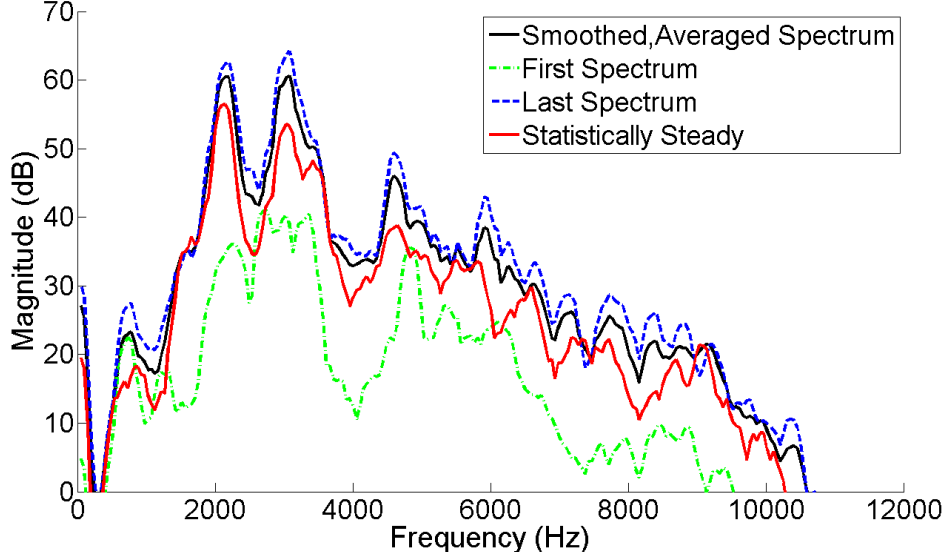


Figure 3.3: Determining a statistically steady simulation

3.3 Results

Before examining the details of the spectra from the various simulations, it is interesting to compare instances of the flow from 2D and 3D simulations. First, a pressure snapshot with the corresponding velocity snapshot from a 2D simulation is shown in Figure 3.4. One can observe the non-reflecting boundaries washing out the acoustic waves and the unphysically high pressures of the sound waves. The snapshots from the 3D simulations are shown in Figure 3.5. The flow is noticeably different, and the acoustic waves have more physically realistic values.

We can consider the results from 2D simulations, which are shown in Figure 3.6. From the first simulations, it quickly became clear that neither the mass flow inlet nor RANS simulations gave reasonable results. Thus we focused upon using a pressure inlet at a constant 800Pa rather than a mass flow inlet, and we used either LES or no turbulence model rather than RANS. We also investigated the acoustic analogy in the case of an incompressible flow. The acoustic analogy results appear much better than the direct measurement results, which are far from Shadle's result, but they

3.3. Results

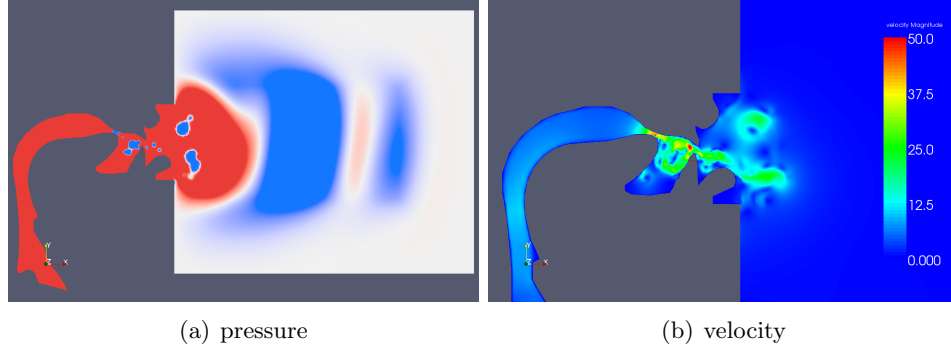


Figure 3.4: A 2D flow snapshot (pressure ranged from -15pa to 15pa so that acoustic waves may be seen, velocity ranged from 0 to 50m/s)

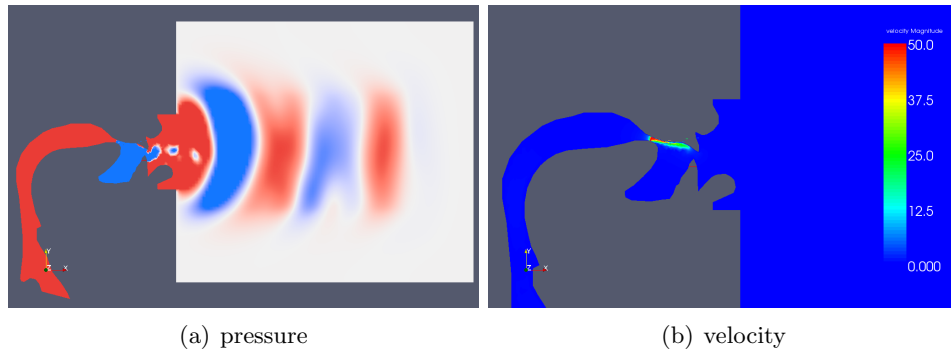


Figure 3.5: A 3D flow snapshot (pressure ranged from -1pa to 1pa so that acoustic waves may be seen, velocity ranged from 0 to 50m/s)

3.3. Results

appear to be a legitimate broadband noise signal unlike the RANS and mass-flow inlet simulations. However, the pressure can vary at a mass flow inlet, and vice versa, thus one cannot express an inlet of one type that is exactly equivalent to the other. Therefore, while 800Pa is a reasonable pressure in speech, it cannot be considered an exact comparison to Shadle’s test case. The amplitudes will be discussed later in more detail.

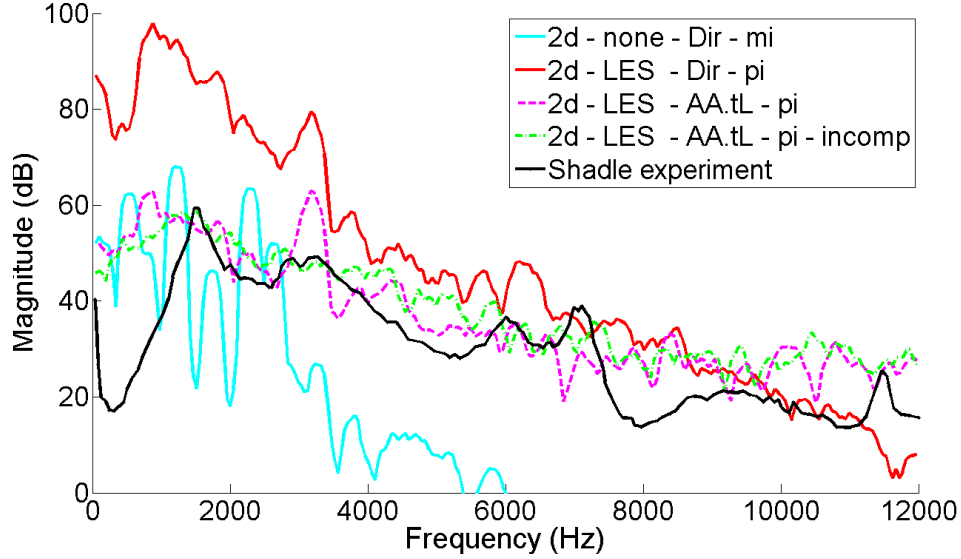


Figure 3.6: 2D simulations. None = no turbulence model, Dir = direct method, AA.tL = acoustic analogy from lower tooth; pi = pressure inlet; mi = mass inlet; incomp = incompressible

Next we can consider the results from the 3D simulations, as shown in Figure 3.7. In general, one may notice that the acoustic analogy and direct measurements match each other much better than in the 2D case, and also match Shadle’s experiments better, though they are still significantly different. One may also observe that, like the RANS in 2D simulations, the 3D RANS yields a signal that is unphysical. While the 2D results are largely limited by a geometry that is unrealistic in 2D and 2D turbulence, which is unlike the true physics of fricatives, the 3D simulations are limited by the mesh being too coarse. We therefore sought to refine the mesh in the areas critical to the flow and observe how such refinements alter the spectra. This

3.3. Results

is not a proper mesh refinement study with which to observe the convergence of the numerical solutions [13] as such a grid refinement requires all parts of the mesh to be refined. However, this refinement does indicate how the spectrum changes with better flow resolution. Not surprisingly, the refined and unrefined grid vary more in the high frequencies. These results are included in Figure 3.7.

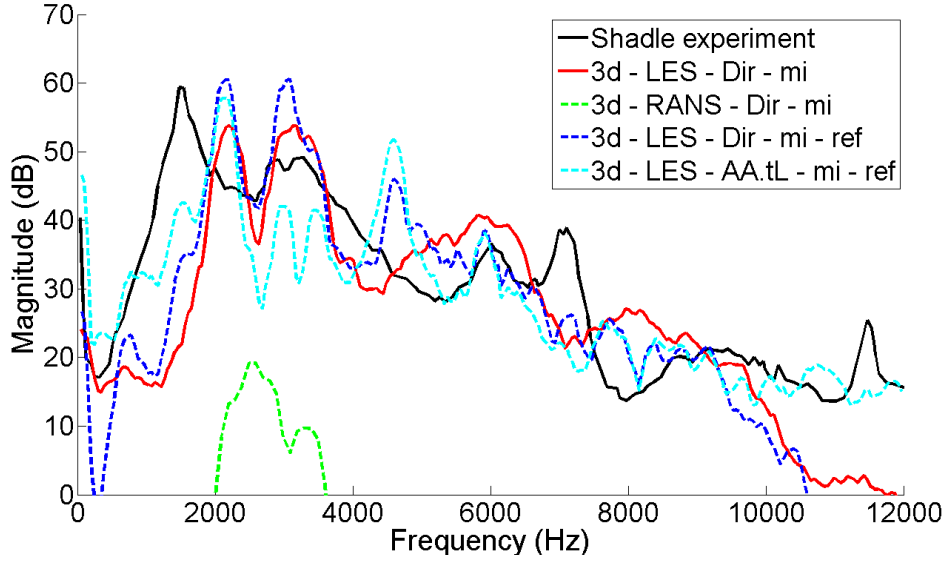


Figure 3.7: 3D simulations. Ref = refined

One may consider the acoustic analogy on two levels. First, one may compare the results from the AA to the results using the direct method, as a measure of error. Second, one may assume that the AA results are perfect, and use the results to study the coherence between the source surface and the final sound heard in the far field. We know this assumption isn't true, but one may still consider it to get a general feeling for the contributions of each sound source. To these ends, we present the acoustic analogy results from the same simulation but different source surfaces, and compare them with the sound recorded by the direct method, which can be seen in Figure 3.8.

Finally, Figure 3.9 compares the best 2D simulation results with the best

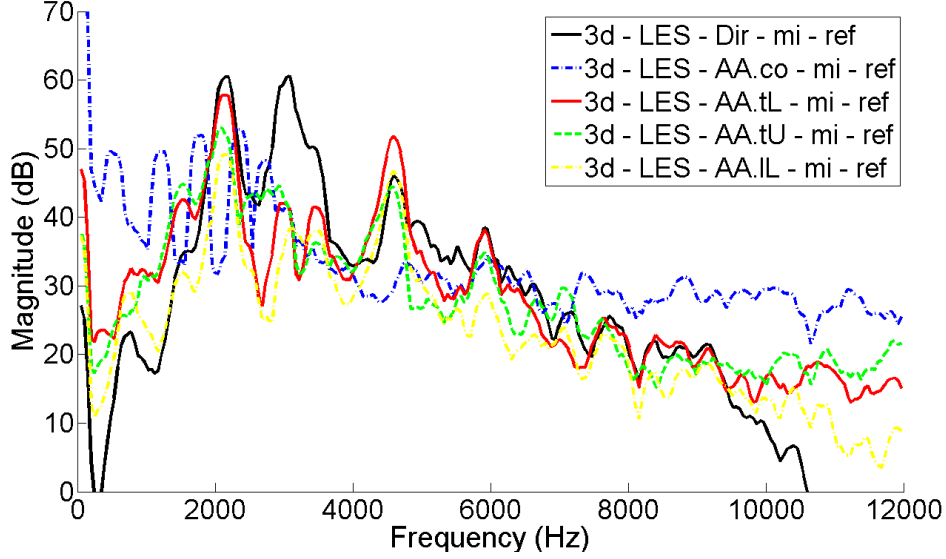


Figure 3.8: Acoustic analogy results. The acoustic analogy locations are: co = constriction, tL = lower tooth, tU = upper tooth, iL = lower lip

3D simulation results, as a side by side comparison.

3.4 Discussion

The results from the RANS simulations and the 2D mass-flow inlet simulations appear to be unphysical upon first glance, but it is good to have an objective reason why they should be discarded. When sound propagates through the vocal tract, it encounters resonators such as the cavity below the tongue and the cavity between the lips and teeth, which will resonate with a frequency range, and thus cause a distinctive peak in the spectrum [1]. The spectra from all the 2D mass-flow inlet and RANS simulations don't have the characteristics of broadband noise with a few distinguished peaks which might correspond to cavities in the vocal tract, thus they are discarded as clearly violating the physics of this flow. The 2D simulations with a pressure inlet do have peaks which may represent resonance with the cavities, and the 3D results clearly do, though it is questionable how well

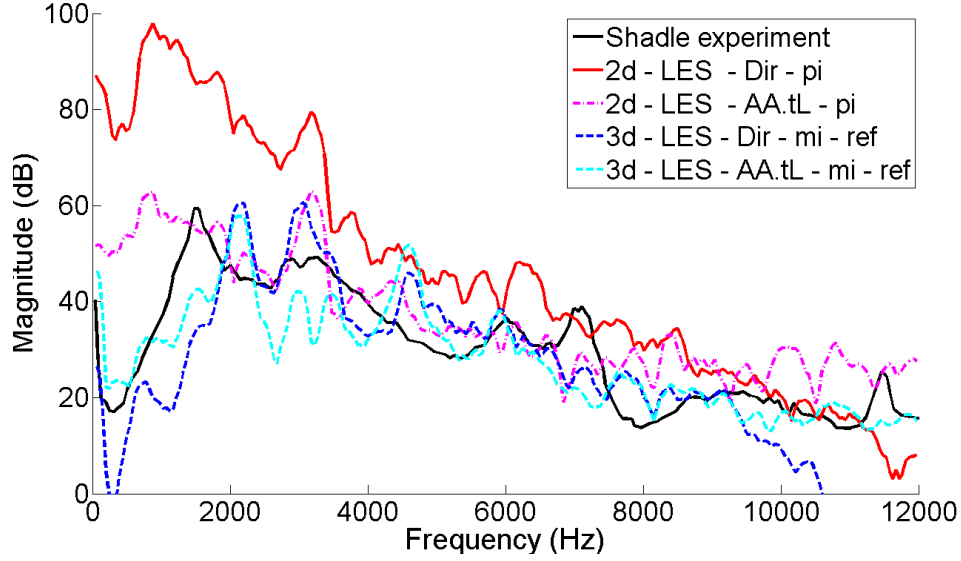


Figure 3.9: Comparison of best results

they match the experimental results.

Both 2D and 3D simulations measure the sound at the same location, and Shadle's data is scaled for distance, but the amplitudes should be viewed with some caution. The 3D mass flow inlet was designed to match Shadle's $670 \text{ [cm}^3/\text{s}]$ volume flow rate by assuming incompressibility at the inlet. The 2D mass flow inlet was scaled to match the 3D rate by dividing the third dimension out (2.54cm deep), but most simulations used a pressure inlet instead. Some ambiguity also comes from the constriction. Shadle formed a narrow constriction by filling the third-dimension with clay. The constriction shape of the 3D geometry was estimated from Shadle's description, but the 2D simulation cannot include the narrowing in the third dimension at the constriction. As a consequence the 3D constriction area is about 0.4% of the inlet area, while the 2D constriction area is about 9.6% of the inlet area (in 2D, it is really a length rather than an area). Thus the velocity increase at the constriction is not expected to be the same, and there remains some ambiguity between the 3D simulation geometry constriction shape and the experiment. As one might expect, the velocity at the 3D constriction was

observed to be much higher than at the 2D constriction, yet the sound from the 2D simulations is much higher amplitudes. This is attributed to the inability of the 2D equations to describe the energy dissipation that occurs in 3D turbulence [14], and thus is a fundamental shortcoming of 2D simulations.

In this study we are considering a static geometry, just as Shadle did in her experiment, yet it is worthwhile to consider the implications of such an assumption. First we can ask whether a static geometry is a reasonable assumption in the case of fricative generation. This is likely a safe assumption as the defining sound production mechanism comes from turbulent flow interacting with a static obstacle. Secondly, from the broad perspective of speech modeling, one must include vocal tract dynamics, as a fricative shape is just one position in a constantly transitioning vocal tract. However, successful static methods must be developed before one can hope to succeed in simulations with a dynamic vocal tract.

One issue that is difficult for simulations to handle is the material properties. Shadle’s experiments used plexiglass and clay to form the constriction, while simulations using basic wall boundary conditions will treat all walls as an acoustically hard surface. This may have caused discrepancies between our results and Shadle’s, making this validation less certain. However, this will be a bigger concern when trying to simulate a true fricative, because the flesh walls of a true vocal tract will increase the bandwidth of the cavities and cause energy losses as a function of the frequency [1]. To simulate this properly would require specialized boundary conditions at all of the walls.

From Figure 3.8 one may observe that the acoustic analogy using either the upper tooth, lower tooth, or the lower lip as the source surface matches the direct recordings quite well. From this one might draw some useful conclusions. First, the sound is very close to its final form at the teeth and the acoustic analogy is able to capture this. Second, because the acoustic analogy can replicate the sound from direct measurement (in 3D simulations), there is little need to extend the domain far beyond the lips. One can consider a fictitious source surface just outside of the lips and propagate the sound to the far field. This allows the domain and nonreflecting boundaries

to be much smaller. One should be cautious in choosing the acoustic analogy surface because the method of propagation in Fluent only propagates to free space and will not consider any obstacles that lie between the source surface and the receiver (as mentioned in methods).

Though the acoustic analogy does fit the direct measurement quite well, there are two exceptions worth noting. First, from about 9500Hz and above the direct method spectrum drops in amplitude while the acoustic analogy stays roughly the same. This is quite likely an indication that the mesh and time step were too coarse to adequately resolve those frequencies, thus one should trust the acoustic analogy results more. This failure in the higher frequencies was forecast and discussed in the methods. Second, there is a distinct peak at 3000Hz which the direct method picked up but none of the acoustic analogy surfaces recorded. This peak may represent a quadrupole sound source, which is sound created by stresses within turbulence rather than sound created by turbulence interacting with a surface. The AA source surfaces will not account for this quadrupole noise because they are calculated from an impermeable source surface [17]. However, in such a flow the quadrupole contributions are expected to be small, and it may be that this peak is a numerical artifact, such as domain resonance.

In acoustics one often defines the acoustic far field as one wavelength from the source. Assuming that the non-linear flow has little effect in the far field, one may improve a CAA simulation by using more efficient equations (such as a high order linearized Euler equation scheme) to calculate sound propagation, or one may end the domain and use a theoretical equation to find the sound at a point deeper in the far field, thus decreasing computational expense.

In this study the recordings were taken at 20cm, but for the sake of investigating the far field in the simulation results, Figure 3.10 shows a measurement at 10cm compared with the measurement at 20cm. One may note that from about 4500Hz and above, the two spectra stay close to parallel, an indication that at 10cm from the mouth, this frequency and higher ones can be considered in the far field. The frequency 4500Hz has wavelength $\lambda = 7.5$ cm, thus this is slightly more than one wavelength distant from the

source, but still not an unreasonable estimate of the far field. If treated as a point source, the decibel amplitude of sound pressure should decrease 6dB for each doubling of distance from the source. In Figure 3.10 the difference is around 7dB which is still reasonably close to the expected value.

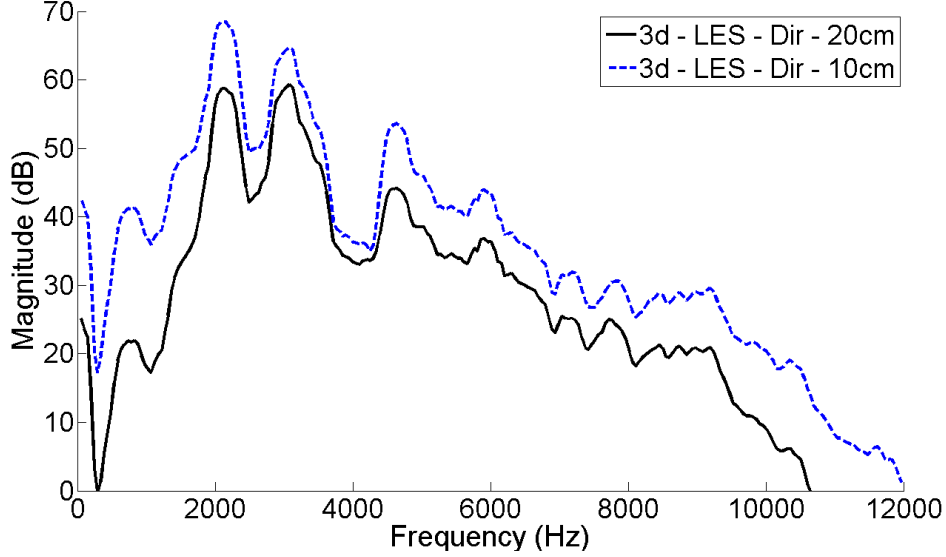


Figure 3.10: Investigation of far field location

In all the methods used, the $k - \omega$ SST model failed to find a reasonable spectrum. Also, an initial simulation with the $k - \epsilon$ turbulence model showed a similar behavior to the $k - \omega$, which speaks against the usefulness of turbulence models in fricative modeling, although there are other models to consider. In the case that the RANS model is run on the same grid as the LES, the RANS model actually runs slower. The benefit of a turbulence model is that it should give reasonable results on a coarser grid and thus run faster, but in these cases the RANS model was run on the same mesh as the LES and it ran slightly slower and yielded worse results.

To give an idea of simulation runtimes, the 3D RANS simulation took 338s per time step and the LES took 304s per time step (both parallel processing on 3 cores). The 2D LES on a single core took 15.9s per time step. Running a 2D simulation without a turbulence model offered a large increase

in speed, while changing the flow from compressible to incompressible offered a smaller speed increase.

In 2D we ran simulations using the LES turbulence model, and with no turbulence model. A comparison of these simulations is shown in Figure 3.11. Using no turbulence model is presumably a DNS, which requires a very fine mesh. While the 2D mesh was not fine enough for a DNS, it is worthwhile to note that the spectra between these two simulations are very similar. This implies that the subgrid turbulence model, and the wall model that LES uses, have very minimal contributions to this flow. For the sake of a faster simulation, one may consider not running a turbulence model at all in 2D simulations.

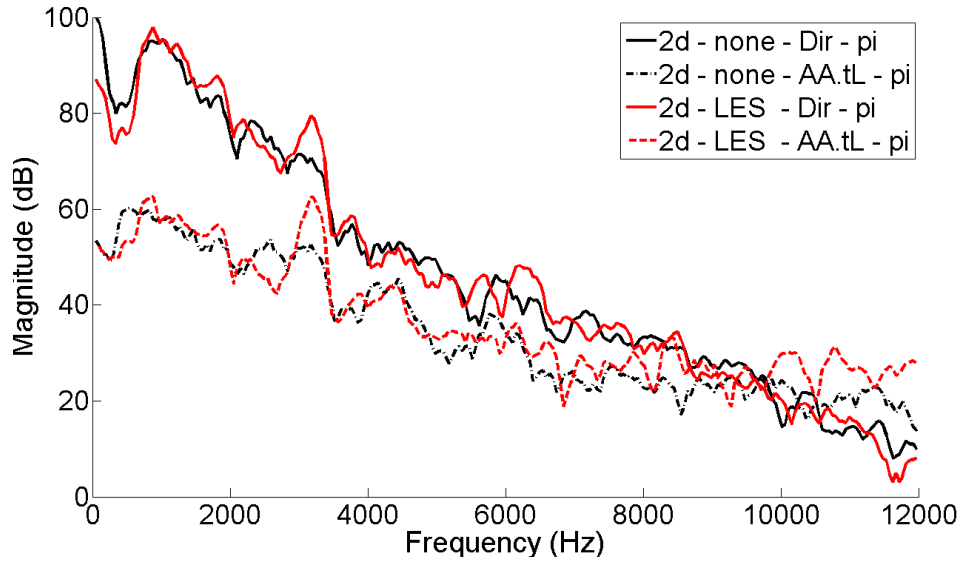


Figure 3.11: Comparison of LES with no turbulence model

While these simulations are primarily compared to Shadle’s experiments, we can still make a statement concerning 2D simulations of the true geometry. Because the 2D geometry is derived from a midsagittal X-ray of the vocal tract, and because a 2D simulation can never include the true 3-dimensionality of the vocal tract, we may consider the 2D simulation geometry as good as it can get. Thus it is reasonable to compare these results

not just with Shadle’s experiment, but with a true /f/, which is shown in Figure 3.12. The simulation results bear little resemblance to the spoken /f/; however, one might also note how little resemblance Shadle’s spoken /f/ has to the experimental /f/ and in comparison with Fant’s /f/ (as given in [15]).

The great variability that occurs in the spoken /f/ accentuates the need to understand what features of /f/ cause it to be understood properly. In other words, amidst this great variation, how do the listeners properly perceive /f/ to be such? Simulating and modeling /f/ will certainly be limited until we understand the defining characteristics of /f/. Furthermore, until this is understood one doesn’t know what features to look for in a simulation to compare how close the result is to a real /f/, thus validation through listening has an important role.

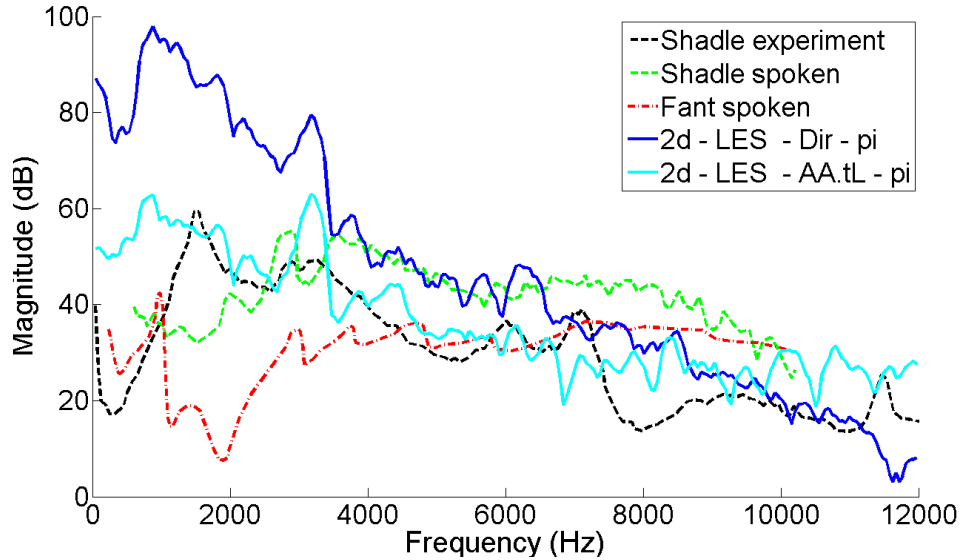


Figure 3.12: Comparison of 2D simulation with spoken /f/

3.5 Conclusion

While these simulations don't provide as close a fit with Shadle's experimental data as hoped, numerous observations have been made concerning the strengths and weaknesses of the simulations which may be applied in future attempts to model fricatives computationally. First, they demonstrate the inability of the $k - \omega$ sst model (and quite likely all RANS models) to find a reasonable spectrum. Second, they demonstrate the superiority of 3D simulations to find a physically reasonable spectrum. Third, they demonstrate how suitable non-reflecting boundaries may be created in Fluent. They also show that an acoustic analogy may offer reasonable results, and can likely be used to simplify the computational domain in future simulations.

From the observations of these simulations, we can make recommendations for future fricative simulations. First, a 3D geometry should be used, but the domain can be significantly truncated. The crucial flow features occur at the constriction in the vocal tract, thus one might start the domain further up the from the vocal chord allowing just enough distance between the inlet and the constriction for the flow to fully develop. The domain can also be truncated beyond the lips. Soon after the lips the sound can be considered to come from a simple sound source and can be propagated to a further distance using a theoretical approach. Truncating the domain will save many mesh cells; however, those mesh cells should be used to obtain better flow resolution around the constriction and the teeth. Because the constriction creates a strong jet down the midsagittal plane, the highest flow gradients and important flow features occur here. Thus one should concentrate more cells in the midplane of the domain. The wall should be meshed in much finer detail, preferably enough to resolve the boundary layer without a wall function (see [16, 17] for further discussion). Ideally, such a simulation would be done with specialized CAA code. Non-reflecting inlet and outlet boundaries using a sophisticated method such as those discussed in [9, 10] should be implemented, and will be of smaller computational expense than the large damper used in this study.

In such a simulation, one might consider numerous AA surfaces. Rather

3.6. *Acknowledgements*

than including the whole tooth or lip as a source surface, one might divide these surfaces into small sections to investigate the dipole source locations in finer detail. Also, it would be helpful to place a permeable source surface in front of the lips to account for quadrupole sources. A carefully developed simulation with these characteristics should improve upon the simulations presented in this study, and is a recommended next step.

3.6 Acknowledgements

Thanks to Sid Fels and the Artisynth project for support. Thanks to Donald Derrick for discussions.

3.7 Bibliography

- [1] Kenneth Stevens. *Acoustic Phonetics*. The MIT Press, Cambridge, 2000.
- [2] M. J. Lighthill. The Bakerian Lecture, 1961. Sound Generated Aerodynamically. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 267(1329):147–182, 1962.
- [3] M. S. Howe and R. S. McGowan. Aeroacoustics of [s]. *Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences*, 461(2056):1005–1028–, 2005.
- [4] D. G. Crighton. Acoustics As A Branch Of Fluid-Mechanics. *Journal of Fluid Mechanics*, 106(MAY):261–298–, 1981.
- [5] J. E. F. Williams. Hydrodynamic Noise. *Annual Review of Fluid Mechanics*, 1:197–&–, 1969.
- [6] M. J. Lighthill. On Sound Generated Aerodynamically. I. General Theory. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 211(1107):564–587–, 1952.
- [7] A. T. Fedorchenko. On some fundamental flaws in present aeroacoustic theory. *Journal of Sound and Vibration*, 232(4):719–782–, 2000.
- [8] C. K. W. Tam. Computational aeroacoustics examples showing the failure of the acoustic analogy theory to identify the correct noise sources. *Journal of Computational Acoustics*, 10(4):387–405–, 2002.
- [9] C. K. W. Tam. Computational aeroacoustics: An overview of computational challenges and applications. *International Journal of Computational Fluid Dynamics*, 18(6):547–567–, 2004.
- [10] T. Colonius and S. K. Lele. Computational aeroacoustics: progress on nonlinear problems of sound generation. *Progress in Aerospace Sciences*, 40(6):345–416–, 2004.

3.7. Bibliography

- [11] M. Wang, J. B. Freund, and S. K. Lele. Computational prediction of flow-generated sound. *Annual Review of Fluid Mechanics*, 38:483–512–, 2006.
- [12] Christine H. Shadle. Articulatory-Acoustic Relationships In Fricative Consonants. In *Speech Production and Speech Modelling*, pages 187–209–. Kluwer Academic, Netherlands, 1990.
- [13] P. J. Roache. Quantification of uncertainty in computational fluid dynamics. *Annu. Rev. Fluid Mech.*, 29:123–160, 1997.
- [14] Pijush K. Kundu and Ira M. Cohen. *Fluid Mechanics*, volume 2. Academic Press, San Diego, 2002.
- [15] Gunnar Fant. *Acoustic Theory of Speech Production*, volume 2. Mouton, The Hague, 1970.
- [16] Stephen B. Pope. *Turbulent Flows*. Cambridge University Press, 2000.
- [17] Fluent. *Fluent 6.2 Documentation*, 2005.

Chapter 4

Thesis Conclusions

4.1 Introduction

Computational Fluid Dynamics is a rapidly growing field, with applications in many diverse fields. Linguistics is one such field, and we will consider two recent applications of CFD to the field of linguistics. Derrick et al. applied CFD to the English /pa/ in [1], while Anderson et al. applied CFD to the English /sh/ in [2]. On the surface these may seem like two very similar studies, yet in detail one finds that they required surprisingly different approaches with CFD. The differences between the simulation requirements allowed for the relative success and usefulness of the simulations in [1] and the relative failure of the simulations in [2] to replicate the experimental results, though the latter study did offer useful findings as well.

4.2 Comparison of Papers

From the start, one can quickly note the similarities between these two studies. They are both applying CFD to problems in linguistics. They both use Fluent as the solver in the simulations. Both are concerned with turbulent flows, thus requiring an unsteady simulation. Both studies investigated a wide range of simulation settings, including 2D and 3D geometries and RANS and LES for turbulence models. Both studies assumed a static geometry. And both simulations used similar solver settings, such as time step size and numerical integration schemes.

With these similarities, both studies shared many common strengths and weaknesses. One strength possible in all numerical simulations is that a complete data set is gathered of the flow, and without worry of instru-

ments altering the flow. This advantage is used in both studies. In [1], the centerline velocity was extracted at numerous locations spanning the simulated time. Likewise, [2] examines the acoustic analogy of various surfaces. The ability to investigate the data in this manner is a great strength of simulations, and makes possible a deeper exploration of the underlying phenomenon.

On the other hand, both studies suffered under the computational costs of 3D simulations. The mesh must be kept as sparse as possible to keep the runtime down, which already was 2 weeks or longer, but a coarser mesh means poorer resolution of the flow. Thus the computational limitations also limit the quality of the results. As will be seen later, these limitations were acceptable in [1], while perhaps too limiting in [2].

Both of these simulations were performed with Fluent. The advantage here is that one doesn't have to write one's own CFD code, which would be a big project in itself. On the other hand, Fluent is designed to be a robust software used in many applications, and therefore lacks the speed, fine tuning, and specialization for these specific projects. As mentioned above, speed was a big issue for both studies, and specialization was an issue for [2] in particular which would have benefited greatly from high-order wave propagation code and efficient boundary conditions. While using Fluent does provide a method which future researchers can use, it does not provide a code-base which can be applied and enhanced in directions specific to linguistics.

Finally, both studies considered geometrical and flow simplifications, in particular, using 2D simulations is a large simplification of the geometry and the flow, and the use of turbulence models makes assumptions about the flow. A detailed analysis of these assumptions is shown later, but for the purpose of this comparison it is useful to note that all 2D flows failed to give strong results in either study, nor were RANS turbulence models of any use.

While both studies share many broad characteristics, in the details they are drastically different, and test the abilities of CFD in different aspects.

The first difference comes from the linguistics feature being studied. The

first study [1] considers the bilabial plosive /pa/. In /pa/, the speaker builds up pressure behind the lips, and then the burst is released as the lips rapidly open and the acoustic /pa/ is accompanied by a jet escaping the lips which is driven by the built up pressure[3, 4]. The sound is largely generated at the lips, and the flow around the lips is complex [5], but the flow of interest in [1] is beyond the lips, and the lips are not even included in the simulation, which is considered a significant source of error in retrospect. In contrast, the other study [2] considers the English fricative /sh/. The fricative /sh/ starts with a flow driven by pressure behind the vocal chords. The critical region occurs where the flow is channeled into a jet at a narrow constriction formed by the tongue and the roof of the mouth. The jet strikes the roof of the mouth and the teeth, thus generating sound. The sound is also modified by the cavities in the vocal tract, such as the cavity under the tongue (sublingual cavity). The non-acoustic flow that escapes the lips is of little importance, thus the flow of interest in [2] is by and large within the vocal tract. This has a huge implications in simulation design, because the boundary layer at the vocal tract walls requires a fine mesh to resolve the flow, thus adding a large computational expense and introducing wall models that were completely absent in [1].

The most critical difference, however, between the two studies is that [2] seeks to resolve the actual acoustics of the fricative, while the simulation in [1] is only concerned with the non-acoustic flow that results from the bilabial plosive. Therefore, while both simulations need to resolve the turbulence from the flow, [2] seeks to resolve the acoustic waves in addition. The consequences of this can be understood in light of the discussion of Section 1.2.2:

- The fricative study [2] needed a compressible flow solution while [1] did not.
- In [2], the double precision solver of Fluent was used (as recommended in [6]), which wasn't needed in [1].
- In [2], a buffer layer was used to gradually damp out the waves and create non-reflecting boundary conditions. This buffer layer, however,

required many computational cells and thus greatly increased the computational cost, particularly in the 3D cases. In [1], a standard pressure outlet was deemed sufficient.

- While [1] sought to capture the transient expansion of a jet out to a certain distance, [2] sought to attain as long of a statistically steady signal as possible for the sake of signal processing.

Resolving the acoustics in [2] greatly increased the computational burden, which wasn't the case in [1].

4.3 Analysis of Chapter 2 and Chapter 3

4.3.1 Two-Dimensional Flows

Both studies used 2D simulations, and found them largely inadequate, thus it is useful to consider why they failed. Section 1.2.1 considers the theoretical differences between 2D and 3D flows, which are significant, thus one should be careful when approximating a flow as 2D. The flows studied in [1] and [2] are not 2D flows, and such simulations were observed to give unphysical results. In the 2D simulations of [1], the eddies were observed to form into a strong eddy and move in the streamwise direction with an almost constant velocity. This observation agrees very well with the theory presented. Unfortunately, when the penetration rate of the jet is an important characteristic to model well, such a result is useless.

In the fricative study [2], the errors due to 2D turbulence are not as obvious because sound is being investigated and the origin of sound is not known in the NSE. Like 3D turbulence, the 2D turbulence does contain seemingly random pressure fluctuations, thus the turbulent jet will still cause a dipole sound source when it strikes a surface. On one hand, the turbulence is expected to be wrong, yet on the other hand, the 2D turbulence may be workable if it is young and thus its incorrect energy cascade hasn't fully developed. However, one clearly does observe much larger amplitudes in the 2D sound spectrum, particularly at the longer wavelengths, which is likely caused by the energy movement to the large scales.

4.3.2 RANS Simulations

As discussed in Section 1.2.2, RANS simulations like the $k - \omega$ SST time-average the flow, and assume isotropy at large scales. Both of these factors are important to the failure of RANS methods in the studies [1] and [2]. When one wishes to observe instantaneous pressure fluctuations, the averaged equations are insufficient. In the /pa/ study, the turbulent fluctuations which cause the microphone ‘pop’ were not observed, thus rendering the RANS simulations useless. Likewise, in the /sh/ study, the objective was to resolve the acoustic pressure fluctuations, which is defeated by the averaging. The RANS simulations washed out the lower frequencies greatly and completely removed the higher frequencies. As one might predict from the theory of RANS, it is not suitable for these studies.

4.3.3 LES Simulations

A couple of important consequences arise from Fluent’s implementation of LES (which is discussed in Section 1.2.2). First of all, in Fluent’s 2D implementation, C_s seems to always be zero. As a consequence μ_t will always be zero, which means that, though the filtered Navier-Stokes equations will be solved, the subgrid turbulent viscosity will always be zero. It is no surprise that the 2D LES simulations were observed to give results like the 2D simulations without a turbulence model in [2]. Initially this may seem like an error, but viewed in light of the theory of 2D turbulence, energy is cascading to the larger scales, so we do not expect viscous dissipation in the small scales. Therefore, while the 2D LES simulations do a bad job of imitating 3D flow, they are behaving as theory predicts.

In 3D, C_s is not universally zero. The ratio of turbulent viscosity μ_t to laminar viscosity $\eta_0 = 1.83 \cdot 10^{-5} [Pa \cdot s]$ was observed to be as high as 90 in [2], though in most of the domain the ratio was below one. Therefore, the LES model does provide a significant contribution to the viscous effects in locations of high turbulence, which is expected.

4.3.4 Boundary Layers

In the study of /pa/, the boundary layer limitations, as discussed in Section 1.2.2, had no effect because the lips were not included in the simulation thus there were no bounding walls which would have a boundary layer. However, in the /sh/ study, as in most other applications of CFD to linguistics, the flow is bounded by a wall thus the boundary layer becomes an important issue in simulation design and a demanding factor for computational resources. In the /sh/ study, $y_+ = 1$ is estimated to occur around $y = 0.0075\text{mm}$ around the constriction. However, the mesh cells used in this region started at 0.2mm at the wall thus placing mesh points in the logarithmic layer, but not resolving down to the viscous sublayer, therefore the Fluent wall function was used rather than complete resolution. For a stronger simulation, one should solve the boundary layer without an approximate wall function, especially when one potentially has separation and impingement as occurs in the fricative /sh/.

4.3.5 Acoustic Analogy

In the study [2], which employed the FW-H acoustic analogy, various parts of the mouth geometry were used as source surfaces, thus they clearly assumed that the non-linear flow wouldn't disrupt the spectra. In retrospect, it would have been wise to use a permeable source surface just outside of the mouth. This surface would still be in the non-linear flow, but would be better suited for propagation into free space and accounting for the quadrupole sound sources. This would have made a better comparison with the direct recordings made further from the mouth, and in a medium that was nearly still.

4.4 Usefulness to Current Research

This research is primarily of interest to two groups of people: those who care about CFD methods for the sake of future studies of airflow and acoustics in and around the vocal tract, and to those who care about any good flow

results in speech, with which they can better understand speech.

To start off, the simulations covered a wide range of CFD methods, and many of them were found not to work or to have a very limited use, thus one is warned of what does not work. For example, 2D simulations and RANS models don't perform well, as was observed both from theory and practice, and the application of such methods is inadvisable.

On the other hand, both studies did find the 3D, large eddy simulations to be the best approach. Though such an approach has a high computational cost, it is arguably the only method that was useful in either study, thus is a necessary price to pay. Special effort should be given to domain design and meshing to ensure that all that needs to be included is included, but not more.

For those who want to study acoustics, study [2] is quite useful. While it doesn't replicate the experiment very nicely, it does provide numerous lessons about which methods to use or avoid. The acoustic analogy was observed to match the direct method for most parts of the spectra, which means that one might consider running an incompressible simulation and employing only the acoustic analogy. If one does wish to still resolve the sound directly, then the flow must be compressible, but one should consider ending the domain very soon after the lips to keep the computational expense to a minimum and propagating the sound in another method.

However, these studies are not only for the linguist who wants to use CFD, but also for the linguist who wants to learn about these phenomena.

The study of the English /pa/ provides insight into the nature of the jet that is associated with the sound. First, by comparison between simulation and experiment one finds that the initial milliseconds of the burst are crucial. In this time, the shape and motion of the lips have a large influence on the jet. This insight is useful to linguists who are interested in the mechanics of a plosive. Second, one can visualize (both from simulations and experiments) how the jet evolves with time. Those interested in enhanced speech perception from feeling the burst, as well as microphone designers and users wishing to avoid capturing the burst, can find this data useful. In particular, one observes how the burst decreases in strength with distance

from the speaker and with distance away from centerline. However, this study can also interest those beyond linguistics. There is quite an interest in starting jets and puffs (see [7] for a great example), but in [1], the authors investigated a case where the pressure driving the burst faded out over time, thus studying the realm in between a starting jet (which has a constant source) and a puff (in which the source ends quickly). The author is not aware of a study that investigates this intermediate region, which may be of interest to the larger jet community and not just linguists.

The study of the English /sh/ is useful for those interested in the mechanisms of creating the fricative. While in many ways this study sought to validate the CFD findings against the Shadle findings, it may also be used to advance the knowledge of fricatives. It is very difficult to get good data for flow in the vocal tract because any instrumentation would be disruptive to the flow and very uncomfortable for the speaker. However, a CFD simulation provides such data, and in fact one may make observations and answer questions that haven't been noticed yet.

First, one may observe how long the jet is attached to the roof of the mouth. In the 3D simulation the flow stays attached to the roof of the mouth thus causing the jet to largely strike the upper tooth at a rather oblique angle. In comparison, the 2D flow detaches from the roof of the mouth soon after the constriction causing a larger portion of the jet to strike the lower tooth at an angle close to 90 degrees. In this case, due to the significantly better wall resolution of the 2D mesh, the 2D results may be better. This attachment may have important acoustical implications because the angle at which the jet strikes the obstruction effects the strength of the dipole sound source [8, 3]. The simulations in [2], as well as the Shadle experiments, consider the roof of the mouth to be smooth, but in reality there are small ridges which may effect flow attachment much like the dimples on a golf ball [9]. This observation, arising from observations of the simulation data, provides an interesting topic for future research.

Second, one may observe vortices in the sublingual cavity. In 2D there are two fairly distinctive vortices which fill up the cavity, while in 3D these vortices seem to form much more slowly and are much weaker. Such vortices,

however, may have acoustical consequences as suggested by Powell [10] who relates vortices to sound generation in turbulence.

Third, one may observe the flow as it escapes between the teeth. Howe et al suggest that the primary source of sound in /s/ arises from “the ‘diffraction’ of jet turbulence pressure fluctuations by the incisors” [11]. One interesting processing which can be done is to filter the data for individual frequencies to observe the extent to which sound is created by this diffraction at the teeth. In fact, filtering the flow by wavelengths would be very insightful to observe the generation and resonances of various wavelengths. While this would be a very useful application, it is left as a topic of future research. However, just with initial observation one may notice that the lower tooth doesn’t split the main strength of the turbulent flow in the 3D simulation because the flow stays attached to the roof of the mouth and thus doesn’t detach to strike the lower tooth with strength.

4.5 Further Research

Both [1] and [2] demonstrate the feasibility and usefulness of using CFD in linguistics problems, but both also leave many questions yet to be answered by future research.

As mentioned above, creating a finely meshed 3D model of the fricative /sh/ to study the flow characteristics and performing post-processing filtering to observe the sound source mechanisms in detail would be an interesting study. This would be of great theoretical interest. However, such a simulation, which requires weeks of runtime, doesn’t help those interested in a fast fricative model, such as can be used in real time voice simulation. Therefore, once one has a simulation that can replicate the sound produced in a fricative, one should consider how a fast model can be derived which considers the flow and sound generation properties yet solves quickly.

A big leap forward would include a dynamic geometry in the simulations. A dynamic geometry could take on two forms. The simpler would be a geometry that is changing in a predetermined way. An example of this could be including the initial lip separation in the experiments of [1]. However,

the more complex type of dynamic geometry would involve the structure reacting to the pressures within the flow, thus involving fluid-structure interaction, or FSI. One must describe the properties of the walls to simulate how they will react for forces in the fluid, which greatly complicates the simulation.

In either case, a dynamic geometry adds numerous complexities to a simulation. As the geometry changes, the mesh must change as well. One may attempt to stretch the mesh for a mildly deforming geometry, but if the changes are large one will need to remesh. The remeshing should be automated for the sake of consistency and because this will happen numerous times in the simulation. Furthermore, one must effectively transfer the solution from the mesh of the old geometry to the mesh of the new geometry for each time the geometry is updated.

Though the challenges are great with a dynamic geometry, so are the possibilities. Many motions of the vocal tract cannot be modeled without considering a FSI. For example, to model snoring or obstructive sleep apnea one must consider the way the vocal tract reacts to the airflow, and indeed, such simulations have been envisioned [12].

One step in this direction is to closely link a CFD solver with a biomechanical computational model of the vocal tract, like that provided by the Artisynt project. Such a solver can already calculate the reaction of the vocal tract to external forces, thus a large part of the problem is already solved.

Depending on the application, one might even consider dropping a CFD simulation in favor of the mesh-free smoothed-particle hydrodynamics (SPH) method. It is unlikely that SPH would do a good job for resolving sound waves, but it may provide a fast way to find the gross flow characteristics, and perhaps provides a more natural and simple solution to dynamic geometries.

4.6 Conclusion

Thus concludes the studies of /pa/ and /sh/. These two diverse CFD studies are seen to be useful to CFD practitioners and linguists alike. The author suggests that CFD simulations will become an increasingly useful tool in linguistics in the future.

4.7 Thesis Contributions

- Examines the jet of air associated with the bilabial plosive ‘pa’. The initial 40ms of the burst are found to be critical to the jet penetration rate. Simulations are found to accurately model the flow after 40ms.
- Examines the flow and acoustics of the fricative ‘sh’. Three-dimensional, large eddy simulations are found to be the best approach, though not closely matching experimental data. The acoustic analogy is found to agree well with direct measurements.
- Gives details of CFD methods useful to speech, and also CFD methods which should be avoided.

4.8 Bibliography

- [1] D. Derrick, P. Anderson, B. Gick, and S. Green. Characteristics of Air Puffs Produced in English ‘pa’: Data and Simulations. Preprint, 2008.
- [2] P. Anderson, S. Green, and S. Fels. Computational Aeroacoustic Simulations of the English Fricative ‘sh’. Preprint, 2008.
- [3] Kenneth Stevens. *Acoustic Phonetics*. The MIT Press, Cambridge, 2000.
- [4] Osamu Fujimura. Bilabial Stop and Nasal Consonants: a Motion Picture Study and its Acoustical Implications. *Journal of Speech and Hearing Research*, 4(3):233–247, 1961.
- [5] X. Pelorson, G. C. J. Hofmans, M. Ranucci, and R. C. M. Bosch. On the fluid mechanics of bilabial plosives. *Speech Communication*, 22(2-3):155–172–, 1997.
- [6] Fluent. *Fluent 6.2 Documentation*, 2005.
- [7] F. J. Diez, R. Sangras, O. C. Kwon, and G. M. Faeth. Self-preserving properties of unsteady round nonbuoyant turbulent starting jets and puffs in still fluids (vol 124, pg 460, 2002). *Journal of Heat Transfer-Transactions of the Asme*, 125(1):204–205, 2003.
- [8] Christine H. Shadle. Articulatory-Acoustic Relationships In Fricative Consonants. In *Speech Production and Speech Modelling*, pages 187–209–. Kluwer Academic, Netherlands, 1990.
- [9] Pijush K. Kundu and Ira M. Cohen. *Fluid Mechanics*, volume 2. Academic Press, San Diego, 2002.
- [10] Alan Powell. Theory of Vortex Sound. *The Journal Acoustical Society of America*, 36(1):177–195–, 1964.
- [11] M. S. Howe and R. S. McGowan. Aeroacoustics of [s]. *Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences*, 461(2056):1005–1028–, 2005.

4.8. Bibliography

- [12] P. Nithiarasu, O. Hassan, K. Morgan, N. P. Weatherill, C. Fielder, H. Whittet, P. Ebdon, and K. R. Lewis. Steady flow through a realistic human upper airway geometry. *International Journal for Numerical Methods in Fluids*, 57(5):631–651, 2008.